# Frame Level Speech Recognize

February 26, 2021

```python
[1]: import time
     import numpy as np
     import torch
     import matplotlib.pyplot as plt
     import seaborn
     import os
     import sys
     import pandas as pd
     import torch.nn as nn
     import torch.nn.functional as F
     from torch.utils.data import DataLoader,Dataset
     import torch.optim as optim
     from sklearn.metrics import accuracy_score
     %matplotlib inline
```

```python
[2]: X_train_path = r"C:\Users\Michael\Documents\ALY 6020\group_project\final\train.
      ↪npy"
     y_train_path = r"C:\Users\Michael\Documents\ALY␣
      ↪6020\group_project\final\train_labels.npy"

     X_val_path = r"C:\Users\Michael\Documents\ALY 6020\group_project\final\dev.npy"
     y_val_path = r"C:\Users\Michael\Documents\ALY␣
      ↪6020\group_project\final\dev_labels.npy"
```

```python
[3]: cuda = torch.cuda.is_available()
     num_workers = 0 if cuda else 0
     print("Cuda = " +str(cuda)+" with num_workders = "+str(num_workers))
```

```
Cuda = True with num_workders = 0
```

```python
[49]: # Define dataset
      class WJSdataset(Dataset):
          def __init__(self, X_path, y_path,k):
              self.X = np.load(X_path,allow_pickle=True)
              self.X = np.concatenate(self.X,axis = 0)
              self.X = np.pad(self.X, pad_width = ((k,k),(0,0)), mode = 'constant')

              self.y = np.load(y_path,allow_pickle=True)
```

```python
        self.y = np.concatenate(self.y,axis = None)

        #for i in range(self.X.shape[0]):
        #    self.X[i] = np.pad(self.X[i], pad_width =  ((k, k), (0, 0)),
        #                    mode = 'constant')
        #self.X = np.concatenate(self.X,axis = 0).
        #    for j in range(self.X[i].shape[0]):
        #        x = np.concatenate(X_temp[j:j+2*k+1],axis = None)
    def __len__(self):
        return len(self.y)

    def __getitem__(self, index):

        x = self.X[index:index+2*context+1].flatten()
        y = self.y[index]
        x = torch.from_numpy(x).float()
        y = torch.from_numpy(np.array(y)).long()

        return x, y
```

```python
[95]: context = 20
```

```python
[96]: train = WJSdataset(X_train_path,y_train_path,context)
      val = WJSdataset(X_val_path,y_val_path,context)
```

```python
[52]: y = np.load(y_train_path,allow_pickle=True)
      with np.printoptions(threshold=np.inf):
          print(y[155])
```

```
[ 0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.
  0.  0.  0.  0.  0. 20. 20. 20. 20. 20.  9.  9.  9.  9. 58. 58. 58. 58.
 58. 58. 58. 67. 67. 67. 67. 67. 67. 67. 37. 37. 37. 37. 46. 46. 46. 35.
 35. 35. 40. 40. 40. 40. 40. 40. 40. 40. 43. 43. 43. 43. 43. 43. 43. 43.
 43. 43. 31. 31. 31. 31. 31. 31. 31. 31. 31. 31. 31. 31. 31. 31. 45. 45.
 45. 45. 45. 45. 45. 45. 45. 45. 58. 58. 58. 58. 58. 58. 58. 58. 58. 58.
 64. 64. 64. 64. 44. 44. 44. 44. 44. 44. 44. 44. 44. 44. 44. 44. 44. 61.
 61. 61. 61. 61. 61. 61. 61. 61. 61. 61. 61. 61. 61. 61. 61. 61. 61. 61.
 61. 61. 61. 43. 43. 43. 43. 43. 43. 43. 43. 43. 43. 43. 43. 43. 43. 43.
 43. 43.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.
  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.  0.
 23. 23. 23. 23. 23. 23. 23. 25. 25. 25. 25. 55. 55. 55. 55. 55. 55. 67.
 67. 67. 67. 67. 67. 67.  8.  8.  8.  8. 69. 69. 69. 69. 69. 69. 69. 69.
 69. 46. 46. 46. 46. 46. 46. 49. 49. 49. 49. 49. 49. 49. 49. 49. 43. 43.
 43. 43. 43. 43. 43. 43. 43. 43. 43. 40. 40. 40. 40. 40. 40. 40.
 40. 40. 35. 35. 35. 35. 35. 35. 35. 35. 35. 35. 49. 49. 49. 49. 49. 49.
 49. 49. 49. 49. 44. 44. 44. 44. 44. 44. 44. 58. 58. 58. 58. 58. 58. 58.
 58. 58. 64. 64. 64. 64. 64. 64. 64. 64. 64. 64. 64. 64. 36. 36. 36.
 36. 36. 36. 36. 36. 36. 36. 58. 58. 58. 58. 58. 58. 58.  0.  0.  0.  0.
```

```
        0.  0.  0.  0.  0.  0.  0.  0.  0.  0.]
```

[53]: 
```python
X_Viz = np.load(X_train_path, allow_pickle=True)
```

[54]: 
```python
plt.figure(figsize=(12,40))
plt.xlabel('num_frame')
plt.ylabel('Mel scaled frequency')
plt.title('The Mel spectrogram of the first trainning utterance')
plt.imshow(X_Viz[0].T)
print(X_Viz[0].T.shape)
del X_Viz
```

```
(40, 1184)
```



[ ]:
```python
"""import librosa
import librosa.display
import IPython.display as ipd

test_audio = librosa.feature.inverse.mel_to_audio(X_val[1500],window =
    'hamming',\
                                                  sr=8000,n_fft=512)
X_val[155].shape[0]
ipd.Audio(test_audio,rate = 8000)"""
```

[99]: 
```python
len(train)
```

[99]: 
```
18482968
```

[97]: 
```python
train_loader_arg = dict(shuffle = True,
                        batch_size = 4096,
                        pin_memory = True, drop_last = True)
train_loader = DataLoader(train,**train_loader_arg, num_workers = num_workers)

val_loader_arg = dict(shuffle = False, batch_size = 4096, pin_memory = True)
val_loader = DataLoader(val,**val_loader_arg, num_workers = num_workers)
```

[80]: 
```python
class CreateModel(nn.Module):
    def __init__(self, num_node_list):
        super(CreateModel,self).__init__()
        layers = []
```

3

```
        self.num_node_list = num_node_list
        for i in range(len(num_node_list)-2):
            layers.append(nn.Linear(num_node_list[i],num_node_list[i+1]))
            layers.append(nn.Dropout(p = 0.3))
            layers.append(nn.ReLU())
            layers.append(nn.BatchNorm1d(num_node_list[i+1]))
        layers.append(nn.Linear(num_node_list[-2], num_node_list[-1]))
        self.net = nn.Sequential(*layers)

    def forward(self, x):
        return self.net(x)
```

[81]:
```
def one_epoch(model, train_loader, criterion, optimizer):
    model.train()

    running_loss = 0.0
    total_prediction = 0.0
    correct_predictions = 0.0

    start_time = time.time()
    for idx, (data,target) in enumerate(train_loader):

        optimizer.zero_grad()
        data = data.to(device)
        target = target.to(device)

        outputs = model(data)
        loss = criterion(outputs, target)
        running_loss += loss.item()
        #print(outputs.data)
        #print(outputs.data.shape)
        loss.backward()
        optimizer.step()

        _, predicted = torch.max(outputs.data, 1)
        #print(target.size(0))
        #print(predicted.shape)
        total_prediction += target.shape[0]
        correct_predictions += (predicted == target).sum().item()
        #print(correct_predictions)
        print([idx],'Train Loss: ', loss)
    end_time = time.time()

    running_loss /= len(train_loader)
    print('Training Loss: ', running_loss, 'Time: ', end_time - start_time, 's')

    acc = (correct_predictions/total_prediction)*100.0
```

```python
        print("Training Accuaracy: ", acc, '%')
        return running_loss, acc
```

```python
[82]: def val_model(model, val_loader, criterion):
          with torch.no_grad():
              model.eval()
              running_loss = 0.0
              total_prediction = 0.0
              correct_predictions = 0.0

              for idx, (data,target) in enumerate(val_loader):

                  data = data.to(device)
                  target = target.to(device)

                  outputs = model(data)
                  loss = criterion(outputs, target).detach()
                  print([idx],'Valid Loss: ', loss)
                  running_loss += loss.item()
                  #print(outputs.data)
                  _, predicted = torch.max(outputs.data,1)
                  total_prediction += target.size(0)
                  #print(target.size(0))
                  correct_predictions += (predicted == target).sum().item()
                  #print(correct_predictions)

              running_loss /= len(val_loader)
              print('Testing Loss: ', running_loss)

              acc = (correct_predictions/total_prediction)*100.0
              print("Testing Accuaracy: ", acc, '%')
              return running_loss, acc
```

```python
[83]: def training_plot(a, b):
          plt.figure(1)
          plt.plot(a,'b',label = "train")
          plt.plot(b,'g', label = "test")
          plt.xlabel("epochs")
          plt.ylabel("Loss")
          plt.title('Training/Test Loss')
          plt.legend()
          plt.show()
```

```python
[84]: def training_plot_acc(a, b):
          plt.figure(1)
```

```
    plt.plot(a,'b',label = "train")
    plt.plot(b,'g', label = "test")
    plt.xlabel("epochs")
    plt.ylabel("Accuracy")
    plt.title('Training/Test Accuracy')
    plt.legend()
    plt.show()
```

[85]:
```
print('context: ',context)
model = CreateModel([40*(2*context+1),2048,2048,1024,512,256,71])
criterion = nn.CrossEntropyLoss()
optimizer = optim.Adam(model.parameters(),lr=2e-3)
device = torch.device("cuda")
model.to(device)
print(model)
```

```
context:  25
CreateModel(
  (net): Sequential(
    (0): Linear(in_features=2040, out_features=2048, bias=True)
    (1): Dropout(p=0.3, inplace=False)
    (2): ReLU()
    (3): BatchNorm1d(2048, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    (4): Linear(in_features=2048, out_features=2048, bias=True)
    (5): Dropout(p=0.3, inplace=False)
    (6): ReLU()
    (7): BatchNorm1d(2048, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    (8): Linear(in_features=2048, out_features=1024, bias=True)
    (9): Dropout(p=0.3, inplace=False)
    (10): ReLU()
    (11): BatchNorm1d(1024, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    (12): Linear(in_features=1024, out_features=512, bias=True)
    (13): Dropout(p=0.3, inplace=False)
    (14): ReLU()
    (15): BatchNorm1d(512, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    (16): Linear(in_features=512, out_features=256, bias=True)
    (17): Dropout(p=0.3, inplace=False)
    (18): ReLU()
    (19): BatchNorm1d(256, eps=1e-05, momentum=0.1, affine=True,
track_running_stats=True)
    (20): Linear(in_features=256, out_features=71, bias=True)
  )
)
```

```python
[86]: Path = r'C:\Users\Michael\Documents\ALY 6020\group_project\final'
      print(Path)
```

C:\Users\Michael\Documents\ALY 6020\group_project\final

```python
[87]: num_epochs = 10
      Train_loss = []
      Test_loss = []
      Train_acc = []
      Test_acc = []


      for i in range(num_epochs):
          print("Epoch Number: ", i)
          train_loss, train_acc = one_epoch(model,
                                            train_loader,
                                            criterion=criterion,
                                            optimizer=optimizer)
          test_loss, test_acc = val_model(model,val_loader,
                                          criterion=criterion)
          torch.save(model,os.path.join(Path,'{}.pt'.format('model_V3_'+ str(i))))

          Train_loss.append(train_loss)
          Test_loss.append(test_loss)
          Train_acc.append(train_acc)
          Test_acc.append(test_acc)

      torch.save(model.state_dict(), os.path.join(Path,'{}.pt'.
       →format('modelDict_V3')))
      training_plot(Train_loss,Test_loss)
```

```
Epoch Number:  0
[0] Train Loss:  tensor(4.4351, device='cuda:0', grad_fn=<NllLossBackward>)
[1] Train Loss:  tensor(4.0518, device='cuda:0', grad_fn=<NllLossBackward>)
[2] Train Loss:  tensor(3.6510, device='cuda:0', grad_fn=<NllLossBackward>)
[3] Train Loss:  tensor(3.3554, device='cuda:0', grad_fn=<NllLossBackward>)
[4] Train Loss:  tensor(3.1982, device='cuda:0', grad_fn=<NllLossBackward>)
[5] Train Loss:  tensor(3.0103, device='cuda:0', grad_fn=<NllLossBackward>)
[6] Train Loss:  tensor(2.9034, device='cuda:0', grad_fn=<NllLossBackward>)
[7] Train Loss:  tensor(2.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[8] Train Loss:  tensor(2.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[9] Train Loss:  tensor(2.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[10] Train Loss:  tensor(2.6288, device='cuda:0', grad_fn=<NllLossBackward>)
[11] Train Loss:  tensor(2.5719, device='cuda:0', grad_fn=<NllLossBackward>)
[12] Train Loss:  tensor(2.5456, device='cuda:0', grad_fn=<NllLossBackward>)
[13] Train Loss:  tensor(2.4636, device='cuda:0', grad_fn=<NllLossBackward>)
[14] Train Loss:  tensor(2.4513, device='cuda:0', grad_fn=<NllLossBackward>)
[15] Train Loss:  tensor(2.3872, device='cuda:0', grad_fn=<NllLossBackward>)
[16] Train Loss:  tensor(2.3315, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[17] Train Loss:   tensor(2.3269, device='cuda:0', grad_fn=<NllLossBackward>)
[18] Train Loss:   tensor(2.3457, device='cuda:0', grad_fn=<NllLossBackward>)
[19] Train Loss:   tensor(2.3097, device='cuda:0', grad_fn=<NllLossBackward>)
[20] Train Loss:   tensor(2.3058, device='cuda:0', grad_fn=<NllLossBackward>)
[21] Train Loss:   tensor(2.1952, device='cuda:0', grad_fn=<NllLossBackward>)
[22] Train Loss:   tensor(2.2203, device='cuda:0', grad_fn=<NllLossBackward>)
[23] Train Loss:   tensor(2.2442, device='cuda:0', grad_fn=<NllLossBackward>)
[24] Train Loss:   tensor(2.1814, device='cuda:0', grad_fn=<NllLossBackward>)
[25] Train Loss:   tensor(2.1541, device='cuda:0', grad_fn=<NllLossBackward>)
[26] Train Loss:   tensor(2.1494, device='cuda:0', grad_fn=<NllLossBackward>)
[27] Train Loss:   tensor(2.1603, device='cuda:0', grad_fn=<NllLossBackward>)
[28] Train Loss:   tensor(2.1847, device='cuda:0', grad_fn=<NllLossBackward>)
[29] Train Loss:   tensor(2.1427, device='cuda:0', grad_fn=<NllLossBackward>)
[30] Train Loss:   tensor(2.1149, device='cuda:0', grad_fn=<NllLossBackward>)
[31] Train Loss:   tensor(2.0865, device='cuda:0', grad_fn=<NllLossBackward>)
[32] Train Loss:   tensor(2.0605, device='cuda:0', grad_fn=<NllLossBackward>)
[33] Train Loss:   tensor(2.0571, device='cuda:0', grad_fn=<NllLossBackward>)
[34] Train Loss:   tensor(2.0761, device='cuda:0', grad_fn=<NllLossBackward>)
[35] Train Loss:   tensor(2.0683, device='cuda:0', grad_fn=<NllLossBackward>)
[36] Train Loss:   tensor(2.0297, device='cuda:0', grad_fn=<NllLossBackward>)
[37] Train Loss:   tensor(2.0407, device='cuda:0', grad_fn=<NllLossBackward>)
[38] Train Loss:   tensor(2.0184, device='cuda:0', grad_fn=<NllLossBackward>)
[39] Train Loss:   tensor(1.9961, device='cuda:0', grad_fn=<NllLossBackward>)
[40] Train Loss:   tensor(1.9662, device='cuda:0', grad_fn=<NllLossBackward>)
[41] Train Loss:   tensor(2.0023, device='cuda:0', grad_fn=<NllLossBackward>)
[42] Train Loss:   tensor(2.0259, device='cuda:0', grad_fn=<NllLossBackward>)
[43] Train Loss:   tensor(1.9939, device='cuda:0', grad_fn=<NllLossBackward>)
[44] Train Loss:   tensor(1.9839, device='cuda:0', grad_fn=<NllLossBackward>)
[45] Train Loss:   tensor(1.9936, device='cuda:0', grad_fn=<NllLossBackward>)
[46] Train Loss:   tensor(1.9768, device='cuda:0', grad_fn=<NllLossBackward>)
[47] Train Loss:   tensor(1.9656, device='cuda:0', grad_fn=<NllLossBackward>)
[48] Train Loss:   tensor(1.9870, device='cuda:0', grad_fn=<NllLossBackward>)
[49] Train Loss:   tensor(1.9835, device='cuda:0', grad_fn=<NllLossBackward>)
[50] Train Loss:   tensor(2.0045, device='cuda:0', grad_fn=<NllLossBackward>)
[51] Train Loss:   tensor(1.8876, device='cuda:0', grad_fn=<NllLossBackward>)
[52] Train Loss:   tensor(1.9599, device='cuda:0', grad_fn=<NllLossBackward>)
[53] Train Loss:   tensor(1.8695, device='cuda:0', grad_fn=<NllLossBackward>)
[54] Train Loss:   tensor(1.9465, device='cuda:0', grad_fn=<NllLossBackward>)
[55] Train Loss:   tensor(1.8880, device='cuda:0', grad_fn=<NllLossBackward>)
[56] Train Loss:   tensor(1.8743, device='cuda:0', grad_fn=<NllLossBackward>)
[57] Train Loss:   tensor(1.8951, device='cuda:0', grad_fn=<NllLossBackward>)
[58] Train Loss:   tensor(1.9336, device='cuda:0', grad_fn=<NllLossBackward>)
[59] Train Loss:   tensor(1.9358, device='cuda:0', grad_fn=<NllLossBackward>)
[60] Train Loss:   tensor(1.9072, device='cuda:0', grad_fn=<NllLossBackward>)
[61] Train Loss:   tensor(1.9104, device='cuda:0', grad_fn=<NllLossBackward>)
[62] Train Loss:   tensor(1.9028, device='cuda:0', grad_fn=<NllLossBackward>)
[63] Train Loss:   tensor(1.8976, device='cuda:0', grad_fn=<NllLossBackward>)
[64] Train Loss:   tensor(1.9093, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[65] Train Loss:  tensor(1.8599, device='cuda:0', grad_fn=<NllLossBackward>)
[66] Train Loss:  tensor(1.8757, device='cuda:0', grad_fn=<NllLossBackward>)
[67] Train Loss:  tensor(1.8491, device='cuda:0', grad_fn=<NllLossBackward>)
[68] Train Loss:  tensor(1.8971, device='cuda:0', grad_fn=<NllLossBackward>)
[69] Train Loss:  tensor(1.8004, device='cuda:0', grad_fn=<NllLossBackward>)
[70] Train Loss:  tensor(1.9302, device='cuda:0', grad_fn=<NllLossBackward>)
[71] Train Loss:  tensor(1.8584, device='cuda:0', grad_fn=<NllLossBackward>)
[72] Train Loss:  tensor(1.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[73] Train Loss:  tensor(1.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[74] Train Loss:  tensor(1.8535, device='cuda:0', grad_fn=<NllLossBackward>)
[75] Train Loss:  tensor(1.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[76] Train Loss:  tensor(1.8638, device='cuda:0', grad_fn=<NllLossBackward>)
[77] Train Loss:  tensor(1.8547, device='cuda:0', grad_fn=<NllLossBackward>)
[78] Train Loss:  tensor(1.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[79] Train Loss:  tensor(1.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[80] Train Loss:  tensor(1.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[81] Train Loss:  tensor(1.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[82] Train Loss:  tensor(1.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[83] Train Loss:  tensor(1.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[84] Train Loss:  tensor(1.8183, device='cuda:0', grad_fn=<NllLossBackward>)
[85] Train Loss:  tensor(1.8267, device='cuda:0', grad_fn=<NllLossBackward>)
[86] Train Loss:  tensor(1.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[87] Train Loss:  tensor(1.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[88] Train Loss:  tensor(1.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[89] Train Loss:  tensor(1.8036, device='cuda:0', grad_fn=<NllLossBackward>)
[90] Train Loss:  tensor(1.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[91] Train Loss:  tensor(1.7980, device='cuda:0', grad_fn=<NllLossBackward>)
[92] Train Loss:  tensor(1.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[93] Train Loss:  tensor(1.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[94] Train Loss:  tensor(1.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[95] Train Loss:  tensor(1.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[96] Train Loss:  tensor(1.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[97] Train Loss:  tensor(1.7844, device='cuda:0', grad_fn=<NllLossBackward>)
[98] Train Loss:  tensor(1.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[99] Train Loss:  tensor(1.7976, device='cuda:0', grad_fn=<NllLossBackward>)
[100] Train Loss:  tensor(1.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[101] Train Loss:  tensor(1.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[102] Train Loss:  tensor(1.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[103] Train Loss:  tensor(1.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[104] Train Loss:  tensor(1.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[105] Train Loss:  tensor(1.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[106] Train Loss:  tensor(1.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[107] Train Loss:  tensor(1.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[108] Train Loss:  tensor(1.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[109] Train Loss:  tensor(1.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[110] Train Loss:  tensor(1.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[111] Train Loss:  tensor(1.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[112] Train Loss:  tensor(1.7378, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[113] Train Loss:   tensor(1.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[114] Train Loss:   tensor(1.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[115] Train Loss:   tensor(1.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[116] Train Loss:   tensor(1.6492, device='cuda:0', grad_fn=<NllLossBackward>)
[117] Train Loss:   tensor(1.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[118] Train Loss:   tensor(1.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[119] Train Loss:   tensor(1.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[120] Train Loss:   tensor(1.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[121] Train Loss:   tensor(1.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[122] Train Loss:   tensor(1.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[123] Train Loss:   tensor(1.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[124] Train Loss:   tensor(1.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[125] Train Loss:   tensor(1.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[126] Train Loss:   tensor(1.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[127] Train Loss:   tensor(1.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[128] Train Loss:   tensor(1.6499, device='cuda:0', grad_fn=<NllLossBackward>)
[129] Train Loss:   tensor(1.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[130] Train Loss:   tensor(1.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[131] Train Loss:   tensor(1.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[132] Train Loss:   tensor(1.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[133] Train Loss:   tensor(1.6684, device='cuda:0', grad_fn=<NllLossBackward>)
[134] Train Loss:   tensor(1.6499, device='cuda:0', grad_fn=<NllLossBackward>)
[135] Train Loss:   tensor(1.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[136] Train Loss:   tensor(1.6127, device='cuda:0', grad_fn=<NllLossBackward>)
[137] Train Loss:   tensor(1.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[138] Train Loss:   tensor(1.6692, device='cuda:0', grad_fn=<NllLossBackward>)
[139] Train Loss:   tensor(1.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[140] Train Loss:   tensor(1.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[141] Train Loss:   tensor(1.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[142] Train Loss:   tensor(1.6635, device='cuda:0', grad_fn=<NllLossBackward>)
[143] Train Loss:   tensor(1.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[144] Train Loss:   tensor(1.6219, device='cuda:0', grad_fn=<NllLossBackward>)
[145] Train Loss:   tensor(1.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[146] Train Loss:   tensor(1.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[147] Train Loss:   tensor(1.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[148] Train Loss:   tensor(1.6684, device='cuda:0', grad_fn=<NllLossBackward>)
[149] Train Loss:   tensor(1.6565, device='cuda:0', grad_fn=<NllLossBackward>)
[150] Train Loss:   tensor(1.6328, device='cuda:0', grad_fn=<NllLossBackward>)
[151] Train Loss:   tensor(1.6423, device='cuda:0', grad_fn=<NllLossBackward>)
[152] Train Loss:   tensor(1.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[153] Train Loss:   tensor(1.6581, device='cuda:0', grad_fn=<NllLossBackward>)
[154] Train Loss:   tensor(1.6278, device='cuda:0', grad_fn=<NllLossBackward>)
[155] Train Loss:   tensor(1.6570, device='cuda:0', grad_fn=<NllLossBackward>)
[156] Train Loss:   tensor(1.6492, device='cuda:0', grad_fn=<NllLossBackward>)
[157] Train Loss:   tensor(1.6474, device='cuda:0', grad_fn=<NllLossBackward>)
[158] Train Loss:   tensor(1.6250, device='cuda:0', grad_fn=<NllLossBackward>)
[159] Train Loss:   tensor(1.6113, device='cuda:0', grad_fn=<NllLossBackward>)
[160] Train Loss:   tensor(1.6386, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[161] Train Loss:   tensor(1.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[162] Train Loss:   tensor(1.6056, device='cuda:0', grad_fn=<NllLossBackward>)
[163] Train Loss:   tensor(1.6172, device='cuda:0', grad_fn=<NllLossBackward>)
[164] Train Loss:   tensor(1.6179, device='cuda:0', grad_fn=<NllLossBackward>)
[165] Train Loss:   tensor(1.6362, device='cuda:0', grad_fn=<NllLossBackward>)
[166] Train Loss:   tensor(1.6157, device='cuda:0', grad_fn=<NllLossBackward>)
[167] Train Loss:   tensor(1.6236, device='cuda:0', grad_fn=<NllLossBackward>)
[168] Train Loss:   tensor(1.6357, device='cuda:0', grad_fn=<NllLossBackward>)
[169] Train Loss:   tensor(1.6253, device='cuda:0', grad_fn=<NllLossBackward>)
[170] Train Loss:   tensor(1.5957, device='cuda:0', grad_fn=<NllLossBackward>)
[171] Train Loss:   tensor(1.5882, device='cuda:0', grad_fn=<NllLossBackward>)
[172] Train Loss:   tensor(1.6283, device='cuda:0', grad_fn=<NllLossBackward>)
[173] Train Loss:   tensor(1.6189, device='cuda:0', grad_fn=<NllLossBackward>)
[174] Train Loss:   tensor(1.6339, device='cuda:0', grad_fn=<NllLossBackward>)
[175] Train Loss:   tensor(1.6503, device='cuda:0', grad_fn=<NllLossBackward>)
[176] Train Loss:   tensor(1.6205, device='cuda:0', grad_fn=<NllLossBackward>)
[177] Train Loss:   tensor(1.6006, device='cuda:0', grad_fn=<NllLossBackward>)
[178] Train Loss:   tensor(1.6159, device='cuda:0', grad_fn=<NllLossBackward>)
[179] Train Loss:   tensor(1.6073, device='cuda:0', grad_fn=<NllLossBackward>)
[180] Train Loss:   tensor(1.5813, device='cuda:0', grad_fn=<NllLossBackward>)
[181] Train Loss:   tensor(1.5907, device='cuda:0', grad_fn=<NllLossBackward>)
[182] Train Loss:   tensor(1.5781, device='cuda:0', grad_fn=<NllLossBackward>)
[183] Train Loss:   tensor(1.6208, device='cuda:0', grad_fn=<NllLossBackward>)
[184] Train Loss:   tensor(1.5989, device='cuda:0', grad_fn=<NllLossBackward>)
[185] Train Loss:   tensor(1.5591, device='cuda:0', grad_fn=<NllLossBackward>)
[186] Train Loss:   tensor(1.5667, device='cuda:0', grad_fn=<NllLossBackward>)
[187] Train Loss:   tensor(1.5681, device='cuda:0', grad_fn=<NllLossBackward>)
[188] Train Loss:   tensor(1.6141, device='cuda:0', grad_fn=<NllLossBackward>)
[189] Train Loss:   tensor(1.5845, device='cuda:0', grad_fn=<NllLossBackward>)
[190] Train Loss:   tensor(1.5572, device='cuda:0', grad_fn=<NllLossBackward>)
[191] Train Loss:   tensor(1.5773, device='cuda:0', grad_fn=<NllLossBackward>)
[192] Train Loss:   tensor(1.5591, device='cuda:0', grad_fn=<NllLossBackward>)
[193] Train Loss:   tensor(1.5379, device='cuda:0', grad_fn=<NllLossBackward>)
[194] Train Loss:   tensor(1.5884, device='cuda:0', grad_fn=<NllLossBackward>)
[195] Train Loss:   tensor(1.5845, device='cuda:0', grad_fn=<NllLossBackward>)
[196] Train Loss:   tensor(1.6329, device='cuda:0', grad_fn=<NllLossBackward>)
[197] Train Loss:   tensor(1.5852, device='cuda:0', grad_fn=<NllLossBackward>)
[198] Train Loss:   tensor(1.5561, device='cuda:0', grad_fn=<NllLossBackward>)
[199] Train Loss:   tensor(1.5547, device='cuda:0', grad_fn=<NllLossBackward>)
[200] Train Loss:   tensor(1.5515, device='cuda:0', grad_fn=<NllLossBackward>)
[201] Train Loss:   tensor(1.5849, device='cuda:0', grad_fn=<NllLossBackward>)
[202] Train Loss:   tensor(1.5640, device='cuda:0', grad_fn=<NllLossBackward>)
[203] Train Loss:   tensor(1.5603, device='cuda:0', grad_fn=<NllLossBackward>)
[204] Train Loss:   tensor(1.5798, device='cuda:0', grad_fn=<NllLossBackward>)
[205] Train Loss:   tensor(1.5901, device='cuda:0', grad_fn=<NllLossBackward>)
[206] Train Loss:   tensor(1.5414, device='cuda:0', grad_fn=<NllLossBackward>)
[207] Train Loss:   tensor(1.5786, device='cuda:0', grad_fn=<NllLossBackward>)
[208] Train Loss:   tensor(1.5594, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[209] Train Loss:  tensor(1.5558, device='cuda:0', grad_fn=<NllLossBackward>)
[210] Train Loss:  tensor(1.5646, device='cuda:0', grad_fn=<NllLossBackward>)
[211] Train Loss:  tensor(1.5529, device='cuda:0', grad_fn=<NllLossBackward>)
[212] Train Loss:  tensor(1.5719, device='cuda:0', grad_fn=<NllLossBackward>)
[213] Train Loss:  tensor(1.6007, device='cuda:0', grad_fn=<NllLossBackward>)
[214] Train Loss:  tensor(1.4760, device='cuda:0', grad_fn=<NllLossBackward>)
[215] Train Loss:  tensor(1.5429, device='cuda:0', grad_fn=<NllLossBackward>)
[216] Train Loss:  tensor(1.5105, device='cuda:0', grad_fn=<NllLossBackward>)
[217] Train Loss:  tensor(1.5388, device='cuda:0', grad_fn=<NllLossBackward>)
[218] Train Loss:  tensor(1.5689, device='cuda:0', grad_fn=<NllLossBackward>)
[219] Train Loss:  tensor(1.5590, device='cuda:0', grad_fn=<NllLossBackward>)
[220] Train Loss:  tensor(1.5523, device='cuda:0', grad_fn=<NllLossBackward>)
[221] Train Loss:  tensor(1.5048, device='cuda:0', grad_fn=<NllLossBackward>)
[222] Train Loss:  tensor(1.4954, device='cuda:0', grad_fn=<NllLossBackward>)
[223] Train Loss:  tensor(1.5295, device='cuda:0', grad_fn=<NllLossBackward>)
[224] Train Loss:  tensor(1.5442, device='cuda:0', grad_fn=<NllLossBackward>)
[225] Train Loss:  tensor(1.5746, device='cuda:0', grad_fn=<NllLossBackward>)
[226] Train Loss:  tensor(1.5459, device='cuda:0', grad_fn=<NllLossBackward>)
[227] Train Loss:  tensor(1.4938, device='cuda:0', grad_fn=<NllLossBackward>)
[228] Train Loss:  tensor(1.5538, device='cuda:0', grad_fn=<NllLossBackward>)
[229] Train Loss:  tensor(1.5156, device='cuda:0', grad_fn=<NllLossBackward>)
[230] Train Loss:  tensor(1.5119, device='cuda:0', grad_fn=<NllLossBackward>)
[231] Train Loss:  tensor(1.5601, device='cuda:0', grad_fn=<NllLossBackward>)
[232] Train Loss:  tensor(1.5290, device='cuda:0', grad_fn=<NllLossBackward>)
[233] Train Loss:  tensor(1.5175, device='cuda:0', grad_fn=<NllLossBackward>)
[234] Train Loss:  tensor(1.5130, device='cuda:0', grad_fn=<NllLossBackward>)
[235] Train Loss:  tensor(1.5675, device='cuda:0', grad_fn=<NllLossBackward>)
[236] Train Loss:  tensor(1.5061, device='cuda:0', grad_fn=<NllLossBackward>)
[237] Train Loss:  tensor(1.5417, device='cuda:0', grad_fn=<NllLossBackward>)
[238] Train Loss:  tensor(1.5434, device='cuda:0', grad_fn=<NllLossBackward>)
[239] Train Loss:  tensor(1.5316, device='cuda:0', grad_fn=<NllLossBackward>)
[240] Train Loss:  tensor(1.5583, device='cuda:0', grad_fn=<NllLossBackward>)
[241] Train Loss:  tensor(1.4800, device='cuda:0', grad_fn=<NllLossBackward>)
[242] Train Loss:  tensor(1.5713, device='cuda:0', grad_fn=<NllLossBackward>)
[243] Train Loss:  tensor(1.5142, device='cuda:0', grad_fn=<NllLossBackward>)
[244] Train Loss:  tensor(1.5200, device='cuda:0', grad_fn=<NllLossBackward>)
[245] Train Loss:  tensor(1.5134, device='cuda:0', grad_fn=<NllLossBackward>)
[246] Train Loss:  tensor(1.5324, device='cuda:0', grad_fn=<NllLossBackward>)
[247] Train Loss:  tensor(1.5033, device='cuda:0', grad_fn=<NllLossBackward>)
[248] Train Loss:  tensor(1.4413, device='cuda:0', grad_fn=<NllLossBackward>)
[249] Train Loss:  tensor(1.5297, device='cuda:0', grad_fn=<NllLossBackward>)
[250] Train Loss:  tensor(1.5465, device='cuda:0', grad_fn=<NllLossBackward>)
[251] Train Loss:  tensor(1.4536, device='cuda:0', grad_fn=<NllLossBackward>)
[252] Train Loss:  tensor(1.5013, device='cuda:0', grad_fn=<NllLossBackward>)
[253] Train Loss:  tensor(1.5065, device='cuda:0', grad_fn=<NllLossBackward>)
[254] Train Loss:  tensor(1.4707, device='cuda:0', grad_fn=<NllLossBackward>)
[255] Train Loss:  tensor(1.5017, device='cuda:0', grad_fn=<NllLossBackward>)
[256] Train Loss:  tensor(1.4477, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[257] Train Loss:   tensor(1.5547, device='cuda:0', grad_fn=<NllLossBackward>)
[258] Train Loss:   tensor(1.4890, device='cuda:0', grad_fn=<NllLossBackward>)
[259] Train Loss:   tensor(1.4740, device='cuda:0', grad_fn=<NllLossBackward>)
[260] Train Loss:   tensor(1.4885, device='cuda:0', grad_fn=<NllLossBackward>)
[261] Train Loss:   tensor(1.4991, device='cuda:0', grad_fn=<NllLossBackward>)
[262] Train Loss:   tensor(1.5135, device='cuda:0', grad_fn=<NllLossBackward>)
[263] Train Loss:   tensor(1.5002, device='cuda:0', grad_fn=<NllLossBackward>)
[264] Train Loss:   tensor(1.5071, device='cuda:0', grad_fn=<NllLossBackward>)
[265] Train Loss:   tensor(1.4695, device='cuda:0', grad_fn=<NllLossBackward>)
[266] Train Loss:   tensor(1.4958, device='cuda:0', grad_fn=<NllLossBackward>)
[267] Train Loss:   tensor(1.4867, device='cuda:0', grad_fn=<NllLossBackward>)
[268] Train Loss:   tensor(1.4760, device='cuda:0', grad_fn=<NllLossBackward>)
[269] Train Loss:   tensor(1.4649, device='cuda:0', grad_fn=<NllLossBackward>)
[270] Train Loss:   tensor(1.4871, device='cuda:0', grad_fn=<NllLossBackward>)
[271] Train Loss:   tensor(1.5170, device='cuda:0', grad_fn=<NllLossBackward>)
[272] Train Loss:   tensor(1.4618, device='cuda:0', grad_fn=<NllLossBackward>)
[273] Train Loss:   tensor(1.4944, device='cuda:0', grad_fn=<NllLossBackward>)
[274] Train Loss:   tensor(1.4783, device='cuda:0', grad_fn=<NllLossBackward>)
[275] Train Loss:   tensor(1.5016, device='cuda:0', grad_fn=<NllLossBackward>)
[276] Train Loss:   tensor(1.4549, device='cuda:0', grad_fn=<NllLossBackward>)
[277] Train Loss:   tensor(1.4922, device='cuda:0', grad_fn=<NllLossBackward>)
[278] Train Loss:   tensor(1.4316, device='cuda:0', grad_fn=<NllLossBackward>)
[279] Train Loss:   tensor(1.4733, device='cuda:0', grad_fn=<NllLossBackward>)
[280] Train Loss:   tensor(1.4727, device='cuda:0', grad_fn=<NllLossBackward>)
[281] Train Loss:   tensor(1.4994, device='cuda:0', grad_fn=<NllLossBackward>)
[282] Train Loss:   tensor(1.4952, device='cuda:0', grad_fn=<NllLossBackward>)
[283] Train Loss:   tensor(1.5071, device='cuda:0', grad_fn=<NllLossBackward>)
[284] Train Loss:   tensor(1.4968, device='cuda:0', grad_fn=<NllLossBackward>)
[285] Train Loss:   tensor(1.4784, device='cuda:0', grad_fn=<NllLossBackward>)
[286] Train Loss:   tensor(1.4916, device='cuda:0', grad_fn=<NllLossBackward>)
[287] Train Loss:   tensor(1.4686, device='cuda:0', grad_fn=<NllLossBackward>)
[288] Train Loss:   tensor(1.4855, device='cuda:0', grad_fn=<NllLossBackward>)
[289] Train Loss:   tensor(1.4916, device='cuda:0', grad_fn=<NllLossBackward>)
[290] Train Loss:   tensor(1.4914, device='cuda:0', grad_fn=<NllLossBackward>)
[291] Train Loss:   tensor(1.4823, device='cuda:0', grad_fn=<NllLossBackward>)
[292] Train Loss:   tensor(1.4743, device='cuda:0', grad_fn=<NllLossBackward>)
[293] Train Loss:   tensor(1.4430, device='cuda:0', grad_fn=<NllLossBackward>)
[294] Train Loss:   tensor(1.4724, device='cuda:0', grad_fn=<NllLossBackward>)
[295] Train Loss:   tensor(1.4634, device='cuda:0', grad_fn=<NllLossBackward>)
[296] Train Loss:   tensor(1.4408, device='cuda:0', grad_fn=<NllLossBackward>)
[297] Train Loss:   tensor(1.4937, device='cuda:0', grad_fn=<NllLossBackward>)
[298] Train Loss:   tensor(1.4472, device='cuda:0', grad_fn=<NllLossBackward>)
[299] Train Loss:   tensor(1.4329, device='cuda:0', grad_fn=<NllLossBackward>)
[300] Train Loss:   tensor(1.4903, device='cuda:0', grad_fn=<NllLossBackward>)
[301] Train Loss:   tensor(1.4844, device='cuda:0', grad_fn=<NllLossBackward>)
[302] Train Loss:   tensor(1.4676, device='cuda:0', grad_fn=<NllLossBackward>)
[303] Train Loss:   tensor(1.5325, device='cuda:0', grad_fn=<NllLossBackward>)
[304] Train Loss:   tensor(1.4583, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[305] Train Loss:    tensor(1.4568, device='cuda:0', grad_fn=<NllLossBackward>)
[306] Train Loss:    tensor(1.4377, device='cuda:0', grad_fn=<NllLossBackward>)
[307] Train Loss:    tensor(1.4135, device='cuda:0', grad_fn=<NllLossBackward>)
[308] Train Loss:    tensor(1.4394, device='cuda:0', grad_fn=<NllLossBackward>)
[309] Train Loss:    tensor(1.4731, device='cuda:0', grad_fn=<NllLossBackward>)
[310] Train Loss:    tensor(1.4367, device='cuda:0', grad_fn=<NllLossBackward>)
[311] Train Loss:    tensor(1.4326, device='cuda:0', grad_fn=<NllLossBackward>)
[312] Train Loss:    tensor(1.4462, device='cuda:0', grad_fn=<NllLossBackward>)
[313] Train Loss:    tensor(1.4552, device='cuda:0', grad_fn=<NllLossBackward>)
[314] Train Loss:    tensor(1.4436, device='cuda:0', grad_fn=<NllLossBackward>)
[315] Train Loss:    tensor(1.4204, device='cuda:0', grad_fn=<NllLossBackward>)
[316] Train Loss:    tensor(1.4228, device='cuda:0', grad_fn=<NllLossBackward>)
[317] Train Loss:    tensor(1.4572, device='cuda:0', grad_fn=<NllLossBackward>)
[318] Train Loss:    tensor(1.4406, device='cuda:0', grad_fn=<NllLossBackward>)
[319] Train Loss:    tensor(1.4836, device='cuda:0', grad_fn=<NllLossBackward>)
[320] Train Loss:    tensor(1.4038, device='cuda:0', grad_fn=<NllLossBackward>)
[321] Train Loss:    tensor(1.3925, device='cuda:0', grad_fn=<NllLossBackward>)
[322] Train Loss:    tensor(1.4936, device='cuda:0', grad_fn=<NllLossBackward>)
[323] Train Loss:    tensor(1.4708, device='cuda:0', grad_fn=<NllLossBackward>)
[324] Train Loss:    tensor(1.4279, device='cuda:0', grad_fn=<NllLossBackward>)
[325] Train Loss:    tensor(1.4367, device='cuda:0', grad_fn=<NllLossBackward>)
[326] Train Loss:    tensor(1.4701, device='cuda:0', grad_fn=<NllLossBackward>)
[327] Train Loss:    tensor(1.4088, device='cuda:0', grad_fn=<NllLossBackward>)
[328] Train Loss:    tensor(1.4452, device='cuda:0', grad_fn=<NllLossBackward>)
[329] Train Loss:    tensor(1.4293, device='cuda:0', grad_fn=<NllLossBackward>)
[330] Train Loss:    tensor(1.4346, device='cuda:0', grad_fn=<NllLossBackward>)
[331] Train Loss:    tensor(1.4340, device='cuda:0', grad_fn=<NllLossBackward>)
[332] Train Loss:    tensor(1.4320, device='cuda:0', grad_fn=<NllLossBackward>)
[333] Train Loss:    tensor(1.4435, device='cuda:0', grad_fn=<NllLossBackward>)
[334] Train Loss:    tensor(1.4135, device='cuda:0', grad_fn=<NllLossBackward>)
[335] Train Loss:    tensor(1.3855, device='cuda:0', grad_fn=<NllLossBackward>)
[336] Train Loss:    tensor(1.4292, device='cuda:0', grad_fn=<NllLossBackward>)
[337] Train Loss:    tensor(1.4029, device='cuda:0', grad_fn=<NllLossBackward>)
[338] Train Loss:    tensor(1.4035, device='cuda:0', grad_fn=<NllLossBackward>)
[339] Train Loss:    tensor(1.4333, device='cuda:0', grad_fn=<NllLossBackward>)
[340] Train Loss:    tensor(1.4503, device='cuda:0', grad_fn=<NllLossBackward>)
[341] Train Loss:    tensor(1.4735, device='cuda:0', grad_fn=<NllLossBackward>)
[342] Train Loss:    tensor(1.4370, device='cuda:0', grad_fn=<NllLossBackward>)
[343] Train Loss:    tensor(1.4255, device='cuda:0', grad_fn=<NllLossBackward>)
[344] Train Loss:    tensor(1.4127, device='cuda:0', grad_fn=<NllLossBackward>)
[345] Train Loss:    tensor(1.4046, device='cuda:0', grad_fn=<NllLossBackward>)
[346] Train Loss:    tensor(1.3795, device='cuda:0', grad_fn=<NllLossBackward>)
[347] Train Loss:    tensor(1.3920, device='cuda:0', grad_fn=<NllLossBackward>)
[348] Train Loss:    tensor(1.4607, device='cuda:0', grad_fn=<NllLossBackward>)
[349] Train Loss:    tensor(1.4012, device='cuda:0', grad_fn=<NllLossBackward>)
[350] Train Loss:    tensor(1.4060, device='cuda:0', grad_fn=<NllLossBackward>)
[351] Train Loss:    tensor(1.4135, device='cuda:0', grad_fn=<NllLossBackward>)
[352] Train Loss:    tensor(1.4070, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[353] Train Loss:    tensor(1.4227, device='cuda:0', grad_fn=<NllLossBackward>)
[354] Train Loss:    tensor(1.4189, device='cuda:0', grad_fn=<NllLossBackward>)
[355] Train Loss:    tensor(1.4792, device='cuda:0', grad_fn=<NllLossBackward>)
[356] Train Loss:    tensor(1.3974, device='cuda:0', grad_fn=<NllLossBackward>)
[357] Train Loss:    tensor(1.3674, device='cuda:0', grad_fn=<NllLossBackward>)
[358] Train Loss:    tensor(1.4059, device='cuda:0', grad_fn=<NllLossBackward>)
[359] Train Loss:    tensor(1.4093, device='cuda:0', grad_fn=<NllLossBackward>)
[360] Train Loss:    tensor(1.4245, device='cuda:0', grad_fn=<NllLossBackward>)
[361] Train Loss:    tensor(1.4260, device='cuda:0', grad_fn=<NllLossBackward>)
[362] Train Loss:    tensor(1.4325, device='cuda:0', grad_fn=<NllLossBackward>)
[363] Train Loss:    tensor(1.4114, device='cuda:0', grad_fn=<NllLossBackward>)
[364] Train Loss:    tensor(1.3925, device='cuda:0', grad_fn=<NllLossBackward>)
[365] Train Loss:    tensor(1.4387, device='cuda:0', grad_fn=<NllLossBackward>)
[366] Train Loss:    tensor(1.3713, device='cuda:0', grad_fn=<NllLossBackward>)
[367] Train Loss:    tensor(1.3774, device='cuda:0', grad_fn=<NllLossBackward>)
[368] Train Loss:    tensor(1.4079, device='cuda:0', grad_fn=<NllLossBackward>)
[369] Train Loss:    tensor(1.3962, device='cuda:0', grad_fn=<NllLossBackward>)
[370] Train Loss:    tensor(1.3638, device='cuda:0', grad_fn=<NllLossBackward>)
[371] Train Loss:    tensor(1.4197, device='cuda:0', grad_fn=<NllLossBackward>)
[372] Train Loss:    tensor(1.4074, device='cuda:0', grad_fn=<NllLossBackward>)
[373] Train Loss:    tensor(1.3938, device='cuda:0', grad_fn=<NllLossBackward>)
[374] Train Loss:    tensor(1.4134, device='cuda:0', grad_fn=<NllLossBackward>)
[375] Train Loss:    tensor(1.3799, device='cuda:0', grad_fn=<NllLossBackward>)
[376] Train Loss:    tensor(1.3979, device='cuda:0', grad_fn=<NllLossBackward>)
[377] Train Loss:    tensor(1.4032, device='cuda:0', grad_fn=<NllLossBackward>)
[378] Train Loss:    tensor(1.3758, device='cuda:0', grad_fn=<NllLossBackward>)
[379] Train Loss:    tensor(1.4030, device='cuda:0', grad_fn=<NllLossBackward>)
[380] Train Loss:    tensor(1.4077, device='cuda:0', grad_fn=<NllLossBackward>)
[381] Train Loss:    tensor(1.4074, device='cuda:0', grad_fn=<NllLossBackward>)
[382] Train Loss:    tensor(1.3886, device='cuda:0', grad_fn=<NllLossBackward>)
[383] Train Loss:    tensor(1.3543, device='cuda:0', grad_fn=<NllLossBackward>)
[384] Train Loss:    tensor(1.4129, device='cuda:0', grad_fn=<NllLossBackward>)
[385] Train Loss:    tensor(1.3853, device='cuda:0', grad_fn=<NllLossBackward>)
[386] Train Loss:    tensor(1.3577, device='cuda:0', grad_fn=<NllLossBackward>)
[387] Train Loss:    tensor(1.4261, device='cuda:0', grad_fn=<NllLossBackward>)
[388] Train Loss:    tensor(1.4141, device='cuda:0', grad_fn=<NllLossBackward>)
[389] Train Loss:    tensor(1.3797, device='cuda:0', grad_fn=<NllLossBackward>)
[390] Train Loss:    tensor(1.4015, device='cuda:0', grad_fn=<NllLossBackward>)
[391] Train Loss:    tensor(1.4006, device='cuda:0', grad_fn=<NllLossBackward>)
[392] Train Loss:    tensor(1.3751, device='cuda:0', grad_fn=<NllLossBackward>)
[393] Train Loss:    tensor(1.4106, device='cuda:0', grad_fn=<NllLossBackward>)
[394] Train Loss:    tensor(1.3841, device='cuda:0', grad_fn=<NllLossBackward>)
[395] Train Loss:    tensor(1.3830, device='cuda:0', grad_fn=<NllLossBackward>)
[396] Train Loss:    tensor(1.4183, device='cuda:0', grad_fn=<NllLossBackward>)
[397] Train Loss:    tensor(1.3923, device='cuda:0', grad_fn=<NllLossBackward>)
[398] Train Loss:    tensor(1.4091, device='cuda:0', grad_fn=<NllLossBackward>)
[399] Train Loss:    tensor(1.3982, device='cuda:0', grad_fn=<NllLossBackward>)
[400] Train Loss:    tensor(1.3496, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[401] Train Loss:    tensor(1.3623, device='cuda:0', grad_fn=<NllLossBackward>)
[402] Train Loss:    tensor(1.3863, device='cuda:0', grad_fn=<NllLossBackward>)
[403] Train Loss:    tensor(1.3518, device='cuda:0', grad_fn=<NllLossBackward>)
[404] Train Loss:    tensor(1.3833, device='cuda:0', grad_fn=<NllLossBackward>)
[405] Train Loss:    tensor(1.4078, device='cuda:0', grad_fn=<NllLossBackward>)
[406] Train Loss:    tensor(1.3918, device='cuda:0', grad_fn=<NllLossBackward>)
[407] Train Loss:    tensor(1.4080, device='cuda:0', grad_fn=<NllLossBackward>)
[408] Train Loss:    tensor(1.3505, device='cuda:0', grad_fn=<NllLossBackward>)
[409] Train Loss:    tensor(1.3457, device='cuda:0', grad_fn=<NllLossBackward>)
[410] Train Loss:    tensor(1.3559, device='cuda:0', grad_fn=<NllLossBackward>)
[411] Train Loss:    tensor(1.3637, device='cuda:0', grad_fn=<NllLossBackward>)
[412] Train Loss:    tensor(1.4142, device='cuda:0', grad_fn=<NllLossBackward>)
[413] Train Loss:    tensor(1.3943, device='cuda:0', grad_fn=<NllLossBackward>)
[414] Train Loss:    tensor(1.4183, device='cuda:0', grad_fn=<NllLossBackward>)
[415] Train Loss:    tensor(1.3954, device='cuda:0', grad_fn=<NllLossBackward>)
[416] Train Loss:    tensor(1.4293, device='cuda:0', grad_fn=<NllLossBackward>)
[417] Train Loss:    tensor(1.3734, device='cuda:0', grad_fn=<NllLossBackward>)
[418] Train Loss:    tensor(1.3538, device='cuda:0', grad_fn=<NllLossBackward>)
[419] Train Loss:    tensor(1.3777, device='cuda:0', grad_fn=<NllLossBackward>)
[420] Train Loss:    tensor(1.3999, device='cuda:0', grad_fn=<NllLossBackward>)
[421] Train Loss:    tensor(1.4052, device='cuda:0', grad_fn=<NllLossBackward>)
[422] Train Loss:    tensor(1.3465, device='cuda:0', grad_fn=<NllLossBackward>)
[423] Train Loss:    tensor(1.3672, device='cuda:0', grad_fn=<NllLossBackward>)
[424] Train Loss:    tensor(1.3462, device='cuda:0', grad_fn=<NllLossBackward>)
[425] Train Loss:    tensor(1.3447, device='cuda:0', grad_fn=<NllLossBackward>)
[426] Train Loss:    tensor(1.3822, device='cuda:0', grad_fn=<NllLossBackward>)
[427] Train Loss:    tensor(1.4006, device='cuda:0', grad_fn=<NllLossBackward>)
[428] Train Loss:    tensor(1.3384, device='cuda:0', grad_fn=<NllLossBackward>)
[429] Train Loss:    tensor(1.3596, device='cuda:0', grad_fn=<NllLossBackward>)
[430] Train Loss:    tensor(1.3774, device='cuda:0', grad_fn=<NllLossBackward>)
[431] Train Loss:    tensor(1.3667, device='cuda:0', grad_fn=<NllLossBackward>)
[432] Train Loss:    tensor(1.3604, device='cuda:0', grad_fn=<NllLossBackward>)
[433] Train Loss:    tensor(1.3827, device='cuda:0', grad_fn=<NllLossBackward>)
[434] Train Loss:    tensor(1.3851, device='cuda:0', grad_fn=<NllLossBackward>)
[435] Train Loss:    tensor(1.3560, device='cuda:0', grad_fn=<NllLossBackward>)
[436] Train Loss:    tensor(1.3802, device='cuda:0', grad_fn=<NllLossBackward>)
[437] Train Loss:    tensor(1.3524, device='cuda:0', grad_fn=<NllLossBackward>)
[438] Train Loss:    tensor(1.3894, device='cuda:0', grad_fn=<NllLossBackward>)
[439] Train Loss:    tensor(1.3830, device='cuda:0', grad_fn=<NllLossBackward>)
[440] Train Loss:    tensor(1.3596, device='cuda:0', grad_fn=<NllLossBackward>)
[441] Train Loss:    tensor(1.3728, device='cuda:0', grad_fn=<NllLossBackward>)
[442] Train Loss:    tensor(1.3579, device='cuda:0', grad_fn=<NllLossBackward>)
[443] Train Loss:    tensor(1.4200, device='cuda:0', grad_fn=<NllLossBackward>)
[444] Train Loss:    tensor(1.3768, device='cuda:0', grad_fn=<NllLossBackward>)
[445] Train Loss:    tensor(1.3824, device='cuda:0', grad_fn=<NllLossBackward>)
[446] Train Loss:    tensor(1.3434, device='cuda:0', grad_fn=<NllLossBackward>)
[447] Train Loss:    tensor(1.3642, device='cuda:0', grad_fn=<NllLossBackward>)
[448] Train Loss:    tensor(1.3426, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[449] Train Loss:   tensor(1.3203, device='cuda:0', grad_fn=<NllLossBackward>)
[450] Train Loss:   tensor(1.3477, device='cuda:0', grad_fn=<NllLossBackward>)
[451] Train Loss:   tensor(1.3313, device='cuda:0', grad_fn=<NllLossBackward>)
[452] Train Loss:   tensor(1.3395, device='cuda:0', grad_fn=<NllLossBackward>)
[453] Train Loss:   tensor(1.3960, device='cuda:0', grad_fn=<NllLossBackward>)
[454] Train Loss:   tensor(1.3657, device='cuda:0', grad_fn=<NllLossBackward>)
[455] Train Loss:   tensor(1.3490, device='cuda:0', grad_fn=<NllLossBackward>)
[456] Train Loss:   tensor(1.3406, device='cuda:0', grad_fn=<NllLossBackward>)
[457] Train Loss:   tensor(1.3523, device='cuda:0', grad_fn=<NllLossBackward>)
[458] Train Loss:   tensor(1.3316, device='cuda:0', grad_fn=<NllLossBackward>)
[459] Train Loss:   tensor(1.3802, device='cuda:0', grad_fn=<NllLossBackward>)
[460] Train Loss:   tensor(1.3507, device='cuda:0', grad_fn=<NllLossBackward>)
[461] Train Loss:   tensor(1.4025, device='cuda:0', grad_fn=<NllLossBackward>)
[462] Train Loss:   tensor(1.3636, device='cuda:0', grad_fn=<NllLossBackward>)
[463] Train Loss:   tensor(1.3388, device='cuda:0', grad_fn=<NllLossBackward>)
[464] Train Loss:   tensor(1.3197, device='cuda:0', grad_fn=<NllLossBackward>)
[465] Train Loss:   tensor(1.3363, device='cuda:0', grad_fn=<NllLossBackward>)
[466] Train Loss:   tensor(1.3692, device='cuda:0', grad_fn=<NllLossBackward>)
[467] Train Loss:   tensor(1.3875, device='cuda:0', grad_fn=<NllLossBackward>)
[468] Train Loss:   tensor(1.4145, device='cuda:0', grad_fn=<NllLossBackward>)
[469] Train Loss:   tensor(1.3324, device='cuda:0', grad_fn=<NllLossBackward>)
[470] Train Loss:   tensor(1.3716, device='cuda:0', grad_fn=<NllLossBackward>)
[471] Train Loss:   tensor(1.3469, device='cuda:0', grad_fn=<NllLossBackward>)
[472] Train Loss:   tensor(1.3727, device='cuda:0', grad_fn=<NllLossBackward>)
[473] Train Loss:   tensor(1.3753, device='cuda:0', grad_fn=<NllLossBackward>)
[474] Train Loss:   tensor(1.3076, device='cuda:0', grad_fn=<NllLossBackward>)
[475] Train Loss:   tensor(1.3319, device='cuda:0', grad_fn=<NllLossBackward>)
[476] Train Loss:   tensor(1.3688, device='cuda:0', grad_fn=<NllLossBackward>)
[477] Train Loss:   tensor(1.3062, device='cuda:0', grad_fn=<NllLossBackward>)
[478] Train Loss:   tensor(1.3281, device='cuda:0', grad_fn=<NllLossBackward>)
[479] Train Loss:   tensor(1.3799, device='cuda:0', grad_fn=<NllLossBackward>)
[480] Train Loss:   tensor(1.3389, device='cuda:0', grad_fn=<NllLossBackward>)
[481] Train Loss:   tensor(1.3129, device='cuda:0', grad_fn=<NllLossBackward>)
[482] Train Loss:   tensor(1.3803, device='cuda:0', grad_fn=<NllLossBackward>)
[483] Train Loss:   tensor(1.3145, device='cuda:0', grad_fn=<NllLossBackward>)
[484] Train Loss:   tensor(1.3675, device='cuda:0', grad_fn=<NllLossBackward>)
[485] Train Loss:   tensor(1.3293, device='cuda:0', grad_fn=<NllLossBackward>)
[486] Train Loss:   tensor(1.3448, device='cuda:0', grad_fn=<NllLossBackward>)
[487] Train Loss:   tensor(1.3201, device='cuda:0', grad_fn=<NllLossBackward>)
[488] Train Loss:   tensor(1.2727, device='cuda:0', grad_fn=<NllLossBackward>)
[489] Train Loss:   tensor(1.3395, device='cuda:0', grad_fn=<NllLossBackward>)
[490] Train Loss:   tensor(1.3425, device='cuda:0', grad_fn=<NllLossBackward>)
[491] Train Loss:   tensor(1.2908, device='cuda:0', grad_fn=<NllLossBackward>)
[492] Train Loss:   tensor(1.2934, device='cuda:0', grad_fn=<NllLossBackward>)
[493] Train Loss:   tensor(1.3466, device='cuda:0', grad_fn=<NllLossBackward>)
[494] Train Loss:   tensor(1.3359, device='cuda:0', grad_fn=<NllLossBackward>)
[495] Train Loss:   tensor(1.3290, device='cuda:0', grad_fn=<NllLossBackward>)
[496] Train Loss:   tensor(1.3011, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[497] Train Loss:   tensor(1.3488, device='cuda:0', grad_fn=<NllLossBackward>)
[498] Train Loss:   tensor(1.3407, device='cuda:0', grad_fn=<NllLossBackward>)
[499] Train Loss:   tensor(1.3621, device='cuda:0', grad_fn=<NllLossBackward>)
[500] Train Loss:   tensor(1.3336, device='cuda:0', grad_fn=<NllLossBackward>)
[501] Train Loss:   tensor(1.2941, device='cuda:0', grad_fn=<NllLossBackward>)
[502] Train Loss:   tensor(1.3442, device='cuda:0', grad_fn=<NllLossBackward>)
[503] Train Loss:   tensor(1.3412, device='cuda:0', grad_fn=<NllLossBackward>)
[504] Train Loss:   tensor(1.3160, device='cuda:0', grad_fn=<NllLossBackward>)
[505] Train Loss:   tensor(1.2864, device='cuda:0', grad_fn=<NllLossBackward>)
[506] Train Loss:   tensor(1.3375, device='cuda:0', grad_fn=<NllLossBackward>)
[507] Train Loss:   tensor(1.3349, device='cuda:0', grad_fn=<NllLossBackward>)
[508] Train Loss:   tensor(1.3066, device='cuda:0', grad_fn=<NllLossBackward>)
[509] Train Loss:   tensor(1.2927, device='cuda:0', grad_fn=<NllLossBackward>)
[510] Train Loss:   tensor(1.3515, device='cuda:0', grad_fn=<NllLossBackward>)
[511] Train Loss:   tensor(1.3240, device='cuda:0', grad_fn=<NllLossBackward>)
[512] Train Loss:   tensor(1.3111, device='cuda:0', grad_fn=<NllLossBackward>)
[513] Train Loss:   tensor(1.3251, device='cuda:0', grad_fn=<NllLossBackward>)
[514] Train Loss:   tensor(1.3516, device='cuda:0', grad_fn=<NllLossBackward>)
[515] Train Loss:   tensor(1.3603, device='cuda:0', grad_fn=<NllLossBackward>)
[516] Train Loss:   tensor(1.3379, device='cuda:0', grad_fn=<NllLossBackward>)
[517] Train Loss:   tensor(1.3015, device='cuda:0', grad_fn=<NllLossBackward>)
[518] Train Loss:   tensor(1.2772, device='cuda:0', grad_fn=<NllLossBackward>)
[519] Train Loss:   tensor(1.3217, device='cuda:0', grad_fn=<NllLossBackward>)
[520] Train Loss:   tensor(1.3003, device='cuda:0', grad_fn=<NllLossBackward>)
[521] Train Loss:   tensor(1.3400, device='cuda:0', grad_fn=<NllLossBackward>)
[522] Train Loss:   tensor(1.3118, device='cuda:0', grad_fn=<NllLossBackward>)
[523] Train Loss:   tensor(1.3377, device='cuda:0', grad_fn=<NllLossBackward>)
[524] Train Loss:   tensor(1.3308, device='cuda:0', grad_fn=<NllLossBackward>)
[525] Train Loss:   tensor(1.3105, device='cuda:0', grad_fn=<NllLossBackward>)
[526] Train Loss:   tensor(1.3141, device='cuda:0', grad_fn=<NllLossBackward>)
[527] Train Loss:   tensor(1.2817, device='cuda:0', grad_fn=<NllLossBackward>)
[528] Train Loss:   tensor(1.2814, device='cuda:0', grad_fn=<NllLossBackward>)
[529] Train Loss:   tensor(1.3301, device='cuda:0', grad_fn=<NllLossBackward>)
[530] Train Loss:   tensor(1.3410, device='cuda:0', grad_fn=<NllLossBackward>)
[531] Train Loss:   tensor(1.3063, device='cuda:0', grad_fn=<NllLossBackward>)
[532] Train Loss:   tensor(1.3207, device='cuda:0', grad_fn=<NllLossBackward>)
[533] Train Loss:   tensor(1.2920, device='cuda:0', grad_fn=<NllLossBackward>)
[534] Train Loss:   tensor(1.3099, device='cuda:0', grad_fn=<NllLossBackward>)
[535] Train Loss:   tensor(1.3249, device='cuda:0', grad_fn=<NllLossBackward>)
[536] Train Loss:   tensor(1.3252, device='cuda:0', grad_fn=<NllLossBackward>)
[537] Train Loss:   tensor(1.2832, device='cuda:0', grad_fn=<NllLossBackward>)
[538] Train Loss:   tensor(1.3449, device='cuda:0', grad_fn=<NllLossBackward>)
[539] Train Loss:   tensor(1.2811, device='cuda:0', grad_fn=<NllLossBackward>)
[540] Train Loss:   tensor(1.3423, device='cuda:0', grad_fn=<NllLossBackward>)
[541] Train Loss:   tensor(1.3231, device='cuda:0', grad_fn=<NllLossBackward>)
[542] Train Loss:   tensor(1.3467, device='cuda:0', grad_fn=<NllLossBackward>)
[543] Train Loss:   tensor(1.3151, device='cuda:0', grad_fn=<NllLossBackward>)
[544] Train Loss:   tensor(1.3225, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[545] Train Loss:   tensor(1.2780, device='cuda:0', grad_fn=<NllLossBackward>)
[546] Train Loss:   tensor(1.2882, device='cuda:0', grad_fn=<NllLossBackward>)
[547] Train Loss:   tensor(1.3535, device='cuda:0', grad_fn=<NllLossBackward>)
[548] Train Loss:   tensor(1.2876, device='cuda:0', grad_fn=<NllLossBackward>)
[549] Train Loss:   tensor(1.3083, device='cuda:0', grad_fn=<NllLossBackward>)
[550] Train Loss:   tensor(1.3287, device='cuda:0', grad_fn=<NllLossBackward>)
[551] Train Loss:   tensor(1.3081, device='cuda:0', grad_fn=<NllLossBackward>)
[552] Train Loss:   tensor(1.2964, device='cuda:0', grad_fn=<NllLossBackward>)
[553] Train Loss:   tensor(1.3402, device='cuda:0', grad_fn=<NllLossBackward>)
[554] Train Loss:   tensor(1.3149, device='cuda:0', grad_fn=<NllLossBackward>)
[555] Train Loss:   tensor(1.3523, device='cuda:0', grad_fn=<NllLossBackward>)
[556] Train Loss:   tensor(1.2793, device='cuda:0', grad_fn=<NllLossBackward>)
[557] Train Loss:   tensor(1.3165, device='cuda:0', grad_fn=<NllLossBackward>)
[558] Train Loss:   tensor(1.2983, device='cuda:0', grad_fn=<NllLossBackward>)
[559] Train Loss:   tensor(1.3275, device='cuda:0', grad_fn=<NllLossBackward>)
[560] Train Loss:   tensor(1.3086, device='cuda:0', grad_fn=<NllLossBackward>)
[561] Train Loss:   tensor(1.2898, device='cuda:0', grad_fn=<NllLossBackward>)
[562] Train Loss:   tensor(1.2987, device='cuda:0', grad_fn=<NllLossBackward>)
[563] Train Loss:   tensor(1.2853, device='cuda:0', grad_fn=<NllLossBackward>)
[564] Train Loss:   tensor(1.3223, device='cuda:0', grad_fn=<NllLossBackward>)
[565] Train Loss:   tensor(1.3213, device='cuda:0', grad_fn=<NllLossBackward>)
[566] Train Loss:   tensor(1.2977, device='cuda:0', grad_fn=<NllLossBackward>)
[567] Train Loss:   tensor(1.3195, device='cuda:0', grad_fn=<NllLossBackward>)
[568] Train Loss:   tensor(1.2906, device='cuda:0', grad_fn=<NllLossBackward>)
[569] Train Loss:   tensor(1.3251, device='cuda:0', grad_fn=<NllLossBackward>)
[570] Train Loss:   tensor(1.2788, device='cuda:0', grad_fn=<NllLossBackward>)
[571] Train Loss:   tensor(1.3076, device='cuda:0', grad_fn=<NllLossBackward>)
[572] Train Loss:   tensor(1.3050, device='cuda:0', grad_fn=<NllLossBackward>)
[573] Train Loss:   tensor(1.2986, device='cuda:0', grad_fn=<NllLossBackward>)
[574] Train Loss:   tensor(1.3150, device='cuda:0', grad_fn=<NllLossBackward>)
[575] Train Loss:   tensor(1.3136, device='cuda:0', grad_fn=<NllLossBackward>)
[576] Train Loss:   tensor(1.2871, device='cuda:0', grad_fn=<NllLossBackward>)
[577] Train Loss:   tensor(1.3172, device='cuda:0', grad_fn=<NllLossBackward>)
[578] Train Loss:   tensor(1.2561, device='cuda:0', grad_fn=<NllLossBackward>)
[579] Train Loss:   tensor(1.3255, device='cuda:0', grad_fn=<NllLossBackward>)
[580] Train Loss:   tensor(1.3189, device='cuda:0', grad_fn=<NllLossBackward>)
[581] Train Loss:   tensor(1.2796, device='cuda:0', grad_fn=<NllLossBackward>)
[582] Train Loss:   tensor(1.2771, device='cuda:0', grad_fn=<NllLossBackward>)
[583] Train Loss:   tensor(1.3146, device='cuda:0', grad_fn=<NllLossBackward>)
[584] Train Loss:   tensor(1.2843, device='cuda:0', grad_fn=<NllLossBackward>)
[585] Train Loss:   tensor(1.3071, device='cuda:0', grad_fn=<NllLossBackward>)
[586] Train Loss:   tensor(1.3262, device='cuda:0', grad_fn=<NllLossBackward>)
[587] Train Loss:   tensor(1.2793, device='cuda:0', grad_fn=<NllLossBackward>)
[588] Train Loss:   tensor(1.3168, device='cuda:0', grad_fn=<NllLossBackward>)
[589] Train Loss:   tensor(1.2662, device='cuda:0', grad_fn=<NllLossBackward>)
[590] Train Loss:   tensor(1.2713, device='cuda:0', grad_fn=<NllLossBackward>)
[591] Train Loss:   tensor(1.2936, device='cuda:0', grad_fn=<NllLossBackward>)
[592] Train Loss:   tensor(1.2617, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[593] Train Loss:    tensor(1.2759, device='cuda:0', grad_fn=<NllLossBackward>)
[594] Train Loss:    tensor(1.2568, device='cuda:0', grad_fn=<NllLossBackward>)
[595] Train Loss:    tensor(1.3174, device='cuda:0', grad_fn=<NllLossBackward>)
[596] Train Loss:    tensor(1.3526, device='cuda:0', grad_fn=<NllLossBackward>)
[597] Train Loss:    tensor(1.3083, device='cuda:0', grad_fn=<NllLossBackward>)
[598] Train Loss:    tensor(1.3011, device='cuda:0', grad_fn=<NllLossBackward>)
[599] Train Loss:    tensor(1.2851, device='cuda:0', grad_fn=<NllLossBackward>)
[600] Train Loss:    tensor(1.2819, device='cuda:0', grad_fn=<NllLossBackward>)
[601] Train Loss:    tensor(1.2484, device='cuda:0', grad_fn=<NllLossBackward>)
[602] Train Loss:    tensor(1.3008, device='cuda:0', grad_fn=<NllLossBackward>)
[603] Train Loss:    tensor(1.3062, device='cuda:0', grad_fn=<NllLossBackward>)
[604] Train Loss:    tensor(1.2589, device='cuda:0', grad_fn=<NllLossBackward>)
[605] Train Loss:    tensor(1.2785, device='cuda:0', grad_fn=<NllLossBackward>)
[606] Train Loss:    tensor(1.2739, device='cuda:0', grad_fn=<NllLossBackward>)
[607] Train Loss:    tensor(1.2840, device='cuda:0', grad_fn=<NllLossBackward>)
[608] Train Loss:    tensor(1.2592, device='cuda:0', grad_fn=<NllLossBackward>)
[609] Train Loss:    tensor(1.2440, device='cuda:0', grad_fn=<NllLossBackward>)
[610] Train Loss:    tensor(1.2653, device='cuda:0', grad_fn=<NllLossBackward>)
[611] Train Loss:    tensor(1.2602, device='cuda:0', grad_fn=<NllLossBackward>)
[612] Train Loss:    tensor(1.3501, device='cuda:0', grad_fn=<NllLossBackward>)
[613] Train Loss:    tensor(1.2551, device='cuda:0', grad_fn=<NllLossBackward>)
[614] Train Loss:    tensor(1.2940, device='cuda:0', grad_fn=<NllLossBackward>)
[615] Train Loss:    tensor(1.2864, device='cuda:0', grad_fn=<NllLossBackward>)
[616] Train Loss:    tensor(1.2992, device='cuda:0', grad_fn=<NllLossBackward>)
[617] Train Loss:    tensor(1.3040, device='cuda:0', grad_fn=<NllLossBackward>)
[618] Train Loss:    tensor(1.2826, device='cuda:0', grad_fn=<NllLossBackward>)
[619] Train Loss:    tensor(1.3094, device='cuda:0', grad_fn=<NllLossBackward>)
[620] Train Loss:    tensor(1.2448, device='cuda:0', grad_fn=<NllLossBackward>)
[621] Train Loss:    tensor(1.2946, device='cuda:0', grad_fn=<NllLossBackward>)
[622] Train Loss:    tensor(1.2590, device='cuda:0', grad_fn=<NllLossBackward>)
[623] Train Loss:    tensor(1.2912, device='cuda:0', grad_fn=<NllLossBackward>)
[624] Train Loss:    tensor(1.2703, device='cuda:0', grad_fn=<NllLossBackward>)
[625] Train Loss:    tensor(1.2803, device='cuda:0', grad_fn=<NllLossBackward>)
[626] Train Loss:    tensor(1.2688, device='cuda:0', grad_fn=<NllLossBackward>)
[627] Train Loss:    tensor(1.3099, device='cuda:0', grad_fn=<NllLossBackward>)
[628] Train Loss:    tensor(1.2974, device='cuda:0', grad_fn=<NllLossBackward>)
[629] Train Loss:    tensor(1.2946, device='cuda:0', grad_fn=<NllLossBackward>)
[630] Train Loss:    tensor(1.2775, device='cuda:0', grad_fn=<NllLossBackward>)
[631] Train Loss:    tensor(1.2955, device='cuda:0', grad_fn=<NllLossBackward>)
[632] Train Loss:    tensor(1.2432, device='cuda:0', grad_fn=<NllLossBackward>)
[633] Train Loss:    tensor(1.2937, device='cuda:0', grad_fn=<NllLossBackward>)
[634] Train Loss:    tensor(1.2858, device='cuda:0', grad_fn=<NllLossBackward>)
[635] Train Loss:    tensor(1.2595, device='cuda:0', grad_fn=<NllLossBackward>)
[636] Train Loss:    tensor(1.2624, device='cuda:0', grad_fn=<NllLossBackward>)
[637] Train Loss:    tensor(1.2471, device='cuda:0', grad_fn=<NllLossBackward>)
[638] Train Loss:    tensor(1.2748, device='cuda:0', grad_fn=<NllLossBackward>)
[639] Train Loss:    tensor(1.2500, device='cuda:0', grad_fn=<NllLossBackward>)
[640] Train Loss:    tensor(1.2788, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[641] Train Loss:   tensor(1.2674, device='cuda:0', grad_fn=<NllLossBackward>)
[642] Train Loss:   tensor(1.2792, device='cuda:0', grad_fn=<NllLossBackward>)
[643] Train Loss:   tensor(1.2757, device='cuda:0', grad_fn=<NllLossBackward>)
[644] Train Loss:   tensor(1.2744, device='cuda:0', grad_fn=<NllLossBackward>)
[645] Train Loss:   tensor(1.2417, device='cuda:0', grad_fn=<NllLossBackward>)
[646] Train Loss:   tensor(1.3076, device='cuda:0', grad_fn=<NllLossBackward>)
[647] Train Loss:   tensor(1.3058, device='cuda:0', grad_fn=<NllLossBackward>)
[648] Train Loss:   tensor(1.2823, device='cuda:0', grad_fn=<NllLossBackward>)
[649] Train Loss:   tensor(1.2489, device='cuda:0', grad_fn=<NllLossBackward>)
[650] Train Loss:   tensor(1.2811, device='cuda:0', grad_fn=<NllLossBackward>)
[651] Train Loss:   tensor(1.2494, device='cuda:0', grad_fn=<NllLossBackward>)
[652] Train Loss:   tensor(1.2370, device='cuda:0', grad_fn=<NllLossBackward>)
[653] Train Loss:   tensor(1.2811, device='cuda:0', grad_fn=<NllLossBackward>)
[654] Train Loss:   tensor(1.2358, device='cuda:0', grad_fn=<NllLossBackward>)
[655] Train Loss:   tensor(1.2310, device='cuda:0', grad_fn=<NllLossBackward>)
[656] Train Loss:   tensor(1.2345, device='cuda:0', grad_fn=<NllLossBackward>)
[657] Train Loss:   tensor(1.2818, device='cuda:0', grad_fn=<NllLossBackward>)
[658] Train Loss:   tensor(1.2618, device='cuda:0', grad_fn=<NllLossBackward>)
[659] Train Loss:   tensor(1.2772, device='cuda:0', grad_fn=<NllLossBackward>)
[660] Train Loss:   tensor(1.2626, device='cuda:0', grad_fn=<NllLossBackward>)
[661] Train Loss:   tensor(1.3000, device='cuda:0', grad_fn=<NllLossBackward>)
[662] Train Loss:   tensor(1.2397, device='cuda:0', grad_fn=<NllLossBackward>)
[663] Train Loss:   tensor(1.2174, device='cuda:0', grad_fn=<NllLossBackward>)
[664] Train Loss:   tensor(1.2306, device='cuda:0', grad_fn=<NllLossBackward>)
[665] Train Loss:   tensor(1.2530, device='cuda:0', grad_fn=<NllLossBackward>)
[666] Train Loss:   tensor(1.2603, device='cuda:0', grad_fn=<NllLossBackward>)
[667] Train Loss:   tensor(1.2975, device='cuda:0', grad_fn=<NllLossBackward>)
[668] Train Loss:   tensor(1.2398, device='cuda:0', grad_fn=<NllLossBackward>)
[669] Train Loss:   tensor(1.2553, device='cuda:0', grad_fn=<NllLossBackward>)
[670] Train Loss:   tensor(1.2574, device='cuda:0', grad_fn=<NllLossBackward>)
[671] Train Loss:   tensor(1.2402, device='cuda:0', grad_fn=<NllLossBackward>)
[672] Train Loss:   tensor(1.2442, device='cuda:0', grad_fn=<NllLossBackward>)
[673] Train Loss:   tensor(1.2746, device='cuda:0', grad_fn=<NllLossBackward>)
[674] Train Loss:   tensor(1.2245, device='cuda:0', grad_fn=<NllLossBackward>)
[675] Train Loss:   tensor(1.2889, device='cuda:0', grad_fn=<NllLossBackward>)
[676] Train Loss:   tensor(1.2618, device='cuda:0', grad_fn=<NllLossBackward>)
[677] Train Loss:   tensor(1.2239, device='cuda:0', grad_fn=<NllLossBackward>)
[678] Train Loss:   tensor(1.2606, device='cuda:0', grad_fn=<NllLossBackward>)
[679] Train Loss:   tensor(1.2330, device='cuda:0', grad_fn=<NllLossBackward>)
[680] Train Loss:   tensor(1.2980, device='cuda:0', grad_fn=<NllLossBackward>)
[681] Train Loss:   tensor(1.2527, device='cuda:0', grad_fn=<NllLossBackward>)
[682] Train Loss:   tensor(1.2332, device='cuda:0', grad_fn=<NllLossBackward>)
[683] Train Loss:   tensor(1.2668, device='cuda:0', grad_fn=<NllLossBackward>)
[684] Train Loss:   tensor(1.2312, device='cuda:0', grad_fn=<NllLossBackward>)
[685] Train Loss:   tensor(1.2743, device='cuda:0', grad_fn=<NllLossBackward>)
[686] Train Loss:   tensor(1.2604, device='cuda:0', grad_fn=<NllLossBackward>)
[687] Train Loss:   tensor(1.2441, device='cuda:0', grad_fn=<NllLossBackward>)
[688] Train Loss:   tensor(1.2725, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[689] Train Loss:   tensor(1.2642, device='cuda:0', grad_fn=<NllLossBackward>)
[690] Train Loss:   tensor(1.2414, device='cuda:0', grad_fn=<NllLossBackward>)
[691] Train Loss:   tensor(1.2444, device='cuda:0', grad_fn=<NllLossBackward>)
[692] Train Loss:   tensor(1.2607, device='cuda:0', grad_fn=<NllLossBackward>)
[693] Train Loss:   tensor(1.2391, device='cuda:0', grad_fn=<NllLossBackward>)
[694] Train Loss:   tensor(1.2433, device='cuda:0', grad_fn=<NllLossBackward>)
[695] Train Loss:   tensor(1.2208, device='cuda:0', grad_fn=<NllLossBackward>)
[696] Train Loss:   tensor(1.2579, device='cuda:0', grad_fn=<NllLossBackward>)
[697] Train Loss:   tensor(1.2399, device='cuda:0', grad_fn=<NllLossBackward>)
[698] Train Loss:   tensor(1.2852, device='cuda:0', grad_fn=<NllLossBackward>)
[699] Train Loss:   tensor(1.2894, device='cuda:0', grad_fn=<NllLossBackward>)
[700] Train Loss:   tensor(1.2331, device='cuda:0', grad_fn=<NllLossBackward>)
[701] Train Loss:   tensor(1.2933, device='cuda:0', grad_fn=<NllLossBackward>)
[702] Train Loss:   tensor(1.2482, device='cuda:0', grad_fn=<NllLossBackward>)
[703] Train Loss:   tensor(1.2794, device='cuda:0', grad_fn=<NllLossBackward>)
[704] Train Loss:   tensor(1.2559, device='cuda:0', grad_fn=<NllLossBackward>)
[705] Train Loss:   tensor(1.2580, device='cuda:0', grad_fn=<NllLossBackward>)
[706] Train Loss:   tensor(1.2387, device='cuda:0', grad_fn=<NllLossBackward>)
[707] Train Loss:   tensor(1.2917, device='cuda:0', grad_fn=<NllLossBackward>)
[708] Train Loss:   tensor(1.2869, device='cuda:0', grad_fn=<NllLossBackward>)
[709] Train Loss:   tensor(1.2921, device='cuda:0', grad_fn=<NllLossBackward>)
[710] Train Loss:   tensor(1.2343, device='cuda:0', grad_fn=<NllLossBackward>)
[711] Train Loss:   tensor(1.2696, device='cuda:0', grad_fn=<NllLossBackward>)
[712] Train Loss:   tensor(1.2699, device='cuda:0', grad_fn=<NllLossBackward>)
[713] Train Loss:   tensor(1.2350, device='cuda:0', grad_fn=<NllLossBackward>)
[714] Train Loss:   tensor(1.3021, device='cuda:0', grad_fn=<NllLossBackward>)
[715] Train Loss:   tensor(1.2638, device='cuda:0', grad_fn=<NllLossBackward>)
[716] Train Loss:   tensor(1.2484, device='cuda:0', grad_fn=<NllLossBackward>)
[717] Train Loss:   tensor(1.2469, device='cuda:0', grad_fn=<NllLossBackward>)
[718] Train Loss:   tensor(1.2570, device='cuda:0', grad_fn=<NllLossBackward>)
[719] Train Loss:   tensor(1.2228, device='cuda:0', grad_fn=<NllLossBackward>)
[720] Train Loss:   tensor(1.2079, device='cuda:0', grad_fn=<NllLossBackward>)
[721] Train Loss:   tensor(1.2432, device='cuda:0', grad_fn=<NllLossBackward>)
[722] Train Loss:   tensor(1.2640, device='cuda:0', grad_fn=<NllLossBackward>)
[723] Train Loss:   tensor(1.2211, device='cuda:0', grad_fn=<NllLossBackward>)
[724] Train Loss:   tensor(1.2206, device='cuda:0', grad_fn=<NllLossBackward>)
[725] Train Loss:   tensor(1.2826, device='cuda:0', grad_fn=<NllLossBackward>)
[726] Train Loss:   tensor(1.2576, device='cuda:0', grad_fn=<NllLossBackward>)
[727] Train Loss:   tensor(1.2436, device='cuda:0', grad_fn=<NllLossBackward>)
[728] Train Loss:   tensor(1.2305, device='cuda:0', grad_fn=<NllLossBackward>)
[729] Train Loss:   tensor(1.2485, device='cuda:0', grad_fn=<NllLossBackward>)
[730] Train Loss:   tensor(1.2821, device='cuda:0', grad_fn=<NllLossBackward>)
[731] Train Loss:   tensor(1.2416, device='cuda:0', grad_fn=<NllLossBackward>)
[732] Train Loss:   tensor(1.2409, device='cuda:0', grad_fn=<NllLossBackward>)
[733] Train Loss:   tensor(1.2731, device='cuda:0', grad_fn=<NllLossBackward>)
[734] Train Loss:   tensor(1.2430, device='cuda:0', grad_fn=<NllLossBackward>)
[735] Train Loss:   tensor(1.2535, device='cuda:0', grad_fn=<NllLossBackward>)
[736] Train Loss:   tensor(1.2769, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[737] Train Loss:    tensor(1.2331, device='cuda:0', grad_fn=<NllLossBackward>)
[738] Train Loss:    tensor(1.2437, device='cuda:0', grad_fn=<NllLossBackward>)
[739] Train Loss:    tensor(1.2215, device='cuda:0', grad_fn=<NllLossBackward>)
[740] Train Loss:    tensor(1.2387, device='cuda:0', grad_fn=<NllLossBackward>)
[741] Train Loss:    tensor(1.2636, device='cuda:0', grad_fn=<NllLossBackward>)
[742] Train Loss:    tensor(1.2360, device='cuda:0', grad_fn=<NllLossBackward>)
[743] Train Loss:    tensor(1.2670, device='cuda:0', grad_fn=<NllLossBackward>)
[744] Train Loss:    tensor(1.2398, device='cuda:0', grad_fn=<NllLossBackward>)
[745] Train Loss:    tensor(1.2487, device='cuda:0', grad_fn=<NllLossBackward>)
[746] Train Loss:    tensor(1.2624, device='cuda:0', grad_fn=<NllLossBackward>)
[747] Train Loss:    tensor(1.2293, device='cuda:0', grad_fn=<NllLossBackward>)
[748] Train Loss:    tensor(1.2039, device='cuda:0', grad_fn=<NllLossBackward>)
[749] Train Loss:    tensor(1.2465, device='cuda:0', grad_fn=<NllLossBackward>)
[750] Train Loss:    tensor(1.2460, device='cuda:0', grad_fn=<NllLossBackward>)
[751] Train Loss:    tensor(1.2547, device='cuda:0', grad_fn=<NllLossBackward>)
[752] Train Loss:    tensor(1.2157, device='cuda:0', grad_fn=<NllLossBackward>)
[753] Train Loss:    tensor(1.1940, device='cuda:0', grad_fn=<NllLossBackward>)
[754] Train Loss:    tensor(1.2322, device='cuda:0', grad_fn=<NllLossBackward>)
[755] Train Loss:    tensor(1.2205, device='cuda:0', grad_fn=<NllLossBackward>)
[756] Train Loss:    tensor(1.1772, device='cuda:0', grad_fn=<NllLossBackward>)
[757] Train Loss:    tensor(1.2749, device='cuda:0', grad_fn=<NllLossBackward>)
[758] Train Loss:    tensor(1.2145, device='cuda:0', grad_fn=<NllLossBackward>)
[759] Train Loss:    tensor(1.2207, device='cuda:0', grad_fn=<NllLossBackward>)
[760] Train Loss:    tensor(1.2788, device='cuda:0', grad_fn=<NllLossBackward>)
[761] Train Loss:    tensor(1.2745, device='cuda:0', grad_fn=<NllLossBackward>)
[762] Train Loss:    tensor(1.2138, device='cuda:0', grad_fn=<NllLossBackward>)
[763] Train Loss:    tensor(1.1737, device='cuda:0', grad_fn=<NllLossBackward>)
[764] Train Loss:    tensor(1.2341, device='cuda:0', grad_fn=<NllLossBackward>)
[765] Train Loss:    tensor(1.2756, device='cuda:0', grad_fn=<NllLossBackward>)
[766] Train Loss:    tensor(1.2250, device='cuda:0', grad_fn=<NllLossBackward>)
[767] Train Loss:    tensor(1.2316, device='cuda:0', grad_fn=<NllLossBackward>)
[768] Train Loss:    tensor(1.2401, device='cuda:0', grad_fn=<NllLossBackward>)
[769] Train Loss:    tensor(1.2573, device='cuda:0', grad_fn=<NllLossBackward>)
[770] Train Loss:    tensor(1.2461, device='cuda:0', grad_fn=<NllLossBackward>)
[771] Train Loss:    tensor(1.2071, device='cuda:0', grad_fn=<NllLossBackward>)
[772] Train Loss:    tensor(1.2581, device='cuda:0', grad_fn=<NllLossBackward>)
[773] Train Loss:    tensor(1.2281, device='cuda:0', grad_fn=<NllLossBackward>)
[774] Train Loss:    tensor(1.2314, device='cuda:0', grad_fn=<NllLossBackward>)
[775] Train Loss:    tensor(1.2080, device='cuda:0', grad_fn=<NllLossBackward>)
[776] Train Loss:    tensor(1.1911, device='cuda:0', grad_fn=<NllLossBackward>)
[777] Train Loss:    tensor(1.2299, device='cuda:0', grad_fn=<NllLossBackward>)
[778] Train Loss:    tensor(1.2349, device='cuda:0', grad_fn=<NllLossBackward>)
[779] Train Loss:    tensor(1.2386, device='cuda:0', grad_fn=<NllLossBackward>)
[780] Train Loss:    tensor(1.2255, device='cuda:0', grad_fn=<NllLossBackward>)
[781] Train Loss:    tensor(1.2298, device='cuda:0', grad_fn=<NllLossBackward>)
[782] Train Loss:    tensor(1.2626, device='cuda:0', grad_fn=<NllLossBackward>)
[783] Train Loss:    tensor(1.2366, device='cuda:0', grad_fn=<NllLossBackward>)
[784] Train Loss:    tensor(1.2303, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[785] Train Loss:   tensor(1.2541, device='cuda:0', grad_fn=<NllLossBackward>)
[786] Train Loss:   tensor(1.2407, device='cuda:0', grad_fn=<NllLossBackward>)
[787] Train Loss:   tensor(1.2320, device='cuda:0', grad_fn=<NllLossBackward>)
[788] Train Loss:   tensor(1.2191, device='cuda:0', grad_fn=<NllLossBackward>)
[789] Train Loss:   tensor(1.2346, device='cuda:0', grad_fn=<NllLossBackward>)
[790] Train Loss:   tensor(1.2263, device='cuda:0', grad_fn=<NllLossBackward>)
[791] Train Loss:   tensor(1.2235, device='cuda:0', grad_fn=<NllLossBackward>)
[792] Train Loss:   tensor(1.2222, device='cuda:0', grad_fn=<NllLossBackward>)
[793] Train Loss:   tensor(1.2737, device='cuda:0', grad_fn=<NllLossBackward>)
[794] Train Loss:   tensor(1.2283, device='cuda:0', grad_fn=<NllLossBackward>)
[795] Train Loss:   tensor(1.2719, device='cuda:0', grad_fn=<NllLossBackward>)
[796] Train Loss:   tensor(1.2093, device='cuda:0', grad_fn=<NllLossBackward>)
[797] Train Loss:   tensor(1.2109, device='cuda:0', grad_fn=<NllLossBackward>)
[798] Train Loss:   tensor(1.2148, device='cuda:0', grad_fn=<NllLossBackward>)
[799] Train Loss:   tensor(1.2462, device='cuda:0', grad_fn=<NllLossBackward>)
[800] Train Loss:   tensor(1.2144, device='cuda:0', grad_fn=<NllLossBackward>)
[801] Train Loss:   tensor(1.2231, device='cuda:0', grad_fn=<NllLossBackward>)
[802] Train Loss:   tensor(1.2168, device='cuda:0', grad_fn=<NllLossBackward>)
[803] Train Loss:   tensor(1.2017, device='cuda:0', grad_fn=<NllLossBackward>)
[804] Train Loss:   tensor(1.2264, device='cuda:0', grad_fn=<NllLossBackward>)
[805] Train Loss:   tensor(1.2291, device='cuda:0', grad_fn=<NllLossBackward>)
[806] Train Loss:   tensor(1.2491, device='cuda:0', grad_fn=<NllLossBackward>)
[807] Train Loss:   tensor(1.2350, device='cuda:0', grad_fn=<NllLossBackward>)
[808] Train Loss:   tensor(1.2015, device='cuda:0', grad_fn=<NllLossBackward>)
[809] Train Loss:   tensor(1.2572, device='cuda:0', grad_fn=<NllLossBackward>)
[810] Train Loss:   tensor(1.2124, device='cuda:0', grad_fn=<NllLossBackward>)
[811] Train Loss:   tensor(1.2404, device='cuda:0', grad_fn=<NllLossBackward>)
[812] Train Loss:   tensor(1.2479, device='cuda:0', grad_fn=<NllLossBackward>)
[813] Train Loss:   tensor(1.2165, device='cuda:0', grad_fn=<NllLossBackward>)
[814] Train Loss:   tensor(1.2326, device='cuda:0', grad_fn=<NllLossBackward>)
[815] Train Loss:   tensor(1.2351, device='cuda:0', grad_fn=<NllLossBackward>)
[816] Train Loss:   tensor(1.2200, device='cuda:0', grad_fn=<NllLossBackward>)
[817] Train Loss:   tensor(1.2205, device='cuda:0', grad_fn=<NllLossBackward>)
[818] Train Loss:   tensor(1.2063, device='cuda:0', grad_fn=<NllLossBackward>)
[819] Train Loss:   tensor(1.2215, device='cuda:0', grad_fn=<NllLossBackward>)
[820] Train Loss:   tensor(1.2109, device='cuda:0', grad_fn=<NllLossBackward>)
[821] Train Loss:   tensor(1.2139, device='cuda:0', grad_fn=<NllLossBackward>)
[822] Train Loss:   tensor(1.2077, device='cuda:0', grad_fn=<NllLossBackward>)
[823] Train Loss:   tensor(1.2346, device='cuda:0', grad_fn=<NllLossBackward>)
[824] Train Loss:   tensor(1.2255, device='cuda:0', grad_fn=<NllLossBackward>)
[825] Train Loss:   tensor(1.1980, device='cuda:0', grad_fn=<NllLossBackward>)
[826] Train Loss:   tensor(1.2456, device='cuda:0', grad_fn=<NllLossBackward>)
[827] Train Loss:   tensor(1.2132, device='cuda:0', grad_fn=<NllLossBackward>)
[828] Train Loss:   tensor(1.2641, device='cuda:0', grad_fn=<NllLossBackward>)
[829] Train Loss:   tensor(1.2084, device='cuda:0', grad_fn=<NllLossBackward>)
[830] Train Loss:   tensor(1.2116, device='cuda:0', grad_fn=<NllLossBackward>)
[831] Train Loss:   tensor(1.2356, device='cuda:0', grad_fn=<NllLossBackward>)
[832] Train Loss:   tensor(1.2260, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[833] Train Loss:    tensor(1.2252, device='cuda:0', grad_fn=<NllLossBackward>)
[834] Train Loss:    tensor(1.1860, device='cuda:0', grad_fn=<NllLossBackward>)
[835] Train Loss:    tensor(1.2165, device='cuda:0', grad_fn=<NllLossBackward>)
[836] Train Loss:    tensor(1.2349, device='cuda:0', grad_fn=<NllLossBackward>)
[837] Train Loss:    tensor(1.2302, device='cuda:0', grad_fn=<NllLossBackward>)
[838] Train Loss:    tensor(1.2338, device='cuda:0', grad_fn=<NllLossBackward>)
[839] Train Loss:    tensor(1.2593, device='cuda:0', grad_fn=<NllLossBackward>)
[840] Train Loss:    tensor(1.1981, device='cuda:0', grad_fn=<NllLossBackward>)
[841] Train Loss:    tensor(1.2394, device='cuda:0', grad_fn=<NllLossBackward>)
[842] Train Loss:    tensor(1.2127, device='cuda:0', grad_fn=<NllLossBackward>)
[843] Train Loss:    tensor(1.2310, device='cuda:0', grad_fn=<NllLossBackward>)
[844] Train Loss:    tensor(1.2401, device='cuda:0', grad_fn=<NllLossBackward>)
[845] Train Loss:    tensor(1.2372, device='cuda:0', grad_fn=<NllLossBackward>)
[846] Train Loss:    tensor(1.2265, device='cuda:0', grad_fn=<NllLossBackward>)
[847] Train Loss:    tensor(1.1726, device='cuda:0', grad_fn=<NllLossBackward>)
[848] Train Loss:    tensor(1.1778, device='cuda:0', grad_fn=<NllLossBackward>)
[849] Train Loss:    tensor(1.2291, device='cuda:0', grad_fn=<NllLossBackward>)
[850] Train Loss:    tensor(1.1963, device='cuda:0', grad_fn=<NllLossBackward>)
[851] Train Loss:    tensor(1.1957, device='cuda:0', grad_fn=<NllLossBackward>)
[852] Train Loss:    tensor(1.2198, device='cuda:0', grad_fn=<NllLossBackward>)
[853] Train Loss:    tensor(1.2232, device='cuda:0', grad_fn=<NllLossBackward>)
[854] Train Loss:    tensor(1.2097, device='cuda:0', grad_fn=<NllLossBackward>)
[855] Train Loss:    tensor(1.2246, device='cuda:0', grad_fn=<NllLossBackward>)
[856] Train Loss:    tensor(1.1797, device='cuda:0', grad_fn=<NllLossBackward>)
[857] Train Loss:    tensor(1.1970, device='cuda:0', grad_fn=<NllLossBackward>)
[858] Train Loss:    tensor(1.2036, device='cuda:0', grad_fn=<NllLossBackward>)
[859] Train Loss:    tensor(1.2094, device='cuda:0', grad_fn=<NllLossBackward>)
[860] Train Loss:    tensor(1.2072, device='cuda:0', grad_fn=<NllLossBackward>)
[861] Train Loss:    tensor(1.2101, device='cuda:0', grad_fn=<NllLossBackward>)
[862] Train Loss:    tensor(1.2437, device='cuda:0', grad_fn=<NllLossBackward>)
[863] Train Loss:    tensor(1.2066, device='cuda:0', grad_fn=<NllLossBackward>)
[864] Train Loss:    tensor(1.2189, device='cuda:0', grad_fn=<NllLossBackward>)
[865] Train Loss:    tensor(1.2120, device='cuda:0', grad_fn=<NllLossBackward>)
[866] Train Loss:    tensor(1.2418, device='cuda:0', grad_fn=<NllLossBackward>)
[867] Train Loss:    tensor(1.2176, device='cuda:0', grad_fn=<NllLossBackward>)
[868] Train Loss:    tensor(1.2377, device='cuda:0', grad_fn=<NllLossBackward>)
[869] Train Loss:    tensor(1.1908, device='cuda:0', grad_fn=<NllLossBackward>)
[870] Train Loss:    tensor(1.1798, device='cuda:0', grad_fn=<NllLossBackward>)
[871] Train Loss:    tensor(1.2254, device='cuda:0', grad_fn=<NllLossBackward>)
[872] Train Loss:    tensor(1.1979, device='cuda:0', grad_fn=<NllLossBackward>)
[873] Train Loss:    tensor(1.2226, device='cuda:0', grad_fn=<NllLossBackward>)
[874] Train Loss:    tensor(1.1753, device='cuda:0', grad_fn=<NllLossBackward>)
[875] Train Loss:    tensor(1.2108, device='cuda:0', grad_fn=<NllLossBackward>)
[876] Train Loss:    tensor(1.2574, device='cuda:0', grad_fn=<NllLossBackward>)
[877] Train Loss:    tensor(1.1668, device='cuda:0', grad_fn=<NllLossBackward>)
[878] Train Loss:    tensor(1.2104, device='cuda:0', grad_fn=<NllLossBackward>)
[879] Train Loss:    tensor(1.2102, device='cuda:0', grad_fn=<NllLossBackward>)
[880] Train Loss:    tensor(1.2497, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[881] Train Loss:   tensor(1.1962, device='cuda:0', grad_fn=<NllLossBackward>)
[882] Train Loss:   tensor(1.2159, device='cuda:0', grad_fn=<NllLossBackward>)
[883] Train Loss:   tensor(1.2233, device='cuda:0', grad_fn=<NllLossBackward>)
[884] Train Loss:   tensor(1.2207, device='cuda:0', grad_fn=<NllLossBackward>)
[885] Train Loss:   tensor(1.2341, device='cuda:0', grad_fn=<NllLossBackward>)
[886] Train Loss:   tensor(1.2119, device='cuda:0', grad_fn=<NllLossBackward>)
[887] Train Loss:   tensor(1.2010, device='cuda:0', grad_fn=<NllLossBackward>)
[888] Train Loss:   tensor(1.1595, device='cuda:0', grad_fn=<NllLossBackward>)
[889] Train Loss:   tensor(1.2046, device='cuda:0', grad_fn=<NllLossBackward>)
[890] Train Loss:   tensor(1.2171, device='cuda:0', grad_fn=<NllLossBackward>)
[891] Train Loss:   tensor(1.2227, device='cuda:0', grad_fn=<NllLossBackward>)
[892] Train Loss:   tensor(1.2390, device='cuda:0', grad_fn=<NllLossBackward>)
[893] Train Loss:   tensor(1.1805, device='cuda:0', grad_fn=<NllLossBackward>)
[894] Train Loss:   tensor(1.2064, device='cuda:0', grad_fn=<NllLossBackward>)
[895] Train Loss:   tensor(1.2331, device='cuda:0', grad_fn=<NllLossBackward>)
[896] Train Loss:   tensor(1.2154, device='cuda:0', grad_fn=<NllLossBackward>)
[897] Train Loss:   tensor(1.2186, device='cuda:0', grad_fn=<NllLossBackward>)
[898] Train Loss:   tensor(1.2166, device='cuda:0', grad_fn=<NllLossBackward>)
[899] Train Loss:   tensor(1.2413, device='cuda:0', grad_fn=<NllLossBackward>)
[900] Train Loss:   tensor(1.2327, device='cuda:0', grad_fn=<NllLossBackward>)
[901] Train Loss:   tensor(1.1792, device='cuda:0', grad_fn=<NllLossBackward>)
[902] Train Loss:   tensor(1.2360, device='cuda:0', grad_fn=<NllLossBackward>)
[903] Train Loss:   tensor(1.2065, device='cuda:0', grad_fn=<NllLossBackward>)
[904] Train Loss:   tensor(1.1887, device='cuda:0', grad_fn=<NllLossBackward>)
[905] Train Loss:   tensor(1.2089, device='cuda:0', grad_fn=<NllLossBackward>)
[906] Train Loss:   tensor(1.2149, device='cuda:0', grad_fn=<NllLossBackward>)
[907] Train Loss:   tensor(1.1918, device='cuda:0', grad_fn=<NllLossBackward>)
[908] Train Loss:   tensor(1.1963, device='cuda:0', grad_fn=<NllLossBackward>)
[909] Train Loss:   tensor(1.1886, device='cuda:0', grad_fn=<NllLossBackward>)
[910] Train Loss:   tensor(1.1957, device='cuda:0', grad_fn=<NllLossBackward>)
[911] Train Loss:   tensor(1.2171, device='cuda:0', grad_fn=<NllLossBackward>)
[912] Train Loss:   tensor(1.1810, device='cuda:0', grad_fn=<NllLossBackward>)
[913] Train Loss:   tensor(1.2377, device='cuda:0', grad_fn=<NllLossBackward>)
[914] Train Loss:   tensor(1.2159, device='cuda:0', grad_fn=<NllLossBackward>)
[915] Train Loss:   tensor(1.1948, device='cuda:0', grad_fn=<NllLossBackward>)
[916] Train Loss:   tensor(1.2176, device='cuda:0', grad_fn=<NllLossBackward>)
[917] Train Loss:   tensor(1.1806, device='cuda:0', grad_fn=<NllLossBackward>)
[918] Train Loss:   tensor(1.2122, device='cuda:0', grad_fn=<NllLossBackward>)
[919] Train Loss:   tensor(1.1861, device='cuda:0', grad_fn=<NllLossBackward>)
[920] Train Loss:   tensor(1.2183, device='cuda:0', grad_fn=<NllLossBackward>)
[921] Train Loss:   tensor(1.1980, device='cuda:0', grad_fn=<NllLossBackward>)
[922] Train Loss:   tensor(1.2134, device='cuda:0', grad_fn=<NllLossBackward>)
[923] Train Loss:   tensor(1.1810, device='cuda:0', grad_fn=<NllLossBackward>)
[924] Train Loss:   tensor(1.1926, device='cuda:0', grad_fn=<NllLossBackward>)
[925] Train Loss:   tensor(1.2044, device='cuda:0', grad_fn=<NllLossBackward>)
[926] Train Loss:   tensor(1.2073, device='cuda:0', grad_fn=<NllLossBackward>)
[927] Train Loss:   tensor(1.2172, device='cuda:0', grad_fn=<NllLossBackward>)
[928] Train Loss:   tensor(1.2032, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[929] Train Loss:    tensor(1.1973, device='cuda:0', grad_fn=<NllLossBackward>)
[930] Train Loss:    tensor(1.2098, device='cuda:0', grad_fn=<NllLossBackward>)
[931] Train Loss:    tensor(1.1895, device='cuda:0', grad_fn=<NllLossBackward>)
[932] Train Loss:    tensor(1.2007, device='cuda:0', grad_fn=<NllLossBackward>)
[933] Train Loss:    tensor(1.2165, device='cuda:0', grad_fn=<NllLossBackward>)
[934] Train Loss:    tensor(1.2305, device='cuda:0', grad_fn=<NllLossBackward>)
[935] Train Loss:    tensor(1.1873, device='cuda:0', grad_fn=<NllLossBackward>)
[936] Train Loss:    tensor(1.2220, device='cuda:0', grad_fn=<NllLossBackward>)
[937] Train Loss:    tensor(1.2161, device='cuda:0', grad_fn=<NllLossBackward>)
[938] Train Loss:    tensor(1.2169, device='cuda:0', grad_fn=<NllLossBackward>)
[939] Train Loss:    tensor(1.1940, device='cuda:0', grad_fn=<NllLossBackward>)
[940] Train Loss:    tensor(1.1813, device='cuda:0', grad_fn=<NllLossBackward>)
[941] Train Loss:    tensor(1.2261, device='cuda:0', grad_fn=<NllLossBackward>)
[942] Train Loss:    tensor(1.1746, device='cuda:0', grad_fn=<NllLossBackward>)
[943] Train Loss:    tensor(1.1981, device='cuda:0', grad_fn=<NllLossBackward>)
[944] Train Loss:    tensor(1.2273, device='cuda:0', grad_fn=<NllLossBackward>)
[945] Train Loss:    tensor(1.1897, device='cuda:0', grad_fn=<NllLossBackward>)
[946] Train Loss:    tensor(1.1968, device='cuda:0', grad_fn=<NllLossBackward>)
[947] Train Loss:    tensor(1.1884, device='cuda:0', grad_fn=<NllLossBackward>)
[948] Train Loss:    tensor(1.1971, device='cuda:0', grad_fn=<NllLossBackward>)
[949] Train Loss:    tensor(1.2302, device='cuda:0', grad_fn=<NllLossBackward>)
[950] Train Loss:    tensor(1.1645, device='cuda:0', grad_fn=<NllLossBackward>)
[951] Train Loss:    tensor(1.2445, device='cuda:0', grad_fn=<NllLossBackward>)
[952] Train Loss:    tensor(1.1925, device='cuda:0', grad_fn=<NllLossBackward>)
[953] Train Loss:    tensor(1.1847, device='cuda:0', grad_fn=<NllLossBackward>)
[954] Train Loss:    tensor(1.1962, device='cuda:0', grad_fn=<NllLossBackward>)
[955] Train Loss:    tensor(1.1791, device='cuda:0', grad_fn=<NllLossBackward>)
[956] Train Loss:    tensor(1.2146, device='cuda:0', grad_fn=<NllLossBackward>)
[957] Train Loss:    tensor(1.2217, device='cuda:0', grad_fn=<NllLossBackward>)
[958] Train Loss:    tensor(1.1832, device='cuda:0', grad_fn=<NllLossBackward>)
[959] Train Loss:    tensor(1.1843, device='cuda:0', grad_fn=<NllLossBackward>)
[960] Train Loss:    tensor(1.1860, device='cuda:0', grad_fn=<NllLossBackward>)
[961] Train Loss:    tensor(1.2042, device='cuda:0', grad_fn=<NllLossBackward>)
[962] Train Loss:    tensor(1.1955, device='cuda:0', grad_fn=<NllLossBackward>)
[963] Train Loss:    tensor(1.1627, device='cuda:0', grad_fn=<NllLossBackward>)
[964] Train Loss:    tensor(1.1957, device='cuda:0', grad_fn=<NllLossBackward>)
[965] Train Loss:    tensor(1.1857, device='cuda:0', grad_fn=<NllLossBackward>)
[966] Train Loss:    tensor(1.1946, device='cuda:0', grad_fn=<NllLossBackward>)
[967] Train Loss:    tensor(1.2195, device='cuda:0', grad_fn=<NllLossBackward>)
[968] Train Loss:    tensor(1.2025, device='cuda:0', grad_fn=<NllLossBackward>)
[969] Train Loss:    tensor(1.1865, device='cuda:0', grad_fn=<NllLossBackward>)
[970] Train Loss:    tensor(1.1659, device='cuda:0', grad_fn=<NllLossBackward>)
[971] Train Loss:    tensor(1.1862, device='cuda:0', grad_fn=<NllLossBackward>)
[972] Train Loss:    tensor(1.1683, device='cuda:0', grad_fn=<NllLossBackward>)
[973] Train Loss:    tensor(1.2204, device='cuda:0', grad_fn=<NllLossBackward>)
[974] Train Loss:    tensor(1.1559, device='cuda:0', grad_fn=<NllLossBackward>)
[975] Train Loss:    tensor(1.1868, device='cuda:0', grad_fn=<NllLossBackward>)
[976] Train Loss:    tensor(1.1996, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[977]  Train Loss:  tensor(1.2217, device='cuda:0', grad_fn=<NllLossBackward>)
[978]  Train Loss:  tensor(1.1870, device='cuda:0', grad_fn=<NllLossBackward>)
[979]  Train Loss:  tensor(1.1557, device='cuda:0', grad_fn=<NllLossBackward>)
[980]  Train Loss:  tensor(1.1342, device='cuda:0', grad_fn=<NllLossBackward>)
[981]  Train Loss:  tensor(1.2000, device='cuda:0', grad_fn=<NllLossBackward>)
[982]  Train Loss:  tensor(1.2016, device='cuda:0', grad_fn=<NllLossBackward>)
[983]  Train Loss:  tensor(1.1979, device='cuda:0', grad_fn=<NllLossBackward>)
[984]  Train Loss:  tensor(1.2242, device='cuda:0', grad_fn=<NllLossBackward>)
[985]  Train Loss:  tensor(1.2047, device='cuda:0', grad_fn=<NllLossBackward>)
[986]  Train Loss:  tensor(1.2024, device='cuda:0', grad_fn=<NllLossBackward>)
[987]  Train Loss:  tensor(1.1812, device='cuda:0', grad_fn=<NllLossBackward>)
[988]  Train Loss:  tensor(1.1782, device='cuda:0', grad_fn=<NllLossBackward>)
[989]  Train Loss:  tensor(1.2202, device='cuda:0', grad_fn=<NllLossBackward>)
[990]  Train Loss:  tensor(1.1722, device='cuda:0', grad_fn=<NllLossBackward>)
[991]  Train Loss:  tensor(1.1995, device='cuda:0', grad_fn=<NllLossBackward>)
[992]  Train Loss:  tensor(1.1904, device='cuda:0', grad_fn=<NllLossBackward>)
[993]  Train Loss:  tensor(1.1311, device='cuda:0', grad_fn=<NllLossBackward>)
[994]  Train Loss:  tensor(1.1748, device='cuda:0', grad_fn=<NllLossBackward>)
[995]  Train Loss:  tensor(1.1744, device='cuda:0', grad_fn=<NllLossBackward>)
[996]  Train Loss:  tensor(1.1861, device='cuda:0', grad_fn=<NllLossBackward>)
[997]  Train Loss:  tensor(1.1607, device='cuda:0', grad_fn=<NllLossBackward>)
[998]  Train Loss:  tensor(1.1694, device='cuda:0', grad_fn=<NllLossBackward>)
[999]  Train Loss:  tensor(1.1883, device='cuda:0', grad_fn=<NllLossBackward>)
[1000] Train Loss:  tensor(1.1560, device='cuda:0', grad_fn=<NllLossBackward>)
[1001] Train Loss:  tensor(1.1420, device='cuda:0', grad_fn=<NllLossBackward>)
[1002] Train Loss:  tensor(1.2262, device='cuda:0', grad_fn=<NllLossBackward>)
[1003] Train Loss:  tensor(1.2124, device='cuda:0', grad_fn=<NllLossBackward>)
[1004] Train Loss:  tensor(1.2050, device='cuda:0', grad_fn=<NllLossBackward>)
[1005] Train Loss:  tensor(1.1581, device='cuda:0', grad_fn=<NllLossBackward>)
[1006] Train Loss:  tensor(1.1682, device='cuda:0', grad_fn=<NllLossBackward>)
[1007] Train Loss:  tensor(1.1949, device='cuda:0', grad_fn=<NllLossBackward>)
[1008] Train Loss:  tensor(1.1967, device='cuda:0', grad_fn=<NllLossBackward>)
[1009] Train Loss:  tensor(1.1401, device='cuda:0', grad_fn=<NllLossBackward>)
[1010] Train Loss:  tensor(1.1885, device='cuda:0', grad_fn=<NllLossBackward>)
[1011] Train Loss:  tensor(1.2039, device='cuda:0', grad_fn=<NllLossBackward>)
[1012] Train Loss:  tensor(1.1782, device='cuda:0', grad_fn=<NllLossBackward>)
[1013] Train Loss:  tensor(1.2150, device='cuda:0', grad_fn=<NllLossBackward>)
[1014] Train Loss:  tensor(1.1401, device='cuda:0', grad_fn=<NllLossBackward>)
[1015] Train Loss:  tensor(1.2222, device='cuda:0', grad_fn=<NllLossBackward>)
[1016] Train Loss:  tensor(1.1707, device='cuda:0', grad_fn=<NllLossBackward>)
[1017] Train Loss:  tensor(1.1386, device='cuda:0', grad_fn=<NllLossBackward>)
[1018] Train Loss:  tensor(1.1846, device='cuda:0', grad_fn=<NllLossBackward>)
[1019] Train Loss:  tensor(1.1607, device='cuda:0', grad_fn=<NllLossBackward>)
[1020] Train Loss:  tensor(1.1585, device='cuda:0', grad_fn=<NllLossBackward>)
[1021] Train Loss:  tensor(1.1441, device='cuda:0', grad_fn=<NllLossBackward>)
[1022] Train Loss:  tensor(1.1784, device='cuda:0', grad_fn=<NllLossBackward>)
[1023] Train Loss:  tensor(1.1360, device='cuda:0', grad_fn=<NllLossBackward>)
[1024] Train Loss:  tensor(1.1820, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1025] Train Loss:    tensor(1.1689, device='cuda:0', grad_fn=<NllLossBackward>)
[1026] Train Loss:    tensor(1.1476, device='cuda:0', grad_fn=<NllLossBackward>)
[1027] Train Loss:    tensor(1.2022, device='cuda:0', grad_fn=<NllLossBackward>)
[1028] Train Loss:    tensor(1.1829, device='cuda:0', grad_fn=<NllLossBackward>)
[1029] Train Loss:    tensor(1.1719, device='cuda:0', grad_fn=<NllLossBackward>)
[1030] Train Loss:    tensor(1.1313, device='cuda:0', grad_fn=<NllLossBackward>)
[1031] Train Loss:    tensor(1.1892, device='cuda:0', grad_fn=<NllLossBackward>)
[1032] Train Loss:    tensor(1.1487, device='cuda:0', grad_fn=<NllLossBackward>)
[1033] Train Loss:    tensor(1.1986, device='cuda:0', grad_fn=<NllLossBackward>)
[1034] Train Loss:    tensor(1.1472, device='cuda:0', grad_fn=<NllLossBackward>)
[1035] Train Loss:    tensor(1.1857, device='cuda:0', grad_fn=<NllLossBackward>)
[1036] Train Loss:    tensor(1.2270, device='cuda:0', grad_fn=<NllLossBackward>)
[1037] Train Loss:    tensor(1.1796, device='cuda:0', grad_fn=<NllLossBackward>)
[1038] Train Loss:    tensor(1.1779, device='cuda:0', grad_fn=<NllLossBackward>)
[1039] Train Loss:    tensor(1.1755, device='cuda:0', grad_fn=<NllLossBackward>)
[1040] Train Loss:    tensor(1.1525, device='cuda:0', grad_fn=<NllLossBackward>)
[1041] Train Loss:    tensor(1.2104, device='cuda:0', grad_fn=<NllLossBackward>)
[1042] Train Loss:    tensor(1.1679, device='cuda:0', grad_fn=<NllLossBackward>)
[1043] Train Loss:    tensor(1.1829, device='cuda:0', grad_fn=<NllLossBackward>)
[1044] Train Loss:    tensor(1.1917, device='cuda:0', grad_fn=<NllLossBackward>)
[1045] Train Loss:    tensor(1.1582, device='cuda:0', grad_fn=<NllLossBackward>)
[1046] Train Loss:    tensor(1.1820, device='cuda:0', grad_fn=<NllLossBackward>)
[1047] Train Loss:    tensor(1.1593, device='cuda:0', grad_fn=<NllLossBackward>)
[1048] Train Loss:    tensor(1.1861, device='cuda:0', grad_fn=<NllLossBackward>)
[1049] Train Loss:    tensor(1.1611, device='cuda:0', grad_fn=<NllLossBackward>)
[1050] Train Loss:    tensor(1.1866, device='cuda:0', grad_fn=<NllLossBackward>)
[1051] Train Loss:    tensor(1.1571, device='cuda:0', grad_fn=<NllLossBackward>)
[1052] Train Loss:    tensor(1.1874, device='cuda:0', grad_fn=<NllLossBackward>)
[1053] Train Loss:    tensor(1.1799, device='cuda:0', grad_fn=<NllLossBackward>)
[1054] Train Loss:    tensor(1.1786, device='cuda:0', grad_fn=<NllLossBackward>)
[1055] Train Loss:    tensor(1.1762, device='cuda:0', grad_fn=<NllLossBackward>)
[1056] Train Loss:    tensor(1.1500, device='cuda:0', grad_fn=<NllLossBackward>)
[1057] Train Loss:    tensor(1.1447, device='cuda:0', grad_fn=<NllLossBackward>)
[1058] Train Loss:    tensor(1.2009, device='cuda:0', grad_fn=<NllLossBackward>)
[1059] Train Loss:    tensor(1.1601, device='cuda:0', grad_fn=<NllLossBackward>)
[1060] Train Loss:    tensor(1.1978, device='cuda:0', grad_fn=<NllLossBackward>)
[1061] Train Loss:    tensor(1.1605, device='cuda:0', grad_fn=<NllLossBackward>)
[1062] Train Loss:    tensor(1.1842, device='cuda:0', grad_fn=<NllLossBackward>)
[1063] Train Loss:    tensor(1.1751, device='cuda:0', grad_fn=<NllLossBackward>)
[1064] Train Loss:    tensor(1.1476, device='cuda:0', grad_fn=<NllLossBackward>)
[1065] Train Loss:    tensor(1.1854, device='cuda:0', grad_fn=<NllLossBackward>)
[1066] Train Loss:    tensor(1.1488, device='cuda:0', grad_fn=<NllLossBackward>)
[1067] Train Loss:    tensor(1.1902, device='cuda:0', grad_fn=<NllLossBackward>)
[1068] Train Loss:    tensor(1.1547, device='cuda:0', grad_fn=<NllLossBackward>)
[1069] Train Loss:    tensor(1.1343, device='cuda:0', grad_fn=<NllLossBackward>)
[1070] Train Loss:    tensor(1.1638, device='cuda:0', grad_fn=<NllLossBackward>)
[1071] Train Loss:    tensor(1.1666, device='cuda:0', grad_fn=<NllLossBackward>)
[1072] Train Loss:    tensor(1.1741, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1073] Train Loss:    tensor(1.2282, device='cuda:0', grad_fn=<NllLossBackward>)
[1074] Train Loss:    tensor(1.1815, device='cuda:0', grad_fn=<NllLossBackward>)
[1075] Train Loss:    tensor(1.1774, device='cuda:0', grad_fn=<NllLossBackward>)
[1076] Train Loss:    tensor(1.2186, device='cuda:0', grad_fn=<NllLossBackward>)
[1077] Train Loss:    tensor(1.1719, device='cuda:0', grad_fn=<NllLossBackward>)
[1078] Train Loss:    tensor(1.1606, device='cuda:0', grad_fn=<NllLossBackward>)
[1079] Train Loss:    tensor(1.1523, device='cuda:0', grad_fn=<NllLossBackward>)
[1080] Train Loss:    tensor(1.1644, device='cuda:0', grad_fn=<NllLossBackward>)
[1081] Train Loss:    tensor(1.2028, device='cuda:0', grad_fn=<NllLossBackward>)
[1082] Train Loss:    tensor(1.1961, device='cuda:0', grad_fn=<NllLossBackward>)
[1083] Train Loss:    tensor(1.1642, device='cuda:0', grad_fn=<NllLossBackward>)
[1084] Train Loss:    tensor(1.1448, device='cuda:0', grad_fn=<NllLossBackward>)
[1085] Train Loss:    tensor(1.1375, device='cuda:0', grad_fn=<NllLossBackward>)
[1086] Train Loss:    tensor(1.1528, device='cuda:0', grad_fn=<NllLossBackward>)
[1087] Train Loss:    tensor(1.1669, device='cuda:0', grad_fn=<NllLossBackward>)
[1088] Train Loss:    tensor(1.2102, device='cuda:0', grad_fn=<NllLossBackward>)
[1089] Train Loss:    tensor(1.1710, device='cuda:0', grad_fn=<NllLossBackward>)
[1090] Train Loss:    tensor(1.1954, device='cuda:0', grad_fn=<NllLossBackward>)
[1091] Train Loss:    tensor(1.2005, device='cuda:0', grad_fn=<NllLossBackward>)
[1092] Train Loss:    tensor(1.1095, device='cuda:0', grad_fn=<NllLossBackward>)
[1093] Train Loss:    tensor(1.1843, device='cuda:0', grad_fn=<NllLossBackward>)
[1094] Train Loss:    tensor(1.1892, device='cuda:0', grad_fn=<NllLossBackward>)
[1095] Train Loss:    tensor(1.1630, device='cuda:0', grad_fn=<NllLossBackward>)
[1096] Train Loss:    tensor(1.1471, device='cuda:0', grad_fn=<NllLossBackward>)
[1097] Train Loss:    tensor(1.2047, device='cuda:0', grad_fn=<NllLossBackward>)
[1098] Train Loss:    tensor(1.1787, device='cuda:0', grad_fn=<NllLossBackward>)
[1099] Train Loss:    tensor(1.1599, device='cuda:0', grad_fn=<NllLossBackward>)
[1100] Train Loss:    tensor(1.1683, device='cuda:0', grad_fn=<NllLossBackward>)
[1101] Train Loss:    tensor(1.1694, device='cuda:0', grad_fn=<NllLossBackward>)
[1102] Train Loss:    tensor(1.1301, device='cuda:0', grad_fn=<NllLossBackward>)
[1103] Train Loss:    tensor(1.1881, device='cuda:0', grad_fn=<NllLossBackward>)
[1104] Train Loss:    tensor(1.1666, device='cuda:0', grad_fn=<NllLossBackward>)
[1105] Train Loss:    tensor(1.1210, device='cuda:0', grad_fn=<NllLossBackward>)
[1106] Train Loss:    tensor(1.1294, device='cuda:0', grad_fn=<NllLossBackward>)
[1107] Train Loss:    tensor(1.2153, device='cuda:0', grad_fn=<NllLossBackward>)
[1108] Train Loss:    tensor(1.1751, device='cuda:0', grad_fn=<NllLossBackward>)
[1109] Train Loss:    tensor(1.1497, device='cuda:0', grad_fn=<NllLossBackward>)
[1110] Train Loss:    tensor(1.1668, device='cuda:0', grad_fn=<NllLossBackward>)
[1111] Train Loss:    tensor(1.1369, device='cuda:0', grad_fn=<NllLossBackward>)
[1112] Train Loss:    tensor(1.1459, device='cuda:0', grad_fn=<NllLossBackward>)
[1113] Train Loss:    tensor(1.1551, device='cuda:0', grad_fn=<NllLossBackward>)
[1114] Train Loss:    tensor(1.1900, device='cuda:0', grad_fn=<NllLossBackward>)
[1115] Train Loss:    tensor(1.1765, device='cuda:0', grad_fn=<NllLossBackward>)
[1116] Train Loss:    tensor(1.1535, device='cuda:0', grad_fn=<NllLossBackward>)
[1117] Train Loss:    tensor(1.1764, device='cuda:0', grad_fn=<NllLossBackward>)
[1118] Train Loss:    tensor(1.1739, device='cuda:0', grad_fn=<NllLossBackward>)
[1119] Train Loss:    tensor(1.1579, device='cuda:0', grad_fn=<NllLossBackward>)
[1120] Train Loss:    tensor(1.1935, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1121] Train Loss:   tensor(1.1308, device='cuda:0', grad_fn=<NllLossBackward>)
[1122] Train Loss:   tensor(1.1525, device='cuda:0', grad_fn=<NllLossBackward>)
[1123] Train Loss:   tensor(1.1185, device='cuda:0', grad_fn=<NllLossBackward>)
[1124] Train Loss:   tensor(1.1642, device='cuda:0', grad_fn=<NllLossBackward>)
[1125] Train Loss:   tensor(1.1863, device='cuda:0', grad_fn=<NllLossBackward>)
[1126] Train Loss:   tensor(1.1752, device='cuda:0', grad_fn=<NllLossBackward>)
[1127] Train Loss:   tensor(1.1580, device='cuda:0', grad_fn=<NllLossBackward>)
[1128] Train Loss:   tensor(1.1739, device='cuda:0', grad_fn=<NllLossBackward>)
[1129] Train Loss:   tensor(1.1634, device='cuda:0', grad_fn=<NllLossBackward>)
[1130] Train Loss:   tensor(1.1841, device='cuda:0', grad_fn=<NllLossBackward>)
[1131] Train Loss:   tensor(1.1229, device='cuda:0', grad_fn=<NllLossBackward>)
[1132] Train Loss:   tensor(1.1458, device='cuda:0', grad_fn=<NllLossBackward>)
[1133] Train Loss:   tensor(1.1669, device='cuda:0', grad_fn=<NllLossBackward>)
[1134] Train Loss:   tensor(1.1410, device='cuda:0', grad_fn=<NllLossBackward>)
[1135] Train Loss:   tensor(1.1676, device='cuda:0', grad_fn=<NllLossBackward>)
[1136] Train Loss:   tensor(1.1556, device='cuda:0', grad_fn=<NllLossBackward>)
[1137] Train Loss:   tensor(1.1722, device='cuda:0', grad_fn=<NllLossBackward>)
[1138] Train Loss:   tensor(1.1609, device='cuda:0', grad_fn=<NllLossBackward>)
[1139] Train Loss:   tensor(1.1713, device='cuda:0', grad_fn=<NllLossBackward>)
[1140] Train Loss:   tensor(1.1699, device='cuda:0', grad_fn=<NllLossBackward>)
[1141] Train Loss:   tensor(1.1555, device='cuda:0', grad_fn=<NllLossBackward>)
[1142] Train Loss:   tensor(1.1543, device='cuda:0', grad_fn=<NllLossBackward>)
[1143] Train Loss:   tensor(1.1838, device='cuda:0', grad_fn=<NllLossBackward>)
[1144] Train Loss:   tensor(1.1759, device='cuda:0', grad_fn=<NllLossBackward>)
[1145] Train Loss:   tensor(1.1575, device='cuda:0', grad_fn=<NllLossBackward>)
[1146] Train Loss:   tensor(1.1509, device='cuda:0', grad_fn=<NllLossBackward>)
[1147] Train Loss:   tensor(1.1471, device='cuda:0', grad_fn=<NllLossBackward>)
[1148] Train Loss:   tensor(1.1954, device='cuda:0', grad_fn=<NllLossBackward>)
[1149] Train Loss:   tensor(1.1562, device='cuda:0', grad_fn=<NllLossBackward>)
[1150] Train Loss:   tensor(1.1662, device='cuda:0', grad_fn=<NllLossBackward>)
[1151] Train Loss:   tensor(1.1430, device='cuda:0', grad_fn=<NllLossBackward>)
[1152] Train Loss:   tensor(1.1594, device='cuda:0', grad_fn=<NllLossBackward>)
[1153] Train Loss:   tensor(1.1481, device='cuda:0', grad_fn=<NllLossBackward>)
[1154] Train Loss:   tensor(1.1768, device='cuda:0', grad_fn=<NllLossBackward>)
[1155] Train Loss:   tensor(1.1557, device='cuda:0', grad_fn=<NllLossBackward>)
[1156] Train Loss:   tensor(1.1492, device='cuda:0', grad_fn=<NllLossBackward>)
[1157] Train Loss:   tensor(1.1794, device='cuda:0', grad_fn=<NllLossBackward>)
[1158] Train Loss:   tensor(1.1993, device='cuda:0', grad_fn=<NllLossBackward>)
[1159] Train Loss:   tensor(1.1495, device='cuda:0', grad_fn=<NllLossBackward>)
[1160] Train Loss:   tensor(1.1451, device='cuda:0', grad_fn=<NllLossBackward>)
[1161] Train Loss:   tensor(1.1675, device='cuda:0', grad_fn=<NllLossBackward>)
[1162] Train Loss:   tensor(1.1533, device='cuda:0', grad_fn=<NllLossBackward>)
[1163] Train Loss:   tensor(1.1264, device='cuda:0', grad_fn=<NllLossBackward>)
[1164] Train Loss:   tensor(1.1361, device='cuda:0', grad_fn=<NllLossBackward>)
[1165] Train Loss:   tensor(1.1780, device='cuda:0', grad_fn=<NllLossBackward>)
[1166] Train Loss:   tensor(1.1390, device='cuda:0', grad_fn=<NllLossBackward>)
[1167] Train Loss:   tensor(1.1582, device='cuda:0', grad_fn=<NllLossBackward>)
[1168] Train Loss:   tensor(1.1800, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1169] Train Loss:    tensor(1.1578, device='cuda:0', grad_fn=<NllLossBackward>)
[1170] Train Loss:    tensor(1.1451, device='cuda:0', grad_fn=<NllLossBackward>)
[1171] Train Loss:    tensor(1.2046, device='cuda:0', grad_fn=<NllLossBackward>)
[1172] Train Loss:    tensor(1.1624, device='cuda:0', grad_fn=<NllLossBackward>)
[1173] Train Loss:    tensor(1.1728, device='cuda:0', grad_fn=<NllLossBackward>)
[1174] Train Loss:    tensor(1.1232, device='cuda:0', grad_fn=<NllLossBackward>)
[1175] Train Loss:    tensor(1.1469, device='cuda:0', grad_fn=<NllLossBackward>)
[1176] Train Loss:    tensor(1.1916, device='cuda:0', grad_fn=<NllLossBackward>)
[1177] Train Loss:    tensor(1.1720, device='cuda:0', grad_fn=<NllLossBackward>)
[1178] Train Loss:    tensor(1.1527, device='cuda:0', grad_fn=<NllLossBackward>)
[1179] Train Loss:    tensor(1.1408, device='cuda:0', grad_fn=<NllLossBackward>)
[1180] Train Loss:    tensor(1.1679, device='cuda:0', grad_fn=<NllLossBackward>)
[1181] Train Loss:    tensor(1.1846, device='cuda:0', grad_fn=<NllLossBackward>)
[1182] Train Loss:    tensor(1.1237, device='cuda:0', grad_fn=<NllLossBackward>)
[1183] Train Loss:    tensor(1.1418, device='cuda:0', grad_fn=<NllLossBackward>)
[1184] Train Loss:    tensor(1.1341, device='cuda:0', grad_fn=<NllLossBackward>)
[1185] Train Loss:    tensor(1.1877, device='cuda:0', grad_fn=<NllLossBackward>)
[1186] Train Loss:    tensor(1.1467, device='cuda:0', grad_fn=<NllLossBackward>)
[1187] Train Loss:    tensor(1.1549, device='cuda:0', grad_fn=<NllLossBackward>)
[1188] Train Loss:    tensor(1.1409, device='cuda:0', grad_fn=<NllLossBackward>)
[1189] Train Loss:    tensor(1.1732, device='cuda:0', grad_fn=<NllLossBackward>)
[1190] Train Loss:    tensor(1.1659, device='cuda:0', grad_fn=<NllLossBackward>)
[1191] Train Loss:    tensor(1.1222, device='cuda:0', grad_fn=<NllLossBackward>)
[1192] Train Loss:    tensor(1.1332, device='cuda:0', grad_fn=<NllLossBackward>)
[1193] Train Loss:    tensor(1.1520, device='cuda:0', grad_fn=<NllLossBackward>)
[1194] Train Loss:    tensor(1.1436, device='cuda:0', grad_fn=<NllLossBackward>)
[1195] Train Loss:    tensor(1.1719, device='cuda:0', grad_fn=<NllLossBackward>)
[1196] Train Loss:    tensor(1.1570, device='cuda:0', grad_fn=<NllLossBackward>)
[1197] Train Loss:    tensor(1.1761, device='cuda:0', grad_fn=<NllLossBackward>)
[1198] Train Loss:    tensor(1.1466, device='cuda:0', grad_fn=<NllLossBackward>)
[1199] Train Loss:    tensor(1.1333, device='cuda:0', grad_fn=<NllLossBackward>)
[1200] Train Loss:    tensor(1.1416, device='cuda:0', grad_fn=<NllLossBackward>)
[1201] Train Loss:    tensor(1.1535, device='cuda:0', grad_fn=<NllLossBackward>)
[1202] Train Loss:    tensor(1.1393, device='cuda:0', grad_fn=<NllLossBackward>)
[1203] Train Loss:    tensor(1.1686, device='cuda:0', grad_fn=<NllLossBackward>)
[1204] Train Loss:    tensor(1.1450, device='cuda:0', grad_fn=<NllLossBackward>)
[1205] Train Loss:    tensor(1.1566, device='cuda:0', grad_fn=<NllLossBackward>)
[1206] Train Loss:    tensor(1.1941, device='cuda:0', grad_fn=<NllLossBackward>)
[1207] Train Loss:    tensor(1.1569, device='cuda:0', grad_fn=<NllLossBackward>)
[1208] Train Loss:    tensor(1.1228, device='cuda:0', grad_fn=<NllLossBackward>)
[1209] Train Loss:    tensor(1.1673, device='cuda:0', grad_fn=<NllLossBackward>)
[1210] Train Loss:    tensor(1.1842, device='cuda:0', grad_fn=<NllLossBackward>)
[1211] Train Loss:    tensor(1.1719, device='cuda:0', grad_fn=<NllLossBackward>)
[1212] Train Loss:    tensor(1.1549, device='cuda:0', grad_fn=<NllLossBackward>)
[1213] Train Loss:    tensor(1.1452, device='cuda:0', grad_fn=<NllLossBackward>)
[1214] Train Loss:    tensor(1.1547, device='cuda:0', grad_fn=<NllLossBackward>)
[1215] Train Loss:    tensor(1.1860, device='cuda:0', grad_fn=<NllLossBackward>)
[1216] Train Loss:    tensor(1.1355, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1217] Train Loss:    tensor(1.1750, device='cuda:0', grad_fn=<NllLossBackward>)
[1218] Train Loss:    tensor(1.1381, device='cuda:0', grad_fn=<NllLossBackward>)
[1219] Train Loss:    tensor(1.1604, device='cuda:0', grad_fn=<NllLossBackward>)
[1220] Train Loss:    tensor(1.1572, device='cuda:0', grad_fn=<NllLossBackward>)
[1221] Train Loss:    tensor(1.1565, device='cuda:0', grad_fn=<NllLossBackward>)
[1222] Train Loss:    tensor(1.1830, device='cuda:0', grad_fn=<NllLossBackward>)
[1223] Train Loss:    tensor(1.1506, device='cuda:0', grad_fn=<NllLossBackward>)
[1224] Train Loss:    tensor(1.1600, device='cuda:0', grad_fn=<NllLossBackward>)
[1225] Train Loss:    tensor(1.1380, device='cuda:0', grad_fn=<NllLossBackward>)
[1226] Train Loss:    tensor(1.1577, device='cuda:0', grad_fn=<NllLossBackward>)
[1227] Train Loss:    tensor(1.1388, device='cuda:0', grad_fn=<NllLossBackward>)
[1228] Train Loss:    tensor(1.1424, device='cuda:0', grad_fn=<NllLossBackward>)
[1229] Train Loss:    tensor(1.1526, device='cuda:0', grad_fn=<NllLossBackward>)
[1230] Train Loss:    tensor(1.1301, device='cuda:0', grad_fn=<NllLossBackward>)
[1231] Train Loss:    tensor(1.1227, device='cuda:0', grad_fn=<NllLossBackward>)
[1232] Train Loss:    tensor(1.1616, device='cuda:0', grad_fn=<NllLossBackward>)
[1233] Train Loss:    tensor(1.1366, device='cuda:0', grad_fn=<NllLossBackward>)
[1234] Train Loss:    tensor(1.1702, device='cuda:0', grad_fn=<NllLossBackward>)
[1235] Train Loss:    tensor(1.1591, device='cuda:0', grad_fn=<NllLossBackward>)
[1236] Train Loss:    tensor(1.1564, device='cuda:0', grad_fn=<NllLossBackward>)
[1237] Train Loss:    tensor(1.2081, device='cuda:0', grad_fn=<NllLossBackward>)
[1238] Train Loss:    tensor(1.1494, device='cuda:0', grad_fn=<NllLossBackward>)
[1239] Train Loss:    tensor(1.1516, device='cuda:0', grad_fn=<NllLossBackward>)
[1240] Train Loss:    tensor(1.1639, device='cuda:0', grad_fn=<NllLossBackward>)
[1241] Train Loss:    tensor(1.1172, device='cuda:0', grad_fn=<NllLossBackward>)
[1242] Train Loss:    tensor(1.1878, device='cuda:0', grad_fn=<NllLossBackward>)
[1243] Train Loss:    tensor(1.1470, device='cuda:0', grad_fn=<NllLossBackward>)
[1244] Train Loss:    tensor(1.1294, device='cuda:0', grad_fn=<NllLossBackward>)
[1245] Train Loss:    tensor(1.1203, device='cuda:0', grad_fn=<NllLossBackward>)
[1246] Train Loss:    tensor(1.1224, device='cuda:0', grad_fn=<NllLossBackward>)
[1247] Train Loss:    tensor(1.1718, device='cuda:0', grad_fn=<NllLossBackward>)
[1248] Train Loss:    tensor(1.1497, device='cuda:0', grad_fn=<NllLossBackward>)
[1249] Train Loss:    tensor(1.1795, device='cuda:0', grad_fn=<NllLossBackward>)
[1250] Train Loss:    tensor(1.1509, device='cuda:0', grad_fn=<NllLossBackward>)
[1251] Train Loss:    tensor(1.1746, device='cuda:0', grad_fn=<NllLossBackward>)
[1252] Train Loss:    tensor(1.1049, device='cuda:0', grad_fn=<NllLossBackward>)
[1253] Train Loss:    tensor(1.1190, device='cuda:0', grad_fn=<NllLossBackward>)
[1254] Train Loss:    tensor(1.1446, device='cuda:0', grad_fn=<NllLossBackward>)
[1255] Train Loss:    tensor(1.1318, device='cuda:0', grad_fn=<NllLossBackward>)
[1256] Train Loss:    tensor(1.1614, device='cuda:0', grad_fn=<NllLossBackward>)
[1257] Train Loss:    tensor(1.1342, device='cuda:0', grad_fn=<NllLossBackward>)
[1258] Train Loss:    tensor(1.1353, device='cuda:0', grad_fn=<NllLossBackward>)
[1259] Train Loss:    tensor(1.1688, device='cuda:0', grad_fn=<NllLossBackward>)
[1260] Train Loss:    tensor(1.1486, device='cuda:0', grad_fn=<NllLossBackward>)
[1261] Train Loss:    tensor(1.1099, device='cuda:0', grad_fn=<NllLossBackward>)
[1262] Train Loss:    tensor(1.1018, device='cuda:0', grad_fn=<NllLossBackward>)
[1263] Train Loss:    tensor(1.1372, device='cuda:0', grad_fn=<NllLossBackward>)
[1264] Train Loss:    tensor(1.1602, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1265] Train Loss:    tensor(1.1645, device='cuda:0', grad_fn=<NllLossBackward>)
[1266] Train Loss:    tensor(1.1158, device='cuda:0', grad_fn=<NllLossBackward>)
[1267] Train Loss:    tensor(1.1349, device='cuda:0', grad_fn=<NllLossBackward>)
[1268] Train Loss:    tensor(1.0837, device='cuda:0', grad_fn=<NllLossBackward>)
[1269] Train Loss:    tensor(1.1174, device='cuda:0', grad_fn=<NllLossBackward>)
[1270] Train Loss:    tensor(1.1391, device='cuda:0', grad_fn=<NllLossBackward>)
[1271] Train Loss:    tensor(1.1205, device='cuda:0', grad_fn=<NllLossBackward>)
[1272] Train Loss:    tensor(1.1560, device='cuda:0', grad_fn=<NllLossBackward>)
[1273] Train Loss:    tensor(1.1637, device='cuda:0', grad_fn=<NllLossBackward>)
[1274] Train Loss:    tensor(1.1245, device='cuda:0', grad_fn=<NllLossBackward>)
[1275] Train Loss:    tensor(1.1399, device='cuda:0', grad_fn=<NllLossBackward>)
[1276] Train Loss:    tensor(1.1595, device='cuda:0', grad_fn=<NllLossBackward>)
[1277] Train Loss:    tensor(1.0701, device='cuda:0', grad_fn=<NllLossBackward>)
[1278] Train Loss:    tensor(1.1600, device='cuda:0', grad_fn=<NllLossBackward>)
[1279] Train Loss:    tensor(1.1299, device='cuda:0', grad_fn=<NllLossBackward>)
[1280] Train Loss:    tensor(1.1525, device='cuda:0', grad_fn=<NllLossBackward>)
[1281] Train Loss:    tensor(1.1579, device='cuda:0', grad_fn=<NllLossBackward>)
[1282] Train Loss:    tensor(1.1229, device='cuda:0', grad_fn=<NllLossBackward>)
[1283] Train Loss:    tensor(1.1881, device='cuda:0', grad_fn=<NllLossBackward>)
[1284] Train Loss:    tensor(1.1244, device='cuda:0', grad_fn=<NllLossBackward>)
[1285] Train Loss:    tensor(1.1574, device='cuda:0', grad_fn=<NllLossBackward>)
[1286] Train Loss:    tensor(1.1335, device='cuda:0', grad_fn=<NllLossBackward>)
[1287] Train Loss:    tensor(1.1675, device='cuda:0', grad_fn=<NllLossBackward>)
[1288] Train Loss:    tensor(1.1428, device='cuda:0', grad_fn=<NllLossBackward>)
[1289] Train Loss:    tensor(1.1581, device='cuda:0', grad_fn=<NllLossBackward>)
[1290] Train Loss:    tensor(1.1199, device='cuda:0', grad_fn=<NllLossBackward>)
[1291] Train Loss:    tensor(1.1559, device='cuda:0', grad_fn=<NllLossBackward>)
[1292] Train Loss:    tensor(1.1248, device='cuda:0', grad_fn=<NllLossBackward>)
[1293] Train Loss:    tensor(1.1252, device='cuda:0', grad_fn=<NllLossBackward>)
[1294] Train Loss:    tensor(1.1620, device='cuda:0', grad_fn=<NllLossBackward>)
[1295] Train Loss:    tensor(1.1147, device='cuda:0', grad_fn=<NllLossBackward>)
[1296] Train Loss:    tensor(1.1507, device='cuda:0', grad_fn=<NllLossBackward>)
[1297] Train Loss:    tensor(1.1202, device='cuda:0', grad_fn=<NllLossBackward>)
[1298] Train Loss:    tensor(1.1759, device='cuda:0', grad_fn=<NllLossBackward>)
[1299] Train Loss:    tensor(1.1404, device='cuda:0', grad_fn=<NllLossBackward>)
[1300] Train Loss:    tensor(1.1191, device='cuda:0', grad_fn=<NllLossBackward>)
[1301] Train Loss:    tensor(1.1460, device='cuda:0', grad_fn=<NllLossBackward>)
[1302] Train Loss:    tensor(1.1558, device='cuda:0', grad_fn=<NllLossBackward>)
[1303] Train Loss:    tensor(1.1244, device='cuda:0', grad_fn=<NllLossBackward>)
[1304] Train Loss:    tensor(1.1698, device='cuda:0', grad_fn=<NllLossBackward>)
[1305] Train Loss:    tensor(1.1176, device='cuda:0', grad_fn=<NllLossBackward>)
[1306] Train Loss:    tensor(1.1023, device='cuda:0', grad_fn=<NllLossBackward>)
[1307] Train Loss:    tensor(1.1510, device='cuda:0', grad_fn=<NllLossBackward>)
[1308] Train Loss:    tensor(1.1396, device='cuda:0', grad_fn=<NllLossBackward>)
[1309] Train Loss:    tensor(1.1318, device='cuda:0', grad_fn=<NllLossBackward>)
[1310] Train Loss:    tensor(1.1092, device='cuda:0', grad_fn=<NllLossBackward>)
[1311] Train Loss:    tensor(1.1228, device='cuda:0', grad_fn=<NllLossBackward>)
[1312] Train Loss:    tensor(1.1311, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1313] Train Loss:   tensor(1.1853, device='cuda:0', grad_fn=<NllLossBackward>)
[1314] Train Loss:   tensor(1.1543, device='cuda:0', grad_fn=<NllLossBackward>)
[1315] Train Loss:   tensor(1.1132, device='cuda:0', grad_fn=<NllLossBackward>)
[1316] Train Loss:   tensor(1.1763, device='cuda:0', grad_fn=<NllLossBackward>)
[1317] Train Loss:   tensor(1.1115, device='cuda:0', grad_fn=<NllLossBackward>)
[1318] Train Loss:   tensor(1.1434, device='cuda:0', grad_fn=<NllLossBackward>)
[1319] Train Loss:   tensor(1.1401, device='cuda:0', grad_fn=<NllLossBackward>)
[1320] Train Loss:   tensor(1.1376, device='cuda:0', grad_fn=<NllLossBackward>)
[1321] Train Loss:   tensor(1.1382, device='cuda:0', grad_fn=<NllLossBackward>)
[1322] Train Loss:   tensor(1.1228, device='cuda:0', grad_fn=<NllLossBackward>)
[1323] Train Loss:   tensor(1.1408, device='cuda:0', grad_fn=<NllLossBackward>)
[1324] Train Loss:   tensor(1.1706, device='cuda:0', grad_fn=<NllLossBackward>)
[1325] Train Loss:   tensor(1.1098, device='cuda:0', grad_fn=<NllLossBackward>)
[1326] Train Loss:   tensor(1.1514, device='cuda:0', grad_fn=<NllLossBackward>)
[1327] Train Loss:   tensor(1.1907, device='cuda:0', grad_fn=<NllLossBackward>)
[1328] Train Loss:   tensor(1.1368, device='cuda:0', grad_fn=<NllLossBackward>)
[1329] Train Loss:   tensor(1.0943, device='cuda:0', grad_fn=<NllLossBackward>)
[1330] Train Loss:   tensor(1.1719, device='cuda:0', grad_fn=<NllLossBackward>)
[1331] Train Loss:   tensor(1.1294, device='cuda:0', grad_fn=<NllLossBackward>)
[1332] Train Loss:   tensor(1.1132, device='cuda:0', grad_fn=<NllLossBackward>)
[1333] Train Loss:   tensor(1.1773, device='cuda:0', grad_fn=<NllLossBackward>)
[1334] Train Loss:   tensor(1.1231, device='cuda:0', grad_fn=<NllLossBackward>)
[1335] Train Loss:   tensor(1.1314, device='cuda:0', grad_fn=<NllLossBackward>)
[1336] Train Loss:   tensor(1.1465, device='cuda:0', grad_fn=<NllLossBackward>)
[1337] Train Loss:   tensor(1.1162, device='cuda:0', grad_fn=<NllLossBackward>)
[1338] Train Loss:   tensor(1.1619, device='cuda:0', grad_fn=<NllLossBackward>)
[1339] Train Loss:   tensor(1.1389, device='cuda:0', grad_fn=<NllLossBackward>)
[1340] Train Loss:   tensor(1.1454, device='cuda:0', grad_fn=<NllLossBackward>)
[1341] Train Loss:   tensor(1.1390, device='cuda:0', grad_fn=<NllLossBackward>)
[1342] Train Loss:   tensor(1.1503, device='cuda:0', grad_fn=<NllLossBackward>)
[1343] Train Loss:   tensor(1.1002, device='cuda:0', grad_fn=<NllLossBackward>)
[1344] Train Loss:   tensor(1.1354, device='cuda:0', grad_fn=<NllLossBackward>)
[1345] Train Loss:   tensor(1.1044, device='cuda:0', grad_fn=<NllLossBackward>)
[1346] Train Loss:   tensor(1.1605, device='cuda:0', grad_fn=<NllLossBackward>)
[1347] Train Loss:   tensor(1.1310, device='cuda:0', grad_fn=<NllLossBackward>)
[1348] Train Loss:   tensor(1.1113, device='cuda:0', grad_fn=<NllLossBackward>)
[1349] Train Loss:   tensor(1.1695, device='cuda:0', grad_fn=<NllLossBackward>)
[1350] Train Loss:   tensor(1.1341, device='cuda:0', grad_fn=<NllLossBackward>)
[1351] Train Loss:   tensor(1.1504, device='cuda:0', grad_fn=<NllLossBackward>)
[1352] Train Loss:   tensor(1.1300, device='cuda:0', grad_fn=<NllLossBackward>)
[1353] Train Loss:   tensor(1.1624, device='cuda:0', grad_fn=<NllLossBackward>)
[1354] Train Loss:   tensor(1.1059, device='cuda:0', grad_fn=<NllLossBackward>)
[1355] Train Loss:   tensor(1.1043, device='cuda:0', grad_fn=<NllLossBackward>)
[1356] Train Loss:   tensor(1.1560, device='cuda:0', grad_fn=<NllLossBackward>)
[1357] Train Loss:   tensor(1.0980, device='cuda:0', grad_fn=<NllLossBackward>)
[1358] Train Loss:   tensor(1.1360, device='cuda:0', grad_fn=<NllLossBackward>)
[1359] Train Loss:   tensor(1.1530, device='cuda:0', grad_fn=<NllLossBackward>)
[1360] Train Loss:   tensor(1.1445, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1361] Train Loss:    tensor(1.1105, device='cuda:0', grad_fn=<NllLossBackward>)
[1362] Train Loss:    tensor(1.1327, device='cuda:0', grad_fn=<NllLossBackward>)
[1363] Train Loss:    tensor(1.1773, device='cuda:0', grad_fn=<NllLossBackward>)
[1364] Train Loss:    tensor(1.1097, device='cuda:0', grad_fn=<NllLossBackward>)
[1365] Train Loss:    tensor(1.1357, device='cuda:0', grad_fn=<NllLossBackward>)
[1366] Train Loss:    tensor(1.1164, device='cuda:0', grad_fn=<NllLossBackward>)
[1367] Train Loss:    tensor(1.1271, device='cuda:0', grad_fn=<NllLossBackward>)
[1368] Train Loss:    tensor(1.1252, device='cuda:0', grad_fn=<NllLossBackward>)
[1369] Train Loss:    tensor(1.1245, device='cuda:0', grad_fn=<NllLossBackward>)
[1370] Train Loss:    tensor(1.1207, device='cuda:0', grad_fn=<NllLossBackward>)
[1371] Train Loss:    tensor(1.0993, device='cuda:0', grad_fn=<NllLossBackward>)
[1372] Train Loss:    tensor(1.1435, device='cuda:0', grad_fn=<NllLossBackward>)
[1373] Train Loss:    tensor(1.1186, device='cuda:0', grad_fn=<NllLossBackward>)
[1374] Train Loss:    tensor(1.1103, device='cuda:0', grad_fn=<NllLossBackward>)
[1375] Train Loss:    tensor(1.1297, device='cuda:0', grad_fn=<NllLossBackward>)
[1376] Train Loss:    tensor(1.1057, device='cuda:0', grad_fn=<NllLossBackward>)
[1377] Train Loss:    tensor(1.1322, device='cuda:0', grad_fn=<NllLossBackward>)
[1378] Train Loss:    tensor(1.1401, device='cuda:0', grad_fn=<NllLossBackward>)
[1379] Train Loss:    tensor(1.1447, device='cuda:0', grad_fn=<NllLossBackward>)
[1380] Train Loss:    tensor(1.1370, device='cuda:0', grad_fn=<NllLossBackward>)
[1381] Train Loss:    tensor(1.1246, device='cuda:0', grad_fn=<NllLossBackward>)
[1382] Train Loss:    tensor(1.1517, device='cuda:0', grad_fn=<NllLossBackward>)
[1383] Train Loss:    tensor(1.0987, device='cuda:0', grad_fn=<NllLossBackward>)
[1384] Train Loss:    tensor(1.1151, device='cuda:0', grad_fn=<NllLossBackward>)
[1385] Train Loss:    tensor(1.1210, device='cuda:0', grad_fn=<NllLossBackward>)
[1386] Train Loss:    tensor(1.1167, device='cuda:0', grad_fn=<NllLossBackward>)
[1387] Train Loss:    tensor(1.1460, device='cuda:0', grad_fn=<NllLossBackward>)
[1388] Train Loss:    tensor(1.1345, device='cuda:0', grad_fn=<NllLossBackward>)
[1389] Train Loss:    tensor(1.0785, device='cuda:0', grad_fn=<NllLossBackward>)
[1390] Train Loss:    tensor(1.1779, device='cuda:0', grad_fn=<NllLossBackward>)
[1391] Train Loss:    tensor(1.1028, device='cuda:0', grad_fn=<NllLossBackward>)
[1392] Train Loss:    tensor(1.1342, device='cuda:0', grad_fn=<NllLossBackward>)
[1393] Train Loss:    tensor(1.1542, device='cuda:0', grad_fn=<NllLossBackward>)
[1394] Train Loss:    tensor(1.1593, device='cuda:0', grad_fn=<NllLossBackward>)
[1395] Train Loss:    tensor(1.1419, device='cuda:0', grad_fn=<NllLossBackward>)
[1396] Train Loss:    tensor(1.1235, device='cuda:0', grad_fn=<NllLossBackward>)
[1397] Train Loss:    tensor(1.1376, device='cuda:0', grad_fn=<NllLossBackward>)
[1398] Train Loss:    tensor(1.1368, device='cuda:0', grad_fn=<NllLossBackward>)
[1399] Train Loss:    tensor(1.1157, device='cuda:0', grad_fn=<NllLossBackward>)
[1400] Train Loss:    tensor(1.1139, device='cuda:0', grad_fn=<NllLossBackward>)
[1401] Train Loss:    tensor(1.0749, device='cuda:0', grad_fn=<NllLossBackward>)
[1402] Train Loss:    tensor(1.1406, device='cuda:0', grad_fn=<NllLossBackward>)
[1403] Train Loss:    tensor(1.1357, device='cuda:0', grad_fn=<NllLossBackward>)
[1404] Train Loss:    tensor(1.1070, device='cuda:0', grad_fn=<NllLossBackward>)
[1405] Train Loss:    tensor(1.1141, device='cuda:0', grad_fn=<NllLossBackward>)
[1406] Train Loss:    tensor(1.0981, device='cuda:0', grad_fn=<NllLossBackward>)
[1407] Train Loss:    tensor(1.1018, device='cuda:0', grad_fn=<NllLossBackward>)
[1408] Train Loss:    tensor(1.1468, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1409] Train Loss:    tensor(1.1122, device='cuda:0', grad_fn=<NllLossBackward>)
[1410] Train Loss:    tensor(1.1194, device='cuda:0', grad_fn=<NllLossBackward>)
[1411] Train Loss:    tensor(1.1208, device='cuda:0', grad_fn=<NllLossBackward>)
[1412] Train Loss:    tensor(1.1355, device='cuda:0', grad_fn=<NllLossBackward>)
[1413] Train Loss:    tensor(1.1105, device='cuda:0', grad_fn=<NllLossBackward>)
[1414] Train Loss:    tensor(1.1188, device='cuda:0', grad_fn=<NllLossBackward>)
[1415] Train Loss:    tensor(1.0592, device='cuda:0', grad_fn=<NllLossBackward>)
[1416] Train Loss:    tensor(1.1504, device='cuda:0', grad_fn=<NllLossBackward>)
[1417] Train Loss:    tensor(1.1444, device='cuda:0', grad_fn=<NllLossBackward>)
[1418] Train Loss:    tensor(1.1116, device='cuda:0', grad_fn=<NllLossBackward>)
[1419] Train Loss:    tensor(1.1217, device='cuda:0', grad_fn=<NllLossBackward>)
[1420] Train Loss:    tensor(1.1412, device='cuda:0', grad_fn=<NllLossBackward>)
[1421] Train Loss:    tensor(1.1005, device='cuda:0', grad_fn=<NllLossBackward>)
[1422] Train Loss:    tensor(1.1364, device='cuda:0', grad_fn=<NllLossBackward>)
[1423] Train Loss:    tensor(1.0807, device='cuda:0', grad_fn=<NllLossBackward>)
[1424] Train Loss:    tensor(1.0963, device='cuda:0', grad_fn=<NllLossBackward>)
[1425] Train Loss:    tensor(1.1296, device='cuda:0', grad_fn=<NllLossBackward>)
[1426] Train Loss:    tensor(1.1077, device='cuda:0', grad_fn=<NllLossBackward>)
[1427] Train Loss:    tensor(1.1448, device='cuda:0', grad_fn=<NllLossBackward>)
[1428] Train Loss:    tensor(1.1351, device='cuda:0', grad_fn=<NllLossBackward>)
[1429] Train Loss:    tensor(1.0810, device='cuda:0', grad_fn=<NllLossBackward>)
[1430] Train Loss:    tensor(1.0814, device='cuda:0', grad_fn=<NllLossBackward>)
[1431] Train Loss:    tensor(1.1100, device='cuda:0', grad_fn=<NllLossBackward>)
[1432] Train Loss:    tensor(1.1008, device='cuda:0', grad_fn=<NllLossBackward>)
[1433] Train Loss:    tensor(1.0954, device='cuda:0', grad_fn=<NllLossBackward>)
[1434] Train Loss:    tensor(1.1197, device='cuda:0', grad_fn=<NllLossBackward>)
[1435] Train Loss:    tensor(1.1428, device='cuda:0', grad_fn=<NllLossBackward>)
[1436] Train Loss:    tensor(1.1305, device='cuda:0', grad_fn=<NllLossBackward>)
[1437] Train Loss:    tensor(1.1403, device='cuda:0', grad_fn=<NllLossBackward>)
[1438] Train Loss:    tensor(1.1256, device='cuda:0', grad_fn=<NllLossBackward>)
[1439] Train Loss:    tensor(1.1733, device='cuda:0', grad_fn=<NllLossBackward>)
[1440] Train Loss:    tensor(1.0837, device='cuda:0', grad_fn=<NllLossBackward>)
[1441] Train Loss:    tensor(1.1284, device='cuda:0', grad_fn=<NllLossBackward>)
[1442] Train Loss:    tensor(1.0976, device='cuda:0', grad_fn=<NllLossBackward>)
[1443] Train Loss:    tensor(1.1269, device='cuda:0', grad_fn=<NllLossBackward>)
[1444] Train Loss:    tensor(1.1238, device='cuda:0', grad_fn=<NllLossBackward>)
[1445] Train Loss:    tensor(1.1102, device='cuda:0', grad_fn=<NllLossBackward>)
[1446] Train Loss:    tensor(1.0987, device='cuda:0', grad_fn=<NllLossBackward>)
[1447] Train Loss:    tensor(1.1647, device='cuda:0', grad_fn=<NllLossBackward>)
[1448] Train Loss:    tensor(1.0957, device='cuda:0', grad_fn=<NllLossBackward>)
[1449] Train Loss:    tensor(1.0715, device='cuda:0', grad_fn=<NllLossBackward>)
[1450] Train Loss:    tensor(1.1278, device='cuda:0', grad_fn=<NllLossBackward>)
[1451] Train Loss:    tensor(1.0896, device='cuda:0', grad_fn=<NllLossBackward>)
[1452] Train Loss:    tensor(1.1348, device='cuda:0', grad_fn=<NllLossBackward>)
[1453] Train Loss:    tensor(1.1165, device='cuda:0', grad_fn=<NllLossBackward>)
[1454] Train Loss:    tensor(1.1422, device='cuda:0', grad_fn=<NllLossBackward>)
[1455] Train Loss:    tensor(1.1382, device='cuda:0', grad_fn=<NllLossBackward>)
[1456] Train Loss:    tensor(1.0670, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1457] Train Loss:    tensor(1.1174, device='cuda:0', grad_fn=<NllLossBackward>)
[1458] Train Loss:    tensor(1.1286, device='cuda:0', grad_fn=<NllLossBackward>)
[1459] Train Loss:    tensor(1.1354, device='cuda:0', grad_fn=<NllLossBackward>)
[1460] Train Loss:    tensor(1.1383, device='cuda:0', grad_fn=<NllLossBackward>)
[1461] Train Loss:    tensor(1.1277, device='cuda:0', grad_fn=<NllLossBackward>)
[1462] Train Loss:    tensor(1.1406, device='cuda:0', grad_fn=<NllLossBackward>)
[1463] Train Loss:    tensor(1.1425, device='cuda:0', grad_fn=<NllLossBackward>)
[1464] Train Loss:    tensor(1.1074, device='cuda:0', grad_fn=<NllLossBackward>)
[1465] Train Loss:    tensor(1.1007, device='cuda:0', grad_fn=<NllLossBackward>)
[1466] Train Loss:    tensor(1.1469, device='cuda:0', grad_fn=<NllLossBackward>)
[1467] Train Loss:    tensor(1.1224, device='cuda:0', grad_fn=<NllLossBackward>)
[1468] Train Loss:    tensor(1.0971, device='cuda:0', grad_fn=<NllLossBackward>)
[1469] Train Loss:    tensor(1.0929, device='cuda:0', grad_fn=<NllLossBackward>)
[1470] Train Loss:    tensor(1.1049, device='cuda:0', grad_fn=<NllLossBackward>)
[1471] Train Loss:    tensor(1.1312, device='cuda:0', grad_fn=<NllLossBackward>)
[1472] Train Loss:    tensor(1.1287, device='cuda:0', grad_fn=<NllLossBackward>)
[1473] Train Loss:    tensor(1.0910, device='cuda:0', grad_fn=<NllLossBackward>)
[1474] Train Loss:    tensor(1.1519, device='cuda:0', grad_fn=<NllLossBackward>)
[1475] Train Loss:    tensor(1.1354, device='cuda:0', grad_fn=<NllLossBackward>)
[1476] Train Loss:    tensor(1.1348, device='cuda:0', grad_fn=<NllLossBackward>)
[1477] Train Loss:    tensor(1.1368, device='cuda:0', grad_fn=<NllLossBackward>)
[1478] Train Loss:    tensor(1.1502, device='cuda:0', grad_fn=<NllLossBackward>)
[1479] Train Loss:    tensor(1.0915, device='cuda:0', grad_fn=<NllLossBackward>)
[1480] Train Loss:    tensor(1.1242, device='cuda:0', grad_fn=<NllLossBackward>)
[1481] Train Loss:    tensor(1.1131, device='cuda:0', grad_fn=<NllLossBackward>)
[1482] Train Loss:    tensor(1.1123, device='cuda:0', grad_fn=<NllLossBackward>)
[1483] Train Loss:    tensor(1.1046, device='cuda:0', grad_fn=<NllLossBackward>)
[1484] Train Loss:    tensor(1.1156, device='cuda:0', grad_fn=<NllLossBackward>)
[1485] Train Loss:    tensor(1.1299, device='cuda:0', grad_fn=<NllLossBackward>)
[1486] Train Loss:    tensor(1.1345, device='cuda:0', grad_fn=<NllLossBackward>)
[1487] Train Loss:    tensor(1.1309, device='cuda:0', grad_fn=<NllLossBackward>)
[1488] Train Loss:    tensor(1.1144, device='cuda:0', grad_fn=<NllLossBackward>)
[1489] Train Loss:    tensor(1.1224, device='cuda:0', grad_fn=<NllLossBackward>)
[1490] Train Loss:    tensor(1.0915, device='cuda:0', grad_fn=<NllLossBackward>)
[1491] Train Loss:    tensor(1.0857, device='cuda:0', grad_fn=<NllLossBackward>)
[1492] Train Loss:    tensor(1.1562, device='cuda:0', grad_fn=<NllLossBackward>)
[1493] Train Loss:    tensor(1.1287, device='cuda:0', grad_fn=<NllLossBackward>)
[1494] Train Loss:    tensor(1.0936, device='cuda:0', grad_fn=<NllLossBackward>)
[1495] Train Loss:    tensor(1.1244, device='cuda:0', grad_fn=<NllLossBackward>)
[1496] Train Loss:    tensor(1.0761, device='cuda:0', grad_fn=<NllLossBackward>)
[1497] Train Loss:    tensor(1.0884, device='cuda:0', grad_fn=<NllLossBackward>)
[1498] Train Loss:    tensor(1.1020, device='cuda:0', grad_fn=<NllLossBackward>)
[1499] Train Loss:    tensor(1.1104, device='cuda:0', grad_fn=<NllLossBackward>)
[1500] Train Loss:    tensor(1.1156, device='cuda:0', grad_fn=<NllLossBackward>)
[1501] Train Loss:    tensor(1.1425, device='cuda:0', grad_fn=<NllLossBackward>)
[1502] Train Loss:    tensor(1.1349, device='cuda:0', grad_fn=<NllLossBackward>)
[1503] Train Loss:    tensor(1.1110, device='cuda:0', grad_fn=<NllLossBackward>)
[1504] Train Loss:    tensor(1.0916, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1505] Train Loss:    tensor(1.1386, device='cuda:0', grad_fn=<NllLossBackward>)
[1506] Train Loss:    tensor(1.1275, device='cuda:0', grad_fn=<NllLossBackward>)
[1507] Train Loss:    tensor(1.1352, device='cuda:0', grad_fn=<NllLossBackward>)
[1508] Train Loss:    tensor(1.1117, device='cuda:0', grad_fn=<NllLossBackward>)
[1509] Train Loss:    tensor(1.1372, device='cuda:0', grad_fn=<NllLossBackward>)
[1510] Train Loss:    tensor(1.1393, device='cuda:0', grad_fn=<NllLossBackward>)
[1511] Train Loss:    tensor(1.0959, device='cuda:0', grad_fn=<NllLossBackward>)
[1512] Train Loss:    tensor(1.0821, device='cuda:0', grad_fn=<NllLossBackward>)
[1513] Train Loss:    tensor(1.1192, device='cuda:0', grad_fn=<NllLossBackward>)
[1514] Train Loss:    tensor(1.0898, device='cuda:0', grad_fn=<NllLossBackward>)
[1515] Train Loss:    tensor(1.1292, device='cuda:0', grad_fn=<NllLossBackward>)
[1516] Train Loss:    tensor(1.1326, device='cuda:0', grad_fn=<NllLossBackward>)
[1517] Train Loss:    tensor(1.0952, device='cuda:0', grad_fn=<NllLossBackward>)
[1518] Train Loss:    tensor(1.1202, device='cuda:0', grad_fn=<NllLossBackward>)
[1519] Train Loss:    tensor(1.1309, device='cuda:0', grad_fn=<NllLossBackward>)
[1520] Train Loss:    tensor(1.1191, device='cuda:0', grad_fn=<NllLossBackward>)
[1521] Train Loss:    tensor(1.0957, device='cuda:0', grad_fn=<NllLossBackward>)
[1522] Train Loss:    tensor(1.1238, device='cuda:0', grad_fn=<NllLossBackward>)
[1523] Train Loss:    tensor(1.0932, device='cuda:0', grad_fn=<NllLossBackward>)
[1524] Train Loss:    tensor(1.0861, device='cuda:0', grad_fn=<NllLossBackward>)
[1525] Train Loss:    tensor(1.0943, device='cuda:0', grad_fn=<NllLossBackward>)
[1526] Train Loss:    tensor(1.1184, device='cuda:0', grad_fn=<NllLossBackward>)
[1527] Train Loss:    tensor(1.0872, device='cuda:0', grad_fn=<NllLossBackward>)
[1528] Train Loss:    tensor(1.1029, device='cuda:0', grad_fn=<NllLossBackward>)
[1529] Train Loss:    tensor(1.0786, device='cuda:0', grad_fn=<NllLossBackward>)
[1530] Train Loss:    tensor(1.1137, device='cuda:0', grad_fn=<NllLossBackward>)
[1531] Train Loss:    tensor(1.0826, device='cuda:0', grad_fn=<NllLossBackward>)
[1532] Train Loss:    tensor(1.0973, device='cuda:0', grad_fn=<NllLossBackward>)
[1533] Train Loss:    tensor(1.0906, device='cuda:0', grad_fn=<NllLossBackward>)
[1534] Train Loss:    tensor(1.0977, device='cuda:0', grad_fn=<NllLossBackward>)
[1535] Train Loss:    tensor(1.1257, device='cuda:0', grad_fn=<NllLossBackward>)
[1536] Train Loss:    tensor(1.1556, device='cuda:0', grad_fn=<NllLossBackward>)
[1537] Train Loss:    tensor(1.1327, device='cuda:0', grad_fn=<NllLossBackward>)
[1538] Train Loss:    tensor(1.1067, device='cuda:0', grad_fn=<NllLossBackward>)
[1539] Train Loss:    tensor(1.0748, device='cuda:0', grad_fn=<NllLossBackward>)
[1540] Train Loss:    tensor(1.1061, device='cuda:0', grad_fn=<NllLossBackward>)
[1541] Train Loss:    tensor(1.1158, device='cuda:0', grad_fn=<NllLossBackward>)
[1542] Train Loss:    tensor(1.0726, device='cuda:0', grad_fn=<NllLossBackward>)
[1543] Train Loss:    tensor(1.1259, device='cuda:0', grad_fn=<NllLossBackward>)
[1544] Train Loss:    tensor(1.0995, device='cuda:0', grad_fn=<NllLossBackward>)
[1545] Train Loss:    tensor(1.0791, device='cuda:0', grad_fn=<NllLossBackward>)
[1546] Train Loss:    tensor(1.0973, device='cuda:0', grad_fn=<NllLossBackward>)
[1547] Train Loss:    tensor(1.0749, device='cuda:0', grad_fn=<NllLossBackward>)
[1548] Train Loss:    tensor(1.1386, device='cuda:0', grad_fn=<NllLossBackward>)
[1549] Train Loss:    tensor(1.1047, device='cuda:0', grad_fn=<NllLossBackward>)
[1550] Train Loss:    tensor(1.1538, device='cuda:0', grad_fn=<NllLossBackward>)
[1551] Train Loss:    tensor(1.1313, device='cuda:0', grad_fn=<NllLossBackward>)
[1552] Train Loss:    tensor(1.0781, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1553] Train Loss:   tensor(1.1424, device='cuda:0', grad_fn=<NllLossBackward>)
[1554] Train Loss:   tensor(1.1582, device='cuda:0', grad_fn=<NllLossBackward>)
[1555] Train Loss:   tensor(1.1179, device='cuda:0', grad_fn=<NllLossBackward>)
[1556] Train Loss:   tensor(1.0894, device='cuda:0', grad_fn=<NllLossBackward>)
[1557] Train Loss:   tensor(1.0865, device='cuda:0', grad_fn=<NllLossBackward>)
[1558] Train Loss:   tensor(1.1049, device='cuda:0', grad_fn=<NllLossBackward>)
[1559] Train Loss:   tensor(1.1432, device='cuda:0', grad_fn=<NllLossBackward>)
[1560] Train Loss:   tensor(1.0955, device='cuda:0', grad_fn=<NllLossBackward>)
[1561] Train Loss:   tensor(1.0719, device='cuda:0', grad_fn=<NllLossBackward>)
[1562] Train Loss:   tensor(1.0865, device='cuda:0', grad_fn=<NllLossBackward>)
[1563] Train Loss:   tensor(1.1075, device='cuda:0', grad_fn=<NllLossBackward>)
[1564] Train Loss:   tensor(1.1093, device='cuda:0', grad_fn=<NllLossBackward>)
[1565] Train Loss:   tensor(1.1293, device='cuda:0', grad_fn=<NllLossBackward>)
[1566] Train Loss:   tensor(1.1178, device='cuda:0', grad_fn=<NllLossBackward>)
[1567] Train Loss:   tensor(1.1011, device='cuda:0', grad_fn=<NllLossBackward>)
[1568] Train Loss:   tensor(1.0912, device='cuda:0', grad_fn=<NllLossBackward>)
[1569] Train Loss:   tensor(1.0926, device='cuda:0', grad_fn=<NllLossBackward>)
[1570] Train Loss:   tensor(1.1099, device='cuda:0', grad_fn=<NllLossBackward>)
[1571] Train Loss:   tensor(1.1059, device='cuda:0', grad_fn=<NllLossBackward>)
[1572] Train Loss:   tensor(1.0960, device='cuda:0', grad_fn=<NllLossBackward>)
[1573] Train Loss:   tensor(1.1265, device='cuda:0', grad_fn=<NllLossBackward>)
[1574] Train Loss:   tensor(1.0945, device='cuda:0', grad_fn=<NllLossBackward>)
[1575] Train Loss:   tensor(1.1228, device='cuda:0', grad_fn=<NllLossBackward>)
[1576] Train Loss:   tensor(1.0685, device='cuda:0', grad_fn=<NllLossBackward>)
[1577] Train Loss:   tensor(1.0704, device='cuda:0', grad_fn=<NllLossBackward>)
[1578] Train Loss:   tensor(1.1154, device='cuda:0', grad_fn=<NllLossBackward>)
[1579] Train Loss:   tensor(1.1027, device='cuda:0', grad_fn=<NllLossBackward>)
[1580] Train Loss:   tensor(1.1030, device='cuda:0', grad_fn=<NllLossBackward>)
[1581] Train Loss:   tensor(1.0862, device='cuda:0', grad_fn=<NllLossBackward>)
[1582] Train Loss:   tensor(1.0830, device='cuda:0', grad_fn=<NllLossBackward>)
[1583] Train Loss:   tensor(1.0809, device='cuda:0', grad_fn=<NllLossBackward>)
[1584] Train Loss:   tensor(1.1081, device='cuda:0', grad_fn=<NllLossBackward>)
[1585] Train Loss:   tensor(1.0838, device='cuda:0', grad_fn=<NllLossBackward>)
[1586] Train Loss:   tensor(1.0758, device='cuda:0', grad_fn=<NllLossBackward>)
[1587] Train Loss:   tensor(1.0860, device='cuda:0', grad_fn=<NllLossBackward>)
[1588] Train Loss:   tensor(1.0576, device='cuda:0', grad_fn=<NllLossBackward>)
[1589] Train Loss:   tensor(1.1197, device='cuda:0', grad_fn=<NllLossBackward>)
[1590] Train Loss:   tensor(1.1041, device='cuda:0', grad_fn=<NllLossBackward>)
[1591] Train Loss:   tensor(1.1116, device='cuda:0', grad_fn=<NllLossBackward>)
[1592] Train Loss:   tensor(1.1045, device='cuda:0', grad_fn=<NllLossBackward>)
[1593] Train Loss:   tensor(1.1130, device='cuda:0', grad_fn=<NllLossBackward>)
[1594] Train Loss:   tensor(1.1115, device='cuda:0', grad_fn=<NllLossBackward>)
[1595] Train Loss:   tensor(1.0839, device='cuda:0', grad_fn=<NllLossBackward>)
[1596] Train Loss:   tensor(1.1006, device='cuda:0', grad_fn=<NllLossBackward>)
[1597] Train Loss:   tensor(1.1088, device='cuda:0', grad_fn=<NllLossBackward>)
[1598] Train Loss:   tensor(1.0791, device='cuda:0', grad_fn=<NllLossBackward>)
[1599] Train Loss:   tensor(1.0476, device='cuda:0', grad_fn=<NllLossBackward>)
[1600] Train Loss:   tensor(1.0893, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1601] Train Loss:   tensor(1.1286, device='cuda:0', grad_fn=<NllLossBackward>)
[1602] Train Loss:   tensor(1.1189, device='cuda:0', grad_fn=<NllLossBackward>)
[1603] Train Loss:   tensor(1.0550, device='cuda:0', grad_fn=<NllLossBackward>)
[1604] Train Loss:   tensor(1.0579, device='cuda:0', grad_fn=<NllLossBackward>)
[1605] Train Loss:   tensor(1.1357, device='cuda:0', grad_fn=<NllLossBackward>)
[1606] Train Loss:   tensor(1.1147, device='cuda:0', grad_fn=<NllLossBackward>)
[1607] Train Loss:   tensor(1.0816, device='cuda:0', grad_fn=<NllLossBackward>)
[1608] Train Loss:   tensor(1.0952, device='cuda:0', grad_fn=<NllLossBackward>)
[1609] Train Loss:   tensor(1.1033, device='cuda:0', grad_fn=<NllLossBackward>)
[1610] Train Loss:   tensor(1.0738, device='cuda:0', grad_fn=<NllLossBackward>)
[1611] Train Loss:   tensor(1.1364, device='cuda:0', grad_fn=<NllLossBackward>)
[1612] Train Loss:   tensor(1.1100, device='cuda:0', grad_fn=<NllLossBackward>)
[1613] Train Loss:   tensor(1.0987, device='cuda:0', grad_fn=<NllLossBackward>)
[1614] Train Loss:   tensor(1.1217, device='cuda:0', grad_fn=<NllLossBackward>)
[1615] Train Loss:   tensor(1.1167, device='cuda:0', grad_fn=<NllLossBackward>)
[1616] Train Loss:   tensor(1.0925, device='cuda:0', grad_fn=<NllLossBackward>)
[1617] Train Loss:   tensor(1.1080, device='cuda:0', grad_fn=<NllLossBackward>)
[1618] Train Loss:   tensor(1.0794, device='cuda:0', grad_fn=<NllLossBackward>)
[1619] Train Loss:   tensor(1.0830, device='cuda:0', grad_fn=<NllLossBackward>)
[1620] Train Loss:   tensor(1.1339, device='cuda:0', grad_fn=<NllLossBackward>)
[1621] Train Loss:   tensor(1.0867, device='cuda:0', grad_fn=<NllLossBackward>)
[1622] Train Loss:   tensor(1.0817, device='cuda:0', grad_fn=<NllLossBackward>)
[1623] Train Loss:   tensor(1.0993, device='cuda:0', grad_fn=<NllLossBackward>)
[1624] Train Loss:   tensor(1.0951, device='cuda:0', grad_fn=<NllLossBackward>)
[1625] Train Loss:   tensor(1.1220, device='cuda:0', grad_fn=<NllLossBackward>)
[1626] Train Loss:   tensor(1.1086, device='cuda:0', grad_fn=<NllLossBackward>)
[1627] Train Loss:   tensor(1.0998, device='cuda:0', grad_fn=<NllLossBackward>)
[1628] Train Loss:   tensor(1.0953, device='cuda:0', grad_fn=<NllLossBackward>)
[1629] Train Loss:   tensor(1.1394, device='cuda:0', grad_fn=<NllLossBackward>)
[1630] Train Loss:   tensor(1.0931, device='cuda:0', grad_fn=<NllLossBackward>)
[1631] Train Loss:   tensor(1.1184, device='cuda:0', grad_fn=<NllLossBackward>)
[1632] Train Loss:   tensor(1.1128, device='cuda:0', grad_fn=<NllLossBackward>)
[1633] Train Loss:   tensor(1.1021, device='cuda:0', grad_fn=<NllLossBackward>)
[1634] Train Loss:   tensor(1.1116, device='cuda:0', grad_fn=<NllLossBackward>)
[1635] Train Loss:   tensor(1.0454, device='cuda:0', grad_fn=<NllLossBackward>)
[1636] Train Loss:   tensor(1.0687, device='cuda:0', grad_fn=<NllLossBackward>)
[1637] Train Loss:   tensor(1.0932, device='cuda:0', grad_fn=<NllLossBackward>)
[1638] Train Loss:   tensor(1.1103, device='cuda:0', grad_fn=<NllLossBackward>)
[1639] Train Loss:   tensor(1.1017, device='cuda:0', grad_fn=<NllLossBackward>)
[1640] Train Loss:   tensor(1.0738, device='cuda:0', grad_fn=<NllLossBackward>)
[1641] Train Loss:   tensor(1.1062, device='cuda:0', grad_fn=<NllLossBackward>)
[1642] Train Loss:   tensor(1.1231, device='cuda:0', grad_fn=<NllLossBackward>)
[1643] Train Loss:   tensor(1.0714, device='cuda:0', grad_fn=<NllLossBackward>)
[1644] Train Loss:   tensor(1.0678, device='cuda:0', grad_fn=<NllLossBackward>)
[1645] Train Loss:   tensor(1.0971, device='cuda:0', grad_fn=<NllLossBackward>)
[1646] Train Loss:   tensor(1.1094, device='cuda:0', grad_fn=<NllLossBackward>)
[1647] Train Loss:   tensor(1.1187, device='cuda:0', grad_fn=<NllLossBackward>)
[1648] Train Loss:   tensor(1.1010, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1649] Train Loss:    tensor(1.0957, device='cuda:0', grad_fn=<NllLossBackward>)
[1650] Train Loss:    tensor(1.0862, device='cuda:0', grad_fn=<NllLossBackward>)
[1651] Train Loss:    tensor(1.0612, device='cuda:0', grad_fn=<NllLossBackward>)
[1652] Train Loss:    tensor(1.1202, device='cuda:0', grad_fn=<NllLossBackward>)
[1653] Train Loss:    tensor(1.0748, device='cuda:0', grad_fn=<NllLossBackward>)
[1654] Train Loss:    tensor(1.0775, device='cuda:0', grad_fn=<NllLossBackward>)
[1655] Train Loss:    tensor(1.1173, device='cuda:0', grad_fn=<NllLossBackward>)
[1656] Train Loss:    tensor(1.1163, device='cuda:0', grad_fn=<NllLossBackward>)
[1657] Train Loss:    tensor(1.1025, device='cuda:0', grad_fn=<NllLossBackward>)
[1658] Train Loss:    tensor(1.0937, device='cuda:0', grad_fn=<NllLossBackward>)
[1659] Train Loss:    tensor(1.0984, device='cuda:0', grad_fn=<NllLossBackward>)
[1660] Train Loss:    tensor(1.0868, device='cuda:0', grad_fn=<NllLossBackward>)
[1661] Train Loss:    tensor(1.0881, device='cuda:0', grad_fn=<NllLossBackward>)
[1662] Train Loss:    tensor(1.0927, device='cuda:0', grad_fn=<NllLossBackward>)
[1663] Train Loss:    tensor(1.0761, device='cuda:0', grad_fn=<NllLossBackward>)
[1664] Train Loss:    tensor(1.0937, device='cuda:0', grad_fn=<NllLossBackward>)
[1665] Train Loss:    tensor(1.0771, device='cuda:0', grad_fn=<NllLossBackward>)
[1666] Train Loss:    tensor(1.0515, device='cuda:0', grad_fn=<NllLossBackward>)
[1667] Train Loss:    tensor(1.0431, device='cuda:0', grad_fn=<NllLossBackward>)
[1668] Train Loss:    tensor(1.1153, device='cuda:0', grad_fn=<NllLossBackward>)
[1669] Train Loss:    tensor(1.0844, device='cuda:0', grad_fn=<NllLossBackward>)
[1670] Train Loss:    tensor(1.0853, device='cuda:0', grad_fn=<NllLossBackward>)
[1671] Train Loss:    tensor(1.1288, device='cuda:0', grad_fn=<NllLossBackward>)
[1672] Train Loss:    tensor(1.0923, device='cuda:0', grad_fn=<NllLossBackward>)
[1673] Train Loss:    tensor(1.0680, device='cuda:0', grad_fn=<NllLossBackward>)
[1674] Train Loss:    tensor(1.1199, device='cuda:0', grad_fn=<NllLossBackward>)
[1675] Train Loss:    tensor(1.0792, device='cuda:0', grad_fn=<NllLossBackward>)
[1676] Train Loss:    tensor(1.0807, device='cuda:0', grad_fn=<NllLossBackward>)
[1677] Train Loss:    tensor(1.0638, device='cuda:0', grad_fn=<NllLossBackward>)
[1678] Train Loss:    tensor(1.0828, device='cuda:0', grad_fn=<NllLossBackward>)
[1679] Train Loss:    tensor(1.0814, device='cuda:0', grad_fn=<NllLossBackward>)
[1680] Train Loss:    tensor(1.1061, device='cuda:0', grad_fn=<NllLossBackward>)
[1681] Train Loss:    tensor(1.1100, device='cuda:0', grad_fn=<NllLossBackward>)
[1682] Train Loss:    tensor(1.1248, device='cuda:0', grad_fn=<NllLossBackward>)
[1683] Train Loss:    tensor(1.0833, device='cuda:0', grad_fn=<NllLossBackward>)
[1684] Train Loss:    tensor(1.0954, device='cuda:0', grad_fn=<NllLossBackward>)
[1685] Train Loss:    tensor(1.0783, device='cuda:0', grad_fn=<NllLossBackward>)
[1686] Train Loss:    tensor(1.0802, device='cuda:0', grad_fn=<NllLossBackward>)
[1687] Train Loss:    tensor(1.1014, device='cuda:0', grad_fn=<NllLossBackward>)
[1688] Train Loss:    tensor(1.1217, device='cuda:0', grad_fn=<NllLossBackward>)
[1689] Train Loss:    tensor(1.1020, device='cuda:0', grad_fn=<NllLossBackward>)
[1690] Train Loss:    tensor(1.0877, device='cuda:0', grad_fn=<NllLossBackward>)
[1691] Train Loss:    tensor(1.0957, device='cuda:0', grad_fn=<NllLossBackward>)
[1692] Train Loss:    tensor(1.1581, device='cuda:0', grad_fn=<NllLossBackward>)
[1693] Train Loss:    tensor(1.1170, device='cuda:0', grad_fn=<NllLossBackward>)
[1694] Train Loss:    tensor(1.1161, device='cuda:0', grad_fn=<NllLossBackward>)
[1695] Train Loss:    tensor(1.0794, device='cuda:0', grad_fn=<NllLossBackward>)
[1696] Train Loss:    tensor(1.0683, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1697] Train Loss:   tensor(1.0471, device='cuda:0', grad_fn=<NllLossBackward>)
[1698] Train Loss:   tensor(1.0527, device='cuda:0', grad_fn=<NllLossBackward>)
[1699] Train Loss:   tensor(1.1036, device='cuda:0', grad_fn=<NllLossBackward>)
[1700] Train Loss:   tensor(1.0845, device='cuda:0', grad_fn=<NllLossBackward>)
[1701] Train Loss:   tensor(1.0553, device='cuda:0', grad_fn=<NllLossBackward>)
[1702] Train Loss:   tensor(1.0917, device='cuda:0', grad_fn=<NllLossBackward>)
[1703] Train Loss:   tensor(1.0779, device='cuda:0', grad_fn=<NllLossBackward>)
[1704] Train Loss:   tensor(1.0385, device='cuda:0', grad_fn=<NllLossBackward>)
[1705] Train Loss:   tensor(1.0903, device='cuda:0', grad_fn=<NllLossBackward>)
[1706] Train Loss:   tensor(1.0726, device='cuda:0', grad_fn=<NllLossBackward>)
[1707] Train Loss:   tensor(1.0890, device='cuda:0', grad_fn=<NllLossBackward>)
[1708] Train Loss:   tensor(1.0688, device='cuda:0', grad_fn=<NllLossBackward>)
[1709] Train Loss:   tensor(1.0918, device='cuda:0', grad_fn=<NllLossBackward>)
[1710] Train Loss:   tensor(1.0790, device='cuda:0', grad_fn=<NllLossBackward>)
[1711] Train Loss:   tensor(1.1226, device='cuda:0', grad_fn=<NllLossBackward>)
[1712] Train Loss:   tensor(1.1189, device='cuda:0', grad_fn=<NllLossBackward>)
[1713] Train Loss:   tensor(1.0958, device='cuda:0', grad_fn=<NllLossBackward>)
[1714] Train Loss:   tensor(1.1153, device='cuda:0', grad_fn=<NllLossBackward>)
[1715] Train Loss:   tensor(1.0815, device='cuda:0', grad_fn=<NllLossBackward>)
[1716] Train Loss:   tensor(1.0775, device='cuda:0', grad_fn=<NllLossBackward>)
[1717] Train Loss:   tensor(1.0891, device='cuda:0', grad_fn=<NllLossBackward>)
[1718] Train Loss:   tensor(1.0826, device='cuda:0', grad_fn=<NllLossBackward>)
[1719] Train Loss:   tensor(1.0887, device='cuda:0', grad_fn=<NllLossBackward>)
[1720] Train Loss:   tensor(1.0986, device='cuda:0', grad_fn=<NllLossBackward>)
[1721] Train Loss:   tensor(1.1111, device='cuda:0', grad_fn=<NllLossBackward>)
[1722] Train Loss:   tensor(1.1001, device='cuda:0', grad_fn=<NllLossBackward>)
[1723] Train Loss:   tensor(1.0964, device='cuda:0', grad_fn=<NllLossBackward>)
[1724] Train Loss:   tensor(1.0867, device='cuda:0', grad_fn=<NllLossBackward>)
[1725] Train Loss:   tensor(1.1100, device='cuda:0', grad_fn=<NllLossBackward>)
[1726] Train Loss:   tensor(1.0708, device='cuda:0', grad_fn=<NllLossBackward>)
[1727] Train Loss:   tensor(1.0492, device='cuda:0', grad_fn=<NllLossBackward>)
[1728] Train Loss:   tensor(1.0921, device='cuda:0', grad_fn=<NllLossBackward>)
[1729] Train Loss:   tensor(1.0480, device='cuda:0', grad_fn=<NllLossBackward>)
[1730] Train Loss:   tensor(1.0750, device='cuda:0', grad_fn=<NllLossBackward>)
[1731] Train Loss:   tensor(1.0814, device='cuda:0', grad_fn=<NllLossBackward>)
[1732] Train Loss:   tensor(1.0642, device='cuda:0', grad_fn=<NllLossBackward>)
[1733] Train Loss:   tensor(1.0792, device='cuda:0', grad_fn=<NllLossBackward>)
[1734] Train Loss:   tensor(1.1293, device='cuda:0', grad_fn=<NllLossBackward>)
[1735] Train Loss:   tensor(1.0960, device='cuda:0', grad_fn=<NllLossBackward>)
[1736] Train Loss:   tensor(1.1005, device='cuda:0', grad_fn=<NllLossBackward>)
[1737] Train Loss:   tensor(1.0953, device='cuda:0', grad_fn=<NllLossBackward>)
[1738] Train Loss:   tensor(1.0627, device='cuda:0', grad_fn=<NllLossBackward>)
[1739] Train Loss:   tensor(1.1239, device='cuda:0', grad_fn=<NllLossBackward>)
[1740] Train Loss:   tensor(1.0909, device='cuda:0', grad_fn=<NllLossBackward>)
[1741] Train Loss:   tensor(1.0614, device='cuda:0', grad_fn=<NllLossBackward>)
[1742] Train Loss:   tensor(1.0567, device='cuda:0', grad_fn=<NllLossBackward>)
[1743] Train Loss:   tensor(1.1278, device='cuda:0', grad_fn=<NllLossBackward>)
[1744] Train Loss:   tensor(1.0264, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1745] Train Loss:    tensor(1.0971, device='cuda:0', grad_fn=<NllLossBackward>)
[1746] Train Loss:    tensor(1.0879, device='cuda:0', grad_fn=<NllLossBackward>)
[1747] Train Loss:    tensor(1.1145, device='cuda:0', grad_fn=<NllLossBackward>)
[1748] Train Loss:    tensor(1.0909, device='cuda:0', grad_fn=<NllLossBackward>)
[1749] Train Loss:    tensor(1.1064, device='cuda:0', grad_fn=<NllLossBackward>)
[1750] Train Loss:    tensor(1.1096, device='cuda:0', grad_fn=<NllLossBackward>)
[1751] Train Loss:    tensor(1.1048, device='cuda:0', grad_fn=<NllLossBackward>)
[1752] Train Loss:    tensor(1.1062, device='cuda:0', grad_fn=<NllLossBackward>)
[1753] Train Loss:    tensor(1.0726, device='cuda:0', grad_fn=<NllLossBackward>)
[1754] Train Loss:    tensor(1.0972, device='cuda:0', grad_fn=<NllLossBackward>)
[1755] Train Loss:    tensor(1.0754, device='cuda:0', grad_fn=<NllLossBackward>)
[1756] Train Loss:    tensor(1.0838, device='cuda:0', grad_fn=<NllLossBackward>)
[1757] Train Loss:    tensor(1.0755, device='cuda:0', grad_fn=<NllLossBackward>)
[1758] Train Loss:    tensor(1.0720, device='cuda:0', grad_fn=<NllLossBackward>)
[1759] Train Loss:    tensor(1.1032, device='cuda:0', grad_fn=<NllLossBackward>)
[1760] Train Loss:    tensor(1.0841, device='cuda:0', grad_fn=<NllLossBackward>)
[1761] Train Loss:    tensor(1.0782, device='cuda:0', grad_fn=<NllLossBackward>)
[1762] Train Loss:    tensor(1.0515, device='cuda:0', grad_fn=<NllLossBackward>)
[1763] Train Loss:    tensor(1.1098, device='cuda:0', grad_fn=<NllLossBackward>)
[1764] Train Loss:    tensor(1.0818, device='cuda:0', grad_fn=<NllLossBackward>)
[1765] Train Loss:    tensor(1.0782, device='cuda:0', grad_fn=<NllLossBackward>)
[1766] Train Loss:    tensor(1.0658, device='cuda:0', grad_fn=<NllLossBackward>)
[1767] Train Loss:    tensor(1.0494, device='cuda:0', grad_fn=<NllLossBackward>)
[1768] Train Loss:    tensor(1.0944, device='cuda:0', grad_fn=<NllLossBackward>)
[1769] Train Loss:    tensor(1.1196, device='cuda:0', grad_fn=<NllLossBackward>)
[1770] Train Loss:    tensor(1.0827, device='cuda:0', grad_fn=<NllLossBackward>)
[1771] Train Loss:    tensor(1.0620, device='cuda:0', grad_fn=<NllLossBackward>)
[1772] Train Loss:    tensor(1.0520, device='cuda:0', grad_fn=<NllLossBackward>)
[1773] Train Loss:    tensor(1.0865, device='cuda:0', grad_fn=<NllLossBackward>)
[1774] Train Loss:    tensor(1.0883, device='cuda:0', grad_fn=<NllLossBackward>)
[1775] Train Loss:    tensor(1.0793, device='cuda:0', grad_fn=<NllLossBackward>)
[1776] Train Loss:    tensor(1.0900, device='cuda:0', grad_fn=<NllLossBackward>)
[1777] Train Loss:    tensor(1.0978, device='cuda:0', grad_fn=<NllLossBackward>)
[1778] Train Loss:    tensor(1.0961, device='cuda:0', grad_fn=<NllLossBackward>)
[1779] Train Loss:    tensor(1.0708, device='cuda:0', grad_fn=<NllLossBackward>)
[1780] Train Loss:    tensor(1.0807, device='cuda:0', grad_fn=<NllLossBackward>)
[1781] Train Loss:    tensor(1.0937, device='cuda:0', grad_fn=<NllLossBackward>)
[1782] Train Loss:    tensor(1.1516, device='cuda:0', grad_fn=<NllLossBackward>)
[1783] Train Loss:    tensor(1.0493, device='cuda:0', grad_fn=<NllLossBackward>)
[1784] Train Loss:    tensor(1.0915, device='cuda:0', grad_fn=<NllLossBackward>)
[1785] Train Loss:    tensor(1.1072, device='cuda:0', grad_fn=<NllLossBackward>)
[1786] Train Loss:    tensor(1.0552, device='cuda:0', grad_fn=<NllLossBackward>)
[1787] Train Loss:    tensor(1.1064, device='cuda:0', grad_fn=<NllLossBackward>)
[1788] Train Loss:    tensor(1.1002, device='cuda:0', grad_fn=<NllLossBackward>)
[1789] Train Loss:    tensor(1.0671, device='cuda:0', grad_fn=<NllLossBackward>)
[1790] Train Loss:    tensor(1.0668, device='cuda:0', grad_fn=<NllLossBackward>)
[1791] Train Loss:    tensor(1.1355, device='cuda:0', grad_fn=<NllLossBackward>)
[1792] Train Loss:    tensor(1.0731, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1793] Train Loss:    tensor(1.0526, device='cuda:0', grad_fn=<NllLossBackward>)
[1794] Train Loss:    tensor(1.0870, device='cuda:0', grad_fn=<NllLossBackward>)
[1795] Train Loss:    tensor(1.0808, device='cuda:0', grad_fn=<NllLossBackward>)
[1796] Train Loss:    tensor(1.0424, device='cuda:0', grad_fn=<NllLossBackward>)
[1797] Train Loss:    tensor(1.0752, device='cuda:0', grad_fn=<NllLossBackward>)
[1798] Train Loss:    tensor(1.0959, device='cuda:0', grad_fn=<NllLossBackward>)
[1799] Train Loss:    tensor(1.0481, device='cuda:0', grad_fn=<NllLossBackward>)
[1800] Train Loss:    tensor(1.0797, device='cuda:0', grad_fn=<NllLossBackward>)
[1801] Train Loss:    tensor(1.1131, device='cuda:0', grad_fn=<NllLossBackward>)
[1802] Train Loss:    tensor(1.1044, device='cuda:0', grad_fn=<NllLossBackward>)
[1803] Train Loss:    tensor(1.0797, device='cuda:0', grad_fn=<NllLossBackward>)
[1804] Train Loss:    tensor(1.0980, device='cuda:0', grad_fn=<NllLossBackward>)
[1805] Train Loss:    tensor(1.0179, device='cuda:0', grad_fn=<NllLossBackward>)
[1806] Train Loss:    tensor(1.0616, device='cuda:0', grad_fn=<NllLossBackward>)
[1807] Train Loss:    tensor(1.0643, device='cuda:0', grad_fn=<NllLossBackward>)
[1808] Train Loss:    tensor(1.0956, device='cuda:0', grad_fn=<NllLossBackward>)
[1809] Train Loss:    tensor(1.0789, device='cuda:0', grad_fn=<NllLossBackward>)
[1810] Train Loss:    tensor(1.0348, device='cuda:0', grad_fn=<NllLossBackward>)
[1811] Train Loss:    tensor(1.0911, device='cuda:0', grad_fn=<NllLossBackward>)
[1812] Train Loss:    tensor(1.1485, device='cuda:0', grad_fn=<NllLossBackward>)
[1813] Train Loss:    tensor(1.0805, device='cuda:0', grad_fn=<NllLossBackward>)
[1814] Train Loss:    tensor(1.0690, device='cuda:0', grad_fn=<NllLossBackward>)
[1815] Train Loss:    tensor(1.1118, device='cuda:0', grad_fn=<NllLossBackward>)
[1816] Train Loss:    tensor(1.0905, device='cuda:0', grad_fn=<NllLossBackward>)
[1817] Train Loss:    tensor(1.0516, device='cuda:0', grad_fn=<NllLossBackward>)
[1818] Train Loss:    tensor(1.0699, device='cuda:0', grad_fn=<NllLossBackward>)
[1819] Train Loss:    tensor(1.0850, device='cuda:0', grad_fn=<NllLossBackward>)
[1820] Train Loss:    tensor(1.0742, device='cuda:0', grad_fn=<NllLossBackward>)
[1821] Train Loss:    tensor(1.1015, device='cuda:0', grad_fn=<NllLossBackward>)
[1822] Train Loss:    tensor(1.0611, device='cuda:0', grad_fn=<NllLossBackward>)
[1823] Train Loss:    tensor(1.0559, device='cuda:0', grad_fn=<NllLossBackward>)
[1824] Train Loss:    tensor(1.0695, device='cuda:0', grad_fn=<NllLossBackward>)
[1825] Train Loss:    tensor(1.0419, device='cuda:0', grad_fn=<NllLossBackward>)
[1826] Train Loss:    tensor(1.0713, device='cuda:0', grad_fn=<NllLossBackward>)
[1827] Train Loss:    tensor(1.0708, device='cuda:0', grad_fn=<NllLossBackward>)
[1828] Train Loss:    tensor(1.0762, device='cuda:0', grad_fn=<NllLossBackward>)
[1829] Train Loss:    tensor(1.0644, device='cuda:0', grad_fn=<NllLossBackward>)
[1830] Train Loss:    tensor(1.1080, device='cuda:0', grad_fn=<NllLossBackward>)
[1831] Train Loss:    tensor(1.0696, device='cuda:0', grad_fn=<NllLossBackward>)
[1832] Train Loss:    tensor(1.0899, device='cuda:0', grad_fn=<NllLossBackward>)
[1833] Train Loss:    tensor(1.0794, device='cuda:0', grad_fn=<NllLossBackward>)
[1834] Train Loss:    tensor(1.0783, device='cuda:0', grad_fn=<NllLossBackward>)
[1835] Train Loss:    tensor(1.0646, device='cuda:0', grad_fn=<NllLossBackward>)
[1836] Train Loss:    tensor(1.0942, device='cuda:0', grad_fn=<NllLossBackward>)
[1837] Train Loss:    tensor(1.0976, device='cuda:0', grad_fn=<NllLossBackward>)
[1838] Train Loss:    tensor(1.0988, device='cuda:0', grad_fn=<NllLossBackward>)
[1839] Train Loss:    tensor(1.0778, device='cuda:0', grad_fn=<NllLossBackward>)
[1840] Train Loss:    tensor(1.0962, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1841] Train Loss:    tensor(1.0889, device='cuda:0', grad_fn=<NllLossBackward>)
[1842] Train Loss:    tensor(1.0711, device='cuda:0', grad_fn=<NllLossBackward>)
[1843] Train Loss:    tensor(1.0845, device='cuda:0', grad_fn=<NllLossBackward>)
[1844] Train Loss:    tensor(1.0590, device='cuda:0', grad_fn=<NllLossBackward>)
[1845] Train Loss:    tensor(1.0582, device='cuda:0', grad_fn=<NllLossBackward>)
[1846] Train Loss:    tensor(1.0510, device='cuda:0', grad_fn=<NllLossBackward>)
[1847] Train Loss:    tensor(1.1034, device='cuda:0', grad_fn=<NllLossBackward>)
[1848] Train Loss:    tensor(1.0567, device='cuda:0', grad_fn=<NllLossBackward>)
[1849] Train Loss:    tensor(1.0475, device='cuda:0', grad_fn=<NllLossBackward>)
[1850] Train Loss:    tensor(1.0568, device='cuda:0', grad_fn=<NllLossBackward>)
[1851] Train Loss:    tensor(1.0501, device='cuda:0', grad_fn=<NllLossBackward>)
[1852] Train Loss:    tensor(1.0662, device='cuda:0', grad_fn=<NllLossBackward>)
[1853] Train Loss:    tensor(1.0979, device='cuda:0', grad_fn=<NllLossBackward>)
[1854] Train Loss:    tensor(1.0434, device='cuda:0', grad_fn=<NllLossBackward>)
[1855] Train Loss:    tensor(1.0518, device='cuda:0', grad_fn=<NllLossBackward>)
[1856] Train Loss:    tensor(1.0619, device='cuda:0', grad_fn=<NllLossBackward>)
[1857] Train Loss:    tensor(1.0866, device='cuda:0', grad_fn=<NllLossBackward>)
[1858] Train Loss:    tensor(1.0447, device='cuda:0', grad_fn=<NllLossBackward>)
[1859] Train Loss:    tensor(1.1047, device='cuda:0', grad_fn=<NllLossBackward>)
[1860] Train Loss:    tensor(1.0498, device='cuda:0', grad_fn=<NllLossBackward>)
[1861] Train Loss:    tensor(1.1085, device='cuda:0', grad_fn=<NllLossBackward>)
[1862] Train Loss:    tensor(1.0444, device='cuda:0', grad_fn=<NllLossBackward>)
[1863] Train Loss:    tensor(1.1063, device='cuda:0', grad_fn=<NllLossBackward>)
[1864] Train Loss:    tensor(1.1102, device='cuda:0', grad_fn=<NllLossBackward>)
[1865] Train Loss:    tensor(1.0786, device='cuda:0', grad_fn=<NllLossBackward>)
[1866] Train Loss:    tensor(1.0801, device='cuda:0', grad_fn=<NllLossBackward>)
[1867] Train Loss:    tensor(1.0635, device='cuda:0', grad_fn=<NllLossBackward>)
[1868] Train Loss:    tensor(1.0774, device='cuda:0', grad_fn=<NllLossBackward>)
[1869] Train Loss:    tensor(1.0534, device='cuda:0', grad_fn=<NllLossBackward>)
[1870] Train Loss:    tensor(1.0776, device='cuda:0', grad_fn=<NllLossBackward>)
[1871] Train Loss:    tensor(1.0817, device='cuda:0', grad_fn=<NllLossBackward>)
[1872] Train Loss:    tensor(1.0711, device='cuda:0', grad_fn=<NllLossBackward>)
[1873] Train Loss:    tensor(1.0913, device='cuda:0', grad_fn=<NllLossBackward>)
[1874] Train Loss:    tensor(1.1340, device='cuda:0', grad_fn=<NllLossBackward>)
[1875] Train Loss:    tensor(1.0615, device='cuda:0', grad_fn=<NllLossBackward>)
[1876] Train Loss:    tensor(1.1011, device='cuda:0', grad_fn=<NllLossBackward>)
[1877] Train Loss:    tensor(1.0567, device='cuda:0', grad_fn=<NllLossBackward>)
[1878] Train Loss:    tensor(1.0677, device='cuda:0', grad_fn=<NllLossBackward>)
[1879] Train Loss:    tensor(1.0661, device='cuda:0', grad_fn=<NllLossBackward>)
[1880] Train Loss:    tensor(1.0640, device='cuda:0', grad_fn=<NllLossBackward>)
[1881] Train Loss:    tensor(1.0526, device='cuda:0', grad_fn=<NllLossBackward>)
[1882] Train Loss:    tensor(1.1042, device='cuda:0', grad_fn=<NllLossBackward>)
[1883] Train Loss:    tensor(1.0380, device='cuda:0', grad_fn=<NllLossBackward>)
[1884] Train Loss:    tensor(1.0667, device='cuda:0', grad_fn=<NllLossBackward>)
[1885] Train Loss:    tensor(1.1142, device='cuda:0', grad_fn=<NllLossBackward>)
[1886] Train Loss:    tensor(1.0576, device='cuda:0', grad_fn=<NllLossBackward>)
[1887] Train Loss:    tensor(1.1043, device='cuda:0', grad_fn=<NllLossBackward>)
[1888] Train Loss:    tensor(1.0685, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1889] Train Loss:   tensor(1.0824, device='cuda:0', grad_fn=<NllLossBackward>)
[1890] Train Loss:   tensor(1.0889, device='cuda:0', grad_fn=<NllLossBackward>)
[1891] Train Loss:   tensor(1.1047, device='cuda:0', grad_fn=<NllLossBackward>)
[1892] Train Loss:   tensor(1.0800, device='cuda:0', grad_fn=<NllLossBackward>)
[1893] Train Loss:   tensor(1.0905, device='cuda:0', grad_fn=<NllLossBackward>)
[1894] Train Loss:   tensor(1.0662, device='cuda:0', grad_fn=<NllLossBackward>)
[1895] Train Loss:   tensor(1.0656, device='cuda:0', grad_fn=<NllLossBackward>)
[1896] Train Loss:   tensor(1.1000, device='cuda:0', grad_fn=<NllLossBackward>)
[1897] Train Loss:   tensor(1.0934, device='cuda:0', grad_fn=<NllLossBackward>)
[1898] Train Loss:   tensor(1.0916, device='cuda:0', grad_fn=<NllLossBackward>)
[1899] Train Loss:   tensor(1.0936, device='cuda:0', grad_fn=<NllLossBackward>)
[1900] Train Loss:   tensor(1.0830, device='cuda:0', grad_fn=<NllLossBackward>)
[1901] Train Loss:   tensor(1.0684, device='cuda:0', grad_fn=<NllLossBackward>)
[1902] Train Loss:   tensor(1.0891, device='cuda:0', grad_fn=<NllLossBackward>)
[1903] Train Loss:   tensor(1.1175, device='cuda:0', grad_fn=<NllLossBackward>)
[1904] Train Loss:   tensor(1.0844, device='cuda:0', grad_fn=<NllLossBackward>)
[1905] Train Loss:   tensor(1.0865, device='cuda:0', grad_fn=<NllLossBackward>)
[1906] Train Loss:   tensor(1.1240, device='cuda:0', grad_fn=<NllLossBackward>)
[1907] Train Loss:   tensor(1.1117, device='cuda:0', grad_fn=<NllLossBackward>)
[1908] Train Loss:   tensor(1.0468, device='cuda:0', grad_fn=<NllLossBackward>)
[1909] Train Loss:   tensor(1.0393, device='cuda:0', grad_fn=<NllLossBackward>)
[1910] Train Loss:   tensor(1.0858, device='cuda:0', grad_fn=<NllLossBackward>)
[1911] Train Loss:   tensor(1.0608, device='cuda:0', grad_fn=<NllLossBackward>)
[1912] Train Loss:   tensor(1.0662, device='cuda:0', grad_fn=<NllLossBackward>)
[1913] Train Loss:   tensor(1.0462, device='cuda:0', grad_fn=<NllLossBackward>)
[1914] Train Loss:   tensor(1.0908, device='cuda:0', grad_fn=<NllLossBackward>)
[1915] Train Loss:   tensor(1.0863, device='cuda:0', grad_fn=<NllLossBackward>)
[1916] Train Loss:   tensor(1.0536, device='cuda:0', grad_fn=<NllLossBackward>)
[1917] Train Loss:   tensor(1.1063, device='cuda:0', grad_fn=<NllLossBackward>)
[1918] Train Loss:   tensor(1.0575, device='cuda:0', grad_fn=<NllLossBackward>)
[1919] Train Loss:   tensor(1.1076, device='cuda:0', grad_fn=<NllLossBackward>)
[1920] Train Loss:   tensor(1.0789, device='cuda:0', grad_fn=<NllLossBackward>)
[1921] Train Loss:   tensor(1.0389, device='cuda:0', grad_fn=<NllLossBackward>)
[1922] Train Loss:   tensor(1.0907, device='cuda:0', grad_fn=<NllLossBackward>)
[1923] Train Loss:   tensor(1.0801, device='cuda:0', grad_fn=<NllLossBackward>)
[1924] Train Loss:   tensor(1.0779, device='cuda:0', grad_fn=<NllLossBackward>)
[1925] Train Loss:   tensor(1.0870, device='cuda:0', grad_fn=<NllLossBackward>)
[1926] Train Loss:   tensor(1.0977, device='cuda:0', grad_fn=<NllLossBackward>)
[1927] Train Loss:   tensor(1.0462, device='cuda:0', grad_fn=<NllLossBackward>)
[1928] Train Loss:   tensor(1.0969, device='cuda:0', grad_fn=<NllLossBackward>)
[1929] Train Loss:   tensor(1.0709, device='cuda:0', grad_fn=<NllLossBackward>)
[1930] Train Loss:   tensor(1.0876, device='cuda:0', grad_fn=<NllLossBackward>)
[1931] Train Loss:   tensor(1.0787, device='cuda:0', grad_fn=<NllLossBackward>)
[1932] Train Loss:   tensor(1.0623, device='cuda:0', grad_fn=<NllLossBackward>)
[1933] Train Loss:   tensor(1.0864, device='cuda:0', grad_fn=<NllLossBackward>)
[1934] Train Loss:   tensor(1.0506, device='cuda:0', grad_fn=<NllLossBackward>)
[1935] Train Loss:   tensor(1.0995, device='cuda:0', grad_fn=<NllLossBackward>)
[1936] Train Loss:   tensor(1.0751, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1937] Train Loss:    tensor(1.0623, device='cuda:0', grad_fn=<NllLossBackward>)
[1938] Train Loss:    tensor(1.0918, device='cuda:0', grad_fn=<NllLossBackward>)
[1939] Train Loss:    tensor(1.0775, device='cuda:0', grad_fn=<NllLossBackward>)
[1940] Train Loss:    tensor(1.0015, device='cuda:0', grad_fn=<NllLossBackward>)
[1941] Train Loss:    tensor(1.0798, device='cuda:0', grad_fn=<NllLossBackward>)
[1942] Train Loss:    tensor(1.0635, device='cuda:0', grad_fn=<NllLossBackward>)
[1943] Train Loss:    tensor(1.0403, device='cuda:0', grad_fn=<NllLossBackward>)
[1944] Train Loss:    tensor(1.0771, device='cuda:0', grad_fn=<NllLossBackward>)
[1945] Train Loss:    tensor(1.0767, device='cuda:0', grad_fn=<NllLossBackward>)
[1946] Train Loss:    tensor(1.0269, device='cuda:0', grad_fn=<NllLossBackward>)
[1947] Train Loss:    tensor(1.0809, device='cuda:0', grad_fn=<NllLossBackward>)
[1948] Train Loss:    tensor(1.0587, device='cuda:0', grad_fn=<NllLossBackward>)
[1949] Train Loss:    tensor(1.0491, device='cuda:0', grad_fn=<NllLossBackward>)
[1950] Train Loss:    tensor(1.0994, device='cuda:0', grad_fn=<NllLossBackward>)
[1951] Train Loss:    tensor(1.0579, device='cuda:0', grad_fn=<NllLossBackward>)
[1952] Train Loss:    tensor(1.0207, device='cuda:0', grad_fn=<NllLossBackward>)
[1953] Train Loss:    tensor(1.0577, device='cuda:0', grad_fn=<NllLossBackward>)
[1954] Train Loss:    tensor(1.0639, device='cuda:0', grad_fn=<NllLossBackward>)
[1955] Train Loss:    tensor(1.0870, device='cuda:0', grad_fn=<NllLossBackward>)
[1956] Train Loss:    tensor(1.0557, device='cuda:0', grad_fn=<NllLossBackward>)
[1957] Train Loss:    tensor(1.0360, device='cuda:0', grad_fn=<NllLossBackward>)
[1958] Train Loss:    tensor(1.1086, device='cuda:0', grad_fn=<NllLossBackward>)
[1959] Train Loss:    tensor(1.0620, device='cuda:0', grad_fn=<NllLossBackward>)
[1960] Train Loss:    tensor(1.0642, device='cuda:0', grad_fn=<NllLossBackward>)
[1961] Train Loss:    tensor(1.0748, device='cuda:0', grad_fn=<NllLossBackward>)
[1962] Train Loss:    tensor(1.1193, device='cuda:0', grad_fn=<NllLossBackward>)
[1963] Train Loss:    tensor(1.0764, device='cuda:0', grad_fn=<NllLossBackward>)
[1964] Train Loss:    tensor(1.0581, device='cuda:0', grad_fn=<NllLossBackward>)
[1965] Train Loss:    tensor(1.0282, device='cuda:0', grad_fn=<NllLossBackward>)
[1966] Train Loss:    tensor(1.0555, device='cuda:0', grad_fn=<NllLossBackward>)
[1967] Train Loss:    tensor(1.0807, device='cuda:0', grad_fn=<NllLossBackward>)
[1968] Train Loss:    tensor(1.0566, device='cuda:0', grad_fn=<NllLossBackward>)
[1969] Train Loss:    tensor(1.0724, device='cuda:0', grad_fn=<NllLossBackward>)
[1970] Train Loss:    tensor(1.0644, device='cuda:0', grad_fn=<NllLossBackward>)
[1971] Train Loss:    tensor(1.0711, device='cuda:0', grad_fn=<NllLossBackward>)
[1972] Train Loss:    tensor(1.0636, device='cuda:0', grad_fn=<NllLossBackward>)
[1973] Train Loss:    tensor(1.0670, device='cuda:0', grad_fn=<NllLossBackward>)
[1974] Train Loss:    tensor(1.0864, device='cuda:0', grad_fn=<NllLossBackward>)
[1975] Train Loss:    tensor(1.0506, device='cuda:0', grad_fn=<NllLossBackward>)
[1976] Train Loss:    tensor(1.0522, device='cuda:0', grad_fn=<NllLossBackward>)
[1977] Train Loss:    tensor(1.0198, device='cuda:0', grad_fn=<NllLossBackward>)
[1978] Train Loss:    tensor(1.1080, device='cuda:0', grad_fn=<NllLossBackward>)
[1979] Train Loss:    tensor(1.0604, device='cuda:0', grad_fn=<NllLossBackward>)
[1980] Train Loss:    tensor(1.0526, device='cuda:0', grad_fn=<NllLossBackward>)
[1981] Train Loss:    tensor(1.0260, device='cuda:0', grad_fn=<NllLossBackward>)
[1982] Train Loss:    tensor(1.0263, device='cuda:0', grad_fn=<NllLossBackward>)
[1983] Train Loss:    tensor(1.0598, device='cuda:0', grad_fn=<NllLossBackward>)
[1984] Train Loss:    tensor(1.0721, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1985] Train Loss:    tensor(1.0384, device='cuda:0', grad_fn=<NllLossBackward>)
[1986] Train Loss:    tensor(1.0712, device='cuda:0', grad_fn=<NllLossBackward>)
[1987] Train Loss:    tensor(1.0820, device='cuda:0', grad_fn=<NllLossBackward>)
[1988] Train Loss:    tensor(1.0563, device='cuda:0', grad_fn=<NllLossBackward>)
[1989] Train Loss:    tensor(1.0592, device='cuda:0', grad_fn=<NllLossBackward>)
[1990] Train Loss:    tensor(1.0603, device='cuda:0', grad_fn=<NllLossBackward>)
[1991] Train Loss:    tensor(1.0612, device='cuda:0', grad_fn=<NllLossBackward>)
[1992] Train Loss:    tensor(1.0565, device='cuda:0', grad_fn=<NllLossBackward>)
[1993] Train Loss:    tensor(1.1359, device='cuda:0', grad_fn=<NllLossBackward>)
[1994] Train Loss:    tensor(1.1022, device='cuda:0', grad_fn=<NllLossBackward>)
[1995] Train Loss:    tensor(1.0741, device='cuda:0', grad_fn=<NllLossBackward>)
[1996] Train Loss:    tensor(1.0793, device='cuda:0', grad_fn=<NllLossBackward>)
[1997] Train Loss:    tensor(1.0617, device='cuda:0', grad_fn=<NllLossBackward>)
[1998] Train Loss:    tensor(1.0530, device='cuda:0', grad_fn=<NllLossBackward>)
[1999] Train Loss:    tensor(1.0940, device='cuda:0', grad_fn=<NllLossBackward>)
[2000] Train Loss:    tensor(1.0585, device='cuda:0', grad_fn=<NllLossBackward>)
[2001] Train Loss:    tensor(1.0932, device='cuda:0', grad_fn=<NllLossBackward>)
[2002] Train Loss:    tensor(1.0551, device='cuda:0', grad_fn=<NllLossBackward>)
[2003] Train Loss:    tensor(1.0926, device='cuda:0', grad_fn=<NllLossBackward>)
[2004] Train Loss:    tensor(1.0612, device='cuda:0', grad_fn=<NllLossBackward>)
[2005] Train Loss:    tensor(1.0520, device='cuda:0', grad_fn=<NllLossBackward>)
[2006] Train Loss:    tensor(1.0805, device='cuda:0', grad_fn=<NllLossBackward>)
[2007] Train Loss:    tensor(1.0647, device='cuda:0', grad_fn=<NllLossBackward>)
[2008] Train Loss:    tensor(1.0537, device='cuda:0', grad_fn=<NllLossBackward>)
[2009] Train Loss:    tensor(1.0537, device='cuda:0', grad_fn=<NllLossBackward>)
[2010] Train Loss:    tensor(1.0806, device='cuda:0', grad_fn=<NllLossBackward>)
[2011] Train Loss:    tensor(1.0801, device='cuda:0', grad_fn=<NllLossBackward>)
[2012] Train Loss:    tensor(1.0634, device='cuda:0', grad_fn=<NllLossBackward>)
[2013] Train Loss:    tensor(1.0346, device='cuda:0', grad_fn=<NllLossBackward>)
[2014] Train Loss:    tensor(1.0636, device='cuda:0', grad_fn=<NllLossBackward>)
[2015] Train Loss:    tensor(1.0726, device='cuda:0', grad_fn=<NllLossBackward>)
[2016] Train Loss:    tensor(1.0728, device='cuda:0', grad_fn=<NllLossBackward>)
[2017] Train Loss:    tensor(1.0612, device='cuda:0', grad_fn=<NllLossBackward>)
[2018] Train Loss:    tensor(1.0534, device='cuda:0', grad_fn=<NllLossBackward>)
[2019] Train Loss:    tensor(1.0910, device='cuda:0', grad_fn=<NllLossBackward>)
[2020] Train Loss:    tensor(1.0734, device='cuda:0', grad_fn=<NllLossBackward>)
[2021] Train Loss:    tensor(1.0567, device='cuda:0', grad_fn=<NllLossBackward>)
[2022] Train Loss:    tensor(1.0912, device='cuda:0', grad_fn=<NllLossBackward>)
[2023] Train Loss:    tensor(1.0488, device='cuda:0', grad_fn=<NllLossBackward>)
[2024] Train Loss:    tensor(1.0596, device='cuda:0', grad_fn=<NllLossBackward>)
[2025] Train Loss:    tensor(1.0719, device='cuda:0', grad_fn=<NllLossBackward>)
[2026] Train Loss:    tensor(1.0813, device='cuda:0', grad_fn=<NllLossBackward>)
[2027] Train Loss:    tensor(1.0704, device='cuda:0', grad_fn=<NllLossBackward>)
[2028] Train Loss:    tensor(1.0539, device='cuda:0', grad_fn=<NllLossBackward>)
[2029] Train Loss:    tensor(1.0833, device='cuda:0', grad_fn=<NllLossBackward>)
[2030] Train Loss:    tensor(1.0902, device='cuda:0', grad_fn=<NllLossBackward>)
[2031] Train Loss:    tensor(1.0818, device='cuda:0', grad_fn=<NllLossBackward>)
[2032] Train Loss:    tensor(1.0646, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2033] Train Loss:   tensor(1.0553, device='cuda:0', grad_fn=<NllLossBackward>)
[2034] Train Loss:   tensor(1.0575, device='cuda:0', grad_fn=<NllLossBackward>)
[2035] Train Loss:   tensor(1.0943, device='cuda:0', grad_fn=<NllLossBackward>)
[2036] Train Loss:   tensor(1.0170, device='cuda:0', grad_fn=<NllLossBackward>)
[2037] Train Loss:   tensor(1.0961, device='cuda:0', grad_fn=<NllLossBackward>)
[2038] Train Loss:   tensor(1.0854, device='cuda:0', grad_fn=<NllLossBackward>)
[2039] Train Loss:   tensor(1.0433, device='cuda:0', grad_fn=<NllLossBackward>)
[2040] Train Loss:   tensor(1.0630, device='cuda:0', grad_fn=<NllLossBackward>)
[2041] Train Loss:   tensor(1.0907, device='cuda:0', grad_fn=<NllLossBackward>)
[2042] Train Loss:   tensor(1.0667, device='cuda:0', grad_fn=<NllLossBackward>)
[2043] Train Loss:   tensor(1.0705, device='cuda:0', grad_fn=<NllLossBackward>)
[2044] Train Loss:   tensor(1.0753, device='cuda:0', grad_fn=<NllLossBackward>)
[2045] Train Loss:   tensor(1.0331, device='cuda:0', grad_fn=<NllLossBackward>)
[2046] Train Loss:   tensor(1.0584, device='cuda:0', grad_fn=<NllLossBackward>)
[2047] Train Loss:   tensor(1.0943, device='cuda:0', grad_fn=<NllLossBackward>)
[2048] Train Loss:   tensor(1.0425, device='cuda:0', grad_fn=<NllLossBackward>)
[2049] Train Loss:   tensor(1.0601, device='cuda:0', grad_fn=<NllLossBackward>)
[2050] Train Loss:   tensor(1.1117, device='cuda:0', grad_fn=<NllLossBackward>)
[2051] Train Loss:   tensor(1.0773, device='cuda:0', grad_fn=<NllLossBackward>)
[2052] Train Loss:   tensor(1.0338, device='cuda:0', grad_fn=<NllLossBackward>)
[2053] Train Loss:   tensor(1.1014, device='cuda:0', grad_fn=<NllLossBackward>)
[2054] Train Loss:   tensor(1.0707, device='cuda:0', grad_fn=<NllLossBackward>)
[2055] Train Loss:   tensor(1.1034, device='cuda:0', grad_fn=<NllLossBackward>)
[2056] Train Loss:   tensor(1.0597, device='cuda:0', grad_fn=<NllLossBackward>)
[2057] Train Loss:   tensor(1.0688, device='cuda:0', grad_fn=<NllLossBackward>)
[2058] Train Loss:   tensor(1.0659, device='cuda:0', grad_fn=<NllLossBackward>)
[2059] Train Loss:   tensor(1.0934, device='cuda:0', grad_fn=<NllLossBackward>)
[2060] Train Loss:   tensor(1.0407, device='cuda:0', grad_fn=<NllLossBackward>)
[2061] Train Loss:   tensor(1.0256, device='cuda:0', grad_fn=<NllLossBackward>)
[2062] Train Loss:   tensor(1.1051, device='cuda:0', grad_fn=<NllLossBackward>)
[2063] Train Loss:   tensor(1.0822, device='cuda:0', grad_fn=<NllLossBackward>)
[2064] Train Loss:   tensor(1.0437, device='cuda:0', grad_fn=<NllLossBackward>)
[2065] Train Loss:   tensor(1.1065, device='cuda:0', grad_fn=<NllLossBackward>)
[2066] Train Loss:   tensor(1.0938, device='cuda:0', grad_fn=<NllLossBackward>)
[2067] Train Loss:   tensor(1.0787, device='cuda:0', grad_fn=<NllLossBackward>)
[2068] Train Loss:   tensor(1.0639, device='cuda:0', grad_fn=<NllLossBackward>)
[2069] Train Loss:   tensor(1.0750, device='cuda:0', grad_fn=<NllLossBackward>)
[2070] Train Loss:   tensor(1.0629, device='cuda:0', grad_fn=<NllLossBackward>)
[2071] Train Loss:   tensor(1.0617, device='cuda:0', grad_fn=<NllLossBackward>)
[2072] Train Loss:   tensor(1.0850, device='cuda:0', grad_fn=<NllLossBackward>)
[2073] Train Loss:   tensor(1.0630, device='cuda:0', grad_fn=<NllLossBackward>)
[2074] Train Loss:   tensor(1.0832, device='cuda:0', grad_fn=<NllLossBackward>)
[2075] Train Loss:   tensor(1.0614, device='cuda:0', grad_fn=<NllLossBackward>)
[2076] Train Loss:   tensor(1.0531, device='cuda:0', grad_fn=<NllLossBackward>)
[2077] Train Loss:   tensor(1.0712, device='cuda:0', grad_fn=<NllLossBackward>)
[2078] Train Loss:   tensor(1.0611, device='cuda:0', grad_fn=<NllLossBackward>)
[2079] Train Loss:   tensor(1.0688, device='cuda:0', grad_fn=<NllLossBackward>)
[2080] Train Loss:   tensor(1.0647, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2081] Train Loss:  tensor(1.0933, device='cuda:0', grad_fn=<NllLossBackward>)
[2082] Train Loss:  tensor(1.0044, device='cuda:0', grad_fn=<NllLossBackward>)
[2083] Train Loss:  tensor(1.0693, device='cuda:0', grad_fn=<NllLossBackward>)
[2084] Train Loss:  tensor(1.0378, device='cuda:0', grad_fn=<NllLossBackward>)
[2085] Train Loss:  tensor(1.0723, device='cuda:0', grad_fn=<NllLossBackward>)
[2086] Train Loss:  tensor(1.0625, device='cuda:0', grad_fn=<NllLossBackward>)
[2087] Train Loss:  tensor(1.0419, device='cuda:0', grad_fn=<NllLossBackward>)
[2088] Train Loss:  tensor(1.0582, device='cuda:0', grad_fn=<NllLossBackward>)
[2089] Train Loss:  tensor(1.0265, device='cuda:0', grad_fn=<NllLossBackward>)
[2090] Train Loss:  tensor(1.0209, device='cuda:0', grad_fn=<NllLossBackward>)
[2091] Train Loss:  tensor(1.0422, device='cuda:0', grad_fn=<NllLossBackward>)
[2092] Train Loss:  tensor(1.0518, device='cuda:0', grad_fn=<NllLossBackward>)
[2093] Train Loss:  tensor(1.0420, device='cuda:0', grad_fn=<NllLossBackward>)
[2094] Train Loss:  tensor(1.1121, device='cuda:0', grad_fn=<NllLossBackward>)
[2095] Train Loss:  tensor(1.0219, device='cuda:0', grad_fn=<NllLossBackward>)
[2096] Train Loss:  tensor(1.0663, device='cuda:0', grad_fn=<NllLossBackward>)
[2097] Train Loss:  tensor(1.0595, device='cuda:0', grad_fn=<NllLossBackward>)
[2098] Train Loss:  tensor(1.0663, device='cuda:0', grad_fn=<NllLossBackward>)
[2099] Train Loss:  tensor(1.0698, device='cuda:0', grad_fn=<NllLossBackward>)
[2100] Train Loss:  tensor(1.0654, device='cuda:0', grad_fn=<NllLossBackward>)
[2101] Train Loss:  tensor(1.0544, device='cuda:0', grad_fn=<NllLossBackward>)
[2102] Train Loss:  tensor(1.0444, device='cuda:0', grad_fn=<NllLossBackward>)
[2103] Train Loss:  tensor(1.0554, device='cuda:0', grad_fn=<NllLossBackward>)
[2104] Train Loss:  tensor(1.0878, device='cuda:0', grad_fn=<NllLossBackward>)
[2105] Train Loss:  tensor(1.0261, device='cuda:0', grad_fn=<NllLossBackward>)
[2106] Train Loss:  tensor(1.0551, device='cuda:0', grad_fn=<NllLossBackward>)
[2107] Train Loss:  tensor(1.0626, device='cuda:0', grad_fn=<NllLossBackward>)
[2108] Train Loss:  tensor(1.0525, device='cuda:0', grad_fn=<NllLossBackward>)
[2109] Train Loss:  tensor(1.0857, device='cuda:0', grad_fn=<NllLossBackward>)
[2110] Train Loss:  tensor(1.0436, device='cuda:0', grad_fn=<NllLossBackward>)
[2111] Train Loss:  tensor(1.0840, device='cuda:0', grad_fn=<NllLossBackward>)
[2112] Train Loss:  tensor(1.0658, device='cuda:0', grad_fn=<NllLossBackward>)
[2113] Train Loss:  tensor(1.0512, device='cuda:0', grad_fn=<NllLossBackward>)
[2114] Train Loss:  tensor(1.0198, device='cuda:0', grad_fn=<NllLossBackward>)
[2115] Train Loss:  tensor(1.0639, device='cuda:0', grad_fn=<NllLossBackward>)
[2116] Train Loss:  tensor(1.0737, device='cuda:0', grad_fn=<NllLossBackward>)
[2117] Train Loss:  tensor(1.0170, device='cuda:0', grad_fn=<NllLossBackward>)
[2118] Train Loss:  tensor(1.0519, device='cuda:0', grad_fn=<NllLossBackward>)
[2119] Train Loss:  tensor(1.0548, device='cuda:0', grad_fn=<NllLossBackward>)
[2120] Train Loss:  tensor(1.0658, device='cuda:0', grad_fn=<NllLossBackward>)
[2121] Train Loss:  tensor(1.0560, device='cuda:0', grad_fn=<NllLossBackward>)
[2122] Train Loss:  tensor(1.0814, device='cuda:0', grad_fn=<NllLossBackward>)
[2123] Train Loss:  tensor(1.0482, device='cuda:0', grad_fn=<NllLossBackward>)
[2124] Train Loss:  tensor(1.0301, device='cuda:0', grad_fn=<NllLossBackward>)
[2125] Train Loss:  tensor(1.0354, device='cuda:0', grad_fn=<NllLossBackward>)
[2126] Train Loss:  tensor(1.0722, device='cuda:0', grad_fn=<NllLossBackward>)
[2127] Train Loss:  tensor(1.0369, device='cuda:0', grad_fn=<NllLossBackward>)
[2128] Train Loss:  tensor(1.0629, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2129] Train Loss:    tensor(1.0357, device='cuda:0', grad_fn=<NllLossBackward>)
[2130] Train Loss:    tensor(1.0870, device='cuda:0', grad_fn=<NllLossBackward>)
[2131] Train Loss:    tensor(1.0869, device='cuda:0', grad_fn=<NllLossBackward>)
[2132] Train Loss:    tensor(1.0551, device='cuda:0', grad_fn=<NllLossBackward>)
[2133] Train Loss:    tensor(1.0830, device='cuda:0', grad_fn=<NllLossBackward>)
[2134] Train Loss:    tensor(1.0350, device='cuda:0', grad_fn=<NllLossBackward>)
[2135] Train Loss:    tensor(1.0698, device='cuda:0', grad_fn=<NllLossBackward>)
[2136] Train Loss:    tensor(1.0527, device='cuda:0', grad_fn=<NllLossBackward>)
[2137] Train Loss:    tensor(1.0326, device='cuda:0', grad_fn=<NllLossBackward>)
[2138] Train Loss:    tensor(1.0810, device='cuda:0', grad_fn=<NllLossBackward>)
[2139] Train Loss:    tensor(1.0150, device='cuda:0', grad_fn=<NllLossBackward>)
[2140] Train Loss:    tensor(1.0641, device='cuda:0', grad_fn=<NllLossBackward>)
[2141] Train Loss:    tensor(1.0516, device='cuda:0', grad_fn=<NllLossBackward>)
[2142] Train Loss:    tensor(1.0685, device='cuda:0', grad_fn=<NllLossBackward>)
[2143] Train Loss:    tensor(1.0463, device='cuda:0', grad_fn=<NllLossBackward>)
[2144] Train Loss:    tensor(1.0543, device='cuda:0', grad_fn=<NllLossBackward>)
[2145] Train Loss:    tensor(1.0422, device='cuda:0', grad_fn=<NllLossBackward>)
[2146] Train Loss:    tensor(1.0686, device='cuda:0', grad_fn=<NllLossBackward>)
[2147] Train Loss:    tensor(1.0224, device='cuda:0', grad_fn=<NllLossBackward>)
[2148] Train Loss:    tensor(1.0382, device='cuda:0', grad_fn=<NllLossBackward>)
[2149] Train Loss:    tensor(1.0606, device='cuda:0', grad_fn=<NllLossBackward>)
[2150] Train Loss:    tensor(1.0797, device='cuda:0', grad_fn=<NllLossBackward>)
[2151] Train Loss:    tensor(1.0573, device='cuda:0', grad_fn=<NllLossBackward>)
[2152] Train Loss:    tensor(1.0742, device='cuda:0', grad_fn=<NllLossBackward>)
[2153] Train Loss:    tensor(1.0542, device='cuda:0', grad_fn=<NllLossBackward>)
[2154] Train Loss:    tensor(1.0783, device='cuda:0', grad_fn=<NllLossBackward>)
[2155] Train Loss:    tensor(1.0872, device='cuda:0', grad_fn=<NllLossBackward>)
[2156] Train Loss:    tensor(1.0487, device='cuda:0', grad_fn=<NllLossBackward>)
[2157] Train Loss:    tensor(1.0806, device='cuda:0', grad_fn=<NllLossBackward>)
[2158] Train Loss:    tensor(1.0284, device='cuda:0', grad_fn=<NllLossBackward>)
[2159] Train Loss:    tensor(0.9953, device='cuda:0', grad_fn=<NllLossBackward>)
[2160] Train Loss:    tensor(1.0734, device='cuda:0', grad_fn=<NllLossBackward>)
[2161] Train Loss:    tensor(1.0309, device='cuda:0', grad_fn=<NllLossBackward>)
[2162] Train Loss:    tensor(1.0523, device='cuda:0', grad_fn=<NllLossBackward>)
[2163] Train Loss:    tensor(1.0315, device='cuda:0', grad_fn=<NllLossBackward>)
[2164] Train Loss:    tensor(1.0784, device='cuda:0', grad_fn=<NllLossBackward>)
[2165] Train Loss:    tensor(1.0713, device='cuda:0', grad_fn=<NllLossBackward>)
[2166] Train Loss:    tensor(1.0833, device='cuda:0', grad_fn=<NllLossBackward>)
[2167] Train Loss:    tensor(1.0888, device='cuda:0', grad_fn=<NllLossBackward>)
[2168] Train Loss:    tensor(1.0764, device='cuda:0', grad_fn=<NllLossBackward>)
[2169] Train Loss:    tensor(1.0497, device='cuda:0', grad_fn=<NllLossBackward>)
[2170] Train Loss:    tensor(1.0408, device='cuda:0', grad_fn=<NllLossBackward>)
[2171] Train Loss:    tensor(1.0526, device='cuda:0', grad_fn=<NllLossBackward>)
[2172] Train Loss:    tensor(1.0689, device='cuda:0', grad_fn=<NllLossBackward>)
[2173] Train Loss:    tensor(1.0335, device='cuda:0', grad_fn=<NllLossBackward>)
[2174] Train Loss:    tensor(1.0493, device='cuda:0', grad_fn=<NllLossBackward>)
[2175] Train Loss:    tensor(1.0736, device='cuda:0', grad_fn=<NllLossBackward>)
[2176] Train Loss:    tensor(1.0564, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2177] Train Loss:    tensor(1.0425, device='cuda:0', grad_fn=<NllLossBackward>)
[2178] Train Loss:    tensor(1.0176, device='cuda:0', grad_fn=<NllLossBackward>)
[2179] Train Loss:    tensor(1.0363, device='cuda:0', grad_fn=<NllLossBackward>)
[2180] Train Loss:    tensor(1.0898, device='cuda:0', grad_fn=<NllLossBackward>)
[2181] Train Loss:    tensor(1.0688, device='cuda:0', grad_fn=<NllLossBackward>)
[2182] Train Loss:    tensor(1.0625, device='cuda:0', grad_fn=<NllLossBackward>)
[2183] Train Loss:    tensor(1.0564, device='cuda:0', grad_fn=<NllLossBackward>)
[2184] Train Loss:    tensor(1.0311, device='cuda:0', grad_fn=<NllLossBackward>)
[2185] Train Loss:    tensor(1.0208, device='cuda:0', grad_fn=<NllLossBackward>)
[2186] Train Loss:    tensor(1.0470, device='cuda:0', grad_fn=<NllLossBackward>)
[2187] Train Loss:    tensor(1.0849, device='cuda:0', grad_fn=<NllLossBackward>)
[2188] Train Loss:    tensor(1.0061, device='cuda:0', grad_fn=<NllLossBackward>)
[2189] Train Loss:    tensor(1.0362, device='cuda:0', grad_fn=<NllLossBackward>)
[2190] Train Loss:    tensor(1.0370, device='cuda:0', grad_fn=<NllLossBackward>)
[2191] Train Loss:    tensor(1.0728, device='cuda:0', grad_fn=<NllLossBackward>)
[2192] Train Loss:    tensor(1.0373, device='cuda:0', grad_fn=<NllLossBackward>)
[2193] Train Loss:    tensor(1.0424, device='cuda:0', grad_fn=<NllLossBackward>)
[2194] Train Loss:    tensor(1.0570, device='cuda:0', grad_fn=<NllLossBackward>)
[2195] Train Loss:    tensor(1.0340, device='cuda:0', grad_fn=<NllLossBackward>)
[2196] Train Loss:    tensor(1.0882, device='cuda:0', grad_fn=<NllLossBackward>)
[2197] Train Loss:    tensor(1.0499, device='cuda:0', grad_fn=<NllLossBackward>)
[2198] Train Loss:    tensor(1.0351, device='cuda:0', grad_fn=<NllLossBackward>)
[2199] Train Loss:    tensor(1.0428, device='cuda:0', grad_fn=<NllLossBackward>)
[2200] Train Loss:    tensor(1.0509, device='cuda:0', grad_fn=<NllLossBackward>)
[2201] Train Loss:    tensor(1.0198, device='cuda:0', grad_fn=<NllLossBackward>)
[2202] Train Loss:    tensor(1.0321, device='cuda:0', grad_fn=<NllLossBackward>)
[2203] Train Loss:    tensor(1.0722, device='cuda:0', grad_fn=<NllLossBackward>)
[2204] Train Loss:    tensor(1.0696, device='cuda:0', grad_fn=<NllLossBackward>)
[2205] Train Loss:    tensor(1.0634, device='cuda:0', grad_fn=<NllLossBackward>)
[2206] Train Loss:    tensor(1.0661, device='cuda:0', grad_fn=<NllLossBackward>)
[2207] Train Loss:    tensor(1.0654, device='cuda:0', grad_fn=<NllLossBackward>)
[2208] Train Loss:    tensor(1.0683, device='cuda:0', grad_fn=<NllLossBackward>)
[2209] Train Loss:    tensor(1.0780, device='cuda:0', grad_fn=<NllLossBackward>)
[2210] Train Loss:    tensor(1.0612, device='cuda:0', grad_fn=<NllLossBackward>)
[2211] Train Loss:    tensor(1.0520, device='cuda:0', grad_fn=<NllLossBackward>)
[2212] Train Loss:    tensor(1.0468, device='cuda:0', grad_fn=<NllLossBackward>)
[2213] Train Loss:    tensor(1.0339, device='cuda:0', grad_fn=<NllLossBackward>)
[2214] Train Loss:    tensor(1.0739, device='cuda:0', grad_fn=<NllLossBackward>)
[2215] Train Loss:    tensor(1.0248, device='cuda:0', grad_fn=<NllLossBackward>)
[2216] Train Loss:    tensor(1.0253, device='cuda:0', grad_fn=<NllLossBackward>)
[2217] Train Loss:    tensor(1.0468, device='cuda:0', grad_fn=<NllLossBackward>)
[2218] Train Loss:    tensor(1.0673, device='cuda:0', grad_fn=<NllLossBackward>)
[2219] Train Loss:    tensor(1.0423, device='cuda:0', grad_fn=<NllLossBackward>)
[2220] Train Loss:    tensor(1.0169, device='cuda:0', grad_fn=<NllLossBackward>)
[2221] Train Loss:    tensor(0.9870, device='cuda:0', grad_fn=<NllLossBackward>)
[2222] Train Loss:    tensor(1.0360, device='cuda:0', grad_fn=<NllLossBackward>)
[2223] Train Loss:    tensor(1.0317, device='cuda:0', grad_fn=<NllLossBackward>)
[2224] Train Loss:    tensor(1.0401, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2225] Train Loss:    tensor(1.0939, device='cuda:0', grad_fn=<NllLossBackward>)
[2226] Train Loss:    tensor(1.0557, device='cuda:0', grad_fn=<NllLossBackward>)
[2227] Train Loss:    tensor(1.0719, device='cuda:0', grad_fn=<NllLossBackward>)
[2228] Train Loss:    tensor(1.0564, device='cuda:0', grad_fn=<NllLossBackward>)
[2229] Train Loss:    tensor(1.0792, device='cuda:0', grad_fn=<NllLossBackward>)
[2230] Train Loss:    tensor(1.0608, device='cuda:0', grad_fn=<NllLossBackward>)
[2231] Train Loss:    tensor(1.0269, device='cuda:0', grad_fn=<NllLossBackward>)
[2232] Train Loss:    tensor(1.0235, device='cuda:0', grad_fn=<NllLossBackward>)
[2233] Train Loss:    tensor(1.0249, device='cuda:0', grad_fn=<NllLossBackward>)
[2234] Train Loss:    tensor(1.0542, device='cuda:0', grad_fn=<NllLossBackward>)
[2235] Train Loss:    tensor(1.0515, device='cuda:0', grad_fn=<NllLossBackward>)
[2236] Train Loss:    tensor(1.0598, device='cuda:0', grad_fn=<NllLossBackward>)
[2237] Train Loss:    tensor(1.0447, device='cuda:0', grad_fn=<NllLossBackward>)
[2238] Train Loss:    tensor(1.0692, device='cuda:0', grad_fn=<NllLossBackward>)
[2239] Train Loss:    tensor(1.0528, device='cuda:0', grad_fn=<NllLossBackward>)
[2240] Train Loss:    tensor(1.0347, device='cuda:0', grad_fn=<NllLossBackward>)
[2241] Train Loss:    tensor(1.0499, device='cuda:0', grad_fn=<NllLossBackward>)
[2242] Train Loss:    tensor(1.0063, device='cuda:0', grad_fn=<NllLossBackward>)
[2243] Train Loss:    tensor(1.0154, device='cuda:0', grad_fn=<NllLossBackward>)
[2244] Train Loss:    tensor(1.0305, device='cuda:0', grad_fn=<NllLossBackward>)
[2245] Train Loss:    tensor(1.0277, device='cuda:0', grad_fn=<NllLossBackward>)
[2246] Train Loss:    tensor(1.0514, device='cuda:0', grad_fn=<NllLossBackward>)
[2247] Train Loss:    tensor(1.0501, device='cuda:0', grad_fn=<NllLossBackward>)
[2248] Train Loss:    tensor(1.1327, device='cuda:0', grad_fn=<NllLossBackward>)
[2249] Train Loss:    tensor(1.0446, device='cuda:0', grad_fn=<NllLossBackward>)
[2250] Train Loss:    tensor(1.0679, device='cuda:0', grad_fn=<NllLossBackward>)
[2251] Train Loss:    tensor(1.0365, device='cuda:0', grad_fn=<NllLossBackward>)
[2252] Train Loss:    tensor(1.0719, device='cuda:0', grad_fn=<NllLossBackward>)
[2253] Train Loss:    tensor(1.0625, device='cuda:0', grad_fn=<NllLossBackward>)
[2254] Train Loss:    tensor(1.0520, device='cuda:0', grad_fn=<NllLossBackward>)
[2255] Train Loss:    tensor(1.0060, device='cuda:0', grad_fn=<NllLossBackward>)
[2256] Train Loss:    tensor(1.0500, device='cuda:0', grad_fn=<NllLossBackward>)
[2257] Train Loss:    tensor(1.0219, device='cuda:0', grad_fn=<NllLossBackward>)
[2258] Train Loss:    tensor(1.0562, device='cuda:0', grad_fn=<NllLossBackward>)
[2259] Train Loss:    tensor(1.0311, device='cuda:0', grad_fn=<NllLossBackward>)
[2260] Train Loss:    tensor(1.0292, device='cuda:0', grad_fn=<NllLossBackward>)
[2261] Train Loss:    tensor(1.0600, device='cuda:0', grad_fn=<NllLossBackward>)
[2262] Train Loss:    tensor(1.0349, device='cuda:0', grad_fn=<NllLossBackward>)
[2263] Train Loss:    tensor(1.0411, device='cuda:0', grad_fn=<NllLossBackward>)
[2264] Train Loss:    tensor(1.0356, device='cuda:0', grad_fn=<NllLossBackward>)
[2265] Train Loss:    tensor(1.0544, device='cuda:0', grad_fn=<NllLossBackward>)
[2266] Train Loss:    tensor(1.0286, device='cuda:0', grad_fn=<NllLossBackward>)
[2267] Train Loss:    tensor(1.0479, device='cuda:0', grad_fn=<NllLossBackward>)
[2268] Train Loss:    tensor(1.0374, device='cuda:0', grad_fn=<NllLossBackward>)
[2269] Train Loss:    tensor(1.0026, device='cuda:0', grad_fn=<NllLossBackward>)
[2270] Train Loss:    tensor(1.0610, device='cuda:0', grad_fn=<NllLossBackward>)
[2271] Train Loss:    tensor(1.0104, device='cuda:0', grad_fn=<NllLossBackward>)
[2272] Train Loss:    tensor(1.0341, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2273] Train Loss:   tensor(1.0585, device='cuda:0', grad_fn=<NllLossBackward>)
[2274] Train Loss:   tensor(1.0427, device='cuda:0', grad_fn=<NllLossBackward>)
[2275] Train Loss:   tensor(1.0643, device='cuda:0', grad_fn=<NllLossBackward>)
[2276] Train Loss:   tensor(1.0688, device='cuda:0', grad_fn=<NllLossBackward>)
[2277] Train Loss:   tensor(1.0449, device='cuda:0', grad_fn=<NllLossBackward>)
[2278] Train Loss:   tensor(1.0669, device='cuda:0', grad_fn=<NllLossBackward>)
[2279] Train Loss:   tensor(1.0123, device='cuda:0', grad_fn=<NllLossBackward>)
[2280] Train Loss:   tensor(1.0487, device='cuda:0', grad_fn=<NllLossBackward>)
[2281] Train Loss:   tensor(1.0290, device='cuda:0', grad_fn=<NllLossBackward>)
[2282] Train Loss:   tensor(1.0814, device='cuda:0', grad_fn=<NllLossBackward>)
[2283] Train Loss:   tensor(1.0567, device='cuda:0', grad_fn=<NllLossBackward>)
[2284] Train Loss:   tensor(1.0020, device='cuda:0', grad_fn=<NllLossBackward>)
[2285] Train Loss:   tensor(1.1123, device='cuda:0', grad_fn=<NllLossBackward>)
[2286] Train Loss:   tensor(1.0253, device='cuda:0', grad_fn=<NllLossBackward>)
[2287] Train Loss:   tensor(1.0488, device='cuda:0', grad_fn=<NllLossBackward>)
[2288] Train Loss:   tensor(1.0702, device='cuda:0', grad_fn=<NllLossBackward>)
[2289] Train Loss:   tensor(1.0480, device='cuda:0', grad_fn=<NllLossBackward>)
[2290] Train Loss:   tensor(1.0236, device='cuda:0', grad_fn=<NllLossBackward>)
[2291] Train Loss:   tensor(1.0811, device='cuda:0', grad_fn=<NllLossBackward>)
[2292] Train Loss:   tensor(1.0627, device='cuda:0', grad_fn=<NllLossBackward>)
[2293] Train Loss:   tensor(1.0086, device='cuda:0', grad_fn=<NllLossBackward>)
[2294] Train Loss:   tensor(1.0615, device='cuda:0', grad_fn=<NllLossBackward>)
[2295] Train Loss:   tensor(1.0407, device='cuda:0', grad_fn=<NllLossBackward>)
[2296] Train Loss:   tensor(1.0436, device='cuda:0', grad_fn=<NllLossBackward>)
[2297] Train Loss:   tensor(1.0428, device='cuda:0', grad_fn=<NllLossBackward>)
[2298] Train Loss:   tensor(1.0249, device='cuda:0', grad_fn=<NllLossBackward>)
[2299] Train Loss:   tensor(1.0328, device='cuda:0', grad_fn=<NllLossBackward>)
[2300] Train Loss:   tensor(1.0479, device='cuda:0', grad_fn=<NllLossBackward>)
[2301] Train Loss:   tensor(1.0032, device='cuda:0', grad_fn=<NllLossBackward>)
[2302] Train Loss:   tensor(1.0142, device='cuda:0', grad_fn=<NllLossBackward>)
[2303] Train Loss:   tensor(1.0443, device='cuda:0', grad_fn=<NllLossBackward>)
[2304] Train Loss:   tensor(1.0299, device='cuda:0', grad_fn=<NllLossBackward>)
[2305] Train Loss:   tensor(1.0600, device='cuda:0', grad_fn=<NllLossBackward>)
[2306] Train Loss:   tensor(1.0606, device='cuda:0', grad_fn=<NllLossBackward>)
[2307] Train Loss:   tensor(1.0345, device='cuda:0', grad_fn=<NllLossBackward>)
[2308] Train Loss:   tensor(1.0581, device='cuda:0', grad_fn=<NllLossBackward>)
[2309] Train Loss:   tensor(1.0781, device='cuda:0', grad_fn=<NllLossBackward>)
[2310] Train Loss:   tensor(1.0566, device='cuda:0', grad_fn=<NllLossBackward>)
[2311] Train Loss:   tensor(1.0371, device='cuda:0', grad_fn=<NllLossBackward>)
[2312] Train Loss:   tensor(1.0635, device='cuda:0', grad_fn=<NllLossBackward>)
[2313] Train Loss:   tensor(1.0456, device='cuda:0', grad_fn=<NllLossBackward>)
[2314] Train Loss:   tensor(1.0269, device='cuda:0', grad_fn=<NllLossBackward>)
[2315] Train Loss:   tensor(1.0341, device='cuda:0', grad_fn=<NllLossBackward>)
[2316] Train Loss:   tensor(1.0586, device='cuda:0', grad_fn=<NllLossBackward>)
[2317] Train Loss:   tensor(1.0161, device='cuda:0', grad_fn=<NllLossBackward>)
[2318] Train Loss:   tensor(1.0398, device='cuda:0', grad_fn=<NllLossBackward>)
[2319] Train Loss:   tensor(1.0129, device='cuda:0', grad_fn=<NllLossBackward>)
[2320] Train Loss:   tensor(1.0129, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2321] Train Loss:    tensor(1.0356, device='cuda:0', grad_fn=<NllLossBackward>)
[2322] Train Loss:    tensor(1.0743, device='cuda:0', grad_fn=<NllLossBackward>)
[2323] Train Loss:    tensor(1.0503, device='cuda:0', grad_fn=<NllLossBackward>)
[2324] Train Loss:    tensor(1.0386, device='cuda:0', grad_fn=<NllLossBackward>)
[2325] Train Loss:    tensor(1.0839, device='cuda:0', grad_fn=<NllLossBackward>)
[2326] Train Loss:    tensor(1.0233, device='cuda:0', grad_fn=<NllLossBackward>)
[2327] Train Loss:    tensor(1.0759, device='cuda:0', grad_fn=<NllLossBackward>)
[2328] Train Loss:    tensor(1.0548, device='cuda:0', grad_fn=<NllLossBackward>)
[2329] Train Loss:    tensor(1.0988, device='cuda:0', grad_fn=<NllLossBackward>)
[2330] Train Loss:    tensor(1.0660, device='cuda:0', grad_fn=<NllLossBackward>)
[2331] Train Loss:    tensor(1.0346, device='cuda:0', grad_fn=<NllLossBackward>)
[2332] Train Loss:    tensor(1.0334, device='cuda:0', grad_fn=<NllLossBackward>)
[2333] Train Loss:    tensor(1.0046, device='cuda:0', grad_fn=<NllLossBackward>)
[2334] Train Loss:    tensor(1.0605, device='cuda:0', grad_fn=<NllLossBackward>)
[2335] Train Loss:    tensor(1.0537, device='cuda:0', grad_fn=<NllLossBackward>)
[2336] Train Loss:    tensor(1.0643, device='cuda:0', grad_fn=<NllLossBackward>)
[2337] Train Loss:    tensor(1.0531, device='cuda:0', grad_fn=<NllLossBackward>)
[2338] Train Loss:    tensor(1.0306, device='cuda:0', grad_fn=<NllLossBackward>)
[2339] Train Loss:    tensor(1.0891, device='cuda:0', grad_fn=<NllLossBackward>)
[2340] Train Loss:    tensor(1.0780, device='cuda:0', grad_fn=<NllLossBackward>)
[2341] Train Loss:    tensor(1.0598, device='cuda:0', grad_fn=<NllLossBackward>)
[2342] Train Loss:    tensor(1.0870, device='cuda:0', grad_fn=<NllLossBackward>)
[2343] Train Loss:    tensor(1.0480, device='cuda:0', grad_fn=<NllLossBackward>)
[2344] Train Loss:    tensor(1.0405, device='cuda:0', grad_fn=<NllLossBackward>)
[2345] Train Loss:    tensor(1.0419, device='cuda:0', grad_fn=<NllLossBackward>)
[2346] Train Loss:    tensor(1.0403, device='cuda:0', grad_fn=<NllLossBackward>)
[2347] Train Loss:    tensor(1.0164, device='cuda:0', grad_fn=<NllLossBackward>)
[2348] Train Loss:    tensor(1.0561, device='cuda:0', grad_fn=<NllLossBackward>)
[2349] Train Loss:    tensor(1.0375, device='cuda:0', grad_fn=<NllLossBackward>)
[2350] Train Loss:    tensor(1.0278, device='cuda:0', grad_fn=<NllLossBackward>)
[2351] Train Loss:    tensor(1.0093, device='cuda:0', grad_fn=<NllLossBackward>)
[2352] Train Loss:    tensor(1.0223, device='cuda:0', grad_fn=<NllLossBackward>)
[2353] Train Loss:    tensor(1.0117, device='cuda:0', grad_fn=<NllLossBackward>)
[2354] Train Loss:    tensor(1.0206, device='cuda:0', grad_fn=<NllLossBackward>)
[2355] Train Loss:    tensor(1.0686, device='cuda:0', grad_fn=<NllLossBackward>)
[2356] Train Loss:    tensor(1.0116, device='cuda:0', grad_fn=<NllLossBackward>)
[2357] Train Loss:    tensor(1.0368, device='cuda:0', grad_fn=<NllLossBackward>)
[2358] Train Loss:    tensor(1.0387, device='cuda:0', grad_fn=<NllLossBackward>)
[2359] Train Loss:    tensor(1.0367, device='cuda:0', grad_fn=<NllLossBackward>)
[2360] Train Loss:    tensor(1.0246, device='cuda:0', grad_fn=<NllLossBackward>)
[2361] Train Loss:    tensor(1.0463, device='cuda:0', grad_fn=<NllLossBackward>)
[2362] Train Loss:    tensor(1.0579, device='cuda:0', grad_fn=<NllLossBackward>)
[2363] Train Loss:    tensor(1.0411, device='cuda:0', grad_fn=<NllLossBackward>)
[2364] Train Loss:    tensor(1.0798, device='cuda:0', grad_fn=<NllLossBackward>)
[2365] Train Loss:    tensor(1.0701, device='cuda:0', grad_fn=<NllLossBackward>)
[2366] Train Loss:    tensor(1.0200, device='cuda:0', grad_fn=<NllLossBackward>)
[2367] Train Loss:    tensor(1.0368, device='cuda:0', grad_fn=<NllLossBackward>)
[2368] Train Loss:    tensor(1.0678, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2369] Train Loss:    tensor(1.0607, device='cuda:0', grad_fn=<NllLossBackward>)
[2370] Train Loss:    tensor(1.0655, device='cuda:0', grad_fn=<NllLossBackward>)
[2371] Train Loss:    tensor(1.0253, device='cuda:0', grad_fn=<NllLossBackward>)
[2372] Train Loss:    tensor(1.0206, device='cuda:0', grad_fn=<NllLossBackward>)
[2373] Train Loss:    tensor(1.0829, device='cuda:0', grad_fn=<NllLossBackward>)
[2374] Train Loss:    tensor(1.0396, device='cuda:0', grad_fn=<NllLossBackward>)
[2375] Train Loss:    tensor(1.0059, device='cuda:0', grad_fn=<NllLossBackward>)
[2376] Train Loss:    tensor(1.0217, device='cuda:0', grad_fn=<NllLossBackward>)
[2377] Train Loss:    tensor(0.9919, device='cuda:0', grad_fn=<NllLossBackward>)
[2378] Train Loss:    tensor(1.0770, device='cuda:0', grad_fn=<NllLossBackward>)
[2379] Train Loss:    tensor(1.0354, device='cuda:0', grad_fn=<NllLossBackward>)
[2380] Train Loss:    tensor(1.0656, device='cuda:0', grad_fn=<NllLossBackward>)
[2381] Train Loss:    tensor(1.0544, device='cuda:0', grad_fn=<NllLossBackward>)
[2382] Train Loss:    tensor(1.0557, device='cuda:0', grad_fn=<NllLossBackward>)
[2383] Train Loss:    tensor(1.0244, device='cuda:0', grad_fn=<NllLossBackward>)
[2384] Train Loss:    tensor(1.0205, device='cuda:0', grad_fn=<NllLossBackward>)
[2385] Train Loss:    tensor(1.0551, device='cuda:0', grad_fn=<NllLossBackward>)
[2386] Train Loss:    tensor(1.0401, device='cuda:0', grad_fn=<NllLossBackward>)
[2387] Train Loss:    tensor(1.0219, device='cuda:0', grad_fn=<NllLossBackward>)
[2388] Train Loss:    tensor(1.0318, device='cuda:0', grad_fn=<NllLossBackward>)
[2389] Train Loss:    tensor(1.0455, device='cuda:0', grad_fn=<NllLossBackward>)
[2390] Train Loss:    tensor(1.0460, device='cuda:0', grad_fn=<NllLossBackward>)
[2391] Train Loss:    tensor(0.9961, device='cuda:0', grad_fn=<NllLossBackward>)
[2392] Train Loss:    tensor(1.0614, device='cuda:0', grad_fn=<NllLossBackward>)
[2393] Train Loss:    tensor(0.9879, device='cuda:0', grad_fn=<NllLossBackward>)
[2394] Train Loss:    tensor(1.0329, device='cuda:0', grad_fn=<NllLossBackward>)
[2395] Train Loss:    tensor(1.0256, device='cuda:0', grad_fn=<NllLossBackward>)
[2396] Train Loss:    tensor(1.0512, device='cuda:0', grad_fn=<NllLossBackward>)
[2397] Train Loss:    tensor(1.0178, device='cuda:0', grad_fn=<NllLossBackward>)
[2398] Train Loss:    tensor(1.0799, device='cuda:0', grad_fn=<NllLossBackward>)
[2399] Train Loss:    tensor(1.0089, device='cuda:0', grad_fn=<NllLossBackward>)
[2400] Train Loss:    tensor(1.0229, device='cuda:0', grad_fn=<NllLossBackward>)
[2401] Train Loss:    tensor(0.9888, device='cuda:0', grad_fn=<NllLossBackward>)
[2402] Train Loss:    tensor(1.0217, device='cuda:0', grad_fn=<NllLossBackward>)
[2403] Train Loss:    tensor(1.0647, device='cuda:0', grad_fn=<NllLossBackward>)
[2404] Train Loss:    tensor(1.0494, device='cuda:0', grad_fn=<NllLossBackward>)
[2405] Train Loss:    tensor(1.0398, device='cuda:0', grad_fn=<NllLossBackward>)
[2406] Train Loss:    tensor(1.0381, device='cuda:0', grad_fn=<NllLossBackward>)
[2407] Train Loss:    tensor(1.0437, device='cuda:0', grad_fn=<NllLossBackward>)
[2408] Train Loss:    tensor(1.0378, device='cuda:0', grad_fn=<NllLossBackward>)
[2409] Train Loss:    tensor(1.0552, device='cuda:0', grad_fn=<NllLossBackward>)
[2410] Train Loss:    tensor(1.0194, device='cuda:0', grad_fn=<NllLossBackward>)
[2411] Train Loss:    tensor(1.0506, device='cuda:0', grad_fn=<NllLossBackward>)
[2412] Train Loss:    tensor(1.0337, device='cuda:0', grad_fn=<NllLossBackward>)
[2413] Train Loss:    tensor(1.0410, device='cuda:0', grad_fn=<NllLossBackward>)
[2414] Train Loss:    tensor(1.0010, device='cuda:0', grad_fn=<NllLossBackward>)
[2415] Train Loss:    tensor(1.0022, device='cuda:0', grad_fn=<NllLossBackward>)
[2416] Train Loss:    tensor(1.0161, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2417] Train Loss:    tensor(1.0160, device='cuda:0', grad_fn=<NllLossBackward>)
[2418] Train Loss:    tensor(1.0067, device='cuda:0', grad_fn=<NllLossBackward>)
[2419] Train Loss:    tensor(1.0283, device='cuda:0', grad_fn=<NllLossBackward>)
[2420] Train Loss:    tensor(1.0497, device='cuda:0', grad_fn=<NllLossBackward>)
[2421] Train Loss:    tensor(0.9925, device='cuda:0', grad_fn=<NllLossBackward>)
[2422] Train Loss:    tensor(1.0337, device='cuda:0', grad_fn=<NllLossBackward>)
[2423] Train Loss:    tensor(1.0655, device='cuda:0', grad_fn=<NllLossBackward>)
[2424] Train Loss:    tensor(1.0343, device='cuda:0', grad_fn=<NllLossBackward>)
[2425] Train Loss:    tensor(1.0428, device='cuda:0', grad_fn=<NllLossBackward>)
[2426] Train Loss:    tensor(1.0302, device='cuda:0', grad_fn=<NllLossBackward>)
[2427] Train Loss:    tensor(1.0612, device='cuda:0', grad_fn=<NllLossBackward>)
[2428] Train Loss:    tensor(1.0593, device='cuda:0', grad_fn=<NllLossBackward>)
[2429] Train Loss:    tensor(1.0279, device='cuda:0', grad_fn=<NllLossBackward>)
[2430] Train Loss:    tensor(1.0037, device='cuda:0', grad_fn=<NllLossBackward>)
[2431] Train Loss:    tensor(1.0516, device='cuda:0', grad_fn=<NllLossBackward>)
[2432] Train Loss:    tensor(1.0458, device='cuda:0', grad_fn=<NllLossBackward>)
[2433] Train Loss:    tensor(1.0323, device='cuda:0', grad_fn=<NllLossBackward>)
[2434] Train Loss:    tensor(1.0087, device='cuda:0', grad_fn=<NllLossBackward>)
[2435] Train Loss:    tensor(1.0641, device='cuda:0', grad_fn=<NllLossBackward>)
[2436] Train Loss:    tensor(1.0742, device='cuda:0', grad_fn=<NllLossBackward>)
[2437] Train Loss:    tensor(1.0393, device='cuda:0', grad_fn=<NllLossBackward>)
[2438] Train Loss:    tensor(1.0321, device='cuda:0', grad_fn=<NllLossBackward>)
[2439] Train Loss:    tensor(1.0329, device='cuda:0', grad_fn=<NllLossBackward>)
[2440] Train Loss:    tensor(1.0628, device='cuda:0', grad_fn=<NllLossBackward>)
[2441] Train Loss:    tensor(1.0556, device='cuda:0', grad_fn=<NllLossBackward>)
[2442] Train Loss:    tensor(1.0475, device='cuda:0', grad_fn=<NllLossBackward>)
[2443] Train Loss:    tensor(1.0366, device='cuda:0', grad_fn=<NllLossBackward>)
[2444] Train Loss:    tensor(0.9986, device='cuda:0', grad_fn=<NllLossBackward>)
[2445] Train Loss:    tensor(1.0734, device='cuda:0', grad_fn=<NllLossBackward>)
[2446] Train Loss:    tensor(1.0452, device='cuda:0', grad_fn=<NllLossBackward>)
[2447] Train Loss:    tensor(1.0304, device='cuda:0', grad_fn=<NllLossBackward>)
[2448] Train Loss:    tensor(1.0194, device='cuda:0', grad_fn=<NllLossBackward>)
[2449] Train Loss:    tensor(1.0404, device='cuda:0', grad_fn=<NllLossBackward>)
[2450] Train Loss:    tensor(1.0548, device='cuda:0', grad_fn=<NllLossBackward>)
[2451] Train Loss:    tensor(1.0282, device='cuda:0', grad_fn=<NllLossBackward>)
[2452] Train Loss:    tensor(1.0201, device='cuda:0', grad_fn=<NllLossBackward>)
[2453] Train Loss:    tensor(1.0312, device='cuda:0', grad_fn=<NllLossBackward>)
[2454] Train Loss:    tensor(0.9973, device='cuda:0', grad_fn=<NllLossBackward>)
[2455] Train Loss:    tensor(1.0161, device='cuda:0', grad_fn=<NllLossBackward>)
[2456] Train Loss:    tensor(1.0017, device='cuda:0', grad_fn=<NllLossBackward>)
[2457] Train Loss:    tensor(1.0484, device='cuda:0', grad_fn=<NllLossBackward>)
[2458] Train Loss:    tensor(1.0287, device='cuda:0', grad_fn=<NllLossBackward>)
[2459] Train Loss:    tensor(1.0180, device='cuda:0', grad_fn=<NllLossBackward>)
[2460] Train Loss:    tensor(1.0302, device='cuda:0', grad_fn=<NllLossBackward>)
[2461] Train Loss:    tensor(1.0055, device='cuda:0', grad_fn=<NllLossBackward>)
[2462] Train Loss:    tensor(1.0643, device='cuda:0', grad_fn=<NllLossBackward>)
[2463] Train Loss:    tensor(1.0818, device='cuda:0', grad_fn=<NllLossBackward>)
[2464] Train Loss:    tensor(1.0060, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2465] Train Loss:   tensor(1.0042, device='cuda:0', grad_fn=<NllLossBackward>)
[2466] Train Loss:   tensor(1.0054, device='cuda:0', grad_fn=<NllLossBackward>)
[2467] Train Loss:   tensor(1.0506, device='cuda:0', grad_fn=<NllLossBackward>)
[2468] Train Loss:   tensor(1.0186, device='cuda:0', grad_fn=<NllLossBackward>)
[2469] Train Loss:   tensor(1.0238, device='cuda:0', grad_fn=<NllLossBackward>)
[2470] Train Loss:   tensor(1.0177, device='cuda:0', grad_fn=<NllLossBackward>)
[2471] Train Loss:   tensor(1.0694, device='cuda:0', grad_fn=<NllLossBackward>)
[2472] Train Loss:   tensor(1.0461, device='cuda:0', grad_fn=<NllLossBackward>)
[2473] Train Loss:   tensor(1.0462, device='cuda:0', grad_fn=<NllLossBackward>)
[2474] Train Loss:   tensor(1.0060, device='cuda:0', grad_fn=<NllLossBackward>)
[2475] Train Loss:   tensor(1.0120, device='cuda:0', grad_fn=<NllLossBackward>)
[2476] Train Loss:   tensor(1.0381, device='cuda:0', grad_fn=<NllLossBackward>)
[2477] Train Loss:   tensor(1.0673, device='cuda:0', grad_fn=<NllLossBackward>)
[2478] Train Loss:   tensor(1.0476, device='cuda:0', grad_fn=<NllLossBackward>)
[2479] Train Loss:   tensor(1.0303, device='cuda:0', grad_fn=<NllLossBackward>)
[2480] Train Loss:   tensor(1.0219, device='cuda:0', grad_fn=<NllLossBackward>)
[2481] Train Loss:   tensor(1.0377, device='cuda:0', grad_fn=<NllLossBackward>)
[2482] Train Loss:   tensor(1.0289, device='cuda:0', grad_fn=<NllLossBackward>)
[2483] Train Loss:   tensor(1.0402, device='cuda:0', grad_fn=<NllLossBackward>)
[2484] Train Loss:   tensor(1.0251, device='cuda:0', grad_fn=<NllLossBackward>)
[2485] Train Loss:   tensor(1.0474, device='cuda:0', grad_fn=<NllLossBackward>)
[2486] Train Loss:   tensor(1.0320, device='cuda:0', grad_fn=<NllLossBackward>)
[2487] Train Loss:   tensor(1.0166, device='cuda:0', grad_fn=<NllLossBackward>)
[2488] Train Loss:   tensor(1.0449, device='cuda:0', grad_fn=<NllLossBackward>)
[2489] Train Loss:   tensor(1.0507, device='cuda:0', grad_fn=<NllLossBackward>)
[2490] Train Loss:   tensor(1.0652, device='cuda:0', grad_fn=<NllLossBackward>)
[2491] Train Loss:   tensor(1.0429, device='cuda:0', grad_fn=<NllLossBackward>)
[2492] Train Loss:   tensor(1.0510, device='cuda:0', grad_fn=<NllLossBackward>)
[2493] Train Loss:   tensor(1.0550, device='cuda:0', grad_fn=<NllLossBackward>)
[2494] Train Loss:   tensor(1.0464, device='cuda:0', grad_fn=<NllLossBackward>)
[2495] Train Loss:   tensor(0.9929, device='cuda:0', grad_fn=<NllLossBackward>)
[2496] Train Loss:   tensor(1.0259, device='cuda:0', grad_fn=<NllLossBackward>)
[2497] Train Loss:   tensor(1.0511, device='cuda:0', grad_fn=<NllLossBackward>)
[2498] Train Loss:   tensor(1.0309, device='cuda:0', grad_fn=<NllLossBackward>)
[2499] Train Loss:   tensor(1.0465, device='cuda:0', grad_fn=<NllLossBackward>)
[2500] Train Loss:   tensor(1.0142, device='cuda:0', grad_fn=<NllLossBackward>)
[2501] Train Loss:   tensor(1.0536, device='cuda:0', grad_fn=<NllLossBackward>)
[2502] Train Loss:   tensor(1.0473, device='cuda:0', grad_fn=<NllLossBackward>)
[2503] Train Loss:   tensor(1.0655, device='cuda:0', grad_fn=<NllLossBackward>)
[2504] Train Loss:   tensor(0.9914, device='cuda:0', grad_fn=<NllLossBackward>)
[2505] Train Loss:   tensor(1.0574, device='cuda:0', grad_fn=<NllLossBackward>)
[2506] Train Loss:   tensor(0.9985, device='cuda:0', grad_fn=<NllLossBackward>)
[2507] Train Loss:   tensor(1.0508, device='cuda:0', grad_fn=<NllLossBackward>)
[2508] Train Loss:   tensor(1.0258, device='cuda:0', grad_fn=<NllLossBackward>)
[2509] Train Loss:   tensor(1.0399, device='cuda:0', grad_fn=<NllLossBackward>)
[2510] Train Loss:   tensor(1.0251, device='cuda:0', grad_fn=<NllLossBackward>)
[2511] Train Loss:   tensor(1.0175, device='cuda:0', grad_fn=<NllLossBackward>)
[2512] Train Loss:   tensor(0.9989, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2513] Train Loss:   tensor(0.9971, device='cuda:0', grad_fn=<NllLossBackward>)
[2514] Train Loss:   tensor(1.0579, device='cuda:0', grad_fn=<NllLossBackward>)
[2515] Train Loss:   tensor(1.0404, device='cuda:0', grad_fn=<NllLossBackward>)
[2516] Train Loss:   tensor(1.0372, device='cuda:0', grad_fn=<NllLossBackward>)
[2517] Train Loss:   tensor(1.0505, device='cuda:0', grad_fn=<NllLossBackward>)
[2518] Train Loss:   tensor(1.0510, device='cuda:0', grad_fn=<NllLossBackward>)
[2519] Train Loss:   tensor(1.0164, device='cuda:0', grad_fn=<NllLossBackward>)
[2520] Train Loss:   tensor(1.0524, device='cuda:0', grad_fn=<NllLossBackward>)
[2521] Train Loss:   tensor(1.0207, device='cuda:0', grad_fn=<NllLossBackward>)
[2522] Train Loss:   tensor(0.9978, device='cuda:0', grad_fn=<NllLossBackward>)
[2523] Train Loss:   tensor(1.0815, device='cuda:0', grad_fn=<NllLossBackward>)
[2524] Train Loss:   tensor(1.0078, device='cuda:0', grad_fn=<NllLossBackward>)
[2525] Train Loss:   tensor(1.0195, device='cuda:0', grad_fn=<NllLossBackward>)
[2526] Train Loss:   tensor(1.0096, device='cuda:0', grad_fn=<NllLossBackward>)
[2527] Train Loss:   tensor(1.0308, device='cuda:0', grad_fn=<NllLossBackward>)
[2528] Train Loss:   tensor(1.0233, device='cuda:0', grad_fn=<NllLossBackward>)
[2529] Train Loss:   tensor(0.9996, device='cuda:0', grad_fn=<NllLossBackward>)
[2530] Train Loss:   tensor(1.0256, device='cuda:0', grad_fn=<NllLossBackward>)
[2531] Train Loss:   tensor(1.0455, device='cuda:0', grad_fn=<NllLossBackward>)
[2532] Train Loss:   tensor(1.0119, device='cuda:0', grad_fn=<NllLossBackward>)
[2533] Train Loss:   tensor(1.0416, device='cuda:0', grad_fn=<NllLossBackward>)
[2534] Train Loss:   tensor(1.0006, device='cuda:0', grad_fn=<NllLossBackward>)
[2535] Train Loss:   tensor(1.0473, device='cuda:0', grad_fn=<NllLossBackward>)
[2536] Train Loss:   tensor(1.0412, device='cuda:0', grad_fn=<NllLossBackward>)
[2537] Train Loss:   tensor(1.0386, device='cuda:0', grad_fn=<NllLossBackward>)
[2538] Train Loss:   tensor(1.0247, device='cuda:0', grad_fn=<NllLossBackward>)
[2539] Train Loss:   tensor(1.0216, device='cuda:0', grad_fn=<NllLossBackward>)
[2540] Train Loss:   tensor(1.0100, device='cuda:0', grad_fn=<NllLossBackward>)
[2541] Train Loss:   tensor(1.0436, device='cuda:0', grad_fn=<NllLossBackward>)
[2542] Train Loss:   tensor(1.0093, device='cuda:0', grad_fn=<NllLossBackward>)
[2543] Train Loss:   tensor(1.0191, device='cuda:0', grad_fn=<NllLossBackward>)
[2544] Train Loss:   tensor(1.0544, device='cuda:0', grad_fn=<NllLossBackward>)
[2545] Train Loss:   tensor(1.0522, device='cuda:0', grad_fn=<NllLossBackward>)
[2546] Train Loss:   tensor(1.0464, device='cuda:0', grad_fn=<NllLossBackward>)
[2547] Train Loss:   tensor(1.0168, device='cuda:0', grad_fn=<NllLossBackward>)
[2548] Train Loss:   tensor(1.0395, device='cuda:0', grad_fn=<NllLossBackward>)
[2549] Train Loss:   tensor(1.0433, device='cuda:0', grad_fn=<NllLossBackward>)
[2550] Train Loss:   tensor(1.0433, device='cuda:0', grad_fn=<NllLossBackward>)
[2551] Train Loss:   tensor(1.0149, device='cuda:0', grad_fn=<NllLossBackward>)
[2552] Train Loss:   tensor(1.0243, device='cuda:0', grad_fn=<NllLossBackward>)
[2553] Train Loss:   tensor(1.0651, device='cuda:0', grad_fn=<NllLossBackward>)
[2554] Train Loss:   tensor(1.0440, device='cuda:0', grad_fn=<NllLossBackward>)
[2555] Train Loss:   tensor(1.0370, device='cuda:0', grad_fn=<NllLossBackward>)
[2556] Train Loss:   tensor(1.0290, device='cuda:0', grad_fn=<NllLossBackward>)
[2557] Train Loss:   tensor(1.0401, device='cuda:0', grad_fn=<NllLossBackward>)
[2558] Train Loss:   tensor(1.0434, device='cuda:0', grad_fn=<NllLossBackward>)
[2559] Train Loss:   tensor(1.0395, device='cuda:0', grad_fn=<NllLossBackward>)
[2560] Train Loss:   tensor(1.0084, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2561] Train Loss:    tensor(1.0386, device='cuda:0', grad_fn=<NllLossBackward>)
[2562] Train Loss:    tensor(1.0669, device='cuda:0', grad_fn=<NllLossBackward>)
[2563] Train Loss:    tensor(1.0281, device='cuda:0', grad_fn=<NllLossBackward>)
[2564] Train Loss:    tensor(1.0502, device='cuda:0', grad_fn=<NllLossBackward>)
[2565] Train Loss:    tensor(1.0394, device='cuda:0', grad_fn=<NllLossBackward>)
[2566] Train Loss:    tensor(1.0201, device='cuda:0', grad_fn=<NllLossBackward>)
[2567] Train Loss:    tensor(0.9905, device='cuda:0', grad_fn=<NllLossBackward>)
[2568] Train Loss:    tensor(1.0300, device='cuda:0', grad_fn=<NllLossBackward>)
[2569] Train Loss:    tensor(0.9969, device='cuda:0', grad_fn=<NllLossBackward>)
[2570] Train Loss:    tensor(1.0300, device='cuda:0', grad_fn=<NllLossBackward>)
[2571] Train Loss:    tensor(1.0066, device='cuda:0', grad_fn=<NllLossBackward>)
[2572] Train Loss:    tensor(1.0008, device='cuda:0', grad_fn=<NllLossBackward>)
[2573] Train Loss:    tensor(1.0483, device='cuda:0', grad_fn=<NllLossBackward>)
[2574] Train Loss:    tensor(1.0270, device='cuda:0', grad_fn=<NllLossBackward>)
[2575] Train Loss:    tensor(1.0043, device='cuda:0', grad_fn=<NllLossBackward>)
[2576] Train Loss:    tensor(1.0188, device='cuda:0', grad_fn=<NllLossBackward>)
[2577] Train Loss:    tensor(1.0175, device='cuda:0', grad_fn=<NllLossBackward>)
[2578] Train Loss:    tensor(1.0129, device='cuda:0', grad_fn=<NllLossBackward>)
[2579] Train Loss:    tensor(1.0714, device='cuda:0', grad_fn=<NllLossBackward>)
[2580] Train Loss:    tensor(0.9749, device='cuda:0', grad_fn=<NllLossBackward>)
[2581] Train Loss:    tensor(1.0379, device='cuda:0', grad_fn=<NllLossBackward>)
[2582] Train Loss:    tensor(1.0197, device='cuda:0', grad_fn=<NllLossBackward>)
[2583] Train Loss:    tensor(1.0569, device='cuda:0', grad_fn=<NllLossBackward>)
[2584] Train Loss:    tensor(0.9970, device='cuda:0', grad_fn=<NllLossBackward>)
[2585] Train Loss:    tensor(1.0242, device='cuda:0', grad_fn=<NllLossBackward>)
[2586] Train Loss:    tensor(1.0127, device='cuda:0', grad_fn=<NllLossBackward>)
[2587] Train Loss:    tensor(1.0252, device='cuda:0', grad_fn=<NllLossBackward>)
[2588] Train Loss:    tensor(1.0013, device='cuda:0', grad_fn=<NllLossBackward>)
[2589] Train Loss:    tensor(1.0329, device='cuda:0', grad_fn=<NllLossBackward>)
[2590] Train Loss:    tensor(1.0272, device='cuda:0', grad_fn=<NllLossBackward>)
[2591] Train Loss:    tensor(1.0183, device='cuda:0', grad_fn=<NllLossBackward>)
[2592] Train Loss:    tensor(1.0421, device='cuda:0', grad_fn=<NllLossBackward>)
[2593] Train Loss:    tensor(1.0008, device='cuda:0', grad_fn=<NllLossBackward>)
[2594] Train Loss:    tensor(0.9695, device='cuda:0', grad_fn=<NllLossBackward>)
[2595] Train Loss:    tensor(1.0570, device='cuda:0', grad_fn=<NllLossBackward>)
[2596] Train Loss:    tensor(1.0376, device='cuda:0', grad_fn=<NllLossBackward>)
[2597] Train Loss:    tensor(1.0118, device='cuda:0', grad_fn=<NllLossBackward>)
[2598] Train Loss:    tensor(1.0228, device='cuda:0', grad_fn=<NllLossBackward>)
[2599] Train Loss:    tensor(1.0242, device='cuda:0', grad_fn=<NllLossBackward>)
[2600] Train Loss:    tensor(1.0043, device='cuda:0', grad_fn=<NllLossBackward>)
[2601] Train Loss:    tensor(1.0042, device='cuda:0', grad_fn=<NllLossBackward>)
[2602] Train Loss:    tensor(1.0271, device='cuda:0', grad_fn=<NllLossBackward>)
[2603] Train Loss:    tensor(0.9831, device='cuda:0', grad_fn=<NllLossBackward>)
[2604] Train Loss:    tensor(0.9886, device='cuda:0', grad_fn=<NllLossBackward>)
[2605] Train Loss:    tensor(1.0163, device='cuda:0', grad_fn=<NllLossBackward>)
[2606] Train Loss:    tensor(1.0383, device='cuda:0', grad_fn=<NllLossBackward>)
[2607] Train Loss:    tensor(1.0494, device='cuda:0', grad_fn=<NllLossBackward>)
[2608] Train Loss:    tensor(1.0303, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2609] Train Loss:   tensor(1.0453, device='cuda:0', grad_fn=<NllLossBackward>)
[2610] Train Loss:   tensor(1.0537, device='cuda:0', grad_fn=<NllLossBackward>)
[2611] Train Loss:   tensor(1.0482, device='cuda:0', grad_fn=<NllLossBackward>)
[2612] Train Loss:   tensor(1.0631, device='cuda:0', grad_fn=<NllLossBackward>)
[2613] Train Loss:   tensor(1.0211, device='cuda:0', grad_fn=<NllLossBackward>)
[2614] Train Loss:   tensor(1.0627, device='cuda:0', grad_fn=<NllLossBackward>)
[2615] Train Loss:   tensor(1.0189, device='cuda:0', grad_fn=<NllLossBackward>)
[2616] Train Loss:   tensor(1.0186, device='cuda:0', grad_fn=<NllLossBackward>)
[2617] Train Loss:   tensor(1.0210, device='cuda:0', grad_fn=<NllLossBackward>)
[2618] Train Loss:   tensor(1.1106, device='cuda:0', grad_fn=<NllLossBackward>)
[2619] Train Loss:   tensor(1.0105, device='cuda:0', grad_fn=<NllLossBackward>)
[2620] Train Loss:   tensor(1.0311, device='cuda:0', grad_fn=<NllLossBackward>)
[2621] Train Loss:   tensor(1.0154, device='cuda:0', grad_fn=<NllLossBackward>)
[2622] Train Loss:   tensor(1.0382, device='cuda:0', grad_fn=<NllLossBackward>)
[2623] Train Loss:   tensor(1.0613, device='cuda:0', grad_fn=<NllLossBackward>)
[2624] Train Loss:   tensor(1.0444, device='cuda:0', grad_fn=<NllLossBackward>)
[2625] Train Loss:   tensor(1.0378, device='cuda:0', grad_fn=<NllLossBackward>)
[2626] Train Loss:   tensor(1.0196, device='cuda:0', grad_fn=<NllLossBackward>)
[2627] Train Loss:   tensor(1.0006, device='cuda:0', grad_fn=<NllLossBackward>)
[2628] Train Loss:   tensor(0.9978, device='cuda:0', grad_fn=<NllLossBackward>)
[2629] Train Loss:   tensor(1.0287, device='cuda:0', grad_fn=<NllLossBackward>)
[2630] Train Loss:   tensor(0.9993, device='cuda:0', grad_fn=<NllLossBackward>)
[2631] Train Loss:   tensor(1.0182, device='cuda:0', grad_fn=<NllLossBackward>)
[2632] Train Loss:   tensor(1.0029, device='cuda:0', grad_fn=<NllLossBackward>)
[2633] Train Loss:   tensor(1.0495, device='cuda:0', grad_fn=<NllLossBackward>)
[2634] Train Loss:   tensor(1.0272, device='cuda:0', grad_fn=<NllLossBackward>)
[2635] Train Loss:   tensor(1.0440, device='cuda:0', grad_fn=<NllLossBackward>)
[2636] Train Loss:   tensor(1.0466, device='cuda:0', grad_fn=<NllLossBackward>)
[2637] Train Loss:   tensor(0.9710, device='cuda:0', grad_fn=<NllLossBackward>)
[2638] Train Loss:   tensor(1.0794, device='cuda:0', grad_fn=<NllLossBackward>)
[2639] Train Loss:   tensor(0.9888, device='cuda:0', grad_fn=<NllLossBackward>)
[2640] Train Loss:   tensor(1.0311, device='cuda:0', grad_fn=<NllLossBackward>)
[2641] Train Loss:   tensor(1.0329, device='cuda:0', grad_fn=<NllLossBackward>)
[2642] Train Loss:   tensor(1.0295, device='cuda:0', grad_fn=<NllLossBackward>)
[2643] Train Loss:   tensor(1.0041, device='cuda:0', grad_fn=<NllLossBackward>)
[2644] Train Loss:   tensor(1.0531, device='cuda:0', grad_fn=<NllLossBackward>)
[2645] Train Loss:   tensor(1.0467, device='cuda:0', grad_fn=<NllLossBackward>)
[2646] Train Loss:   tensor(1.0539, device='cuda:0', grad_fn=<NllLossBackward>)
[2647] Train Loss:   tensor(1.0486, device='cuda:0', grad_fn=<NllLossBackward>)
[2648] Train Loss:   tensor(1.0231, device='cuda:0', grad_fn=<NllLossBackward>)
[2649] Train Loss:   tensor(1.0542, device='cuda:0', grad_fn=<NllLossBackward>)
[2650] Train Loss:   tensor(1.0374, device='cuda:0', grad_fn=<NllLossBackward>)
[2651] Train Loss:   tensor(1.0400, device='cuda:0', grad_fn=<NllLossBackward>)
[2652] Train Loss:   tensor(1.0275, device='cuda:0', grad_fn=<NllLossBackward>)
[2653] Train Loss:   tensor(1.0461, device='cuda:0', grad_fn=<NllLossBackward>)
[2654] Train Loss:   tensor(1.0398, device='cuda:0', grad_fn=<NllLossBackward>)
[2655] Train Loss:   tensor(1.0248, device='cuda:0', grad_fn=<NllLossBackward>)
[2656] Train Loss:   tensor(0.9969, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2657] Train Loss:    tensor(0.9970, device='cuda:0', grad_fn=<NllLossBackward>)
[2658] Train Loss:    tensor(1.0117, device='cuda:0', grad_fn=<NllLossBackward>)
[2659] Train Loss:    tensor(1.0698, device='cuda:0', grad_fn=<NllLossBackward>)
[2660] Train Loss:    tensor(1.0040, device='cuda:0', grad_fn=<NllLossBackward>)
[2661] Train Loss:    tensor(1.0045, device='cuda:0', grad_fn=<NllLossBackward>)
[2662] Train Loss:    tensor(0.9699, device='cuda:0', grad_fn=<NllLossBackward>)
[2663] Train Loss:    tensor(1.0121, device='cuda:0', grad_fn=<NllLossBackward>)
[2664] Train Loss:    tensor(1.0141, device='cuda:0', grad_fn=<NllLossBackward>)
[2665] Train Loss:    tensor(1.0265, device='cuda:0', grad_fn=<NllLossBackward>)
[2666] Train Loss:    tensor(1.0127, device='cuda:0', grad_fn=<NllLossBackward>)
[2667] Train Loss:    tensor(1.0217, device='cuda:0', grad_fn=<NllLossBackward>)
[2668] Train Loss:    tensor(1.0293, device='cuda:0', grad_fn=<NllLossBackward>)
[2669] Train Loss:    tensor(1.0123, device='cuda:0', grad_fn=<NllLossBackward>)
[2670] Train Loss:    tensor(1.0587, device='cuda:0', grad_fn=<NllLossBackward>)
[2671] Train Loss:    tensor(1.0168, device='cuda:0', grad_fn=<NllLossBackward>)
[2672] Train Loss:    tensor(1.0100, device='cuda:0', grad_fn=<NllLossBackward>)
[2673] Train Loss:    tensor(1.0229, device='cuda:0', grad_fn=<NllLossBackward>)
[2674] Train Loss:    tensor(1.0486, device='cuda:0', grad_fn=<NllLossBackward>)
[2675] Train Loss:    tensor(1.0202, device='cuda:0', grad_fn=<NllLossBackward>)
[2676] Train Loss:    tensor(1.0284, device='cuda:0', grad_fn=<NllLossBackward>)
[2677] Train Loss:    tensor(1.0076, device='cuda:0', grad_fn=<NllLossBackward>)
[2678] Train Loss:    tensor(1.0030, device='cuda:0', grad_fn=<NllLossBackward>)
[2679] Train Loss:    tensor(1.0347, device='cuda:0', grad_fn=<NllLossBackward>)
[2680] Train Loss:    tensor(1.0519, device='cuda:0', grad_fn=<NllLossBackward>)
[2681] Train Loss:    tensor(1.0125, device='cuda:0', grad_fn=<NllLossBackward>)
[2682] Train Loss:    tensor(1.0468, device='cuda:0', grad_fn=<NllLossBackward>)
[2683] Train Loss:    tensor(1.0198, device='cuda:0', grad_fn=<NllLossBackward>)
[2684] Train Loss:    tensor(1.0133, device='cuda:0', grad_fn=<NllLossBackward>)
[2685] Train Loss:    tensor(1.0255, device='cuda:0', grad_fn=<NllLossBackward>)
[2686] Train Loss:    tensor(1.0247, device='cuda:0', grad_fn=<NllLossBackward>)
[2687] Train Loss:    tensor(1.0320, device='cuda:0', grad_fn=<NllLossBackward>)
[2688] Train Loss:    tensor(1.0026, device='cuda:0', grad_fn=<NllLossBackward>)
[2689] Train Loss:    tensor(1.0310, device='cuda:0', grad_fn=<NllLossBackward>)
[2690] Train Loss:    tensor(1.0350, device='cuda:0', grad_fn=<NllLossBackward>)
[2691] Train Loss:    tensor(1.0263, device='cuda:0', grad_fn=<NllLossBackward>)
[2692] Train Loss:    tensor(1.0183, device='cuda:0', grad_fn=<NllLossBackward>)
[2693] Train Loss:    tensor(1.0306, device='cuda:0', grad_fn=<NllLossBackward>)
[2694] Train Loss:    tensor(1.0064, device='cuda:0', grad_fn=<NllLossBackward>)
[2695] Train Loss:    tensor(0.9883, device='cuda:0', grad_fn=<NllLossBackward>)
[2696] Train Loss:    tensor(1.0296, device='cuda:0', grad_fn=<NllLossBackward>)
[2697] Train Loss:    tensor(0.9828, device='cuda:0', grad_fn=<NllLossBackward>)
[2698] Train Loss:    tensor(1.0211, device='cuda:0', grad_fn=<NllLossBackward>)
[2699] Train Loss:    tensor(1.0530, device='cuda:0', grad_fn=<NllLossBackward>)
[2700] Train Loss:    tensor(1.0196, device='cuda:0', grad_fn=<NllLossBackward>)
[2701] Train Loss:    tensor(1.0019, device='cuda:0', grad_fn=<NllLossBackward>)
[2702] Train Loss:    tensor(1.0188, device='cuda:0', grad_fn=<NllLossBackward>)
[2703] Train Loss:    tensor(1.0324, device='cuda:0', grad_fn=<NllLossBackward>)
[2704] Train Loss:    tensor(1.0168, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2705] Train Loss:   tensor(1.0229, device='cuda:0', grad_fn=<NllLossBackward>)
[2706] Train Loss:   tensor(0.9989, device='cuda:0', grad_fn=<NllLossBackward>)
[2707] Train Loss:   tensor(1.0144, device='cuda:0', grad_fn=<NllLossBackward>)
[2708] Train Loss:   tensor(1.0063, device='cuda:0', grad_fn=<NllLossBackward>)
[2709] Train Loss:   tensor(0.9853, device='cuda:0', grad_fn=<NllLossBackward>)
[2710] Train Loss:   tensor(0.9806, device='cuda:0', grad_fn=<NllLossBackward>)
[2711] Train Loss:   tensor(0.9650, device='cuda:0', grad_fn=<NllLossBackward>)
[2712] Train Loss:   tensor(1.0704, device='cuda:0', grad_fn=<NllLossBackward>)
[2713] Train Loss:   tensor(1.0079, device='cuda:0', grad_fn=<NllLossBackward>)
[2714] Train Loss:   tensor(1.0376, device='cuda:0', grad_fn=<NllLossBackward>)
[2715] Train Loss:   tensor(0.9980, device='cuda:0', grad_fn=<NllLossBackward>)
[2716] Train Loss:   tensor(1.0044, device='cuda:0', grad_fn=<NllLossBackward>)
[2717] Train Loss:   tensor(1.0467, device='cuda:0', grad_fn=<NllLossBackward>)
[2718] Train Loss:   tensor(1.0005, device='cuda:0', grad_fn=<NllLossBackward>)
[2719] Train Loss:   tensor(1.0359, device='cuda:0', grad_fn=<NllLossBackward>)
[2720] Train Loss:   tensor(1.0242, device='cuda:0', grad_fn=<NllLossBackward>)
[2721] Train Loss:   tensor(1.0416, device='cuda:0', grad_fn=<NllLossBackward>)
[2722] Train Loss:   tensor(1.0212, device='cuda:0', grad_fn=<NllLossBackward>)
[2723] Train Loss:   tensor(1.0112, device='cuda:0', grad_fn=<NllLossBackward>)
[2724] Train Loss:   tensor(1.0119, device='cuda:0', grad_fn=<NllLossBackward>)
[2725] Train Loss:   tensor(1.0005, device='cuda:0', grad_fn=<NllLossBackward>)
[2726] Train Loss:   tensor(1.0272, device='cuda:0', grad_fn=<NllLossBackward>)
[2727] Train Loss:   tensor(1.0403, device='cuda:0', grad_fn=<NllLossBackward>)
[2728] Train Loss:   tensor(1.0369, device='cuda:0', grad_fn=<NllLossBackward>)
[2729] Train Loss:   tensor(1.0454, device='cuda:0', grad_fn=<NllLossBackward>)
[2730] Train Loss:   tensor(1.0228, device='cuda:0', grad_fn=<NllLossBackward>)
[2731] Train Loss:   tensor(0.9885, device='cuda:0', grad_fn=<NllLossBackward>)
[2732] Train Loss:   tensor(1.0505, device='cuda:0', grad_fn=<NllLossBackward>)
[2733] Train Loss:   tensor(1.0194, device='cuda:0', grad_fn=<NllLossBackward>)
[2734] Train Loss:   tensor(1.0772, device='cuda:0', grad_fn=<NllLossBackward>)
[2735] Train Loss:   tensor(1.0473, device='cuda:0', grad_fn=<NllLossBackward>)
[2736] Train Loss:   tensor(1.0377, device='cuda:0', grad_fn=<NllLossBackward>)
[2737] Train Loss:   tensor(1.0057, device='cuda:0', grad_fn=<NllLossBackward>)
[2738] Train Loss:   tensor(1.0113, device='cuda:0', grad_fn=<NllLossBackward>)
[2739] Train Loss:   tensor(1.0464, device='cuda:0', grad_fn=<NllLossBackward>)
[2740] Train Loss:   tensor(1.0409, device='cuda:0', grad_fn=<NllLossBackward>)
[2741] Train Loss:   tensor(1.0003, device='cuda:0', grad_fn=<NllLossBackward>)
[2742] Train Loss:   tensor(1.0282, device='cuda:0', grad_fn=<NllLossBackward>)
[2743] Train Loss:   tensor(1.0312, device='cuda:0', grad_fn=<NllLossBackward>)
[2744] Train Loss:   tensor(1.0082, device='cuda:0', grad_fn=<NllLossBackward>)
[2745] Train Loss:   tensor(1.0391, device='cuda:0', grad_fn=<NllLossBackward>)
[2746] Train Loss:   tensor(1.0573, device='cuda:0', grad_fn=<NllLossBackward>)
[2747] Train Loss:   tensor(0.9871, device='cuda:0', grad_fn=<NllLossBackward>)
[2748] Train Loss:   tensor(1.0139, device='cuda:0', grad_fn=<NllLossBackward>)
[2749] Train Loss:   tensor(1.0037, device='cuda:0', grad_fn=<NllLossBackward>)
[2750] Train Loss:   tensor(1.0145, device='cuda:0', grad_fn=<NllLossBackward>)
[2751] Train Loss:   tensor(1.0213, device='cuda:0', grad_fn=<NllLossBackward>)
[2752] Train Loss:   tensor(1.0427, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2753] Train Loss:    tensor(1.0066, device='cuda:0', grad_fn=<NllLossBackward>)
[2754] Train Loss:    tensor(1.0674, device='cuda:0', grad_fn=<NllLossBackward>)
[2755] Train Loss:    tensor(0.9844, device='cuda:0', grad_fn=<NllLossBackward>)
[2756] Train Loss:    tensor(1.0522, device='cuda:0', grad_fn=<NllLossBackward>)
[2757] Train Loss:    tensor(1.0338, device='cuda:0', grad_fn=<NllLossBackward>)
[2758] Train Loss:    tensor(1.0350, device='cuda:0', grad_fn=<NllLossBackward>)
[2759] Train Loss:    tensor(0.9812, device='cuda:0', grad_fn=<NllLossBackward>)
[2760] Train Loss:    tensor(1.0098, device='cuda:0', grad_fn=<NllLossBackward>)
[2761] Train Loss:    tensor(1.0050, device='cuda:0', grad_fn=<NllLossBackward>)
[2762] Train Loss:    tensor(1.0057, device='cuda:0', grad_fn=<NllLossBackward>)
[2763] Train Loss:    tensor(1.0004, device='cuda:0', grad_fn=<NllLossBackward>)
[2764] Train Loss:    tensor(1.0077, device='cuda:0', grad_fn=<NllLossBackward>)
[2765] Train Loss:    tensor(1.0077, device='cuda:0', grad_fn=<NllLossBackward>)
[2766] Train Loss:    tensor(1.0133, device='cuda:0', grad_fn=<NllLossBackward>)
[2767] Train Loss:    tensor(1.0457, device='cuda:0', grad_fn=<NllLossBackward>)
[2768] Train Loss:    tensor(1.0185, device='cuda:0', grad_fn=<NllLossBackward>)
[2769] Train Loss:    tensor(1.0266, device='cuda:0', grad_fn=<NllLossBackward>)
[2770] Train Loss:    tensor(0.9839, device='cuda:0', grad_fn=<NllLossBackward>)
[2771] Train Loss:    tensor(1.0367, device='cuda:0', grad_fn=<NllLossBackward>)
[2772] Train Loss:    tensor(1.0082, device='cuda:0', grad_fn=<NllLossBackward>)
[2773] Train Loss:    tensor(1.0301, device='cuda:0', grad_fn=<NllLossBackward>)
[2774] Train Loss:    tensor(1.0166, device='cuda:0', grad_fn=<NllLossBackward>)
[2775] Train Loss:    tensor(0.9777, device='cuda:0', grad_fn=<NllLossBackward>)
[2776] Train Loss:    tensor(1.0580, device='cuda:0', grad_fn=<NllLossBackward>)
[2777] Train Loss:    tensor(1.0293, device='cuda:0', grad_fn=<NllLossBackward>)
[2778] Train Loss:    tensor(1.0102, device='cuda:0', grad_fn=<NllLossBackward>)
[2779] Train Loss:    tensor(1.0489, device='cuda:0', grad_fn=<NllLossBackward>)
[2780] Train Loss:    tensor(1.0441, device='cuda:0', grad_fn=<NllLossBackward>)
[2781] Train Loss:    tensor(1.0331, device='cuda:0', grad_fn=<NllLossBackward>)
[2782] Train Loss:    tensor(1.0290, device='cuda:0', grad_fn=<NllLossBackward>)
[2783] Train Loss:    tensor(1.0232, device='cuda:0', grad_fn=<NllLossBackward>)
[2784] Train Loss:    tensor(1.0226, device='cuda:0', grad_fn=<NllLossBackward>)
[2785] Train Loss:    tensor(1.0555, device='cuda:0', grad_fn=<NllLossBackward>)
[2786] Train Loss:    tensor(1.0207, device='cuda:0', grad_fn=<NllLossBackward>)
[2787] Train Loss:    tensor(1.0095, device='cuda:0', grad_fn=<NllLossBackward>)
[2788] Train Loss:    tensor(1.0203, device='cuda:0', grad_fn=<NllLossBackward>)
[2789] Train Loss:    tensor(0.9952, device='cuda:0', grad_fn=<NllLossBackward>)
[2790] Train Loss:    tensor(1.0007, device='cuda:0', grad_fn=<NllLossBackward>)
[2791] Train Loss:    tensor(1.0169, device='cuda:0', grad_fn=<NllLossBackward>)
[2792] Train Loss:    tensor(1.0141, device='cuda:0', grad_fn=<NllLossBackward>)
[2793] Train Loss:    tensor(1.0249, device='cuda:0', grad_fn=<NllLossBackward>)
[2794] Train Loss:    tensor(1.0294, device='cuda:0', grad_fn=<NllLossBackward>)
[2795] Train Loss:    tensor(1.0633, device='cuda:0', grad_fn=<NllLossBackward>)
[2796] Train Loss:    tensor(1.0081, device='cuda:0', grad_fn=<NllLossBackward>)
[2797] Train Loss:    tensor(1.0440, device='cuda:0', grad_fn=<NllLossBackward>)
[2798] Train Loss:    tensor(0.9653, device='cuda:0', grad_fn=<NllLossBackward>)
[2799] Train Loss:    tensor(1.0191, device='cuda:0', grad_fn=<NllLossBackward>)
[2800] Train Loss:    tensor(1.0175, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2801] Train Loss:    tensor(1.0138, device='cuda:0', grad_fn=<NllLossBackward>)
[2802] Train Loss:    tensor(0.9814, device='cuda:0', grad_fn=<NllLossBackward>)
[2803] Train Loss:    tensor(1.0134, device='cuda:0', grad_fn=<NllLossBackward>)
[2804] Train Loss:    tensor(1.0180, device='cuda:0', grad_fn=<NllLossBackward>)
[2805] Train Loss:    tensor(1.0317, device='cuda:0', grad_fn=<NllLossBackward>)
[2806] Train Loss:    tensor(1.0164, device='cuda:0', grad_fn=<NllLossBackward>)
[2807] Train Loss:    tensor(1.0247, device='cuda:0', grad_fn=<NllLossBackward>)
[2808] Train Loss:    tensor(1.0058, device='cuda:0', grad_fn=<NllLossBackward>)
[2809] Train Loss:    tensor(0.9727, device='cuda:0', grad_fn=<NllLossBackward>)
[2810] Train Loss:    tensor(1.0308, device='cuda:0', grad_fn=<NllLossBackward>)
[2811] Train Loss:    tensor(1.0055, device='cuda:0', grad_fn=<NllLossBackward>)
[2812] Train Loss:    tensor(1.0221, device='cuda:0', grad_fn=<NllLossBackward>)
[2813] Train Loss:    tensor(0.9914, device='cuda:0', grad_fn=<NllLossBackward>)
[2814] Train Loss:    tensor(1.0106, device='cuda:0', grad_fn=<NllLossBackward>)
[2815] Train Loss:    tensor(1.0113, device='cuda:0', grad_fn=<NllLossBackward>)
[2816] Train Loss:    tensor(1.0204, device='cuda:0', grad_fn=<NllLossBackward>)
[2817] Train Loss:    tensor(1.0324, device='cuda:0', grad_fn=<NllLossBackward>)
[2818] Train Loss:    tensor(1.0303, device='cuda:0', grad_fn=<NllLossBackward>)
[2819] Train Loss:    tensor(1.0367, device='cuda:0', grad_fn=<NllLossBackward>)
[2820] Train Loss:    tensor(0.9864, device='cuda:0', grad_fn=<NllLossBackward>)
[2821] Train Loss:    tensor(0.9974, device='cuda:0', grad_fn=<NllLossBackward>)
[2822] Train Loss:    tensor(1.0116, device='cuda:0', grad_fn=<NllLossBackward>)
[2823] Train Loss:    tensor(1.0238, device='cuda:0', grad_fn=<NllLossBackward>)
[2824] Train Loss:    tensor(1.0317, device='cuda:0', grad_fn=<NllLossBackward>)
[2825] Train Loss:    tensor(1.0186, device='cuda:0', grad_fn=<NllLossBackward>)
[2826] Train Loss:    tensor(1.0242, device='cuda:0', grad_fn=<NllLossBackward>)
[2827] Train Loss:    tensor(1.0155, device='cuda:0', grad_fn=<NllLossBackward>)
[2828] Train Loss:    tensor(0.9873, device='cuda:0', grad_fn=<NllLossBackward>)
[2829] Train Loss:    tensor(1.0288, device='cuda:0', grad_fn=<NllLossBackward>)
[2830] Train Loss:    tensor(0.9867, device='cuda:0', grad_fn=<NllLossBackward>)
[2831] Train Loss:    tensor(1.0076, device='cuda:0', grad_fn=<NllLossBackward>)
[2832] Train Loss:    tensor(1.0087, device='cuda:0', grad_fn=<NllLossBackward>)
[2833] Train Loss:    tensor(1.0026, device='cuda:0', grad_fn=<NllLossBackward>)
[2834] Train Loss:    tensor(1.0193, device='cuda:0', grad_fn=<NllLossBackward>)
[2835] Train Loss:    tensor(1.0235, device='cuda:0', grad_fn=<NllLossBackward>)
[2836] Train Loss:    tensor(1.0004, device='cuda:0', grad_fn=<NllLossBackward>)
[2837] Train Loss:    tensor(1.0084, device='cuda:0', grad_fn=<NllLossBackward>)
[2838] Train Loss:    tensor(1.0016, device='cuda:0', grad_fn=<NllLossBackward>)
[2839] Train Loss:    tensor(0.9731, device='cuda:0', grad_fn=<NllLossBackward>)
[2840] Train Loss:    tensor(0.9981, device='cuda:0', grad_fn=<NllLossBackward>)
[2841] Train Loss:    tensor(0.9893, device='cuda:0', grad_fn=<NllLossBackward>)
[2842] Train Loss:    tensor(1.0190, device='cuda:0', grad_fn=<NllLossBackward>)
[2843] Train Loss:    tensor(0.9854, device='cuda:0', grad_fn=<NllLossBackward>)
[2844] Train Loss:    tensor(0.9980, device='cuda:0', grad_fn=<NllLossBackward>)
[2845] Train Loss:    tensor(1.0032, device='cuda:0', grad_fn=<NllLossBackward>)
[2846] Train Loss:    tensor(0.9988, device='cuda:0', grad_fn=<NllLossBackward>)
[2847] Train Loss:    tensor(0.9962, device='cuda:0', grad_fn=<NllLossBackward>)
[2848] Train Loss:    tensor(1.0072, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2849] Train Loss:    tensor(1.0429, device='cuda:0', grad_fn=<NllLossBackward>)
[2850] Train Loss:    tensor(1.0312, device='cuda:0', grad_fn=<NllLossBackward>)
[2851] Train Loss:    tensor(0.9692, device='cuda:0', grad_fn=<NllLossBackward>)
[2852] Train Loss:    tensor(1.0321, device='cuda:0', grad_fn=<NllLossBackward>)
[2853] Train Loss:    tensor(1.0245, device='cuda:0', grad_fn=<NllLossBackward>)
[2854] Train Loss:    tensor(0.9823, device='cuda:0', grad_fn=<NllLossBackward>)
[2855] Train Loss:    tensor(1.0137, device='cuda:0', grad_fn=<NllLossBackward>)
[2856] Train Loss:    tensor(1.0134, device='cuda:0', grad_fn=<NllLossBackward>)
[2857] Train Loss:    tensor(1.0102, device='cuda:0', grad_fn=<NllLossBackward>)
[2858] Train Loss:    tensor(0.9752, device='cuda:0', grad_fn=<NllLossBackward>)
[2859] Train Loss:    tensor(1.0313, device='cuda:0', grad_fn=<NllLossBackward>)
[2860] Train Loss:    tensor(1.0515, device='cuda:0', grad_fn=<NllLossBackward>)
[2861] Train Loss:    tensor(1.0429, device='cuda:0', grad_fn=<NllLossBackward>)
[2862] Train Loss:    tensor(1.0205, device='cuda:0', grad_fn=<NllLossBackward>)
[2863] Train Loss:    tensor(0.9719, device='cuda:0', grad_fn=<NllLossBackward>)
[2864] Train Loss:    tensor(0.9976, device='cuda:0', grad_fn=<NllLossBackward>)
[2865] Train Loss:    tensor(0.9729, device='cuda:0', grad_fn=<NllLossBackward>)
[2866] Train Loss:    tensor(0.9761, device='cuda:0', grad_fn=<NllLossBackward>)
[2867] Train Loss:    tensor(1.0131, device='cuda:0', grad_fn=<NllLossBackward>)
[2868] Train Loss:    tensor(0.9958, device='cuda:0', grad_fn=<NllLossBackward>)
[2869] Train Loss:    tensor(1.0210, device='cuda:0', grad_fn=<NllLossBackward>)
[2870] Train Loss:    tensor(1.0112, device='cuda:0', grad_fn=<NllLossBackward>)
[2871] Train Loss:    tensor(0.9710, device='cuda:0', grad_fn=<NllLossBackward>)
[2872] Train Loss:    tensor(1.0100, device='cuda:0', grad_fn=<NllLossBackward>)
[2873] Train Loss:    tensor(0.9932, device='cuda:0', grad_fn=<NllLossBackward>)
[2874] Train Loss:    tensor(1.0094, device='cuda:0', grad_fn=<NllLossBackward>)
[2875] Train Loss:    tensor(1.0382, device='cuda:0', grad_fn=<NllLossBackward>)
[2876] Train Loss:    tensor(1.0054, device='cuda:0', grad_fn=<NllLossBackward>)
[2877] Train Loss:    tensor(1.0041, device='cuda:0', grad_fn=<NllLossBackward>)
[2878] Train Loss:    tensor(0.9755, device='cuda:0', grad_fn=<NllLossBackward>)
[2879] Train Loss:    tensor(1.0398, device='cuda:0', grad_fn=<NllLossBackward>)
[2880] Train Loss:    tensor(1.0173, device='cuda:0', grad_fn=<NllLossBackward>)
[2881] Train Loss:    tensor(0.9930, device='cuda:0', grad_fn=<NllLossBackward>)
[2882] Train Loss:    tensor(0.9815, device='cuda:0', grad_fn=<NllLossBackward>)
[2883] Train Loss:    tensor(1.0110, device='cuda:0', grad_fn=<NllLossBackward>)
[2884] Train Loss:    tensor(0.9894, device='cuda:0', grad_fn=<NllLossBackward>)
[2885] Train Loss:    tensor(1.0031, device='cuda:0', grad_fn=<NllLossBackward>)
[2886] Train Loss:    tensor(1.0081, device='cuda:0', grad_fn=<NllLossBackward>)
[2887] Train Loss:    tensor(1.0436, device='cuda:0', grad_fn=<NllLossBackward>)
[2888] Train Loss:    tensor(1.0185, device='cuda:0', grad_fn=<NllLossBackward>)
[2889] Train Loss:    tensor(1.0218, device='cuda:0', grad_fn=<NllLossBackward>)
[2890] Train Loss:    tensor(1.0227, device='cuda:0', grad_fn=<NllLossBackward>)
[2891] Train Loss:    tensor(1.0041, device='cuda:0', grad_fn=<NllLossBackward>)
[2892] Train Loss:    tensor(0.9963, device='cuda:0', grad_fn=<NllLossBackward>)
[2893] Train Loss:    tensor(1.0272, device='cuda:0', grad_fn=<NllLossBackward>)
[2894] Train Loss:    tensor(0.9931, device='cuda:0', grad_fn=<NllLossBackward>)
[2895] Train Loss:    tensor(1.0120, device='cuda:0', grad_fn=<NllLossBackward>)
[2896] Train Loss:    tensor(0.9998, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2897] Train Loss:    tensor(1.0037, device='cuda:0', grad_fn=<NllLossBackward>)
[2898] Train Loss:    tensor(0.9863, device='cuda:0', grad_fn=<NllLossBackward>)
[2899] Train Loss:    tensor(0.9855, device='cuda:0', grad_fn=<NllLossBackward>)
[2900] Train Loss:    tensor(1.0116, device='cuda:0', grad_fn=<NllLossBackward>)
[2901] Train Loss:    tensor(1.0208, device='cuda:0', grad_fn=<NllLossBackward>)
[2902] Train Loss:    tensor(1.0405, device='cuda:0', grad_fn=<NllLossBackward>)
[2903] Train Loss:    tensor(0.9873, device='cuda:0', grad_fn=<NllLossBackward>)
[2904] Train Loss:    tensor(0.9861, device='cuda:0', grad_fn=<NllLossBackward>)
[2905] Train Loss:    tensor(0.9764, device='cuda:0', grad_fn=<NllLossBackward>)
[2906] Train Loss:    tensor(0.9812, device='cuda:0', grad_fn=<NllLossBackward>)
[2907] Train Loss:    tensor(1.0234, device='cuda:0', grad_fn=<NllLossBackward>)
[2908] Train Loss:    tensor(0.9606, device='cuda:0', grad_fn=<NllLossBackward>)
[2909] Train Loss:    tensor(1.0026, device='cuda:0', grad_fn=<NllLossBackward>)
[2910] Train Loss:    tensor(1.0119, device='cuda:0', grad_fn=<NllLossBackward>)
[2911] Train Loss:    tensor(1.0025, device='cuda:0', grad_fn=<NllLossBackward>)
[2912] Train Loss:    tensor(0.9739, device='cuda:0', grad_fn=<NllLossBackward>)
[2913] Train Loss:    tensor(1.0194, device='cuda:0', grad_fn=<NllLossBackward>)
[2914] Train Loss:    tensor(1.0222, device='cuda:0', grad_fn=<NllLossBackward>)
[2915] Train Loss:    tensor(0.9920, device='cuda:0', grad_fn=<NllLossBackward>)
[2916] Train Loss:    tensor(1.0361, device='cuda:0', grad_fn=<NllLossBackward>)
[2917] Train Loss:    tensor(1.0197, device='cuda:0', grad_fn=<NllLossBackward>)
[2918] Train Loss:    tensor(1.0255, device='cuda:0', grad_fn=<NllLossBackward>)
[2919] Train Loss:    tensor(1.0059, device='cuda:0', grad_fn=<NllLossBackward>)
[2920] Train Loss:    tensor(0.9964, device='cuda:0', grad_fn=<NllLossBackward>)
[2921] Train Loss:    tensor(0.9904, device='cuda:0', grad_fn=<NllLossBackward>)
[2922] Train Loss:    tensor(1.0035, device='cuda:0', grad_fn=<NllLossBackward>)
[2923] Train Loss:    tensor(1.0174, device='cuda:0', grad_fn=<NllLossBackward>)
[2924] Train Loss:    tensor(1.0255, device='cuda:0', grad_fn=<NllLossBackward>)
[2925] Train Loss:    tensor(0.9921, device='cuda:0', grad_fn=<NllLossBackward>)
[2926] Train Loss:    tensor(1.0111, device='cuda:0', grad_fn=<NllLossBackward>)
[2927] Train Loss:    tensor(1.0139, device='cuda:0', grad_fn=<NllLossBackward>)
[2928] Train Loss:    tensor(0.9935, device='cuda:0', grad_fn=<NllLossBackward>)
[2929] Train Loss:    tensor(0.9784, device='cuda:0', grad_fn=<NllLossBackward>)
[2930] Train Loss:    tensor(1.0122, device='cuda:0', grad_fn=<NllLossBackward>)
[2931] Train Loss:    tensor(0.9701, device='cuda:0', grad_fn=<NllLossBackward>)
[2932] Train Loss:    tensor(1.0127, device='cuda:0', grad_fn=<NllLossBackward>)
[2933] Train Loss:    tensor(1.0013, device='cuda:0', grad_fn=<NllLossBackward>)
[2934] Train Loss:    tensor(1.0342, device='cuda:0', grad_fn=<NllLossBackward>)
[2935] Train Loss:    tensor(1.0096, device='cuda:0', grad_fn=<NllLossBackward>)
[2936] Train Loss:    tensor(1.0357, device='cuda:0', grad_fn=<NllLossBackward>)
[2937] Train Loss:    tensor(1.0145, device='cuda:0', grad_fn=<NllLossBackward>)
[2938] Train Loss:    tensor(1.0118, device='cuda:0', grad_fn=<NllLossBackward>)
[2939] Train Loss:    tensor(1.0258, device='cuda:0', grad_fn=<NllLossBackward>)
[2940] Train Loss:    tensor(1.0334, device='cuda:0', grad_fn=<NllLossBackward>)
[2941] Train Loss:    tensor(1.0174, device='cuda:0', grad_fn=<NllLossBackward>)
[2942] Train Loss:    tensor(1.0334, device='cuda:0', grad_fn=<NllLossBackward>)
[2943] Train Loss:    tensor(0.9837, device='cuda:0', grad_fn=<NllLossBackward>)
[2944] Train Loss:    tensor(1.0079, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2945] Train Loss:    tensor(0.9703, device='cuda:0', grad_fn=<NllLossBackward>)
[2946] Train Loss:    tensor(0.9976, device='cuda:0', grad_fn=<NllLossBackward>)
[2947] Train Loss:    tensor(0.9922, device='cuda:0', grad_fn=<NllLossBackward>)
[2948] Train Loss:    tensor(1.0166, device='cuda:0', grad_fn=<NllLossBackward>)
[2949] Train Loss:    tensor(1.0383, device='cuda:0', grad_fn=<NllLossBackward>)
[2950] Train Loss:    tensor(0.9923, device='cuda:0', grad_fn=<NllLossBackward>)
[2951] Train Loss:    tensor(0.9798, device='cuda:0', grad_fn=<NllLossBackward>)
[2952] Train Loss:    tensor(1.0147, device='cuda:0', grad_fn=<NllLossBackward>)
[2953] Train Loss:    tensor(1.0264, device='cuda:0', grad_fn=<NllLossBackward>)
[2954] Train Loss:    tensor(1.0114, device='cuda:0', grad_fn=<NllLossBackward>)
[2955] Train Loss:    tensor(1.0130, device='cuda:0', grad_fn=<NllLossBackward>)
[2956] Train Loss:    tensor(1.0399, device='cuda:0', grad_fn=<NllLossBackward>)
[2957] Train Loss:    tensor(0.9922, device='cuda:0', grad_fn=<NllLossBackward>)
[2958] Train Loss:    tensor(1.0441, device='cuda:0', grad_fn=<NllLossBackward>)
[2959] Train Loss:    tensor(1.0387, device='cuda:0', grad_fn=<NllLossBackward>)
[2960] Train Loss:    tensor(1.0695, device='cuda:0', grad_fn=<NllLossBackward>)
[2961] Train Loss:    tensor(0.9941, device='cuda:0', grad_fn=<NllLossBackward>)
[2962] Train Loss:    tensor(1.0091, device='cuda:0', grad_fn=<NllLossBackward>)
[2963] Train Loss:    tensor(0.9926, device='cuda:0', grad_fn=<NllLossBackward>)
[2964] Train Loss:    tensor(0.9687, device='cuda:0', grad_fn=<NllLossBackward>)
[2965] Train Loss:    tensor(0.9849, device='cuda:0', grad_fn=<NllLossBackward>)
[2966] Train Loss:    tensor(1.0161, device='cuda:0', grad_fn=<NllLossBackward>)
[2967] Train Loss:    tensor(1.0106, device='cuda:0', grad_fn=<NllLossBackward>)
[2968] Train Loss:    tensor(0.9999, device='cuda:0', grad_fn=<NllLossBackward>)
[2969] Train Loss:    tensor(1.0030, device='cuda:0', grad_fn=<NllLossBackward>)
[2970] Train Loss:    tensor(1.0127, device='cuda:0', grad_fn=<NllLossBackward>)
[2971] Train Loss:    tensor(1.0279, device='cuda:0', grad_fn=<NllLossBackward>)
[2972] Train Loss:    tensor(1.0123, device='cuda:0', grad_fn=<NllLossBackward>)
[2973] Train Loss:    tensor(1.0216, device='cuda:0', grad_fn=<NllLossBackward>)
[2974] Train Loss:    tensor(1.0501, device='cuda:0', grad_fn=<NllLossBackward>)
[2975] Train Loss:    tensor(1.0144, device='cuda:0', grad_fn=<NllLossBackward>)
[2976] Train Loss:    tensor(0.9933, device='cuda:0', grad_fn=<NllLossBackward>)
[2977] Train Loss:    tensor(1.0136, device='cuda:0', grad_fn=<NllLossBackward>)
[2978] Train Loss:    tensor(1.0014, device='cuda:0', grad_fn=<NllLossBackward>)
[2979] Train Loss:    tensor(1.0139, device='cuda:0', grad_fn=<NllLossBackward>)
[2980] Train Loss:    tensor(1.0609, device='cuda:0', grad_fn=<NllLossBackward>)
[2981] Train Loss:    tensor(0.9797, device='cuda:0', grad_fn=<NllLossBackward>)
[2982] Train Loss:    tensor(1.0119, device='cuda:0', grad_fn=<NllLossBackward>)
[2983] Train Loss:    tensor(0.9858, device='cuda:0', grad_fn=<NllLossBackward>)
[2984] Train Loss:    tensor(0.9874, device='cuda:0', grad_fn=<NllLossBackward>)
[2985] Train Loss:    tensor(0.9873, device='cuda:0', grad_fn=<NllLossBackward>)
[2986] Train Loss:    tensor(1.0100, device='cuda:0', grad_fn=<NllLossBackward>)
[2987] Train Loss:    tensor(0.9827, device='cuda:0', grad_fn=<NllLossBackward>)
[2988] Train Loss:    tensor(0.9824, device='cuda:0', grad_fn=<NllLossBackward>)
[2989] Train Loss:    tensor(0.9837, device='cuda:0', grad_fn=<NllLossBackward>)
[2990] Train Loss:    tensor(1.0089, device='cuda:0', grad_fn=<NllLossBackward>)
[2991] Train Loss:    tensor(0.9904, device='cuda:0', grad_fn=<NllLossBackward>)
[2992] Train Loss:    tensor(0.9730, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2993] Train Loss:   tensor(1.0307, device='cuda:0', grad_fn=<NllLossBackward>)
[2994] Train Loss:   tensor(0.9780, device='cuda:0', grad_fn=<NllLossBackward>)
[2995] Train Loss:   tensor(0.9927, device='cuda:0', grad_fn=<NllLossBackward>)
[2996] Train Loss:   tensor(0.9924, device='cuda:0', grad_fn=<NllLossBackward>)
[2997] Train Loss:   tensor(0.9820, device='cuda:0', grad_fn=<NllLossBackward>)
[2998] Train Loss:   tensor(1.0156, device='cuda:0', grad_fn=<NllLossBackward>)
[2999] Train Loss:   tensor(0.9946, device='cuda:0', grad_fn=<NllLossBackward>)
[3000] Train Loss:   tensor(0.9758, device='cuda:0', grad_fn=<NllLossBackward>)
[3001] Train Loss:   tensor(0.9841, device='cuda:0', grad_fn=<NllLossBackward>)
[3002] Train Loss:   tensor(0.9788, device='cuda:0', grad_fn=<NllLossBackward>)
[3003] Train Loss:   tensor(0.9981, device='cuda:0', grad_fn=<NllLossBackward>)
[3004] Train Loss:   tensor(1.0154, device='cuda:0', grad_fn=<NllLossBackward>)
[3005] Train Loss:   tensor(0.9695, device='cuda:0', grad_fn=<NllLossBackward>)
[3006] Train Loss:   tensor(0.9959, device='cuda:0', grad_fn=<NllLossBackward>)
[3007] Train Loss:   tensor(0.9998, device='cuda:0', grad_fn=<NllLossBackward>)
[3008] Train Loss:   tensor(1.0222, device='cuda:0', grad_fn=<NllLossBackward>)
[3009] Train Loss:   tensor(0.9772, device='cuda:0', grad_fn=<NllLossBackward>)
[3010] Train Loss:   tensor(0.9790, device='cuda:0', grad_fn=<NllLossBackward>)
[3011] Train Loss:   tensor(0.9992, device='cuda:0', grad_fn=<NllLossBackward>)
[3012] Train Loss:   tensor(1.0104, device='cuda:0', grad_fn=<NllLossBackward>)
[3013] Train Loss:   tensor(0.9719, device='cuda:0', grad_fn=<NllLossBackward>)
[3014] Train Loss:   tensor(0.9730, device='cuda:0', grad_fn=<NllLossBackward>)
[3015] Train Loss:   tensor(0.9659, device='cuda:0', grad_fn=<NllLossBackward>)
[3016] Train Loss:   tensor(0.9961, device='cuda:0', grad_fn=<NllLossBackward>)
[3017] Train Loss:   tensor(1.0702, device='cuda:0', grad_fn=<NllLossBackward>)
[3018] Train Loss:   tensor(0.9736, device='cuda:0', grad_fn=<NllLossBackward>)
[3019] Train Loss:   tensor(0.9971, device='cuda:0', grad_fn=<NllLossBackward>)
[3020] Train Loss:   tensor(1.0247, device='cuda:0', grad_fn=<NllLossBackward>)
[3021] Train Loss:   tensor(1.0091, device='cuda:0', grad_fn=<NllLossBackward>)
[3022] Train Loss:   tensor(1.0159, device='cuda:0', grad_fn=<NllLossBackward>)
[3023] Train Loss:   tensor(1.0422, device='cuda:0', grad_fn=<NllLossBackward>)
[3024] Train Loss:   tensor(1.0011, device='cuda:0', grad_fn=<NllLossBackward>)
[3025] Train Loss:   tensor(0.9968, device='cuda:0', grad_fn=<NllLossBackward>)
[3026] Train Loss:   tensor(1.0040, device='cuda:0', grad_fn=<NllLossBackward>)
[3027] Train Loss:   tensor(1.0197, device='cuda:0', grad_fn=<NllLossBackward>)
[3028] Train Loss:   tensor(1.0257, device='cuda:0', grad_fn=<NllLossBackward>)
[3029] Train Loss:   tensor(0.9931, device='cuda:0', grad_fn=<NllLossBackward>)
[3030] Train Loss:   tensor(0.9897, device='cuda:0', grad_fn=<NllLossBackward>)
[3031] Train Loss:   tensor(1.0359, device='cuda:0', grad_fn=<NllLossBackward>)
[3032] Train Loss:   tensor(1.0254, device='cuda:0', grad_fn=<NllLossBackward>)
[3033] Train Loss:   tensor(0.9957, device='cuda:0', grad_fn=<NllLossBackward>)
[3034] Train Loss:   tensor(1.0648, device='cuda:0', grad_fn=<NllLossBackward>)
[3035] Train Loss:   tensor(0.9929, device='cuda:0', grad_fn=<NllLossBackward>)
[3036] Train Loss:   tensor(1.0155, device='cuda:0', grad_fn=<NllLossBackward>)
[3037] Train Loss:   tensor(0.9886, device='cuda:0', grad_fn=<NllLossBackward>)
[3038] Train Loss:   tensor(1.0132, device='cuda:0', grad_fn=<NllLossBackward>)
[3039] Train Loss:   tensor(0.9686, device='cuda:0', grad_fn=<NllLossBackward>)
[3040] Train Loss:   tensor(0.9885, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3041] Train Loss:   tensor(0.9638, device='cuda:0', grad_fn=<NllLossBackward>)
[3042] Train Loss:   tensor(0.9955, device='cuda:0', grad_fn=<NllLossBackward>)
[3043] Train Loss:   tensor(1.0405, device='cuda:0', grad_fn=<NllLossBackward>)
[3044] Train Loss:   tensor(1.0318, device='cuda:0', grad_fn=<NllLossBackward>)
[3045] Train Loss:   tensor(1.0368, device='cuda:0', grad_fn=<NllLossBackward>)
[3046] Train Loss:   tensor(0.9875, device='cuda:0', grad_fn=<NllLossBackward>)
[3047] Train Loss:   tensor(0.9601, device='cuda:0', grad_fn=<NllLossBackward>)
[3048] Train Loss:   tensor(0.9966, device='cuda:0', grad_fn=<NllLossBackward>)
[3049] Train Loss:   tensor(1.0176, device='cuda:0', grad_fn=<NllLossBackward>)
[3050] Train Loss:   tensor(0.9950, device='cuda:0', grad_fn=<NllLossBackward>)
[3051] Train Loss:   tensor(1.0197, device='cuda:0', grad_fn=<NllLossBackward>)
[3052] Train Loss:   tensor(1.0164, device='cuda:0', grad_fn=<NllLossBackward>)
[3053] Train Loss:   tensor(1.0025, device='cuda:0', grad_fn=<NllLossBackward>)
[3054] Train Loss:   tensor(0.9749, device='cuda:0', grad_fn=<NllLossBackward>)
[3055] Train Loss:   tensor(0.9766, device='cuda:0', grad_fn=<NllLossBackward>)
[3056] Train Loss:   tensor(0.9964, device='cuda:0', grad_fn=<NllLossBackward>)
[3057] Train Loss:   tensor(1.0225, device='cuda:0', grad_fn=<NllLossBackward>)
[3058] Train Loss:   tensor(0.9844, device='cuda:0', grad_fn=<NllLossBackward>)
[3059] Train Loss:   tensor(0.9877, device='cuda:0', grad_fn=<NllLossBackward>)
[3060] Train Loss:   tensor(0.9917, device='cuda:0', grad_fn=<NllLossBackward>)
[3061] Train Loss:   tensor(0.9775, device='cuda:0', grad_fn=<NllLossBackward>)
[3062] Train Loss:   tensor(1.0047, device='cuda:0', grad_fn=<NllLossBackward>)
[3063] Train Loss:   tensor(1.0289, device='cuda:0', grad_fn=<NllLossBackward>)
[3064] Train Loss:   tensor(1.0100, device='cuda:0', grad_fn=<NllLossBackward>)
[3065] Train Loss:   tensor(1.0153, device='cuda:0', grad_fn=<NllLossBackward>)
[3066] Train Loss:   tensor(0.9973, device='cuda:0', grad_fn=<NllLossBackward>)
[3067] Train Loss:   tensor(0.9859, device='cuda:0', grad_fn=<NllLossBackward>)
[3068] Train Loss:   tensor(1.0030, device='cuda:0', grad_fn=<NllLossBackward>)
[3069] Train Loss:   tensor(1.0117, device='cuda:0', grad_fn=<NllLossBackward>)
[3070] Train Loss:   tensor(1.0416, device='cuda:0', grad_fn=<NllLossBackward>)
[3071] Train Loss:   tensor(1.0354, device='cuda:0', grad_fn=<NllLossBackward>)
[3072] Train Loss:   tensor(1.0168, device='cuda:0', grad_fn=<NllLossBackward>)
[3073] Train Loss:   tensor(0.9880, device='cuda:0', grad_fn=<NllLossBackward>)
[3074] Train Loss:   tensor(0.9958, device='cuda:0', grad_fn=<NllLossBackward>)
[3075] Train Loss:   tensor(1.0005, device='cuda:0', grad_fn=<NllLossBackward>)
[3076] Train Loss:   tensor(0.9536, device='cuda:0', grad_fn=<NllLossBackward>)
[3077] Train Loss:   tensor(0.9633, device='cuda:0', grad_fn=<NllLossBackward>)
[3078] Train Loss:   tensor(1.0482, device='cuda:0', grad_fn=<NllLossBackward>)
[3079] Train Loss:   tensor(0.9995, device='cuda:0', grad_fn=<NllLossBackward>)
[3080] Train Loss:   tensor(1.0421, device='cuda:0', grad_fn=<NllLossBackward>)
[3081] Train Loss:   tensor(1.0129, device='cuda:0', grad_fn=<NllLossBackward>)
[3082] Train Loss:   tensor(0.9773, device='cuda:0', grad_fn=<NllLossBackward>)
[3083] Train Loss:   tensor(0.9667, device='cuda:0', grad_fn=<NllLossBackward>)
[3084] Train Loss:   tensor(0.9670, device='cuda:0', grad_fn=<NllLossBackward>)
[3085] Train Loss:   tensor(1.0227, device='cuda:0', grad_fn=<NllLossBackward>)
[3086] Train Loss:   tensor(0.9977, device='cuda:0', grad_fn=<NllLossBackward>)
[3087] Train Loss:   tensor(1.0151, device='cuda:0', grad_fn=<NllLossBackward>)
[3088] Train Loss:   tensor(1.0053, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3089] Train Loss:    tensor(0.9909, device='cuda:0', grad_fn=<NllLossBackward>)
[3090] Train Loss:    tensor(1.0173, device='cuda:0', grad_fn=<NllLossBackward>)
[3091] Train Loss:    tensor(0.9952, device='cuda:0', grad_fn=<NllLossBackward>)
[3092] Train Loss:    tensor(0.9739, device='cuda:0', grad_fn=<NllLossBackward>)
[3093] Train Loss:    tensor(1.0053, device='cuda:0', grad_fn=<NllLossBackward>)
[3094] Train Loss:    tensor(1.0318, device='cuda:0', grad_fn=<NllLossBackward>)
[3095] Train Loss:    tensor(0.9934, device='cuda:0', grad_fn=<NllLossBackward>)
[3096] Train Loss:    tensor(0.9752, device='cuda:0', grad_fn=<NllLossBackward>)
[3097] Train Loss:    tensor(1.0258, device='cuda:0', grad_fn=<NllLossBackward>)
[3098] Train Loss:    tensor(1.0084, device='cuda:0', grad_fn=<NllLossBackward>)
[3099] Train Loss:    tensor(1.0004, device='cuda:0', grad_fn=<NllLossBackward>)
[3100] Train Loss:    tensor(0.9711, device='cuda:0', grad_fn=<NllLossBackward>)
[3101] Train Loss:    tensor(0.9733, device='cuda:0', grad_fn=<NllLossBackward>)
[3102] Train Loss:    tensor(1.0017, device='cuda:0', grad_fn=<NllLossBackward>)
[3103] Train Loss:    tensor(1.0076, device='cuda:0', grad_fn=<NllLossBackward>)
[3104] Train Loss:    tensor(0.9662, device='cuda:0', grad_fn=<NllLossBackward>)
[3105] Train Loss:    tensor(1.0053, device='cuda:0', grad_fn=<NllLossBackward>)
[3106] Train Loss:    tensor(1.0005, device='cuda:0', grad_fn=<NllLossBackward>)
[3107] Train Loss:    tensor(0.9964, device='cuda:0', grad_fn=<NllLossBackward>)
[3108] Train Loss:    tensor(1.0085, device='cuda:0', grad_fn=<NllLossBackward>)
[3109] Train Loss:    tensor(1.0268, device='cuda:0', grad_fn=<NllLossBackward>)
[3110] Train Loss:    tensor(0.9975, device='cuda:0', grad_fn=<NllLossBackward>)
[3111] Train Loss:    tensor(0.9932, device='cuda:0', grad_fn=<NllLossBackward>)
[3112] Train Loss:    tensor(1.0209, device='cuda:0', grad_fn=<NllLossBackward>)
[3113] Train Loss:    tensor(1.0235, device='cuda:0', grad_fn=<NllLossBackward>)
[3114] Train Loss:    tensor(0.9944, device='cuda:0', grad_fn=<NllLossBackward>)
[3115] Train Loss:    tensor(1.0435, device='cuda:0', grad_fn=<NllLossBackward>)
[3116] Train Loss:    tensor(1.0214, device='cuda:0', grad_fn=<NllLossBackward>)
[3117] Train Loss:    tensor(1.0430, device='cuda:0', grad_fn=<NllLossBackward>)
[3118] Train Loss:    tensor(0.9687, device='cuda:0', grad_fn=<NllLossBackward>)
[3119] Train Loss:    tensor(1.0021, device='cuda:0', grad_fn=<NllLossBackward>)
[3120] Train Loss:    tensor(0.9727, device='cuda:0', grad_fn=<NllLossBackward>)
[3121] Train Loss:    tensor(1.0252, device='cuda:0', grad_fn=<NllLossBackward>)
[3122] Train Loss:    tensor(1.0209, device='cuda:0', grad_fn=<NllLossBackward>)
[3123] Train Loss:    tensor(1.0069, device='cuda:0', grad_fn=<NllLossBackward>)
[3124] Train Loss:    tensor(1.0029, device='cuda:0', grad_fn=<NllLossBackward>)
[3125] Train Loss:    tensor(1.0047, device='cuda:0', grad_fn=<NllLossBackward>)
[3126] Train Loss:    tensor(0.9878, device='cuda:0', grad_fn=<NllLossBackward>)
[3127] Train Loss:    tensor(1.0148, device='cuda:0', grad_fn=<NllLossBackward>)
[3128] Train Loss:    tensor(0.9908, device='cuda:0', grad_fn=<NllLossBackward>)
[3129] Train Loss:    tensor(1.0398, device='cuda:0', grad_fn=<NllLossBackward>)
[3130] Train Loss:    tensor(1.0211, device='cuda:0', grad_fn=<NllLossBackward>)
[3131] Train Loss:    tensor(0.9938, device='cuda:0', grad_fn=<NllLossBackward>)
[3132] Train Loss:    tensor(0.9872, device='cuda:0', grad_fn=<NllLossBackward>)
[3133] Train Loss:    tensor(1.0333, device='cuda:0', grad_fn=<NllLossBackward>)
[3134] Train Loss:    tensor(0.9999, device='cuda:0', grad_fn=<NllLossBackward>)
[3135] Train Loss:    tensor(1.0130, device='cuda:0', grad_fn=<NllLossBackward>)
[3136] Train Loss:    tensor(0.9868, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3137] Train Loss:    tensor(0.9758, device='cuda:0', grad_fn=<NllLossBackward>)
[3138] Train Loss:    tensor(0.9971, device='cuda:0', grad_fn=<NllLossBackward>)
[3139] Train Loss:    tensor(0.9764, device='cuda:0', grad_fn=<NllLossBackward>)
[3140] Train Loss:    tensor(1.0071, device='cuda:0', grad_fn=<NllLossBackward>)
[3141] Train Loss:    tensor(0.9744, device='cuda:0', grad_fn=<NllLossBackward>)
[3142] Train Loss:    tensor(1.0047, device='cuda:0', grad_fn=<NllLossBackward>)
[3143] Train Loss:    tensor(1.0124, device='cuda:0', grad_fn=<NllLossBackward>)
[3144] Train Loss:    tensor(0.9744, device='cuda:0', grad_fn=<NllLossBackward>)
[3145] Train Loss:    tensor(0.9986, device='cuda:0', grad_fn=<NllLossBackward>)
[3146] Train Loss:    tensor(1.0201, device='cuda:0', grad_fn=<NllLossBackward>)
[3147] Train Loss:    tensor(0.9440, device='cuda:0', grad_fn=<NllLossBackward>)
[3148] Train Loss:    tensor(0.9780, device='cuda:0', grad_fn=<NllLossBackward>)
[3149] Train Loss:    tensor(1.0393, device='cuda:0', grad_fn=<NllLossBackward>)
[3150] Train Loss:    tensor(0.9845, device='cuda:0', grad_fn=<NllLossBackward>)
[3151] Train Loss:    tensor(0.9397, device='cuda:0', grad_fn=<NllLossBackward>)
[3152] Train Loss:    tensor(0.9976, device='cuda:0', grad_fn=<NllLossBackward>)
[3153] Train Loss:    tensor(1.0026, device='cuda:0', grad_fn=<NllLossBackward>)
[3154] Train Loss:    tensor(1.0226, device='cuda:0', grad_fn=<NllLossBackward>)
[3155] Train Loss:    tensor(0.9953, device='cuda:0', grad_fn=<NllLossBackward>)
[3156] Train Loss:    tensor(0.9996, device='cuda:0', grad_fn=<NllLossBackward>)
[3157] Train Loss:    tensor(1.0192, device='cuda:0', grad_fn=<NllLossBackward>)
[3158] Train Loss:    tensor(1.0003, device='cuda:0', grad_fn=<NllLossBackward>)
[3159] Train Loss:    tensor(0.9736, device='cuda:0', grad_fn=<NllLossBackward>)
[3160] Train Loss:    tensor(1.0075, device='cuda:0', grad_fn=<NllLossBackward>)
[3161] Train Loss:    tensor(0.9910, device='cuda:0', grad_fn=<NllLossBackward>)
[3162] Train Loss:    tensor(0.9645, device='cuda:0', grad_fn=<NllLossBackward>)
[3163] Train Loss:    tensor(0.9962, device='cuda:0', grad_fn=<NllLossBackward>)
[3164] Train Loss:    tensor(0.9711, device='cuda:0', grad_fn=<NllLossBackward>)
[3165] Train Loss:    tensor(0.9884, device='cuda:0', grad_fn=<NllLossBackward>)
[3166] Train Loss:    tensor(1.0025, device='cuda:0', grad_fn=<NllLossBackward>)
[3167] Train Loss:    tensor(1.0001, device='cuda:0', grad_fn=<NllLossBackward>)
[3168] Train Loss:    tensor(1.0138, device='cuda:0', grad_fn=<NllLossBackward>)
[3169] Train Loss:    tensor(1.0183, device='cuda:0', grad_fn=<NllLossBackward>)
[3170] Train Loss:    tensor(0.9824, device='cuda:0', grad_fn=<NllLossBackward>)
[3171] Train Loss:    tensor(0.9756, device='cuda:0', grad_fn=<NllLossBackward>)
[3172] Train Loss:    tensor(0.9941, device='cuda:0', grad_fn=<NllLossBackward>)
[3173] Train Loss:    tensor(1.0053, device='cuda:0', grad_fn=<NllLossBackward>)
[3174] Train Loss:    tensor(0.9598, device='cuda:0', grad_fn=<NllLossBackward>)
[3175] Train Loss:    tensor(0.9969, device='cuda:0', grad_fn=<NllLossBackward>)
[3176] Train Loss:    tensor(0.9875, device='cuda:0', grad_fn=<NllLossBackward>)
[3177] Train Loss:    tensor(1.0191, device='cuda:0', grad_fn=<NllLossBackward>)
[3178] Train Loss:    tensor(1.0187, device='cuda:0', grad_fn=<NllLossBackward>)
[3179] Train Loss:    tensor(0.9785, device='cuda:0', grad_fn=<NllLossBackward>)
[3180] Train Loss:    tensor(0.9902, device='cuda:0', grad_fn=<NllLossBackward>)
[3181] Train Loss:    tensor(1.0146, device='cuda:0', grad_fn=<NllLossBackward>)
[3182] Train Loss:    tensor(1.0015, device='cuda:0', grad_fn=<NllLossBackward>)
[3183] Train Loss:    tensor(0.9559, device='cuda:0', grad_fn=<NllLossBackward>)
[3184] Train Loss:    tensor(0.9668, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3185] Train Loss:    tensor(0.9593, device='cuda:0', grad_fn=<NllLossBackward>)
[3186] Train Loss:    tensor(0.9894, device='cuda:0', grad_fn=<NllLossBackward>)
[3187] Train Loss:    tensor(0.9976, device='cuda:0', grad_fn=<NllLossBackward>)
[3188] Train Loss:    tensor(0.9557, device='cuda:0', grad_fn=<NllLossBackward>)
[3189] Train Loss:    tensor(1.0022, device='cuda:0', grad_fn=<NllLossBackward>)
[3190] Train Loss:    tensor(0.9963, device='cuda:0', grad_fn=<NllLossBackward>)
[3191] Train Loss:    tensor(1.0046, device='cuda:0', grad_fn=<NllLossBackward>)
[3192] Train Loss:    tensor(0.9799, device='cuda:0', grad_fn=<NllLossBackward>)
[3193] Train Loss:    tensor(0.9800, device='cuda:0', grad_fn=<NllLossBackward>)
[3194] Train Loss:    tensor(1.0108, device='cuda:0', grad_fn=<NllLossBackward>)
[3195] Train Loss:    tensor(0.9714, device='cuda:0', grad_fn=<NllLossBackward>)
[3196] Train Loss:    tensor(0.9846, device='cuda:0', grad_fn=<NllLossBackward>)
[3197] Train Loss:    tensor(1.0037, device='cuda:0', grad_fn=<NllLossBackward>)
[3198] Train Loss:    tensor(0.9840, device='cuda:0', grad_fn=<NllLossBackward>)
[3199] Train Loss:    tensor(1.0461, device='cuda:0', grad_fn=<NllLossBackward>)
[3200] Train Loss:    tensor(0.9914, device='cuda:0', grad_fn=<NllLossBackward>)
[3201] Train Loss:    tensor(0.9871, device='cuda:0', grad_fn=<NllLossBackward>)
[3202] Train Loss:    tensor(1.0319, device='cuda:0', grad_fn=<NllLossBackward>)
[3203] Train Loss:    tensor(0.9744, device='cuda:0', grad_fn=<NllLossBackward>)
[3204] Train Loss:    tensor(0.9710, device='cuda:0', grad_fn=<NllLossBackward>)
[3205] Train Loss:    tensor(0.9751, device='cuda:0', grad_fn=<NllLossBackward>)
[3206] Train Loss:    tensor(0.9760, device='cuda:0', grad_fn=<NllLossBackward>)
[3207] Train Loss:    tensor(1.0150, device='cuda:0', grad_fn=<NllLossBackward>)
[3208] Train Loss:    tensor(0.9998, device='cuda:0', grad_fn=<NllLossBackward>)
[3209] Train Loss:    tensor(0.9770, device='cuda:0', grad_fn=<NllLossBackward>)
[3210] Train Loss:    tensor(0.9706, device='cuda:0', grad_fn=<NllLossBackward>)
[3211] Train Loss:    tensor(0.9840, device='cuda:0', grad_fn=<NllLossBackward>)
[3212] Train Loss:    tensor(0.9984, device='cuda:0', grad_fn=<NllLossBackward>)
[3213] Train Loss:    tensor(1.0020, device='cuda:0', grad_fn=<NllLossBackward>)
[3214] Train Loss:    tensor(0.9902, device='cuda:0', grad_fn=<NllLossBackward>)
[3215] Train Loss:    tensor(0.9998, device='cuda:0', grad_fn=<NllLossBackward>)
[3216] Train Loss:    tensor(1.0029, device='cuda:0', grad_fn=<NllLossBackward>)
[3217] Train Loss:    tensor(0.9968, device='cuda:0', grad_fn=<NllLossBackward>)
[3218] Train Loss:    tensor(0.9861, device='cuda:0', grad_fn=<NllLossBackward>)
[3219] Train Loss:    tensor(0.9868, device='cuda:0', grad_fn=<NllLossBackward>)
[3220] Train Loss:    tensor(1.0024, device='cuda:0', grad_fn=<NllLossBackward>)
[3221] Train Loss:    tensor(0.9946, device='cuda:0', grad_fn=<NllLossBackward>)
[3222] Train Loss:    tensor(1.0266, device='cuda:0', grad_fn=<NllLossBackward>)
[3223] Train Loss:    tensor(1.0022, device='cuda:0', grad_fn=<NllLossBackward>)
[3224] Train Loss:    tensor(0.9966, device='cuda:0', grad_fn=<NllLossBackward>)
[3225] Train Loss:    tensor(0.9958, device='cuda:0', grad_fn=<NllLossBackward>)
[3226] Train Loss:    tensor(1.0000, device='cuda:0', grad_fn=<NllLossBackward>)
[3227] Train Loss:    tensor(1.0148, device='cuda:0', grad_fn=<NllLossBackward>)
[3228] Train Loss:    tensor(1.0044, device='cuda:0', grad_fn=<NllLossBackward>)
[3229] Train Loss:    tensor(0.9871, device='cuda:0', grad_fn=<NllLossBackward>)
[3230] Train Loss:    tensor(0.9495, device='cuda:0', grad_fn=<NllLossBackward>)
[3231] Train Loss:    tensor(1.0153, device='cuda:0', grad_fn=<NllLossBackward>)
[3232] Train Loss:    tensor(0.9781, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3233] Train Loss:    tensor(0.9934, device='cuda:0', grad_fn=<NllLossBackward>)
[3234] Train Loss:    tensor(0.9854, device='cuda:0', grad_fn=<NllLossBackward>)
[3235] Train Loss:    tensor(1.0097, device='cuda:0', grad_fn=<NllLossBackward>)
[3236] Train Loss:    tensor(0.9928, device='cuda:0', grad_fn=<NllLossBackward>)
[3237] Train Loss:    tensor(0.9657, device='cuda:0', grad_fn=<NllLossBackward>)
[3238] Train Loss:    tensor(0.9889, device='cuda:0', grad_fn=<NllLossBackward>)
[3239] Train Loss:    tensor(1.0106, device='cuda:0', grad_fn=<NllLossBackward>)
[3240] Train Loss:    tensor(0.9614, device='cuda:0', grad_fn=<NllLossBackward>)
[3241] Train Loss:    tensor(0.9949, device='cuda:0', grad_fn=<NllLossBackward>)
[3242] Train Loss:    tensor(1.0128, device='cuda:0', grad_fn=<NllLossBackward>)
[3243] Train Loss:    tensor(0.9886, device='cuda:0', grad_fn=<NllLossBackward>)
[3244] Train Loss:    tensor(0.9545, device='cuda:0', grad_fn=<NllLossBackward>)
[3245] Train Loss:    tensor(0.9861, device='cuda:0', grad_fn=<NllLossBackward>)
[3246] Train Loss:    tensor(0.9828, device='cuda:0', grad_fn=<NllLossBackward>)
[3247] Train Loss:    tensor(0.9688, device='cuda:0', grad_fn=<NllLossBackward>)
[3248] Train Loss:    tensor(1.0067, device='cuda:0', grad_fn=<NllLossBackward>)
[3249] Train Loss:    tensor(0.9937, device='cuda:0', grad_fn=<NllLossBackward>)
[3250] Train Loss:    tensor(0.9925, device='cuda:0', grad_fn=<NllLossBackward>)
[3251] Train Loss:    tensor(0.9588, device='cuda:0', grad_fn=<NllLossBackward>)
[3252] Train Loss:    tensor(0.9953, device='cuda:0', grad_fn=<NllLossBackward>)
[3253] Train Loss:    tensor(0.9716, device='cuda:0', grad_fn=<NllLossBackward>)
[3254] Train Loss:    tensor(0.9754, device='cuda:0', grad_fn=<NllLossBackward>)
[3255] Train Loss:    tensor(1.0161, device='cuda:0', grad_fn=<NllLossBackward>)
[3256] Train Loss:    tensor(0.9860, device='cuda:0', grad_fn=<NllLossBackward>)
[3257] Train Loss:    tensor(0.9737, device='cuda:0', grad_fn=<NllLossBackward>)
[3258] Train Loss:    tensor(1.0363, device='cuda:0', grad_fn=<NllLossBackward>)
[3259] Train Loss:    tensor(0.9984, device='cuda:0', grad_fn=<NllLossBackward>)
[3260] Train Loss:    tensor(0.9646, device='cuda:0', grad_fn=<NllLossBackward>)
[3261] Train Loss:    tensor(0.9764, device='cuda:0', grad_fn=<NllLossBackward>)
[3262] Train Loss:    tensor(0.9786, device='cuda:0', grad_fn=<NllLossBackward>)
[3263] Train Loss:    tensor(1.0164, device='cuda:0', grad_fn=<NllLossBackward>)
[3264] Train Loss:    tensor(1.0091, device='cuda:0', grad_fn=<NllLossBackward>)
[3265] Train Loss:    tensor(0.9990, device='cuda:0', grad_fn=<NllLossBackward>)
[3266] Train Loss:    tensor(1.0147, device='cuda:0', grad_fn=<NllLossBackward>)
[3267] Train Loss:    tensor(1.0191, device='cuda:0', grad_fn=<NllLossBackward>)
[3268] Train Loss:    tensor(0.9836, device='cuda:0', grad_fn=<NllLossBackward>)
[3269] Train Loss:    tensor(1.0008, device='cuda:0', grad_fn=<NllLossBackward>)
[3270] Train Loss:    tensor(0.9772, device='cuda:0', grad_fn=<NllLossBackward>)
[3271] Train Loss:    tensor(1.0299, device='cuda:0', grad_fn=<NllLossBackward>)
[3272] Train Loss:    tensor(1.0035, device='cuda:0', grad_fn=<NllLossBackward>)
[3273] Train Loss:    tensor(1.0081, device='cuda:0', grad_fn=<NllLossBackward>)
[3274] Train Loss:    tensor(0.9699, device='cuda:0', grad_fn=<NllLossBackward>)
[3275] Train Loss:    tensor(1.0148, device='cuda:0', grad_fn=<NllLossBackward>)
[3276] Train Loss:    tensor(0.9771, device='cuda:0', grad_fn=<NllLossBackward>)
[3277] Train Loss:    tensor(0.9922, device='cuda:0', grad_fn=<NllLossBackward>)
[3278] Train Loss:    tensor(1.0222, device='cuda:0', grad_fn=<NllLossBackward>)
[3279] Train Loss:    tensor(0.9616, device='cuda:0', grad_fn=<NllLossBackward>)
[3280] Train Loss:    tensor(0.9928, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3281] Train Loss:    tensor(0.9930, device='cuda:0', grad_fn=<NllLossBackward>)
[3282] Train Loss:    tensor(0.9911, device='cuda:0', grad_fn=<NllLossBackward>)
[3283] Train Loss:    tensor(1.0034, device='cuda:0', grad_fn=<NllLossBackward>)
[3284] Train Loss:    tensor(0.9589, device='cuda:0', grad_fn=<NllLossBackward>)
[3285] Train Loss:    tensor(1.0203, device='cuda:0', grad_fn=<NllLossBackward>)
[3286] Train Loss:    tensor(1.0022, device='cuda:0', grad_fn=<NllLossBackward>)
[3287] Train Loss:    tensor(0.9799, device='cuda:0', grad_fn=<NllLossBackward>)
[3288] Train Loss:    tensor(0.9817, device='cuda:0', grad_fn=<NllLossBackward>)
[3289] Train Loss:    tensor(0.9753, device='cuda:0', grad_fn=<NllLossBackward>)
[3290] Train Loss:    tensor(0.9919, device='cuda:0', grad_fn=<NllLossBackward>)
[3291] Train Loss:    tensor(1.0093, device='cuda:0', grad_fn=<NllLossBackward>)
[3292] Train Loss:    tensor(0.9769, device='cuda:0', grad_fn=<NllLossBackward>)
[3293] Train Loss:    tensor(0.9817, device='cuda:0', grad_fn=<NllLossBackward>)
[3294] Train Loss:    tensor(0.9918, device='cuda:0', grad_fn=<NllLossBackward>)
[3295] Train Loss:    tensor(0.9714, device='cuda:0', grad_fn=<NllLossBackward>)
[3296] Train Loss:    tensor(0.9922, device='cuda:0', grad_fn=<NllLossBackward>)
[3297] Train Loss:    tensor(0.9721, device='cuda:0', grad_fn=<NllLossBackward>)
[3298] Train Loss:    tensor(0.9870, device='cuda:0', grad_fn=<NllLossBackward>)
[3299] Train Loss:    tensor(1.0113, device='cuda:0', grad_fn=<NllLossBackward>)
[3300] Train Loss:    tensor(0.9931, device='cuda:0', grad_fn=<NllLossBackward>)
[3301] Train Loss:    tensor(0.9731, device='cuda:0', grad_fn=<NllLossBackward>)
[3302] Train Loss:    tensor(0.9807, device='cuda:0', grad_fn=<NllLossBackward>)
[3303] Train Loss:    tensor(0.9526, device='cuda:0', grad_fn=<NllLossBackward>)
[3304] Train Loss:    tensor(1.0010, device='cuda:0', grad_fn=<NllLossBackward>)
[3305] Train Loss:    tensor(0.9862, device='cuda:0', grad_fn=<NllLossBackward>)
[3306] Train Loss:    tensor(1.0002, device='cuda:0', grad_fn=<NllLossBackward>)
[3307] Train Loss:    tensor(0.9786, device='cuda:0', grad_fn=<NllLossBackward>)
[3308] Train Loss:    tensor(0.9743, device='cuda:0', grad_fn=<NllLossBackward>)
[3309] Train Loss:    tensor(0.9801, device='cuda:0', grad_fn=<NllLossBackward>)
[3310] Train Loss:    tensor(0.9820, device='cuda:0', grad_fn=<NllLossBackward>)
[3311] Train Loss:    tensor(0.9824, device='cuda:0', grad_fn=<NllLossBackward>)
[3312] Train Loss:    tensor(0.9920, device='cuda:0', grad_fn=<NllLossBackward>)
[3313] Train Loss:    tensor(1.0068, device='cuda:0', grad_fn=<NllLossBackward>)
[3314] Train Loss:    tensor(1.0207, device='cuda:0', grad_fn=<NllLossBackward>)
[3315] Train Loss:    tensor(1.0245, device='cuda:0', grad_fn=<NllLossBackward>)
[3316] Train Loss:    tensor(1.0248, device='cuda:0', grad_fn=<NllLossBackward>)
[3317] Train Loss:    tensor(1.0061, device='cuda:0', grad_fn=<NllLossBackward>)
[3318] Train Loss:    tensor(1.0097, device='cuda:0', grad_fn=<NllLossBackward>)
[3319] Train Loss:    tensor(0.9507, device='cuda:0', grad_fn=<NllLossBackward>)
[3320] Train Loss:    tensor(0.9928, device='cuda:0', grad_fn=<NllLossBackward>)
[3321] Train Loss:    tensor(0.9777, device='cuda:0', grad_fn=<NllLossBackward>)
[3322] Train Loss:    tensor(1.0041, device='cuda:0', grad_fn=<NllLossBackward>)
[3323] Train Loss:    tensor(0.9923, device='cuda:0', grad_fn=<NllLossBackward>)
[3324] Train Loss:    tensor(0.9837, device='cuda:0', grad_fn=<NllLossBackward>)
[3325] Train Loss:    tensor(0.9928, device='cuda:0', grad_fn=<NllLossBackward>)
[3326] Train Loss:    tensor(0.9984, device='cuda:0', grad_fn=<NllLossBackward>)
[3327] Train Loss:    tensor(1.0143, device='cuda:0', grad_fn=<NllLossBackward>)
[3328] Train Loss:    tensor(0.9745, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3329] Train Loss:    tensor(0.9877, device='cuda:0', grad_fn=<NllLossBackward>)
[3330] Train Loss:    tensor(0.9949, device='cuda:0', grad_fn=<NllLossBackward>)
[3331] Train Loss:    tensor(0.9750, device='cuda:0', grad_fn=<NllLossBackward>)
[3332] Train Loss:    tensor(1.0016, device='cuda:0', grad_fn=<NllLossBackward>)
[3333] Train Loss:    tensor(0.9419, device='cuda:0', grad_fn=<NllLossBackward>)
[3334] Train Loss:    tensor(0.9995, device='cuda:0', grad_fn=<NllLossBackward>)
[3335] Train Loss:    tensor(0.9625, device='cuda:0', grad_fn=<NllLossBackward>)
[3336] Train Loss:    tensor(0.9629, device='cuda:0', grad_fn=<NllLossBackward>)
[3337] Train Loss:    tensor(1.0152, device='cuda:0', grad_fn=<NllLossBackward>)
[3338] Train Loss:    tensor(0.9730, device='cuda:0', grad_fn=<NllLossBackward>)
[3339] Train Loss:    tensor(0.9914, device='cuda:0', grad_fn=<NllLossBackward>)
[3340] Train Loss:    tensor(1.0072, device='cuda:0', grad_fn=<NllLossBackward>)
[3341] Train Loss:    tensor(0.9938, device='cuda:0', grad_fn=<NllLossBackward>)
[3342] Train Loss:    tensor(0.9879, device='cuda:0', grad_fn=<NllLossBackward>)
[3343] Train Loss:    tensor(1.0072, device='cuda:0', grad_fn=<NllLossBackward>)
[3344] Train Loss:    tensor(0.9545, device='cuda:0', grad_fn=<NllLossBackward>)
[3345] Train Loss:    tensor(1.0125, device='cuda:0', grad_fn=<NllLossBackward>)
[3346] Train Loss:    tensor(0.9865, device='cuda:0', grad_fn=<NllLossBackward>)
[3347] Train Loss:    tensor(0.9563, device='cuda:0', grad_fn=<NllLossBackward>)
[3348] Train Loss:    tensor(0.9675, device='cuda:0', grad_fn=<NllLossBackward>)
[3349] Train Loss:    tensor(1.0030, device='cuda:0', grad_fn=<NllLossBackward>)
[3350] Train Loss:    tensor(0.9882, device='cuda:0', grad_fn=<NllLossBackward>)
[3351] Train Loss:    tensor(1.0132, device='cuda:0', grad_fn=<NllLossBackward>)
[3352] Train Loss:    tensor(0.9752, device='cuda:0', grad_fn=<NllLossBackward>)
[3353] Train Loss:    tensor(0.9655, device='cuda:0', grad_fn=<NllLossBackward>)
[3354] Train Loss:    tensor(1.0063, device='cuda:0', grad_fn=<NllLossBackward>)
[3355] Train Loss:    tensor(0.9821, device='cuda:0', grad_fn=<NllLossBackward>)
[3356] Train Loss:    tensor(0.9902, device='cuda:0', grad_fn=<NllLossBackward>)
[3357] Train Loss:    tensor(0.9973, device='cuda:0', grad_fn=<NllLossBackward>)
[3358] Train Loss:    tensor(1.0622, device='cuda:0', grad_fn=<NllLossBackward>)
[3359] Train Loss:    tensor(0.9810, device='cuda:0', grad_fn=<NllLossBackward>)
[3360] Train Loss:    tensor(0.9736, device='cuda:0', grad_fn=<NllLossBackward>)
[3361] Train Loss:    tensor(0.9734, device='cuda:0', grad_fn=<NllLossBackward>)
[3362] Train Loss:    tensor(0.9738, device='cuda:0', grad_fn=<NllLossBackward>)
[3363] Train Loss:    tensor(0.9842, device='cuda:0', grad_fn=<NllLossBackward>)
[3364] Train Loss:    tensor(0.9998, device='cuda:0', grad_fn=<NllLossBackward>)
[3365] Train Loss:    tensor(0.9763, device='cuda:0', grad_fn=<NllLossBackward>)
[3366] Train Loss:    tensor(0.9584, device='cuda:0', grad_fn=<NllLossBackward>)
[3367] Train Loss:    tensor(0.9865, device='cuda:0', grad_fn=<NllLossBackward>)
[3368] Train Loss:    tensor(0.9868, device='cuda:0', grad_fn=<NllLossBackward>)
[3369] Train Loss:    tensor(0.9814, device='cuda:0', grad_fn=<NllLossBackward>)
[3370] Train Loss:    tensor(1.0023, device='cuda:0', grad_fn=<NllLossBackward>)
[3371] Train Loss:    tensor(0.9502, device='cuda:0', grad_fn=<NllLossBackward>)
[3372] Train Loss:    tensor(0.9938, device='cuda:0', grad_fn=<NllLossBackward>)
[3373] Train Loss:    tensor(1.0102, device='cuda:0', grad_fn=<NllLossBackward>)
[3374] Train Loss:    tensor(0.9969, device='cuda:0', grad_fn=<NllLossBackward>)
[3375] Train Loss:    tensor(0.9578, device='cuda:0', grad_fn=<NllLossBackward>)
[3376] Train Loss:    tensor(0.9978, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3377] Train Loss:    tensor(0.9786, device='cuda:0', grad_fn=<NllLossBackward>)
[3378] Train Loss:    tensor(1.0104, device='cuda:0', grad_fn=<NllLossBackward>)
[3379] Train Loss:    tensor(1.0069, device='cuda:0', grad_fn=<NllLossBackward>)
[3380] Train Loss:    tensor(0.9568, device='cuda:0', grad_fn=<NllLossBackward>)
[3381] Train Loss:    tensor(1.0462, device='cuda:0', grad_fn=<NllLossBackward>)
[3382] Train Loss:    tensor(0.9785, device='cuda:0', grad_fn=<NllLossBackward>)
[3383] Train Loss:    tensor(0.9601, device='cuda:0', grad_fn=<NllLossBackward>)
[3384] Train Loss:    tensor(0.9655, device='cuda:0', grad_fn=<NllLossBackward>)
[3385] Train Loss:    tensor(0.9969, device='cuda:0', grad_fn=<NllLossBackward>)
[3386] Train Loss:    tensor(0.9777, device='cuda:0', grad_fn=<NllLossBackward>)
[3387] Train Loss:    tensor(0.9909, device='cuda:0', grad_fn=<NllLossBackward>)
[3388] Train Loss:    tensor(0.9629, device='cuda:0', grad_fn=<NllLossBackward>)
[3389] Train Loss:    tensor(0.9755, device='cuda:0', grad_fn=<NllLossBackward>)
[3390] Train Loss:    tensor(0.9672, device='cuda:0', grad_fn=<NllLossBackward>)
[3391] Train Loss:    tensor(0.9531, device='cuda:0', grad_fn=<NllLossBackward>)
[3392] Train Loss:    tensor(0.9806, device='cuda:0', grad_fn=<NllLossBackward>)
[3393] Train Loss:    tensor(0.9770, device='cuda:0', grad_fn=<NllLossBackward>)
[3394] Train Loss:    tensor(1.0117, device='cuda:0', grad_fn=<NllLossBackward>)
[3395] Train Loss:    tensor(0.9357, device='cuda:0', grad_fn=<NllLossBackward>)
[3396] Train Loss:    tensor(0.9948, device='cuda:0', grad_fn=<NllLossBackward>)
[3397] Train Loss:    tensor(0.9725, device='cuda:0', grad_fn=<NllLossBackward>)
[3398] Train Loss:    tensor(0.9909, device='cuda:0', grad_fn=<NllLossBackward>)
[3399] Train Loss:    tensor(0.9874, device='cuda:0', grad_fn=<NllLossBackward>)
[3400] Train Loss:    tensor(0.9464, device='cuda:0', grad_fn=<NllLossBackward>)
[3401] Train Loss:    tensor(0.9768, device='cuda:0', grad_fn=<NllLossBackward>)
[3402] Train Loss:    tensor(0.9801, device='cuda:0', grad_fn=<NllLossBackward>)
[3403] Train Loss:    tensor(0.9882, device='cuda:0', grad_fn=<NllLossBackward>)
[3404] Train Loss:    tensor(0.9864, device='cuda:0', grad_fn=<NllLossBackward>)
[3405] Train Loss:    tensor(0.9781, device='cuda:0', grad_fn=<NllLossBackward>)
[3406] Train Loss:    tensor(0.9869, device='cuda:0', grad_fn=<NllLossBackward>)
[3407] Train Loss:    tensor(0.9722, device='cuda:0', grad_fn=<NllLossBackward>)
[3408] Train Loss:    tensor(1.0135, device='cuda:0', grad_fn=<NllLossBackward>)
[3409] Train Loss:    tensor(1.0103, device='cuda:0', grad_fn=<NllLossBackward>)
[3410] Train Loss:    tensor(0.9708, device='cuda:0', grad_fn=<NllLossBackward>)
[3411] Train Loss:    tensor(0.9933, device='cuda:0', grad_fn=<NllLossBackward>)
[3412] Train Loss:    tensor(0.9923, device='cuda:0', grad_fn=<NllLossBackward>)
[3413] Train Loss:    tensor(0.9696, device='cuda:0', grad_fn=<NllLossBackward>)
[3414] Train Loss:    tensor(0.9898, device='cuda:0', grad_fn=<NllLossBackward>)
[3415] Train Loss:    tensor(1.0053, device='cuda:0', grad_fn=<NllLossBackward>)
[3416] Train Loss:    tensor(0.9498, device='cuda:0', grad_fn=<NllLossBackward>)
[3417] Train Loss:    tensor(0.9878, device='cuda:0', grad_fn=<NllLossBackward>)
[3418] Train Loss:    tensor(0.9743, device='cuda:0', grad_fn=<NllLossBackward>)
[3419] Train Loss:    tensor(1.0009, device='cuda:0', grad_fn=<NllLossBackward>)
[3420] Train Loss:    tensor(1.0114, device='cuda:0', grad_fn=<NllLossBackward>)
[3421] Train Loss:    tensor(0.9338, device='cuda:0', grad_fn=<NllLossBackward>)
[3422] Train Loss:    tensor(0.9969, device='cuda:0', grad_fn=<NllLossBackward>)
[3423] Train Loss:    tensor(0.9777, device='cuda:0', grad_fn=<NllLossBackward>)
[3424] Train Loss:    tensor(0.9706, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3425] Train Loss:   tensor(0.9857, device='cuda:0', grad_fn=<NllLossBackward>)
[3426] Train Loss:   tensor(0.9938, device='cuda:0', grad_fn=<NllLossBackward>)
[3427] Train Loss:   tensor(0.9995, device='cuda:0', grad_fn=<NllLossBackward>)
[3428] Train Loss:   tensor(0.9906, device='cuda:0', grad_fn=<NllLossBackward>)
[3429] Train Loss:   tensor(1.0157, device='cuda:0', grad_fn=<NllLossBackward>)
[3430] Train Loss:   tensor(1.0004, device='cuda:0', grad_fn=<NllLossBackward>)
[3431] Train Loss:   tensor(0.9938, device='cuda:0', grad_fn=<NllLossBackward>)
[3432] Train Loss:   tensor(0.9720, device='cuda:0', grad_fn=<NllLossBackward>)
[3433] Train Loss:   tensor(0.9717, device='cuda:0', grad_fn=<NllLossBackward>)
[3434] Train Loss:   tensor(0.9885, device='cuda:0', grad_fn=<NllLossBackward>)
[3435] Train Loss:   tensor(0.9759, device='cuda:0', grad_fn=<NllLossBackward>)
[3436] Train Loss:   tensor(0.9820, device='cuda:0', grad_fn=<NllLossBackward>)
[3437] Train Loss:   tensor(1.0248, device='cuda:0', grad_fn=<NllLossBackward>)
[3438] Train Loss:   tensor(0.9650, device='cuda:0', grad_fn=<NllLossBackward>)
[3439] Train Loss:   tensor(0.9792, device='cuda:0', grad_fn=<NllLossBackward>)
[3440] Train Loss:   tensor(1.0250, device='cuda:0', grad_fn=<NllLossBackward>)
[3441] Train Loss:   tensor(0.9961, device='cuda:0', grad_fn=<NllLossBackward>)
[3442] Train Loss:   tensor(0.9731, device='cuda:0', grad_fn=<NllLossBackward>)
[3443] Train Loss:   tensor(0.9810, device='cuda:0', grad_fn=<NllLossBackward>)
[3444] Train Loss:   tensor(0.9812, device='cuda:0', grad_fn=<NllLossBackward>)
[3445] Train Loss:   tensor(1.0070, device='cuda:0', grad_fn=<NllLossBackward>)
[3446] Train Loss:   tensor(0.9401, device='cuda:0', grad_fn=<NllLossBackward>)
[3447] Train Loss:   tensor(0.9709, device='cuda:0', grad_fn=<NllLossBackward>)
[3448] Train Loss:   tensor(0.9785, device='cuda:0', grad_fn=<NllLossBackward>)
[3449] Train Loss:   tensor(1.0083, device='cuda:0', grad_fn=<NllLossBackward>)
[3450] Train Loss:   tensor(0.9732, device='cuda:0', grad_fn=<NllLossBackward>)
[3451] Train Loss:   tensor(0.9422, device='cuda:0', grad_fn=<NllLossBackward>)
[3452] Train Loss:   tensor(1.0059, device='cuda:0', grad_fn=<NllLossBackward>)
[3453] Train Loss:   tensor(0.9601, device='cuda:0', grad_fn=<NllLossBackward>)
[3454] Train Loss:   tensor(1.0201, device='cuda:0', grad_fn=<NllLossBackward>)
[3455] Train Loss:   tensor(0.9838, device='cuda:0', grad_fn=<NllLossBackward>)
[3456] Train Loss:   tensor(0.9897, device='cuda:0', grad_fn=<NllLossBackward>)
[3457] Train Loss:   tensor(0.9585, device='cuda:0', grad_fn=<NllLossBackward>)
[3458] Train Loss:   tensor(0.9812, device='cuda:0', grad_fn=<NllLossBackward>)
[3459] Train Loss:   tensor(1.0286, device='cuda:0', grad_fn=<NllLossBackward>)
[3460] Train Loss:   tensor(0.9780, device='cuda:0', grad_fn=<NllLossBackward>)
[3461] Train Loss:   tensor(0.9536, device='cuda:0', grad_fn=<NllLossBackward>)
[3462] Train Loss:   tensor(0.9979, device='cuda:0', grad_fn=<NllLossBackward>)
[3463] Train Loss:   tensor(0.9465, device='cuda:0', grad_fn=<NllLossBackward>)
[3464] Train Loss:   tensor(1.0051, device='cuda:0', grad_fn=<NllLossBackward>)
[3465] Train Loss:   tensor(0.9907, device='cuda:0', grad_fn=<NllLossBackward>)
[3466] Train Loss:   tensor(1.0151, device='cuda:0', grad_fn=<NllLossBackward>)
[3467] Train Loss:   tensor(0.9642, device='cuda:0', grad_fn=<NllLossBackward>)
[3468] Train Loss:   tensor(0.9472, device='cuda:0', grad_fn=<NllLossBackward>)
[3469] Train Loss:   tensor(0.9970, device='cuda:0', grad_fn=<NllLossBackward>)
[3470] Train Loss:   tensor(0.9849, device='cuda:0', grad_fn=<NllLossBackward>)
[3471] Train Loss:   tensor(0.9796, device='cuda:0', grad_fn=<NllLossBackward>)
[3472] Train Loss:   tensor(0.9757, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3473] Train Loss:    tensor(0.9925, device='cuda:0', grad_fn=<NllLossBackward>)
[3474] Train Loss:    tensor(0.9830, device='cuda:0', grad_fn=<NllLossBackward>)
[3475] Train Loss:    tensor(0.9641, device='cuda:0', grad_fn=<NllLossBackward>)
[3476] Train Loss:    tensor(0.9855, device='cuda:0', grad_fn=<NllLossBackward>)
[3477] Train Loss:    tensor(0.9416, device='cuda:0', grad_fn=<NllLossBackward>)
[3478] Train Loss:    tensor(0.9627, device='cuda:0', grad_fn=<NllLossBackward>)
[3479] Train Loss:    tensor(0.9885, device='cuda:0', grad_fn=<NllLossBackward>)
[3480] Train Loss:    tensor(0.9710, device='cuda:0', grad_fn=<NllLossBackward>)
[3481] Train Loss:    tensor(0.9669, device='cuda:0', grad_fn=<NllLossBackward>)
[3482] Train Loss:    tensor(1.0249, device='cuda:0', grad_fn=<NllLossBackward>)
[3483] Train Loss:    tensor(0.9984, device='cuda:0', grad_fn=<NllLossBackward>)
[3484] Train Loss:    tensor(0.9958, device='cuda:0', grad_fn=<NllLossBackward>)
[3485] Train Loss:    tensor(0.9955, device='cuda:0', grad_fn=<NllLossBackward>)
[3486] Train Loss:    tensor(0.9777, device='cuda:0', grad_fn=<NllLossBackward>)
[3487] Train Loss:    tensor(0.9674, device='cuda:0', grad_fn=<NllLossBackward>)
[3488] Train Loss:    tensor(0.9630, device='cuda:0', grad_fn=<NllLossBackward>)
[3489] Train Loss:    tensor(0.9488, device='cuda:0', grad_fn=<NllLossBackward>)
[3490] Train Loss:    tensor(1.0033, device='cuda:0', grad_fn=<NllLossBackward>)
[3491] Train Loss:    tensor(0.9695, device='cuda:0', grad_fn=<NllLossBackward>)
[3492] Train Loss:    tensor(1.0101, device='cuda:0', grad_fn=<NllLossBackward>)
[3493] Train Loss:    tensor(0.9822, device='cuda:0', grad_fn=<NllLossBackward>)
[3494] Train Loss:    tensor(0.9894, device='cuda:0', grad_fn=<NllLossBackward>)
[3495] Train Loss:    tensor(0.9567, device='cuda:0', grad_fn=<NllLossBackward>)
[3496] Train Loss:    tensor(0.9831, device='cuda:0', grad_fn=<NllLossBackward>)
[3497] Train Loss:    tensor(0.9405, device='cuda:0', grad_fn=<NllLossBackward>)
[3498] Train Loss:    tensor(0.9961, device='cuda:0', grad_fn=<NllLossBackward>)
[3499] Train Loss:    tensor(0.9924, device='cuda:0', grad_fn=<NllLossBackward>)
[3500] Train Loss:    tensor(0.9631, device='cuda:0', grad_fn=<NllLossBackward>)
[3501] Train Loss:    tensor(1.0029, device='cuda:0', grad_fn=<NllLossBackward>)
[3502] Train Loss:    tensor(0.9616, device='cuda:0', grad_fn=<NllLossBackward>)
[3503] Train Loss:    tensor(0.9878, device='cuda:0', grad_fn=<NllLossBackward>)
[3504] Train Loss:    tensor(0.9859, device='cuda:0', grad_fn=<NllLossBackward>)
[3505] Train Loss:    tensor(1.0004, device='cuda:0', grad_fn=<NllLossBackward>)
[3506] Train Loss:    tensor(0.9405, device='cuda:0', grad_fn=<NllLossBackward>)
[3507] Train Loss:    tensor(0.9721, device='cuda:0', grad_fn=<NllLossBackward>)
[3508] Train Loss:    tensor(0.9887, device='cuda:0', grad_fn=<NllLossBackward>)
[3509] Train Loss:    tensor(0.9494, device='cuda:0', grad_fn=<NllLossBackward>)
[3510] Train Loss:    tensor(0.9580, device='cuda:0', grad_fn=<NllLossBackward>)
[3511] Train Loss:    tensor(0.9781, device='cuda:0', grad_fn=<NllLossBackward>)
[3512] Train Loss:    tensor(0.9992, device='cuda:0', grad_fn=<NllLossBackward>)
[3513] Train Loss:    tensor(0.9682, device='cuda:0', grad_fn=<NllLossBackward>)
[3514] Train Loss:    tensor(0.9797, device='cuda:0', grad_fn=<NllLossBackward>)
[3515] Train Loss:    tensor(0.9674, device='cuda:0', grad_fn=<NllLossBackward>)
[3516] Train Loss:    tensor(0.9394, device='cuda:0', grad_fn=<NllLossBackward>)
[3517] Train Loss:    tensor(0.9456, device='cuda:0', grad_fn=<NllLossBackward>)
[3518] Train Loss:    tensor(1.0224, device='cuda:0', grad_fn=<NllLossBackward>)
[3519] Train Loss:    tensor(0.9769, device='cuda:0', grad_fn=<NllLossBackward>)
[3520] Train Loss:    tensor(1.0055, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3521] Train Loss:    tensor(0.9727, device='cuda:0', grad_fn=<NllLossBackward>)
[3522] Train Loss:    tensor(1.0018, device='cuda:0', grad_fn=<NllLossBackward>)
[3523] Train Loss:    tensor(0.9906, device='cuda:0', grad_fn=<NllLossBackward>)
[3524] Train Loss:    tensor(0.9988, device='cuda:0', grad_fn=<NllLossBackward>)
[3525] Train Loss:    tensor(0.9951, device='cuda:0', grad_fn=<NllLossBackward>)
[3526] Train Loss:    tensor(0.9401, device='cuda:0', grad_fn=<NllLossBackward>)
[3527] Train Loss:    tensor(0.9672, device='cuda:0', grad_fn=<NllLossBackward>)
[3528] Train Loss:    tensor(0.9885, device='cuda:0', grad_fn=<NllLossBackward>)
[3529] Train Loss:    tensor(1.0088, device='cuda:0', grad_fn=<NllLossBackward>)
[3530] Train Loss:    tensor(0.9544, device='cuda:0', grad_fn=<NllLossBackward>)
[3531] Train Loss:    tensor(0.9905, device='cuda:0', grad_fn=<NllLossBackward>)
[3532] Train Loss:    tensor(0.9440, device='cuda:0', grad_fn=<NllLossBackward>)
[3533] Train Loss:    tensor(0.9827, device='cuda:0', grad_fn=<NllLossBackward>)
[3534] Train Loss:    tensor(0.9809, device='cuda:0', grad_fn=<NllLossBackward>)
[3535] Train Loss:    tensor(1.0040, device='cuda:0', grad_fn=<NllLossBackward>)
[3536] Train Loss:    tensor(0.9572, device='cuda:0', grad_fn=<NllLossBackward>)
[3537] Train Loss:    tensor(0.9799, device='cuda:0', grad_fn=<NllLossBackward>)
[3538] Train Loss:    tensor(1.0070, device='cuda:0', grad_fn=<NllLossBackward>)
[3539] Train Loss:    tensor(0.9645, device='cuda:0', grad_fn=<NllLossBackward>)
[3540] Train Loss:    tensor(1.0002, device='cuda:0', grad_fn=<NllLossBackward>)
[3541] Train Loss:    tensor(0.9794, device='cuda:0', grad_fn=<NllLossBackward>)
[3542] Train Loss:    tensor(0.9565, device='cuda:0', grad_fn=<NllLossBackward>)
[3543] Train Loss:    tensor(0.9976, device='cuda:0', grad_fn=<NllLossBackward>)
[3544] Train Loss:    tensor(0.9810, device='cuda:0', grad_fn=<NllLossBackward>)
[3545] Train Loss:    tensor(0.9668, device='cuda:0', grad_fn=<NllLossBackward>)
[3546] Train Loss:    tensor(1.0144, device='cuda:0', grad_fn=<NllLossBackward>)
[3547] Train Loss:    tensor(0.9622, device='cuda:0', grad_fn=<NllLossBackward>)
[3548] Train Loss:    tensor(0.9812, device='cuda:0', grad_fn=<NllLossBackward>)
[3549] Train Loss:    tensor(1.0043, device='cuda:0', grad_fn=<NllLossBackward>)
[3550] Train Loss:    tensor(0.9949, device='cuda:0', grad_fn=<NllLossBackward>)
[3551] Train Loss:    tensor(1.0057, device='cuda:0', grad_fn=<NllLossBackward>)
[3552] Train Loss:    tensor(0.9420, device='cuda:0', grad_fn=<NllLossBackward>)
[3553] Train Loss:    tensor(0.9769, device='cuda:0', grad_fn=<NllLossBackward>)
[3554] Train Loss:    tensor(1.0068, device='cuda:0', grad_fn=<NllLossBackward>)
[3555] Train Loss:    tensor(0.9735, device='cuda:0', grad_fn=<NllLossBackward>)
[3556] Train Loss:    tensor(0.9293, device='cuda:0', grad_fn=<NllLossBackward>)
[3557] Train Loss:    tensor(0.9809, device='cuda:0', grad_fn=<NllLossBackward>)
[3558] Train Loss:    tensor(0.9593, device='cuda:0', grad_fn=<NllLossBackward>)
[3559] Train Loss:    tensor(1.0005, device='cuda:0', grad_fn=<NllLossBackward>)
[3560] Train Loss:    tensor(0.9694, device='cuda:0', grad_fn=<NllLossBackward>)
[3561] Train Loss:    tensor(1.0434, device='cuda:0', grad_fn=<NllLossBackward>)
[3562] Train Loss:    tensor(0.9419, device='cuda:0', grad_fn=<NllLossBackward>)
[3563] Train Loss:    tensor(0.9668, device='cuda:0', grad_fn=<NllLossBackward>)
[3564] Train Loss:    tensor(0.9505, device='cuda:0', grad_fn=<NllLossBackward>)
[3565] Train Loss:    tensor(0.9554, device='cuda:0', grad_fn=<NllLossBackward>)
[3566] Train Loss:    tensor(0.9946, device='cuda:0', grad_fn=<NllLossBackward>)
[3567] Train Loss:    tensor(0.9699, device='cuda:0', grad_fn=<NllLossBackward>)
[3568] Train Loss:    tensor(0.9765, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3569] Train Loss:    tensor(0.9431, device='cuda:0', grad_fn=<NllLossBackward>)
[3570] Train Loss:    tensor(0.9965, device='cuda:0', grad_fn=<NllLossBackward>)
[3571] Train Loss:    tensor(0.9950, device='cuda:0', grad_fn=<NllLossBackward>)
[3572] Train Loss:    tensor(0.9818, device='cuda:0', grad_fn=<NllLossBackward>)
[3573] Train Loss:    tensor(0.9721, device='cuda:0', grad_fn=<NllLossBackward>)
[3574] Train Loss:    tensor(0.9909, device='cuda:0', grad_fn=<NllLossBackward>)
[3575] Train Loss:    tensor(0.9595, device='cuda:0', grad_fn=<NllLossBackward>)
[3576] Train Loss:    tensor(0.9719, device='cuda:0', grad_fn=<NllLossBackward>)
[3577] Train Loss:    tensor(1.0033, device='cuda:0', grad_fn=<NllLossBackward>)
[3578] Train Loss:    tensor(0.9989, device='cuda:0', grad_fn=<NllLossBackward>)
[3579] Train Loss:    tensor(0.9873, device='cuda:0', grad_fn=<NllLossBackward>)
[3580] Train Loss:    tensor(0.9574, device='cuda:0', grad_fn=<NllLossBackward>)
[3581] Train Loss:    tensor(0.9723, device='cuda:0', grad_fn=<NllLossBackward>)
[3582] Train Loss:    tensor(0.9779, device='cuda:0', grad_fn=<NllLossBackward>)
[3583] Train Loss:    tensor(0.9901, device='cuda:0', grad_fn=<NllLossBackward>)
[3584] Train Loss:    tensor(0.9936, device='cuda:0', grad_fn=<NllLossBackward>)
[3585] Train Loss:    tensor(1.0057, device='cuda:0', grad_fn=<NllLossBackward>)
[3586] Train Loss:    tensor(0.9782, device='cuda:0', grad_fn=<NllLossBackward>)
[3587] Train Loss:    tensor(0.9760, device='cuda:0', grad_fn=<NllLossBackward>)
[3588] Train Loss:    tensor(0.9901, device='cuda:0', grad_fn=<NllLossBackward>)
[3589] Train Loss:    tensor(0.9485, device='cuda:0', grad_fn=<NllLossBackward>)
[3590] Train Loss:    tensor(0.9952, device='cuda:0', grad_fn=<NllLossBackward>)
[3591] Train Loss:    tensor(0.9556, device='cuda:0', grad_fn=<NllLossBackward>)
[3592] Train Loss:    tensor(0.9461, device='cuda:0', grad_fn=<NllLossBackward>)
[3593] Train Loss:    tensor(0.9511, device='cuda:0', grad_fn=<NllLossBackward>)
[3594] Train Loss:    tensor(0.9576, device='cuda:0', grad_fn=<NllLossBackward>)
[3595] Train Loss:    tensor(1.0054, device='cuda:0', grad_fn=<NllLossBackward>)
[3596] Train Loss:    tensor(0.9730, device='cuda:0', grad_fn=<NllLossBackward>)
[3597] Train Loss:    tensor(0.9833, device='cuda:0', grad_fn=<NllLossBackward>)
[3598] Train Loss:    tensor(0.9831, device='cuda:0', grad_fn=<NllLossBackward>)
[3599] Train Loss:    tensor(0.9627, device='cuda:0', grad_fn=<NllLossBackward>)
[3600] Train Loss:    tensor(0.9873, device='cuda:0', grad_fn=<NllLossBackward>)
[3601] Train Loss:    tensor(0.9861, device='cuda:0', grad_fn=<NllLossBackward>)
[3602] Train Loss:    tensor(0.9496, device='cuda:0', grad_fn=<NllLossBackward>)
[3603] Train Loss:    tensor(0.9742, device='cuda:0', grad_fn=<NllLossBackward>)
[3604] Train Loss:    tensor(0.9844, device='cuda:0', grad_fn=<NllLossBackward>)
[3605] Train Loss:    tensor(0.9777, device='cuda:0', grad_fn=<NllLossBackward>)
[3606] Train Loss:    tensor(0.9820, device='cuda:0', grad_fn=<NllLossBackward>)
[3607] Train Loss:    tensor(0.9957, device='cuda:0', grad_fn=<NllLossBackward>)
[3608] Train Loss:    tensor(0.9765, device='cuda:0', grad_fn=<NllLossBackward>)
[3609] Train Loss:    tensor(1.0028, device='cuda:0', grad_fn=<NllLossBackward>)
[3610] Train Loss:    tensor(0.9801, device='cuda:0', grad_fn=<NllLossBackward>)
[3611] Train Loss:    tensor(1.0249, device='cuda:0', grad_fn=<NllLossBackward>)
[3612] Train Loss:    tensor(1.0040, device='cuda:0', grad_fn=<NllLossBackward>)
[3613] Train Loss:    tensor(0.9886, device='cuda:0', grad_fn=<NllLossBackward>)
[3614] Train Loss:    tensor(0.9941, device='cuda:0', grad_fn=<NllLossBackward>)
[3615] Train Loss:    tensor(0.9733, device='cuda:0', grad_fn=<NllLossBackward>)
[3616] Train Loss:    tensor(0.9836, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3617] Train Loss:    tensor(0.9698, device='cuda:0', grad_fn=<NllLossBackward>)
[3618] Train Loss:    tensor(0.9581, device='cuda:0', grad_fn=<NllLossBackward>)
[3619] Train Loss:    tensor(0.9779, device='cuda:0', grad_fn=<NllLossBackward>)
[3620] Train Loss:    tensor(1.0009, device='cuda:0', grad_fn=<NllLossBackward>)
[3621] Train Loss:    tensor(0.9764, device='cuda:0', grad_fn=<NllLossBackward>)
[3622] Train Loss:    tensor(0.9898, device='cuda:0', grad_fn=<NllLossBackward>)
[3623] Train Loss:    tensor(0.9539, device='cuda:0', grad_fn=<NllLossBackward>)
[3624] Train Loss:    tensor(0.9533, device='cuda:0', grad_fn=<NllLossBackward>)
[3625] Train Loss:    tensor(0.9867, device='cuda:0', grad_fn=<NllLossBackward>)
[3626] Train Loss:    tensor(0.9593, device='cuda:0', grad_fn=<NllLossBackward>)
[3627] Train Loss:    tensor(0.9689, device='cuda:0', grad_fn=<NllLossBackward>)
[3628] Train Loss:    tensor(0.9412, device='cuda:0', grad_fn=<NllLossBackward>)
[3629] Train Loss:    tensor(0.9802, device='cuda:0', grad_fn=<NllLossBackward>)
[3630] Train Loss:    tensor(0.9681, device='cuda:0', grad_fn=<NllLossBackward>)
[3631] Train Loss:    tensor(0.9790, device='cuda:0', grad_fn=<NllLossBackward>)
[3632] Train Loss:    tensor(0.9444, device='cuda:0', grad_fn=<NllLossBackward>)
[3633] Train Loss:    tensor(0.9745, device='cuda:0', grad_fn=<NllLossBackward>)
[3634] Train Loss:    tensor(0.9561, device='cuda:0', grad_fn=<NllLossBackward>)
[3635] Train Loss:    tensor(0.9804, device='cuda:0', grad_fn=<NllLossBackward>)
[3636] Train Loss:    tensor(0.9811, device='cuda:0', grad_fn=<NllLossBackward>)
[3637] Train Loss:    tensor(0.9710, device='cuda:0', grad_fn=<NllLossBackward>)
[3638] Train Loss:    tensor(0.9856, device='cuda:0', grad_fn=<NllLossBackward>)
[3639] Train Loss:    tensor(0.9743, device='cuda:0', grad_fn=<NllLossBackward>)
[3640] Train Loss:    tensor(0.9240, device='cuda:0', grad_fn=<NllLossBackward>)
[3641] Train Loss:    tensor(0.9617, device='cuda:0', grad_fn=<NllLossBackward>)
[3642] Train Loss:    tensor(0.9866, device='cuda:0', grad_fn=<NllLossBackward>)
[3643] Train Loss:    tensor(0.9609, device='cuda:0', grad_fn=<NllLossBackward>)
[3644] Train Loss:    tensor(0.9850, device='cuda:0', grad_fn=<NllLossBackward>)
[3645] Train Loss:    tensor(0.9630, device='cuda:0', grad_fn=<NllLossBackward>)
[3646] Train Loss:    tensor(0.9906, device='cuda:0', grad_fn=<NllLossBackward>)
[3647] Train Loss:    tensor(0.9819, device='cuda:0', grad_fn=<NllLossBackward>)
[3648] Train Loss:    tensor(0.9808, device='cuda:0', grad_fn=<NllLossBackward>)
[3649] Train Loss:    tensor(0.9585, device='cuda:0', grad_fn=<NllLossBackward>)
[3650] Train Loss:    tensor(0.9830, device='cuda:0', grad_fn=<NllLossBackward>)
[3651] Train Loss:    tensor(0.9949, device='cuda:0', grad_fn=<NllLossBackward>)
[3652] Train Loss:    tensor(0.9695, device='cuda:0', grad_fn=<NllLossBackward>)
[3653] Train Loss:    tensor(1.0105, device='cuda:0', grad_fn=<NllLossBackward>)
[3654] Train Loss:    tensor(0.9655, device='cuda:0', grad_fn=<NllLossBackward>)
[3655] Train Loss:    tensor(0.9482, device='cuda:0', grad_fn=<NllLossBackward>)
[3656] Train Loss:    tensor(1.0341, device='cuda:0', grad_fn=<NllLossBackward>)
[3657] Train Loss:    tensor(0.9742, device='cuda:0', grad_fn=<NllLossBackward>)
[3658] Train Loss:    tensor(0.9594, device='cuda:0', grad_fn=<NllLossBackward>)
[3659] Train Loss:    tensor(0.9381, device='cuda:0', grad_fn=<NllLossBackward>)
[3660] Train Loss:    tensor(0.9779, device='cuda:0', grad_fn=<NllLossBackward>)
[3661] Train Loss:    tensor(0.9510, device='cuda:0', grad_fn=<NllLossBackward>)
[3662] Train Loss:    tensor(0.9966, device='cuda:0', grad_fn=<NllLossBackward>)
[3663] Train Loss:    tensor(0.9839, device='cuda:0', grad_fn=<NllLossBackward>)
[3664] Train Loss:    tensor(0.9713, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3665] Train Loss:   tensor(0.9419, device='cuda:0', grad_fn=<NllLossBackward>)
[3666] Train Loss:   tensor(0.9845, device='cuda:0', grad_fn=<NllLossBackward>)
[3667] Train Loss:   tensor(0.9849, device='cuda:0', grad_fn=<NllLossBackward>)
[3668] Train Loss:   tensor(0.9483, device='cuda:0', grad_fn=<NllLossBackward>)
[3669] Train Loss:   tensor(0.9632, device='cuda:0', grad_fn=<NllLossBackward>)
[3670] Train Loss:   tensor(0.9361, device='cuda:0', grad_fn=<NllLossBackward>)
[3671] Train Loss:   tensor(0.9755, device='cuda:0', grad_fn=<NllLossBackward>)
[3672] Train Loss:   tensor(1.0079, device='cuda:0', grad_fn=<NllLossBackward>)
[3673] Train Loss:   tensor(1.0133, device='cuda:0', grad_fn=<NllLossBackward>)
[3674] Train Loss:   tensor(0.9560, device='cuda:0', grad_fn=<NllLossBackward>)
[3675] Train Loss:   tensor(0.9818, device='cuda:0', grad_fn=<NllLossBackward>)
[3676] Train Loss:   tensor(1.0158, device='cuda:0', grad_fn=<NllLossBackward>)
[3677] Train Loss:   tensor(0.9414, device='cuda:0', grad_fn=<NllLossBackward>)
[3678] Train Loss:   tensor(0.9476, device='cuda:0', grad_fn=<NllLossBackward>)
[3679] Train Loss:   tensor(1.0013, device='cuda:0', grad_fn=<NllLossBackward>)
[3680] Train Loss:   tensor(0.9964, device='cuda:0', grad_fn=<NllLossBackward>)
[3681] Train Loss:   tensor(0.9780, device='cuda:0', grad_fn=<NllLossBackward>)
[3682] Train Loss:   tensor(0.9458, device='cuda:0', grad_fn=<NllLossBackward>)
[3683] Train Loss:   tensor(0.9984, device='cuda:0', grad_fn=<NllLossBackward>)
[3684] Train Loss:   tensor(0.9904, device='cuda:0', grad_fn=<NllLossBackward>)
[3685] Train Loss:   tensor(0.9989, device='cuda:0', grad_fn=<NllLossBackward>)
[3686] Train Loss:   tensor(0.9692, device='cuda:0', grad_fn=<NllLossBackward>)
[3687] Train Loss:   tensor(0.9700, device='cuda:0', grad_fn=<NllLossBackward>)
[3688] Train Loss:   tensor(0.9589, device='cuda:0', grad_fn=<NllLossBackward>)
[3689] Train Loss:   tensor(0.9859, device='cuda:0', grad_fn=<NllLossBackward>)
[3690] Train Loss:   tensor(0.9743, device='cuda:0', grad_fn=<NllLossBackward>)
[3691] Train Loss:   tensor(0.9514, device='cuda:0', grad_fn=<NllLossBackward>)
[3692] Train Loss:   tensor(0.9951, device='cuda:0', grad_fn=<NllLossBackward>)
[3693] Train Loss:   tensor(0.9624, device='cuda:0', grad_fn=<NllLossBackward>)
[3694] Train Loss:   tensor(1.0145, device='cuda:0', grad_fn=<NllLossBackward>)
[3695] Train Loss:   tensor(0.9800, device='cuda:0', grad_fn=<NllLossBackward>)
[3696] Train Loss:   tensor(0.9632, device='cuda:0', grad_fn=<NllLossBackward>)
[3697] Train Loss:   tensor(0.9550, device='cuda:0', grad_fn=<NllLossBackward>)
[3698] Train Loss:   tensor(0.9615, device='cuda:0', grad_fn=<NllLossBackward>)
[3699] Train Loss:   tensor(0.9850, device='cuda:0', grad_fn=<NllLossBackward>)
[3700] Train Loss:   tensor(0.9686, device='cuda:0', grad_fn=<NllLossBackward>)
[3701] Train Loss:   tensor(0.9769, device='cuda:0', grad_fn=<NllLossBackward>)
[3702] Train Loss:   tensor(0.9701, device='cuda:0', grad_fn=<NllLossBackward>)
[3703] Train Loss:   tensor(1.0226, device='cuda:0', grad_fn=<NllLossBackward>)
[3704] Train Loss:   tensor(0.9441, device='cuda:0', grad_fn=<NllLossBackward>)
[3705] Train Loss:   tensor(0.9571, device='cuda:0', grad_fn=<NllLossBackward>)
[3706] Train Loss:   tensor(0.9524, device='cuda:0', grad_fn=<NllLossBackward>)
[3707] Train Loss:   tensor(0.9620, device='cuda:0', grad_fn=<NllLossBackward>)
[3708] Train Loss:   tensor(0.9437, device='cuda:0', grad_fn=<NllLossBackward>)
[3709] Train Loss:   tensor(0.9728, device='cuda:0', grad_fn=<NllLossBackward>)
[3710] Train Loss:   tensor(0.9707, device='cuda:0', grad_fn=<NllLossBackward>)
[3711] Train Loss:   tensor(0.9148, device='cuda:0', grad_fn=<NllLossBackward>)
[3712] Train Loss:   tensor(0.9233, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3713] Train Loss:   tensor(0.9916, device='cuda:0', grad_fn=<NllLossBackward>)
[3714] Train Loss:   tensor(0.9651, device='cuda:0', grad_fn=<NllLossBackward>)
[3715] Train Loss:   tensor(0.9601, device='cuda:0', grad_fn=<NllLossBackward>)
[3716] Train Loss:   tensor(0.9611, device='cuda:0', grad_fn=<NllLossBackward>)
[3717] Train Loss:   tensor(0.9712, device='cuda:0', grad_fn=<NllLossBackward>)
[3718] Train Loss:   tensor(0.9642, device='cuda:0', grad_fn=<NllLossBackward>)
[3719] Train Loss:   tensor(0.9726, device='cuda:0', grad_fn=<NllLossBackward>)
[3720] Train Loss:   tensor(1.0018, device='cuda:0', grad_fn=<NllLossBackward>)
[3721] Train Loss:   tensor(0.9488, device='cuda:0', grad_fn=<NllLossBackward>)
[3722] Train Loss:   tensor(0.9459, device='cuda:0', grad_fn=<NllLossBackward>)
[3723] Train Loss:   tensor(0.9386, device='cuda:0', grad_fn=<NllLossBackward>)
[3724] Train Loss:   tensor(0.9849, device='cuda:0', grad_fn=<NllLossBackward>)
[3725] Train Loss:   tensor(1.0280, device='cuda:0', grad_fn=<NllLossBackward>)
[3726] Train Loss:   tensor(0.9745, device='cuda:0', grad_fn=<NllLossBackward>)
[3727] Train Loss:   tensor(1.0011, device='cuda:0', grad_fn=<NllLossBackward>)
[3728] Train Loss:   tensor(0.9602, device='cuda:0', grad_fn=<NllLossBackward>)
[3729] Train Loss:   tensor(0.9928, device='cuda:0', grad_fn=<NllLossBackward>)
[3730] Train Loss:   tensor(0.9769, device='cuda:0', grad_fn=<NllLossBackward>)
[3731] Train Loss:   tensor(0.9418, device='cuda:0', grad_fn=<NllLossBackward>)
[3732] Train Loss:   tensor(0.9728, device='cuda:0', grad_fn=<NllLossBackward>)
[3733] Train Loss:   tensor(0.9600, device='cuda:0', grad_fn=<NllLossBackward>)
[3734] Train Loss:   tensor(0.9738, device='cuda:0', grad_fn=<NllLossBackward>)
[3735] Train Loss:   tensor(0.9503, device='cuda:0', grad_fn=<NllLossBackward>)
[3736] Train Loss:   tensor(0.9796, device='cuda:0', grad_fn=<NllLossBackward>)
[3737] Train Loss:   tensor(0.9603, device='cuda:0', grad_fn=<NllLossBackward>)
[3738] Train Loss:   tensor(0.9319, device='cuda:0', grad_fn=<NllLossBackward>)
[3739] Train Loss:   tensor(0.9354, device='cuda:0', grad_fn=<NllLossBackward>)
[3740] Train Loss:   tensor(0.9607, device='cuda:0', grad_fn=<NllLossBackward>)
[3741] Train Loss:   tensor(0.9856, device='cuda:0', grad_fn=<NllLossBackward>)
[3742] Train Loss:   tensor(0.9958, device='cuda:0', grad_fn=<NllLossBackward>)
[3743] Train Loss:   tensor(0.9386, device='cuda:0', grad_fn=<NllLossBackward>)
[3744] Train Loss:   tensor(0.9718, device='cuda:0', grad_fn=<NllLossBackward>)
[3745] Train Loss:   tensor(0.9646, device='cuda:0', grad_fn=<NllLossBackward>)
[3746] Train Loss:   tensor(0.9694, device='cuda:0', grad_fn=<NllLossBackward>)
[3747] Train Loss:   tensor(0.9588, device='cuda:0', grad_fn=<NllLossBackward>)
[3748] Train Loss:   tensor(0.9580, device='cuda:0', grad_fn=<NllLossBackward>)
[3749] Train Loss:   tensor(0.9965, device='cuda:0', grad_fn=<NllLossBackward>)
[3750] Train Loss:   tensor(0.9529, device='cuda:0', grad_fn=<NllLossBackward>)
[3751] Train Loss:   tensor(0.9463, device='cuda:0', grad_fn=<NllLossBackward>)
[3752] Train Loss:   tensor(0.9897, device='cuda:0', grad_fn=<NllLossBackward>)
[3753] Train Loss:   tensor(0.9671, device='cuda:0', grad_fn=<NllLossBackward>)
[3754] Train Loss:   tensor(0.9957, device='cuda:0', grad_fn=<NllLossBackward>)
[3755] Train Loss:   tensor(0.9402, device='cuda:0', grad_fn=<NllLossBackward>)
[3756] Train Loss:   tensor(0.9704, device='cuda:0', grad_fn=<NllLossBackward>)
[3757] Train Loss:   tensor(0.9318, device='cuda:0', grad_fn=<NllLossBackward>)
[3758] Train Loss:   tensor(0.9546, device='cuda:0', grad_fn=<NllLossBackward>)
[3759] Train Loss:   tensor(0.9933, device='cuda:0', grad_fn=<NllLossBackward>)
[3760] Train Loss:   tensor(1.0017, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3761] Train Loss:    tensor(0.9696, device='cuda:0', grad_fn=<NllLossBackward>)
[3762] Train Loss:    tensor(0.9447, device='cuda:0', grad_fn=<NllLossBackward>)
[3763] Train Loss:    tensor(0.9811, device='cuda:0', grad_fn=<NllLossBackward>)
[3764] Train Loss:    tensor(0.9767, device='cuda:0', grad_fn=<NllLossBackward>)
[3765] Train Loss:    tensor(0.9739, device='cuda:0', grad_fn=<NllLossBackward>)
[3766] Train Loss:    tensor(0.9584, device='cuda:0', grad_fn=<NllLossBackward>)
[3767] Train Loss:    tensor(0.9761, device='cuda:0', grad_fn=<NllLossBackward>)
[3768] Train Loss:    tensor(0.9966, device='cuda:0', grad_fn=<NllLossBackward>)
[3769] Train Loss:    tensor(0.9842, device='cuda:0', grad_fn=<NllLossBackward>)
[3770] Train Loss:    tensor(0.9638, device='cuda:0', grad_fn=<NllLossBackward>)
[3771] Train Loss:    tensor(0.9265, device='cuda:0', grad_fn=<NllLossBackward>)
[3772] Train Loss:    tensor(0.9385, device='cuda:0', grad_fn=<NllLossBackward>)
[3773] Train Loss:    tensor(0.9517, device='cuda:0', grad_fn=<NllLossBackward>)
[3774] Train Loss:    tensor(0.9595, device='cuda:0', grad_fn=<NllLossBackward>)
[3775] Train Loss:    tensor(0.9593, device='cuda:0', grad_fn=<NllLossBackward>)
[3776] Train Loss:    tensor(0.9668, device='cuda:0', grad_fn=<NllLossBackward>)
[3777] Train Loss:    tensor(0.9727, device='cuda:0', grad_fn=<NllLossBackward>)
[3778] Train Loss:    tensor(0.9599, device='cuda:0', grad_fn=<NllLossBackward>)
[3779] Train Loss:    tensor(0.9726, device='cuda:0', grad_fn=<NllLossBackward>)
[3780] Train Loss:    tensor(0.9811, device='cuda:0', grad_fn=<NllLossBackward>)
[3781] Train Loss:    tensor(0.9814, device='cuda:0', grad_fn=<NllLossBackward>)
[3782] Train Loss:    tensor(1.0105, device='cuda:0', grad_fn=<NllLossBackward>)
[3783] Train Loss:    tensor(0.9351, device='cuda:0', grad_fn=<NllLossBackward>)
[3784] Train Loss:    tensor(0.9587, device='cuda:0', grad_fn=<NllLossBackward>)
[3785] Train Loss:    tensor(0.9843, device='cuda:0', grad_fn=<NllLossBackward>)
[3786] Train Loss:    tensor(1.0049, device='cuda:0', grad_fn=<NllLossBackward>)
[3787] Train Loss:    tensor(0.9657, device='cuda:0', grad_fn=<NllLossBackward>)
[3788] Train Loss:    tensor(0.9616, device='cuda:0', grad_fn=<NllLossBackward>)
[3789] Train Loss:    tensor(1.0017, device='cuda:0', grad_fn=<NllLossBackward>)
[3790] Train Loss:    tensor(0.9448, device='cuda:0', grad_fn=<NllLossBackward>)
[3791] Train Loss:    tensor(0.9921, device='cuda:0', grad_fn=<NllLossBackward>)
[3792] Train Loss:    tensor(0.9392, device='cuda:0', grad_fn=<NllLossBackward>)
[3793] Train Loss:    tensor(0.9557, device='cuda:0', grad_fn=<NllLossBackward>)
[3794] Train Loss:    tensor(0.9493, device='cuda:0', grad_fn=<NllLossBackward>)
[3795] Train Loss:    tensor(0.9843, device='cuda:0', grad_fn=<NllLossBackward>)
[3796] Train Loss:    tensor(0.9457, device='cuda:0', grad_fn=<NllLossBackward>)
[3797] Train Loss:    tensor(0.9841, device='cuda:0', grad_fn=<NllLossBackward>)
[3798] Train Loss:    tensor(0.9677, device='cuda:0', grad_fn=<NllLossBackward>)
[3799] Train Loss:    tensor(1.0161, device='cuda:0', grad_fn=<NllLossBackward>)
[3800] Train Loss:    tensor(0.9628, device='cuda:0', grad_fn=<NllLossBackward>)
[3801] Train Loss:    tensor(0.9668, device='cuda:0', grad_fn=<NllLossBackward>)
[3802] Train Loss:    tensor(0.9750, device='cuda:0', grad_fn=<NllLossBackward>)
[3803] Train Loss:    tensor(0.9767, device='cuda:0', grad_fn=<NllLossBackward>)
[3804] Train Loss:    tensor(0.9732, device='cuda:0', grad_fn=<NllLossBackward>)
[3805] Train Loss:    tensor(0.9672, device='cuda:0', grad_fn=<NllLossBackward>)
[3806] Train Loss:    tensor(0.9975, device='cuda:0', grad_fn=<NllLossBackward>)
[3807] Train Loss:    tensor(0.9945, device='cuda:0', grad_fn=<NllLossBackward>)
[3808] Train Loss:    tensor(0.9384, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3809] Train Loss:   tensor(0.9541, device='cuda:0', grad_fn=<NllLossBackward>)
[3810] Train Loss:   tensor(0.9273, device='cuda:0', grad_fn=<NllLossBackward>)
[3811] Train Loss:   tensor(0.9768, device='cuda:0', grad_fn=<NllLossBackward>)
[3812] Train Loss:   tensor(0.9599, device='cuda:0', grad_fn=<NllLossBackward>)
[3813] Train Loss:   tensor(1.0057, device='cuda:0', grad_fn=<NllLossBackward>)
[3814] Train Loss:   tensor(0.9803, device='cuda:0', grad_fn=<NllLossBackward>)
[3815] Train Loss:   tensor(1.0173, device='cuda:0', grad_fn=<NllLossBackward>)
[3816] Train Loss:   tensor(0.9820, device='cuda:0', grad_fn=<NllLossBackward>)
[3817] Train Loss:   tensor(0.9718, device='cuda:0', grad_fn=<NllLossBackward>)
[3818] Train Loss:   tensor(1.0073, device='cuda:0', grad_fn=<NllLossBackward>)
[3819] Train Loss:   tensor(0.9840, device='cuda:0', grad_fn=<NllLossBackward>)
[3820] Train Loss:   tensor(0.9489, device='cuda:0', grad_fn=<NllLossBackward>)
[3821] Train Loss:   tensor(0.9491, device='cuda:0', grad_fn=<NllLossBackward>)
[3822] Train Loss:   tensor(0.9474, device='cuda:0', grad_fn=<NllLossBackward>)
[3823] Train Loss:   tensor(0.9579, device='cuda:0', grad_fn=<NllLossBackward>)
[3824] Train Loss:   tensor(0.9796, device='cuda:0', grad_fn=<NllLossBackward>)
[3825] Train Loss:   tensor(0.9810, device='cuda:0', grad_fn=<NllLossBackward>)
[3826] Train Loss:   tensor(0.9607, device='cuda:0', grad_fn=<NllLossBackward>)
[3827] Train Loss:   tensor(0.9679, device='cuda:0', grad_fn=<NllLossBackward>)
[3828] Train Loss:   tensor(0.9312, device='cuda:0', grad_fn=<NllLossBackward>)
[3829] Train Loss:   tensor(0.9960, device='cuda:0', grad_fn=<NllLossBackward>)
[3830] Train Loss:   tensor(0.9214, device='cuda:0', grad_fn=<NllLossBackward>)
[3831] Train Loss:   tensor(0.9948, device='cuda:0', grad_fn=<NllLossBackward>)
[3832] Train Loss:   tensor(0.9797, device='cuda:0', grad_fn=<NllLossBackward>)
[3833] Train Loss:   tensor(0.9481, device='cuda:0', grad_fn=<NllLossBackward>)
[3834] Train Loss:   tensor(0.9912, device='cuda:0', grad_fn=<NllLossBackward>)
[3835] Train Loss:   tensor(0.9930, device='cuda:0', grad_fn=<NllLossBackward>)
[3836] Train Loss:   tensor(0.9929, device='cuda:0', grad_fn=<NllLossBackward>)
[3837] Train Loss:   tensor(0.9646, device='cuda:0', grad_fn=<NllLossBackward>)
[3838] Train Loss:   tensor(1.0192, device='cuda:0', grad_fn=<NllLossBackward>)
[3839] Train Loss:   tensor(0.9816, device='cuda:0', grad_fn=<NllLossBackward>)
[3840] Train Loss:   tensor(1.0151, device='cuda:0', grad_fn=<NllLossBackward>)
[3841] Train Loss:   tensor(0.9756, device='cuda:0', grad_fn=<NllLossBackward>)
[3842] Train Loss:   tensor(1.0147, device='cuda:0', grad_fn=<NllLossBackward>)
[3843] Train Loss:   tensor(0.9756, device='cuda:0', grad_fn=<NllLossBackward>)
[3844] Train Loss:   tensor(0.9433, device='cuda:0', grad_fn=<NllLossBackward>)
[3845] Train Loss:   tensor(0.9475, device='cuda:0', grad_fn=<NllLossBackward>)
[3846] Train Loss:   tensor(0.9702, device='cuda:0', grad_fn=<NllLossBackward>)
[3847] Train Loss:   tensor(1.0100, device='cuda:0', grad_fn=<NllLossBackward>)
[3848] Train Loss:   tensor(0.9746, device='cuda:0', grad_fn=<NllLossBackward>)
[3849] Train Loss:   tensor(0.9518, device='cuda:0', grad_fn=<NllLossBackward>)
[3850] Train Loss:   tensor(1.0303, device='cuda:0', grad_fn=<NllLossBackward>)
[3851] Train Loss:   tensor(0.9958, device='cuda:0', grad_fn=<NllLossBackward>)
[3852] Train Loss:   tensor(0.9353, device='cuda:0', grad_fn=<NllLossBackward>)
[3853] Train Loss:   tensor(0.9930, device='cuda:0', grad_fn=<NllLossBackward>)
[3854] Train Loss:   tensor(0.9550, device='cuda:0', grad_fn=<NllLossBackward>)
[3855] Train Loss:   tensor(0.9802, device='cuda:0', grad_fn=<NllLossBackward>)
[3856] Train Loss:   tensor(0.9391, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3857] Train Loss:   tensor(0.9325, device='cuda:0', grad_fn=<NllLossBackward>)
[3858] Train Loss:   tensor(0.9436, device='cuda:0', grad_fn=<NllLossBackward>)
[3859] Train Loss:   tensor(0.9372, device='cuda:0', grad_fn=<NllLossBackward>)
[3860] Train Loss:   tensor(0.9673, device='cuda:0', grad_fn=<NllLossBackward>)
[3861] Train Loss:   tensor(0.9847, device='cuda:0', grad_fn=<NllLossBackward>)
[3862] Train Loss:   tensor(0.9578, device='cuda:0', grad_fn=<NllLossBackward>)
[3863] Train Loss:   tensor(1.0071, device='cuda:0', grad_fn=<NllLossBackward>)
[3864] Train Loss:   tensor(0.9868, device='cuda:0', grad_fn=<NllLossBackward>)
[3865] Train Loss:   tensor(0.9788, device='cuda:0', grad_fn=<NllLossBackward>)
[3866] Train Loss:   tensor(0.9661, device='cuda:0', grad_fn=<NllLossBackward>)
[3867] Train Loss:   tensor(0.9902, device='cuda:0', grad_fn=<NllLossBackward>)
[3868] Train Loss:   tensor(0.9765, device='cuda:0', grad_fn=<NllLossBackward>)
[3869] Train Loss:   tensor(0.9552, device='cuda:0', grad_fn=<NllLossBackward>)
[3870] Train Loss:   tensor(0.9874, device='cuda:0', grad_fn=<NllLossBackward>)
[3871] Train Loss:   tensor(0.9846, device='cuda:0', grad_fn=<NllLossBackward>)
[3872] Train Loss:   tensor(0.9744, device='cuda:0', grad_fn=<NllLossBackward>)
[3873] Train Loss:   tensor(0.9608, device='cuda:0', grad_fn=<NllLossBackward>)
[3874] Train Loss:   tensor(1.0024, device='cuda:0', grad_fn=<NllLossBackward>)
[3875] Train Loss:   tensor(0.9690, device='cuda:0', grad_fn=<NllLossBackward>)
[3876] Train Loss:   tensor(0.9896, device='cuda:0', grad_fn=<NllLossBackward>)
[3877] Train Loss:   tensor(0.9666, device='cuda:0', grad_fn=<NllLossBackward>)
[3878] Train Loss:   tensor(0.9520, device='cuda:0', grad_fn=<NllLossBackward>)
[3879] Train Loss:   tensor(0.9619, device='cuda:0', grad_fn=<NllLossBackward>)
[3880] Train Loss:   tensor(0.9403, device='cuda:0', grad_fn=<NllLossBackward>)
[3881] Train Loss:   tensor(0.9665, device='cuda:0', grad_fn=<NllLossBackward>)
[3882] Train Loss:   tensor(0.9982, device='cuda:0', grad_fn=<NllLossBackward>)
[3883] Train Loss:   tensor(0.9774, device='cuda:0', grad_fn=<NllLossBackward>)
[3884] Train Loss:   tensor(0.9858, device='cuda:0', grad_fn=<NllLossBackward>)
[3885] Train Loss:   tensor(0.9708, device='cuda:0', grad_fn=<NllLossBackward>)
[3886] Train Loss:   tensor(0.9353, device='cuda:0', grad_fn=<NllLossBackward>)
[3887] Train Loss:   tensor(0.9168, device='cuda:0', grad_fn=<NllLossBackward>)
[3888] Train Loss:   tensor(0.9341, device='cuda:0', grad_fn=<NllLossBackward>)
[3889] Train Loss:   tensor(0.9564, device='cuda:0', grad_fn=<NllLossBackward>)
[3890] Train Loss:   tensor(0.9995, device='cuda:0', grad_fn=<NllLossBackward>)
[3891] Train Loss:   tensor(0.9461, device='cuda:0', grad_fn=<NllLossBackward>)
[3892] Train Loss:   tensor(0.9689, device='cuda:0', grad_fn=<NllLossBackward>)
[3893] Train Loss:   tensor(0.9610, device='cuda:0', grad_fn=<NllLossBackward>)
[3894] Train Loss:   tensor(0.9651, device='cuda:0', grad_fn=<NllLossBackward>)
[3895] Train Loss:   tensor(0.9619, device='cuda:0', grad_fn=<NllLossBackward>)
[3896] Train Loss:   tensor(0.9612, device='cuda:0', grad_fn=<NllLossBackward>)
[3897] Train Loss:   tensor(0.9683, device='cuda:0', grad_fn=<NllLossBackward>)
[3898] Train Loss:   tensor(1.0104, device='cuda:0', grad_fn=<NllLossBackward>)
[3899] Train Loss:   tensor(1.0021, device='cuda:0', grad_fn=<NllLossBackward>)
[3900] Train Loss:   tensor(0.9805, device='cuda:0', grad_fn=<NllLossBackward>)
[3901] Train Loss:   tensor(0.9335, device='cuda:0', grad_fn=<NllLossBackward>)
[3902] Train Loss:   tensor(0.9673, device='cuda:0', grad_fn=<NllLossBackward>)
[3903] Train Loss:   tensor(0.9654, device='cuda:0', grad_fn=<NllLossBackward>)
[3904] Train Loss:   tensor(0.9624, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3905] Train Loss:    tensor(0.9493, device='cuda:0', grad_fn=<NllLossBackward>)
[3906] Train Loss:    tensor(0.9780, device='cuda:0', grad_fn=<NllLossBackward>)
[3907] Train Loss:    tensor(0.9465, device='cuda:0', grad_fn=<NllLossBackward>)
[3908] Train Loss:    tensor(0.9504, device='cuda:0', grad_fn=<NllLossBackward>)
[3909] Train Loss:    tensor(0.9434, device='cuda:0', grad_fn=<NllLossBackward>)
[3910] Train Loss:    tensor(0.9527, device='cuda:0', grad_fn=<NllLossBackward>)
[3911] Train Loss:    tensor(0.9726, device='cuda:0', grad_fn=<NllLossBackward>)
[3912] Train Loss:    tensor(0.9997, device='cuda:0', grad_fn=<NllLossBackward>)
[3913] Train Loss:    tensor(0.9560, device='cuda:0', grad_fn=<NllLossBackward>)
[3914] Train Loss:    tensor(0.9584, device='cuda:0', grad_fn=<NllLossBackward>)
[3915] Train Loss:    tensor(0.9704, device='cuda:0', grad_fn=<NllLossBackward>)
[3916] Train Loss:    tensor(0.9824, device='cuda:0', grad_fn=<NllLossBackward>)
[3917] Train Loss:    tensor(0.9817, device='cuda:0', grad_fn=<NllLossBackward>)
[3918] Train Loss:    tensor(0.9205, device='cuda:0', grad_fn=<NllLossBackward>)
[3919] Train Loss:    tensor(0.9708, device='cuda:0', grad_fn=<NllLossBackward>)
[3920] Train Loss:    tensor(1.0137, device='cuda:0', grad_fn=<NllLossBackward>)
[3921] Train Loss:    tensor(0.9651, device='cuda:0', grad_fn=<NllLossBackward>)
[3922] Train Loss:    tensor(0.9815, device='cuda:0', grad_fn=<NllLossBackward>)
[3923] Train Loss:    tensor(0.9699, device='cuda:0', grad_fn=<NllLossBackward>)
[3924] Train Loss:    tensor(0.9385, device='cuda:0', grad_fn=<NllLossBackward>)
[3925] Train Loss:    tensor(0.9800, device='cuda:0', grad_fn=<NllLossBackward>)
[3926] Train Loss:    tensor(0.9478, device='cuda:0', grad_fn=<NllLossBackward>)
[3927] Train Loss:    tensor(0.9373, device='cuda:0', grad_fn=<NllLossBackward>)
[3928] Train Loss:    tensor(0.9791, device='cuda:0', grad_fn=<NllLossBackward>)
[3929] Train Loss:    tensor(0.9556, device='cuda:0', grad_fn=<NllLossBackward>)
[3930] Train Loss:    tensor(0.9761, device='cuda:0', grad_fn=<NllLossBackward>)
[3931] Train Loss:    tensor(0.9610, device='cuda:0', grad_fn=<NllLossBackward>)
[3932] Train Loss:    tensor(0.9684, device='cuda:0', grad_fn=<NllLossBackward>)
[3933] Train Loss:    tensor(0.9611, device='cuda:0', grad_fn=<NllLossBackward>)
[3934] Train Loss:    tensor(1.0192, device='cuda:0', grad_fn=<NllLossBackward>)
[3935] Train Loss:    tensor(0.9748, device='cuda:0', grad_fn=<NllLossBackward>)
[3936] Train Loss:    tensor(0.9766, device='cuda:0', grad_fn=<NllLossBackward>)
[3937] Train Loss:    tensor(0.9158, device='cuda:0', grad_fn=<NllLossBackward>)
[3938] Train Loss:    tensor(0.9745, device='cuda:0', grad_fn=<NllLossBackward>)
[3939] Train Loss:    tensor(0.9786, device='cuda:0', grad_fn=<NllLossBackward>)
[3940] Train Loss:    tensor(0.9590, device='cuda:0', grad_fn=<NllLossBackward>)
[3941] Train Loss:    tensor(0.9488, device='cuda:0', grad_fn=<NllLossBackward>)
[3942] Train Loss:    tensor(0.9821, device='cuda:0', grad_fn=<NllLossBackward>)
[3943] Train Loss:    tensor(0.9644, device='cuda:0', grad_fn=<NllLossBackward>)
[3944] Train Loss:    tensor(0.9488, device='cuda:0', grad_fn=<NllLossBackward>)
[3945] Train Loss:    tensor(0.9855, device='cuda:0', grad_fn=<NllLossBackward>)
[3946] Train Loss:    tensor(0.9251, device='cuda:0', grad_fn=<NllLossBackward>)
[3947] Train Loss:    tensor(0.9922, device='cuda:0', grad_fn=<NllLossBackward>)
[3948] Train Loss:    tensor(0.9447, device='cuda:0', grad_fn=<NllLossBackward>)
[3949] Train Loss:    tensor(0.9669, device='cuda:0', grad_fn=<NllLossBackward>)
[3950] Train Loss:    tensor(0.9816, device='cuda:0', grad_fn=<NllLossBackward>)
[3951] Train Loss:    tensor(0.9519, device='cuda:0', grad_fn=<NllLossBackward>)
[3952] Train Loss:    tensor(0.9692, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3953] Train Loss:    tensor(0.9532, device='cuda:0', grad_fn=<NllLossBackward>)
[3954] Train Loss:    tensor(0.9324, device='cuda:0', grad_fn=<NllLossBackward>)
[3955] Train Loss:    tensor(0.9517, device='cuda:0', grad_fn=<NllLossBackward>)
[3956] Train Loss:    tensor(0.9734, device='cuda:0', grad_fn=<NllLossBackward>)
[3957] Train Loss:    tensor(0.9786, device='cuda:0', grad_fn=<NllLossBackward>)
[3958] Train Loss:    tensor(0.9593, device='cuda:0', grad_fn=<NllLossBackward>)
[3959] Train Loss:    tensor(0.9554, device='cuda:0', grad_fn=<NllLossBackward>)
[3960] Train Loss:    tensor(0.9518, device='cuda:0', grad_fn=<NllLossBackward>)
[3961] Train Loss:    tensor(0.9673, device='cuda:0', grad_fn=<NllLossBackward>)
[3962] Train Loss:    tensor(0.9656, device='cuda:0', grad_fn=<NllLossBackward>)
[3963] Train Loss:    tensor(0.9872, device='cuda:0', grad_fn=<NllLossBackward>)
[3964] Train Loss:    tensor(0.9528, device='cuda:0', grad_fn=<NllLossBackward>)
[3965] Train Loss:    tensor(0.9638, device='cuda:0', grad_fn=<NllLossBackward>)
[3966] Train Loss:    tensor(0.9640, device='cuda:0', grad_fn=<NllLossBackward>)
[3967] Train Loss:    tensor(0.9747, device='cuda:0', grad_fn=<NllLossBackward>)
[3968] Train Loss:    tensor(0.9568, device='cuda:0', grad_fn=<NllLossBackward>)
[3969] Train Loss:    tensor(0.9633, device='cuda:0', grad_fn=<NllLossBackward>)
[3970] Train Loss:    tensor(0.9176, device='cuda:0', grad_fn=<NllLossBackward>)
[3971] Train Loss:    tensor(0.9407, device='cuda:0', grad_fn=<NllLossBackward>)
[3972] Train Loss:    tensor(0.9724, device='cuda:0', grad_fn=<NllLossBackward>)
[3973] Train Loss:    tensor(0.9728, device='cuda:0', grad_fn=<NllLossBackward>)
[3974] Train Loss:    tensor(0.9602, device='cuda:0', grad_fn=<NllLossBackward>)
[3975] Train Loss:    tensor(1.0125, device='cuda:0', grad_fn=<NllLossBackward>)
[3976] Train Loss:    tensor(0.9699, device='cuda:0', grad_fn=<NllLossBackward>)
[3977] Train Loss:    tensor(1.0053, device='cuda:0', grad_fn=<NllLossBackward>)
[3978] Train Loss:    tensor(0.9428, device='cuda:0', grad_fn=<NllLossBackward>)
[3979] Train Loss:    tensor(0.9685, device='cuda:0', grad_fn=<NllLossBackward>)
[3980] Train Loss:    tensor(0.9717, device='cuda:0', grad_fn=<NllLossBackward>)
[3981] Train Loss:    tensor(0.9361, device='cuda:0', grad_fn=<NllLossBackward>)
[3982] Train Loss:    tensor(0.9238, device='cuda:0', grad_fn=<NllLossBackward>)
[3983] Train Loss:    tensor(0.9621, device='cuda:0', grad_fn=<NllLossBackward>)
[3984] Train Loss:    tensor(0.9377, device='cuda:0', grad_fn=<NllLossBackward>)
[3985] Train Loss:    tensor(0.9728, device='cuda:0', grad_fn=<NllLossBackward>)
[3986] Train Loss:    tensor(0.9596, device='cuda:0', grad_fn=<NllLossBackward>)
[3987] Train Loss:    tensor(0.9435, device='cuda:0', grad_fn=<NllLossBackward>)
[3988] Train Loss:    tensor(0.9702, device='cuda:0', grad_fn=<NllLossBackward>)
[3989] Train Loss:    tensor(0.9567, device='cuda:0', grad_fn=<NllLossBackward>)
[3990] Train Loss:    tensor(0.9683, device='cuda:0', grad_fn=<NllLossBackward>)
[3991] Train Loss:    tensor(0.9495, device='cuda:0', grad_fn=<NllLossBackward>)
[3992] Train Loss:    tensor(0.9574, device='cuda:0', grad_fn=<NllLossBackward>)
[3993] Train Loss:    tensor(0.9575, device='cuda:0', grad_fn=<NllLossBackward>)
[3994] Train Loss:    tensor(0.9554, device='cuda:0', grad_fn=<NllLossBackward>)
[3995] Train Loss:    tensor(0.9696, device='cuda:0', grad_fn=<NllLossBackward>)
[3996] Train Loss:    tensor(1.0109, device='cuda:0', grad_fn=<NllLossBackward>)
[3997] Train Loss:    tensor(0.9842, device='cuda:0', grad_fn=<NllLossBackward>)
[3998] Train Loss:    tensor(0.9972, device='cuda:0', grad_fn=<NllLossBackward>)
[3999] Train Loss:    tensor(0.9610, device='cuda:0', grad_fn=<NllLossBackward>)
[4000] Train Loss:    tensor(0.9511, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4001] Train Loss:    tensor(1.0108, device='cuda:0', grad_fn=<NllLossBackward>)
[4002] Train Loss:    tensor(0.9648, device='cuda:0', grad_fn=<NllLossBackward>)
[4003] Train Loss:    tensor(0.9461, device='cuda:0', grad_fn=<NllLossBackward>)
[4004] Train Loss:    tensor(0.9427, device='cuda:0', grad_fn=<NllLossBackward>)
[4005] Train Loss:    tensor(0.9222, device='cuda:0', grad_fn=<NllLossBackward>)
[4006] Train Loss:    tensor(1.0407, device='cuda:0', grad_fn=<NllLossBackward>)
[4007] Train Loss:    tensor(0.9337, device='cuda:0', grad_fn=<NllLossBackward>)
[4008] Train Loss:    tensor(1.0230, device='cuda:0', grad_fn=<NllLossBackward>)
[4009] Train Loss:    tensor(0.9931, device='cuda:0', grad_fn=<NllLossBackward>)
[4010] Train Loss:    tensor(1.0152, device='cuda:0', grad_fn=<NllLossBackward>)
[4011] Train Loss:    tensor(0.9852, device='cuda:0', grad_fn=<NllLossBackward>)
[4012] Train Loss:    tensor(0.9505, device='cuda:0', grad_fn=<NllLossBackward>)
[4013] Train Loss:    tensor(0.9575, device='cuda:0', grad_fn=<NllLossBackward>)
[4014] Train Loss:    tensor(0.9529, device='cuda:0', grad_fn=<NllLossBackward>)
[4015] Train Loss:    tensor(0.9723, device='cuda:0', grad_fn=<NllLossBackward>)
[4016] Train Loss:    tensor(0.9890, device='cuda:0', grad_fn=<NllLossBackward>)
[4017] Train Loss:    tensor(0.9567, device='cuda:0', grad_fn=<NllLossBackward>)
[4018] Train Loss:    tensor(0.9507, device='cuda:0', grad_fn=<NllLossBackward>)
[4019] Train Loss:    tensor(0.9558, device='cuda:0', grad_fn=<NllLossBackward>)
[4020] Train Loss:    tensor(0.9748, device='cuda:0', grad_fn=<NllLossBackward>)
[4021] Train Loss:    tensor(1.0061, device='cuda:0', grad_fn=<NllLossBackward>)
[4022] Train Loss:    tensor(1.0029, device='cuda:0', grad_fn=<NllLossBackward>)
[4023] Train Loss:    tensor(0.9619, device='cuda:0', grad_fn=<NllLossBackward>)
[4024] Train Loss:    tensor(0.9607, device='cuda:0', grad_fn=<NllLossBackward>)
[4025] Train Loss:    tensor(0.9611, device='cuda:0', grad_fn=<NllLossBackward>)
[4026] Train Loss:    tensor(0.9687, device='cuda:0', grad_fn=<NllLossBackward>)
[4027] Train Loss:    tensor(0.9418, device='cuda:0', grad_fn=<NllLossBackward>)
[4028] Train Loss:    tensor(0.9773, device='cuda:0', grad_fn=<NllLossBackward>)
[4029] Train Loss:    tensor(0.9831, device='cuda:0', grad_fn=<NllLossBackward>)
[4030] Train Loss:    tensor(0.9813, device='cuda:0', grad_fn=<NllLossBackward>)
[4031] Train Loss:    tensor(0.9474, device='cuda:0', grad_fn=<NllLossBackward>)
[4032] Train Loss:    tensor(0.9399, device='cuda:0', grad_fn=<NllLossBackward>)
[4033] Train Loss:    tensor(0.9676, device='cuda:0', grad_fn=<NllLossBackward>)
[4034] Train Loss:    tensor(0.9868, device='cuda:0', grad_fn=<NllLossBackward>)
[4035] Train Loss:    tensor(0.9679, device='cuda:0', grad_fn=<NllLossBackward>)
[4036] Train Loss:    tensor(1.0054, device='cuda:0', grad_fn=<NllLossBackward>)
[4037] Train Loss:    tensor(0.9649, device='cuda:0', grad_fn=<NllLossBackward>)
[4038] Train Loss:    tensor(0.9774, device='cuda:0', grad_fn=<NllLossBackward>)
[4039] Train Loss:    tensor(0.9818, device='cuda:0', grad_fn=<NllLossBackward>)
[4040] Train Loss:    tensor(0.9639, device='cuda:0', grad_fn=<NllLossBackward>)
[4041] Train Loss:    tensor(0.9552, device='cuda:0', grad_fn=<NllLossBackward>)
[4042] Train Loss:    tensor(0.9714, device='cuda:0', grad_fn=<NllLossBackward>)
[4043] Train Loss:    tensor(0.9634, device='cuda:0', grad_fn=<NllLossBackward>)
[4044] Train Loss:    tensor(0.9469, device='cuda:0', grad_fn=<NllLossBackward>)
[4045] Train Loss:    tensor(0.9656, device='cuda:0', grad_fn=<NllLossBackward>)
[4046] Train Loss:    tensor(0.9491, device='cuda:0', grad_fn=<NllLossBackward>)
[4047] Train Loss:    tensor(1.0102, device='cuda:0', grad_fn=<NllLossBackward>)
[4048] Train Loss:    tensor(0.9749, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4049] Train Loss:    tensor(0.9677, device='cuda:0', grad_fn=<NllLossBackward>)
[4050] Train Loss:    tensor(0.9465, device='cuda:0', grad_fn=<NllLossBackward>)
[4051] Train Loss:    tensor(0.9308, device='cuda:0', grad_fn=<NllLossBackward>)
[4052] Train Loss:    tensor(0.9585, device='cuda:0', grad_fn=<NllLossBackward>)
[4053] Train Loss:    tensor(0.9715, device='cuda:0', grad_fn=<NllLossBackward>)
[4054] Train Loss:    tensor(0.9784, device='cuda:0', grad_fn=<NllLossBackward>)
[4055] Train Loss:    tensor(0.9561, device='cuda:0', grad_fn=<NllLossBackward>)
[4056] Train Loss:    tensor(0.9289, device='cuda:0', grad_fn=<NllLossBackward>)
[4057] Train Loss:    tensor(0.9672, device='cuda:0', grad_fn=<NllLossBackward>)
[4058] Train Loss:    tensor(1.0003, device='cuda:0', grad_fn=<NllLossBackward>)
[4059] Train Loss:    tensor(0.9666, device='cuda:0', grad_fn=<NllLossBackward>)
[4060] Train Loss:    tensor(0.9407, device='cuda:0', grad_fn=<NllLossBackward>)
[4061] Train Loss:    tensor(0.9664, device='cuda:0', grad_fn=<NllLossBackward>)
[4062] Train Loss:    tensor(0.9417, device='cuda:0', grad_fn=<NllLossBackward>)
[4063] Train Loss:    tensor(0.9713, device='cuda:0', grad_fn=<NllLossBackward>)
[4064] Train Loss:    tensor(0.9628, device='cuda:0', grad_fn=<NllLossBackward>)
[4065] Train Loss:    tensor(0.9319, device='cuda:0', grad_fn=<NllLossBackward>)
[4066] Train Loss:    tensor(0.9485, device='cuda:0', grad_fn=<NllLossBackward>)
[4067] Train Loss:    tensor(0.9342, device='cuda:0', grad_fn=<NllLossBackward>)
[4068] Train Loss:    tensor(0.9418, device='cuda:0', grad_fn=<NllLossBackward>)
[4069] Train Loss:    tensor(0.9530, device='cuda:0', grad_fn=<NllLossBackward>)
[4070] Train Loss:    tensor(0.9929, device='cuda:0', grad_fn=<NllLossBackward>)
[4071] Train Loss:    tensor(0.9958, device='cuda:0', grad_fn=<NllLossBackward>)
[4072] Train Loss:    tensor(0.9399, device='cuda:0', grad_fn=<NllLossBackward>)
[4073] Train Loss:    tensor(0.9615, device='cuda:0', grad_fn=<NllLossBackward>)
[4074] Train Loss:    tensor(0.9953, device='cuda:0', grad_fn=<NllLossBackward>)
[4075] Train Loss:    tensor(0.9788, device='cuda:0', grad_fn=<NllLossBackward>)
[4076] Train Loss:    tensor(0.9676, device='cuda:0', grad_fn=<NllLossBackward>)
[4077] Train Loss:    tensor(0.9479, device='cuda:0', grad_fn=<NllLossBackward>)
[4078] Train Loss:    tensor(0.9911, device='cuda:0', grad_fn=<NllLossBackward>)
[4079] Train Loss:    tensor(0.9869, device='cuda:0', grad_fn=<NllLossBackward>)
[4080] Train Loss:    tensor(0.9313, device='cuda:0', grad_fn=<NllLossBackward>)
[4081] Train Loss:    tensor(0.9370, device='cuda:0', grad_fn=<NllLossBackward>)
[4082] Train Loss:    tensor(0.9669, device='cuda:0', grad_fn=<NllLossBackward>)
[4083] Train Loss:    tensor(0.9327, device='cuda:0', grad_fn=<NllLossBackward>)
[4084] Train Loss:    tensor(0.9500, device='cuda:0', grad_fn=<NllLossBackward>)
[4085] Train Loss:    tensor(0.9871, device='cuda:0', grad_fn=<NllLossBackward>)
[4086] Train Loss:    tensor(0.9774, device='cuda:0', grad_fn=<NllLossBackward>)
[4087] Train Loss:    tensor(0.9693, device='cuda:0', grad_fn=<NllLossBackward>)
[4088] Train Loss:    tensor(0.9819, device='cuda:0', grad_fn=<NllLossBackward>)
[4089] Train Loss:    tensor(0.9365, device='cuda:0', grad_fn=<NllLossBackward>)
[4090] Train Loss:    tensor(0.9489, device='cuda:0', grad_fn=<NllLossBackward>)
[4091] Train Loss:    tensor(0.9780, device='cuda:0', grad_fn=<NllLossBackward>)
[4092] Train Loss:    tensor(1.0164, device='cuda:0', grad_fn=<NllLossBackward>)
[4093] Train Loss:    tensor(0.9581, device='cuda:0', grad_fn=<NllLossBackward>)
[4094] Train Loss:    tensor(0.9734, device='cuda:0', grad_fn=<NllLossBackward>)
[4095] Train Loss:    tensor(0.9668, device='cuda:0', grad_fn=<NllLossBackward>)
[4096] Train Loss:    tensor(0.9603, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4097] Train Loss:    tensor(0.9530, device='cuda:0', grad_fn=<NllLossBackward>)
[4098] Train Loss:    tensor(0.9302, device='cuda:0', grad_fn=<NllLossBackward>)
[4099] Train Loss:    tensor(0.9508, device='cuda:0', grad_fn=<NllLossBackward>)
[4100] Train Loss:    tensor(0.9624, device='cuda:0', grad_fn=<NllLossBackward>)
[4101] Train Loss:    tensor(0.9726, device='cuda:0', grad_fn=<NllLossBackward>)
[4102] Train Loss:    tensor(0.9659, device='cuda:0', grad_fn=<NllLossBackward>)
[4103] Train Loss:    tensor(0.9380, device='cuda:0', grad_fn=<NllLossBackward>)
[4104] Train Loss:    tensor(0.9840, device='cuda:0', grad_fn=<NllLossBackward>)
[4105] Train Loss:    tensor(0.9663, device='cuda:0', grad_fn=<NllLossBackward>)
[4106] Train Loss:    tensor(0.9489, device='cuda:0', grad_fn=<NllLossBackward>)
[4107] Train Loss:    tensor(0.9601, device='cuda:0', grad_fn=<NllLossBackward>)
[4108] Train Loss:    tensor(0.9915, device='cuda:0', grad_fn=<NllLossBackward>)
[4109] Train Loss:    tensor(0.9212, device='cuda:0', grad_fn=<NllLossBackward>)
[4110] Train Loss:    tensor(0.9542, device='cuda:0', grad_fn=<NllLossBackward>)
[4111] Train Loss:    tensor(0.9433, device='cuda:0', grad_fn=<NllLossBackward>)
[4112] Train Loss:    tensor(0.9474, device='cuda:0', grad_fn=<NllLossBackward>)
[4113] Train Loss:    tensor(0.9593, device='cuda:0', grad_fn=<NllLossBackward>)
[4114] Train Loss:    tensor(0.9586, device='cuda:0', grad_fn=<NllLossBackward>)
[4115] Train Loss:    tensor(0.9405, device='cuda:0', grad_fn=<NllLossBackward>)
[4116] Train Loss:    tensor(0.9444, device='cuda:0', grad_fn=<NllLossBackward>)
[4117] Train Loss:    tensor(0.9599, device='cuda:0', grad_fn=<NllLossBackward>)
[4118] Train Loss:    tensor(0.9546, device='cuda:0', grad_fn=<NllLossBackward>)
[4119] Train Loss:    tensor(0.9358, device='cuda:0', grad_fn=<NllLossBackward>)
[4120] Train Loss:    tensor(0.9470, device='cuda:0', grad_fn=<NllLossBackward>)
[4121] Train Loss:    tensor(0.9374, device='cuda:0', grad_fn=<NllLossBackward>)
[4122] Train Loss:    tensor(0.9789, device='cuda:0', grad_fn=<NllLossBackward>)
[4123] Train Loss:    tensor(0.9461, device='cuda:0', grad_fn=<NllLossBackward>)
[4124] Train Loss:    tensor(0.9513, device='cuda:0', grad_fn=<NllLossBackward>)
[4125] Train Loss:    tensor(0.9387, device='cuda:0', grad_fn=<NllLossBackward>)
[4126] Train Loss:    tensor(0.9847, device='cuda:0', grad_fn=<NllLossBackward>)
[4127] Train Loss:    tensor(0.9699, device='cuda:0', grad_fn=<NllLossBackward>)
[4128] Train Loss:    tensor(0.9368, device='cuda:0', grad_fn=<NllLossBackward>)
[4129] Train Loss:    tensor(0.9306, device='cuda:0', grad_fn=<NllLossBackward>)
[4130] Train Loss:    tensor(0.9520, device='cuda:0', grad_fn=<NllLossBackward>)
[4131] Train Loss:    tensor(0.9221, device='cuda:0', grad_fn=<NllLossBackward>)
[4132] Train Loss:    tensor(0.9740, device='cuda:0', grad_fn=<NllLossBackward>)
[4133] Train Loss:    tensor(0.9747, device='cuda:0', grad_fn=<NllLossBackward>)
[4134] Train Loss:    tensor(0.9539, device='cuda:0', grad_fn=<NllLossBackward>)
[4135] Train Loss:    tensor(0.9797, device='cuda:0', grad_fn=<NllLossBackward>)
[4136] Train Loss:    tensor(0.9279, device='cuda:0', grad_fn=<NllLossBackward>)
[4137] Train Loss:    tensor(0.9583, device='cuda:0', grad_fn=<NllLossBackward>)
[4138] Train Loss:    tensor(0.9679, device='cuda:0', grad_fn=<NllLossBackward>)
[4139] Train Loss:    tensor(0.9246, device='cuda:0', grad_fn=<NllLossBackward>)
[4140] Train Loss:    tensor(0.9836, device='cuda:0', grad_fn=<NllLossBackward>)
[4141] Train Loss:    tensor(0.9552, device='cuda:0', grad_fn=<NllLossBackward>)
[4142] Train Loss:    tensor(0.9181, device='cuda:0', grad_fn=<NllLossBackward>)
[4143] Train Loss:    tensor(0.9391, device='cuda:0', grad_fn=<NllLossBackward>)
[4144] Train Loss:    tensor(0.9769, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4145] Train Loss:   tensor(0.9561, device='cuda:0', grad_fn=<NllLossBackward>)
[4146] Train Loss:   tensor(0.9958, device='cuda:0', grad_fn=<NllLossBackward>)
[4147] Train Loss:   tensor(0.9540, device='cuda:0', grad_fn=<NllLossBackward>)
[4148] Train Loss:   tensor(0.9653, device='cuda:0', grad_fn=<NllLossBackward>)
[4149] Train Loss:   tensor(0.9982, device='cuda:0', grad_fn=<NllLossBackward>)
[4150] Train Loss:   tensor(0.9591, device='cuda:0', grad_fn=<NllLossBackward>)
[4151] Train Loss:   tensor(0.9820, device='cuda:0', grad_fn=<NllLossBackward>)
[4152] Train Loss:   tensor(0.9661, device='cuda:0', grad_fn=<NllLossBackward>)
[4153] Train Loss:   tensor(0.9368, device='cuda:0', grad_fn=<NllLossBackward>)
[4154] Train Loss:   tensor(0.9770, device='cuda:0', grad_fn=<NllLossBackward>)
[4155] Train Loss:   tensor(0.9522, device='cuda:0', grad_fn=<NllLossBackward>)
[4156] Train Loss:   tensor(0.9500, device='cuda:0', grad_fn=<NllLossBackward>)
[4157] Train Loss:   tensor(0.9338, device='cuda:0', grad_fn=<NllLossBackward>)
[4158] Train Loss:   tensor(0.9236, device='cuda:0', grad_fn=<NllLossBackward>)
[4159] Train Loss:   tensor(0.9583, device='cuda:0', grad_fn=<NllLossBackward>)
[4160] Train Loss:   tensor(0.9517, device='cuda:0', grad_fn=<NllLossBackward>)
[4161] Train Loss:   tensor(0.9539, device='cuda:0', grad_fn=<NllLossBackward>)
[4162] Train Loss:   tensor(0.9633, device='cuda:0', grad_fn=<NllLossBackward>)
[4163] Train Loss:   tensor(0.9609, device='cuda:0', grad_fn=<NllLossBackward>)
[4164] Train Loss:   tensor(0.9538, device='cuda:0', grad_fn=<NllLossBackward>)
[4165] Train Loss:   tensor(0.9691, device='cuda:0', grad_fn=<NllLossBackward>)
[4166] Train Loss:   tensor(0.9554, device='cuda:0', grad_fn=<NllLossBackward>)
[4167] Train Loss:   tensor(0.9061, device='cuda:0', grad_fn=<NllLossBackward>)
[4168] Train Loss:   tensor(0.9394, device='cuda:0', grad_fn=<NllLossBackward>)
[4169] Train Loss:   tensor(0.9635, device='cuda:0', grad_fn=<NllLossBackward>)
[4170] Train Loss:   tensor(0.9201, device='cuda:0', grad_fn=<NllLossBackward>)
[4171] Train Loss:   tensor(0.9297, device='cuda:0', grad_fn=<NllLossBackward>)
[4172] Train Loss:   tensor(0.9543, device='cuda:0', grad_fn=<NllLossBackward>)
[4173] Train Loss:   tensor(0.9666, device='cuda:0', grad_fn=<NllLossBackward>)
[4174] Train Loss:   tensor(0.9813, device='cuda:0', grad_fn=<NllLossBackward>)
[4175] Train Loss:   tensor(0.9536, device='cuda:0', grad_fn=<NllLossBackward>)
[4176] Train Loss:   tensor(0.9493, device='cuda:0', grad_fn=<NllLossBackward>)
[4177] Train Loss:   tensor(0.9874, device='cuda:0', grad_fn=<NllLossBackward>)
[4178] Train Loss:   tensor(0.9873, device='cuda:0', grad_fn=<NllLossBackward>)
[4179] Train Loss:   tensor(0.9909, device='cuda:0', grad_fn=<NllLossBackward>)
[4180] Train Loss:   tensor(0.9682, device='cuda:0', grad_fn=<NllLossBackward>)
[4181] Train Loss:   tensor(0.9421, device='cuda:0', grad_fn=<NllLossBackward>)
[4182] Train Loss:   tensor(0.9435, device='cuda:0', grad_fn=<NllLossBackward>)
[4183] Train Loss:   tensor(0.9318, device='cuda:0', grad_fn=<NllLossBackward>)
[4184] Train Loss:   tensor(0.9491, device='cuda:0', grad_fn=<NllLossBackward>)
[4185] Train Loss:   tensor(0.9465, device='cuda:0', grad_fn=<NllLossBackward>)
[4186] Train Loss:   tensor(0.9968, device='cuda:0', grad_fn=<NllLossBackward>)
[4187] Train Loss:   tensor(0.9331, device='cuda:0', grad_fn=<NllLossBackward>)
[4188] Train Loss:   tensor(0.9302, device='cuda:0', grad_fn=<NllLossBackward>)
[4189] Train Loss:   tensor(1.0050, device='cuda:0', grad_fn=<NllLossBackward>)
[4190] Train Loss:   tensor(0.9613, device='cuda:0', grad_fn=<NllLossBackward>)
[4191] Train Loss:   tensor(0.9948, device='cuda:0', grad_fn=<NllLossBackward>)
[4192] Train Loss:   tensor(0.9401, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4193] Train Loss:    tensor(0.9373, device='cuda:0', grad_fn=<NllLossBackward>)
[4194] Train Loss:    tensor(0.9755, device='cuda:0', grad_fn=<NllLossBackward>)
[4195] Train Loss:    tensor(1.0007, device='cuda:0', grad_fn=<NllLossBackward>)
[4196] Train Loss:    tensor(0.9834, device='cuda:0', grad_fn=<NllLossBackward>)
[4197] Train Loss:    tensor(0.9395, device='cuda:0', grad_fn=<NllLossBackward>)
[4198] Train Loss:    tensor(0.9572, device='cuda:0', grad_fn=<NllLossBackward>)
[4199] Train Loss:    tensor(0.9716, device='cuda:0', grad_fn=<NllLossBackward>)
[4200] Train Loss:    tensor(0.9597, device='cuda:0', grad_fn=<NllLossBackward>)
[4201] Train Loss:    tensor(0.9552, device='cuda:0', grad_fn=<NllLossBackward>)
[4202] Train Loss:    tensor(0.9459, device='cuda:0', grad_fn=<NllLossBackward>)
[4203] Train Loss:    tensor(0.9415, device='cuda:0', grad_fn=<NllLossBackward>)
[4204] Train Loss:    tensor(0.9710, device='cuda:0', grad_fn=<NllLossBackward>)
[4205] Train Loss:    tensor(0.9739, device='cuda:0', grad_fn=<NllLossBackward>)
[4206] Train Loss:    tensor(0.9641, device='cuda:0', grad_fn=<NllLossBackward>)
[4207] Train Loss:    tensor(0.9616, device='cuda:0', grad_fn=<NllLossBackward>)
[4208] Train Loss:    tensor(0.9826, device='cuda:0', grad_fn=<NllLossBackward>)
[4209] Train Loss:    tensor(0.9628, device='cuda:0', grad_fn=<NllLossBackward>)
[4210] Train Loss:    tensor(0.9555, device='cuda:0', grad_fn=<NllLossBackward>)
[4211] Train Loss:    tensor(0.9674, device='cuda:0', grad_fn=<NllLossBackward>)
[4212] Train Loss:    tensor(0.9748, device='cuda:0', grad_fn=<NllLossBackward>)
[4213] Train Loss:    tensor(0.9473, device='cuda:0', grad_fn=<NllLossBackward>)
[4214] Train Loss:    tensor(0.9399, device='cuda:0', grad_fn=<NllLossBackward>)
[4215] Train Loss:    tensor(0.9452, device='cuda:0', grad_fn=<NllLossBackward>)
[4216] Train Loss:    tensor(0.9995, device='cuda:0', grad_fn=<NllLossBackward>)
[4217] Train Loss:    tensor(0.9467, device='cuda:0', grad_fn=<NllLossBackward>)
[4218] Train Loss:    tensor(0.9763, device='cuda:0', grad_fn=<NllLossBackward>)
[4219] Train Loss:    tensor(0.9659, device='cuda:0', grad_fn=<NllLossBackward>)
[4220] Train Loss:    tensor(0.9501, device='cuda:0', grad_fn=<NllLossBackward>)
[4221] Train Loss:    tensor(0.9419, device='cuda:0', grad_fn=<NllLossBackward>)
[4222] Train Loss:    tensor(0.9433, device='cuda:0', grad_fn=<NllLossBackward>)
[4223] Train Loss:    tensor(0.9530, device='cuda:0', grad_fn=<NllLossBackward>)
[4224] Train Loss:    tensor(0.9513, device='cuda:0', grad_fn=<NllLossBackward>)
[4225] Train Loss:    tensor(0.9627, device='cuda:0', grad_fn=<NllLossBackward>)
[4226] Train Loss:    tensor(0.9564, device='cuda:0', grad_fn=<NllLossBackward>)
[4227] Train Loss:    tensor(0.9573, device='cuda:0', grad_fn=<NllLossBackward>)
[4228] Train Loss:    tensor(0.9148, device='cuda:0', grad_fn=<NllLossBackward>)
[4229] Train Loss:    tensor(0.9234, device='cuda:0', grad_fn=<NllLossBackward>)
[4230] Train Loss:    tensor(0.9713, device='cuda:0', grad_fn=<NllLossBackward>)
[4231] Train Loss:    tensor(0.9331, device='cuda:0', grad_fn=<NllLossBackward>)
[4232] Train Loss:    tensor(0.9523, device='cuda:0', grad_fn=<NllLossBackward>)
[4233] Train Loss:    tensor(0.9478, device='cuda:0', grad_fn=<NllLossBackward>)
[4234] Train Loss:    tensor(0.9499, device='cuda:0', grad_fn=<NllLossBackward>)
[4235] Train Loss:    tensor(0.9537, device='cuda:0', grad_fn=<NllLossBackward>)
[4236] Train Loss:    tensor(0.9213, device='cuda:0', grad_fn=<NllLossBackward>)
[4237] Train Loss:    tensor(0.9724, device='cuda:0', grad_fn=<NllLossBackward>)
[4238] Train Loss:    tensor(0.9558, device='cuda:0', grad_fn=<NllLossBackward>)
[4239] Train Loss:    tensor(0.9860, device='cuda:0', grad_fn=<NllLossBackward>)
[4240] Train Loss:    tensor(0.9696, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4241] Train Loss:    tensor(0.9514, device='cuda:0', grad_fn=<NllLossBackward>)
[4242] Train Loss:    tensor(0.9174, device='cuda:0', grad_fn=<NllLossBackward>)
[4243] Train Loss:    tensor(0.9447, device='cuda:0', grad_fn=<NllLossBackward>)
[4244] Train Loss:    tensor(0.9773, device='cuda:0', grad_fn=<NllLossBackward>)
[4245] Train Loss:    tensor(0.9668, device='cuda:0', grad_fn=<NllLossBackward>)
[4246] Train Loss:    tensor(0.9395, device='cuda:0', grad_fn=<NllLossBackward>)
[4247] Train Loss:    tensor(0.9172, device='cuda:0', grad_fn=<NllLossBackward>)
[4248] Train Loss:    tensor(0.9490, device='cuda:0', grad_fn=<NllLossBackward>)
[4249] Train Loss:    tensor(0.9497, device='cuda:0', grad_fn=<NllLossBackward>)
[4250] Train Loss:    tensor(0.9731, device='cuda:0', grad_fn=<NllLossBackward>)
[4251] Train Loss:    tensor(0.9722, device='cuda:0', grad_fn=<NllLossBackward>)
[4252] Train Loss:    tensor(0.9116, device='cuda:0', grad_fn=<NllLossBackward>)
[4253] Train Loss:    tensor(0.9505, device='cuda:0', grad_fn=<NllLossBackward>)
[4254] Train Loss:    tensor(0.9471, device='cuda:0', grad_fn=<NllLossBackward>)
[4255] Train Loss:    tensor(0.9310, device='cuda:0', grad_fn=<NllLossBackward>)
[4256] Train Loss:    tensor(0.9662, device='cuda:0', grad_fn=<NllLossBackward>)
[4257] Train Loss:    tensor(0.9521, device='cuda:0', grad_fn=<NllLossBackward>)
[4258] Train Loss:    tensor(0.9445, device='cuda:0', grad_fn=<NllLossBackward>)
[4259] Train Loss:    tensor(0.9653, device='cuda:0', grad_fn=<NllLossBackward>)
[4260] Train Loss:    tensor(0.9715, device='cuda:0', grad_fn=<NllLossBackward>)
[4261] Train Loss:    tensor(0.9506, device='cuda:0', grad_fn=<NllLossBackward>)
[4262] Train Loss:    tensor(1.0269, device='cuda:0', grad_fn=<NllLossBackward>)
[4263] Train Loss:    tensor(0.9297, device='cuda:0', grad_fn=<NllLossBackward>)
[4264] Train Loss:    tensor(0.9643, device='cuda:0', grad_fn=<NllLossBackward>)
[4265] Train Loss:    tensor(0.9508, device='cuda:0', grad_fn=<NllLossBackward>)
[4266] Train Loss:    tensor(0.9377, device='cuda:0', grad_fn=<NllLossBackward>)
[4267] Train Loss:    tensor(0.9566, device='cuda:0', grad_fn=<NllLossBackward>)
[4268] Train Loss:    tensor(0.9389, device='cuda:0', grad_fn=<NllLossBackward>)
[4269] Train Loss:    tensor(0.9261, device='cuda:0', grad_fn=<NllLossBackward>)
[4270] Train Loss:    tensor(0.9255, device='cuda:0', grad_fn=<NllLossBackward>)
[4271] Train Loss:    tensor(0.9492, device='cuda:0', grad_fn=<NllLossBackward>)
[4272] Train Loss:    tensor(0.9745, device='cuda:0', grad_fn=<NllLossBackward>)
[4273] Train Loss:    tensor(0.9665, device='cuda:0', grad_fn=<NllLossBackward>)
[4274] Train Loss:    tensor(0.9849, device='cuda:0', grad_fn=<NllLossBackward>)
[4275] Train Loss:    tensor(0.9461, device='cuda:0', grad_fn=<NllLossBackward>)
[4276] Train Loss:    tensor(1.0028, device='cuda:0', grad_fn=<NllLossBackward>)
[4277] Train Loss:    tensor(0.9285, device='cuda:0', grad_fn=<NllLossBackward>)
[4278] Train Loss:    tensor(0.9426, device='cuda:0', grad_fn=<NllLossBackward>)
[4279] Train Loss:    tensor(0.9288, device='cuda:0', grad_fn=<NllLossBackward>)
[4280] Train Loss:    tensor(0.9070, device='cuda:0', grad_fn=<NllLossBackward>)
[4281] Train Loss:    tensor(0.9670, device='cuda:0', grad_fn=<NllLossBackward>)
[4282] Train Loss:    tensor(0.9691, device='cuda:0', grad_fn=<NllLossBackward>)
[4283] Train Loss:    tensor(0.9671, device='cuda:0', grad_fn=<NllLossBackward>)
[4284] Train Loss:    tensor(0.9678, device='cuda:0', grad_fn=<NllLossBackward>)
[4285] Train Loss:    tensor(0.9442, device='cuda:0', grad_fn=<NllLossBackward>)
[4286] Train Loss:    tensor(0.9547, device='cuda:0', grad_fn=<NllLossBackward>)
[4287] Train Loss:    tensor(0.9320, device='cuda:0', grad_fn=<NllLossBackward>)
[4288] Train Loss:    tensor(0.9535, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4289] Train Loss:    tensor(1.0137, device='cuda:0', grad_fn=<NllLossBackward>)
[4290] Train Loss:    tensor(0.9644, device='cuda:0', grad_fn=<NllLossBackward>)
[4291] Train Loss:    tensor(0.9380, device='cuda:0', grad_fn=<NllLossBackward>)
[4292] Train Loss:    tensor(0.9401, device='cuda:0', grad_fn=<NllLossBackward>)
[4293] Train Loss:    tensor(0.9645, device='cuda:0', grad_fn=<NllLossBackward>)
[4294] Train Loss:    tensor(0.9783, device='cuda:0', grad_fn=<NllLossBackward>)
[4295] Train Loss:    tensor(0.9577, device='cuda:0', grad_fn=<NllLossBackward>)
[4296] Train Loss:    tensor(0.9810, device='cuda:0', grad_fn=<NllLossBackward>)
[4297] Train Loss:    tensor(0.9451, device='cuda:0', grad_fn=<NllLossBackward>)
[4298] Train Loss:    tensor(0.9284, device='cuda:0', grad_fn=<NllLossBackward>)
[4299] Train Loss:    tensor(0.9381, device='cuda:0', grad_fn=<NllLossBackward>)
[4300] Train Loss:    tensor(0.9466, device='cuda:0', grad_fn=<NllLossBackward>)
[4301] Train Loss:    tensor(0.9371, device='cuda:0', grad_fn=<NllLossBackward>)
[4302] Train Loss:    tensor(0.9434, device='cuda:0', grad_fn=<NllLossBackward>)
[4303] Train Loss:    tensor(0.9568, device='cuda:0', grad_fn=<NllLossBackward>)
[4304] Train Loss:    tensor(0.9747, device='cuda:0', grad_fn=<NllLossBackward>)
[4305] Train Loss:    tensor(0.9484, device='cuda:0', grad_fn=<NllLossBackward>)
[4306] Train Loss:    tensor(0.9754, device='cuda:0', grad_fn=<NllLossBackward>)
[4307] Train Loss:    tensor(1.0077, device='cuda:0', grad_fn=<NllLossBackward>)
[4308] Train Loss:    tensor(0.9659, device='cuda:0', grad_fn=<NllLossBackward>)
[4309] Train Loss:    tensor(0.9615, device='cuda:0', grad_fn=<NllLossBackward>)
[4310] Train Loss:    tensor(0.9258, device='cuda:0', grad_fn=<NllLossBackward>)
[4311] Train Loss:    tensor(0.9558, device='cuda:0', grad_fn=<NllLossBackward>)
[4312] Train Loss:    tensor(0.9698, device='cuda:0', grad_fn=<NllLossBackward>)
[4313] Train Loss:    tensor(0.9651, device='cuda:0', grad_fn=<NllLossBackward>)
[4314] Train Loss:    tensor(0.9541, device='cuda:0', grad_fn=<NllLossBackward>)
[4315] Train Loss:    tensor(0.9501, device='cuda:0', grad_fn=<NllLossBackward>)
[4316] Train Loss:    tensor(0.9492, device='cuda:0', grad_fn=<NllLossBackward>)
[4317] Train Loss:    tensor(0.9715, device='cuda:0', grad_fn=<NllLossBackward>)
[4318] Train Loss:    tensor(0.9656, device='cuda:0', grad_fn=<NllLossBackward>)
[4319] Train Loss:    tensor(0.9550, device='cuda:0', grad_fn=<NllLossBackward>)
[4320] Train Loss:    tensor(0.9284, device='cuda:0', grad_fn=<NllLossBackward>)
[4321] Train Loss:    tensor(0.9331, device='cuda:0', grad_fn=<NllLossBackward>)
[4322] Train Loss:    tensor(0.9493, device='cuda:0', grad_fn=<NllLossBackward>)
[4323] Train Loss:    tensor(0.9449, device='cuda:0', grad_fn=<NllLossBackward>)
[4324] Train Loss:    tensor(0.9452, device='cuda:0', grad_fn=<NllLossBackward>)
[4325] Train Loss:    tensor(0.9621, device='cuda:0', grad_fn=<NllLossBackward>)
[4326] Train Loss:    tensor(0.9702, device='cuda:0', grad_fn=<NllLossBackward>)
[4327] Train Loss:    tensor(0.9705, device='cuda:0', grad_fn=<NllLossBackward>)
[4328] Train Loss:    tensor(0.9465, device='cuda:0', grad_fn=<NllLossBackward>)
[4329] Train Loss:    tensor(0.9585, device='cuda:0', grad_fn=<NllLossBackward>)
[4330] Train Loss:    tensor(0.9470, device='cuda:0', grad_fn=<NllLossBackward>)
[4331] Train Loss:    tensor(0.9315, device='cuda:0', grad_fn=<NllLossBackward>)
[4332] Train Loss:    tensor(0.9511, device='cuda:0', grad_fn=<NllLossBackward>)
[4333] Train Loss:    tensor(0.9632, device='cuda:0', grad_fn=<NllLossBackward>)
[4334] Train Loss:    tensor(0.9306, device='cuda:0', grad_fn=<NllLossBackward>)
[4335] Train Loss:    tensor(0.9552, device='cuda:0', grad_fn=<NllLossBackward>)
[4336] Train Loss:    tensor(0.9662, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4337] Train Loss:    tensor(0.9758, device='cuda:0', grad_fn=<NllLossBackward>)
[4338] Train Loss:    tensor(0.9876, device='cuda:0', grad_fn=<NllLossBackward>)
[4339] Train Loss:    tensor(0.9617, device='cuda:0', grad_fn=<NllLossBackward>)
[4340] Train Loss:    tensor(0.9705, device='cuda:0', grad_fn=<NllLossBackward>)
[4341] Train Loss:    tensor(0.9637, device='cuda:0', grad_fn=<NllLossBackward>)
[4342] Train Loss:    tensor(0.9320, device='cuda:0', grad_fn=<NllLossBackward>)
[4343] Train Loss:    tensor(0.9600, device='cuda:0', grad_fn=<NllLossBackward>)
[4344] Train Loss:    tensor(0.9758, device='cuda:0', grad_fn=<NllLossBackward>)
[4345] Train Loss:    tensor(0.9376, device='cuda:0', grad_fn=<NllLossBackward>)
[4346] Train Loss:    tensor(0.9579, device='cuda:0', grad_fn=<NllLossBackward>)
[4347] Train Loss:    tensor(0.9391, device='cuda:0', grad_fn=<NllLossBackward>)
[4348] Train Loss:    tensor(0.9747, device='cuda:0', grad_fn=<NllLossBackward>)
[4349] Train Loss:    tensor(0.9410, device='cuda:0', grad_fn=<NllLossBackward>)
[4350] Train Loss:    tensor(0.9656, device='cuda:0', grad_fn=<NllLossBackward>)
[4351] Train Loss:    tensor(0.9595, device='cuda:0', grad_fn=<NllLossBackward>)
[4352] Train Loss:    tensor(0.9311, device='cuda:0', grad_fn=<NllLossBackward>)
[4353] Train Loss:    tensor(0.9538, device='cuda:0', grad_fn=<NllLossBackward>)
[4354] Train Loss:    tensor(0.9512, device='cuda:0', grad_fn=<NllLossBackward>)
[4355] Train Loss:    tensor(0.9584, device='cuda:0', grad_fn=<NllLossBackward>)
[4356] Train Loss:    tensor(0.9593, device='cuda:0', grad_fn=<NllLossBackward>)
[4357] Train Loss:    tensor(0.9618, device='cuda:0', grad_fn=<NllLossBackward>)
[4358] Train Loss:    tensor(0.9298, device='cuda:0', grad_fn=<NllLossBackward>)
[4359] Train Loss:    tensor(0.9331, device='cuda:0', grad_fn=<NllLossBackward>)
[4360] Train Loss:    tensor(0.9865, device='cuda:0', grad_fn=<NllLossBackward>)
[4361] Train Loss:    tensor(0.9277, device='cuda:0', grad_fn=<NllLossBackward>)
[4362] Train Loss:    tensor(0.9631, device='cuda:0', grad_fn=<NllLossBackward>)
[4363] Train Loss:    tensor(0.9453, device='cuda:0', grad_fn=<NllLossBackward>)
[4364] Train Loss:    tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[4365] Train Loss:    tensor(0.9605, device='cuda:0', grad_fn=<NllLossBackward>)
[4366] Train Loss:    tensor(0.9469, device='cuda:0', grad_fn=<NllLossBackward>)
[4367] Train Loss:    tensor(0.9423, device='cuda:0', grad_fn=<NllLossBackward>)
[4368] Train Loss:    tensor(0.9693, device='cuda:0', grad_fn=<NllLossBackward>)
[4369] Train Loss:    tensor(0.9137, device='cuda:0', grad_fn=<NllLossBackward>)
[4370] Train Loss:    tensor(0.9463, device='cuda:0', grad_fn=<NllLossBackward>)
[4371] Train Loss:    tensor(0.9382, device='cuda:0', grad_fn=<NllLossBackward>)
[4372] Train Loss:    tensor(0.9093, device='cuda:0', grad_fn=<NllLossBackward>)
[4373] Train Loss:    tensor(0.9838, device='cuda:0', grad_fn=<NllLossBackward>)
[4374] Train Loss:    tensor(0.9444, device='cuda:0', grad_fn=<NllLossBackward>)
[4375] Train Loss:    tensor(0.9135, device='cuda:0', grad_fn=<NllLossBackward>)
[4376] Train Loss:    tensor(0.9367, device='cuda:0', grad_fn=<NllLossBackward>)
[4377] Train Loss:    tensor(0.9336, device='cuda:0', grad_fn=<NllLossBackward>)
[4378] Train Loss:    tensor(0.9860, device='cuda:0', grad_fn=<NllLossBackward>)
[4379] Train Loss:    tensor(0.9364, device='cuda:0', grad_fn=<NllLossBackward>)
[4380] Train Loss:    tensor(0.9573, device='cuda:0', grad_fn=<NllLossBackward>)
[4381] Train Loss:    tensor(0.9465, device='cuda:0', grad_fn=<NllLossBackward>)
[4382] Train Loss:    tensor(0.9198, device='cuda:0', grad_fn=<NllLossBackward>)
[4383] Train Loss:    tensor(0.9427, device='cuda:0', grad_fn=<NllLossBackward>)
[4384] Train Loss:    tensor(0.9262, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4385] Train Loss:   tensor(0.9591, device='cuda:0', grad_fn=<NllLossBackward>)
[4386] Train Loss:   tensor(1.0044, device='cuda:0', grad_fn=<NllLossBackward>)
[4387] Train Loss:   tensor(0.9819, device='cuda:0', grad_fn=<NllLossBackward>)
[4388] Train Loss:   tensor(0.9342, device='cuda:0', grad_fn=<NllLossBackward>)
[4389] Train Loss:   tensor(0.9348, device='cuda:0', grad_fn=<NllLossBackward>)
[4390] Train Loss:   tensor(0.9414, device='cuda:0', grad_fn=<NllLossBackward>)
[4391] Train Loss:   tensor(0.9681, device='cuda:0', grad_fn=<NllLossBackward>)
[4392] Train Loss:   tensor(0.9647, device='cuda:0', grad_fn=<NllLossBackward>)
[4393] Train Loss:   tensor(0.9697, device='cuda:0', grad_fn=<NllLossBackward>)
[4394] Train Loss:   tensor(0.9763, device='cuda:0', grad_fn=<NllLossBackward>)
[4395] Train Loss:   tensor(0.9415, device='cuda:0', grad_fn=<NllLossBackward>)
[4396] Train Loss:   tensor(0.9649, device='cuda:0', grad_fn=<NllLossBackward>)
[4397] Train Loss:   tensor(0.9528, device='cuda:0', grad_fn=<NllLossBackward>)
[4398] Train Loss:   tensor(0.9374, device='cuda:0', grad_fn=<NllLossBackward>)
[4399] Train Loss:   tensor(0.9342, device='cuda:0', grad_fn=<NllLossBackward>)
[4400] Train Loss:   tensor(0.9893, device='cuda:0', grad_fn=<NllLossBackward>)
[4401] Train Loss:   tensor(0.9173, device='cuda:0', grad_fn=<NllLossBackward>)
[4402] Train Loss:   tensor(0.9367, device='cuda:0', grad_fn=<NllLossBackward>)
[4403] Train Loss:   tensor(0.9788, device='cuda:0', grad_fn=<NllLossBackward>)
[4404] Train Loss:   tensor(0.9589, device='cuda:0', grad_fn=<NllLossBackward>)
[4405] Train Loss:   tensor(0.9748, device='cuda:0', grad_fn=<NllLossBackward>)
[4406] Train Loss:   tensor(0.9608, device='cuda:0', grad_fn=<NllLossBackward>)
[4407] Train Loss:   tensor(0.9660, device='cuda:0', grad_fn=<NllLossBackward>)
[4408] Train Loss:   tensor(0.9347, device='cuda:0', grad_fn=<NllLossBackward>)
[4409] Train Loss:   tensor(0.9417, device='cuda:0', grad_fn=<NllLossBackward>)
[4410] Train Loss:   tensor(0.9781, device='cuda:0', grad_fn=<NllLossBackward>)
[4411] Train Loss:   tensor(0.9303, device='cuda:0', grad_fn=<NllLossBackward>)
[4412] Train Loss:   tensor(0.9316, device='cuda:0', grad_fn=<NllLossBackward>)
[4413] Train Loss:   tensor(0.9342, device='cuda:0', grad_fn=<NllLossBackward>)
[4414] Train Loss:   tensor(0.9554, device='cuda:0', grad_fn=<NllLossBackward>)
[4415] Train Loss:   tensor(0.9392, device='cuda:0', grad_fn=<NllLossBackward>)
[4416] Train Loss:   tensor(0.9601, device='cuda:0', grad_fn=<NllLossBackward>)
[4417] Train Loss:   tensor(0.9898, device='cuda:0', grad_fn=<NllLossBackward>)
[4418] Train Loss:   tensor(0.9250, device='cuda:0', grad_fn=<NllLossBackward>)
[4419] Train Loss:   tensor(0.9642, device='cuda:0', grad_fn=<NllLossBackward>)
[4420] Train Loss:   tensor(0.9515, device='cuda:0', grad_fn=<NllLossBackward>)
[4421] Train Loss:   tensor(0.9461, device='cuda:0', grad_fn=<NllLossBackward>)
[4422] Train Loss:   tensor(0.9499, device='cuda:0', grad_fn=<NllLossBackward>)
[4423] Train Loss:   tensor(0.9774, device='cuda:0', grad_fn=<NllLossBackward>)
[4424] Train Loss:   tensor(0.9367, device='cuda:0', grad_fn=<NllLossBackward>)
[4425] Train Loss:   tensor(0.9284, device='cuda:0', grad_fn=<NllLossBackward>)
[4426] Train Loss:   tensor(0.9273, device='cuda:0', grad_fn=<NllLossBackward>)
[4427] Train Loss:   tensor(0.9493, device='cuda:0', grad_fn=<NllLossBackward>)
[4428] Train Loss:   tensor(0.9541, device='cuda:0', grad_fn=<NllLossBackward>)
[4429] Train Loss:   tensor(0.9608, device='cuda:0', grad_fn=<NllLossBackward>)
[4430] Train Loss:   tensor(0.9621, device='cuda:0', grad_fn=<NllLossBackward>)
[4431] Train Loss:   tensor(0.9446, device='cuda:0', grad_fn=<NllLossBackward>)
[4432] Train Loss:   tensor(0.9474, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4433] Train Loss:    tensor(0.9382, device='cuda:0', grad_fn=<NllLossBackward>)
[4434] Train Loss:    tensor(0.9550, device='cuda:0', grad_fn=<NllLossBackward>)
[4435] Train Loss:    tensor(0.9350, device='cuda:0', grad_fn=<NllLossBackward>)
[4436] Train Loss:    tensor(0.9541, device='cuda:0', grad_fn=<NllLossBackward>)
[4437] Train Loss:    tensor(0.9552, device='cuda:0', grad_fn=<NllLossBackward>)
[4438] Train Loss:    tensor(0.9713, device='cuda:0', grad_fn=<NllLossBackward>)
[4439] Train Loss:    tensor(0.9663, device='cuda:0', grad_fn=<NllLossBackward>)
[4440] Train Loss:    tensor(0.9443, device='cuda:0', grad_fn=<NllLossBackward>)
[4441] Train Loss:    tensor(0.9717, device='cuda:0', grad_fn=<NllLossBackward>)
[4442] Train Loss:    tensor(0.9527, device='cuda:0', grad_fn=<NllLossBackward>)
[4443] Train Loss:    tensor(0.9219, device='cuda:0', grad_fn=<NllLossBackward>)
[4444] Train Loss:    tensor(0.9556, device='cuda:0', grad_fn=<NllLossBackward>)
[4445] Train Loss:    tensor(0.9635, device='cuda:0', grad_fn=<NllLossBackward>)
[4446] Train Loss:    tensor(0.9518, device='cuda:0', grad_fn=<NllLossBackward>)
[4447] Train Loss:    tensor(0.9556, device='cuda:0', grad_fn=<NllLossBackward>)
[4448] Train Loss:    tensor(0.9653, device='cuda:0', grad_fn=<NllLossBackward>)
[4449] Train Loss:    tensor(0.9978, device='cuda:0', grad_fn=<NllLossBackward>)
[4450] Train Loss:    tensor(0.9509, device='cuda:0', grad_fn=<NllLossBackward>)
[4451] Train Loss:    tensor(0.9725, device='cuda:0', grad_fn=<NllLossBackward>)
[4452] Train Loss:    tensor(0.9324, device='cuda:0', grad_fn=<NllLossBackward>)
[4453] Train Loss:    tensor(0.9684, device='cuda:0', grad_fn=<NllLossBackward>)
[4454] Train Loss:    tensor(0.9238, device='cuda:0', grad_fn=<NllLossBackward>)
[4455] Train Loss:    tensor(0.9708, device='cuda:0', grad_fn=<NllLossBackward>)
[4456] Train Loss:    tensor(0.9717, device='cuda:0', grad_fn=<NllLossBackward>)
[4457] Train Loss:    tensor(0.9474, device='cuda:0', grad_fn=<NllLossBackward>)
[4458] Train Loss:    tensor(0.8950, device='cuda:0', grad_fn=<NllLossBackward>)
[4459] Train Loss:    tensor(0.9466, device='cuda:0', grad_fn=<NllLossBackward>)
[4460] Train Loss:    tensor(0.9855, device='cuda:0', grad_fn=<NllLossBackward>)
[4461] Train Loss:    tensor(0.9513, device='cuda:0', grad_fn=<NllLossBackward>)
[4462] Train Loss:    tensor(0.9516, device='cuda:0', grad_fn=<NllLossBackward>)
[4463] Train Loss:    tensor(0.9368, device='cuda:0', grad_fn=<NllLossBackward>)
[4464] Train Loss:    tensor(0.9692, device='cuda:0', grad_fn=<NllLossBackward>)
[4465] Train Loss:    tensor(0.9312, device='cuda:0', grad_fn=<NllLossBackward>)
[4466] Train Loss:    tensor(0.9693, device='cuda:0', grad_fn=<NllLossBackward>)
[4467] Train Loss:    tensor(0.9522, device='cuda:0', grad_fn=<NllLossBackward>)
[4468] Train Loss:    tensor(0.9483, device='cuda:0', grad_fn=<NllLossBackward>)
[4469] Train Loss:    tensor(0.9281, device='cuda:0', grad_fn=<NllLossBackward>)
[4470] Train Loss:    tensor(0.9323, device='cuda:0', grad_fn=<NllLossBackward>)
[4471] Train Loss:    tensor(0.9503, device='cuda:0', grad_fn=<NllLossBackward>)
[4472] Train Loss:    tensor(0.9328, device='cuda:0', grad_fn=<NllLossBackward>)
[4473] Train Loss:    tensor(0.9346, device='cuda:0', grad_fn=<NllLossBackward>)
[4474] Train Loss:    tensor(0.9754, device='cuda:0', grad_fn=<NllLossBackward>)
[4475] Train Loss:    tensor(0.9773, device='cuda:0', grad_fn=<NllLossBackward>)
[4476] Train Loss:    tensor(0.9311, device='cuda:0', grad_fn=<NllLossBackward>)
[4477] Train Loss:    tensor(0.9598, device='cuda:0', grad_fn=<NllLossBackward>)
[4478] Train Loss:    tensor(0.9329, device='cuda:0', grad_fn=<NllLossBackward>)
[4479] Train Loss:    tensor(0.9333, device='cuda:0', grad_fn=<NllLossBackward>)
[4480] Train Loss:    tensor(0.9719, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4481] Train Loss:  tensor(0.9321, device='cuda:0', grad_fn=<NllLossBackward>)
[4482] Train Loss:  tensor(0.9258, device='cuda:0', grad_fn=<NllLossBackward>)
[4483] Train Loss:  tensor(0.9426, device='cuda:0', grad_fn=<NllLossBackward>)
[4484] Train Loss:  tensor(0.9797, device='cuda:0', grad_fn=<NllLossBackward>)
[4485] Train Loss:  tensor(0.9515, device='cuda:0', grad_fn=<NllLossBackward>)
[4486] Train Loss:  tensor(0.9414, device='cuda:0', grad_fn=<NllLossBackward>)
[4487] Train Loss:  tensor(0.9599, device='cuda:0', grad_fn=<NllLossBackward>)
[4488] Train Loss:  tensor(0.9804, device='cuda:0', grad_fn=<NllLossBackward>)
[4489] Train Loss:  tensor(0.9531, device='cuda:0', grad_fn=<NllLossBackward>)
[4490] Train Loss:  tensor(0.9359, device='cuda:0', grad_fn=<NllLossBackward>)
[4491] Train Loss:  tensor(0.9511, device='cuda:0', grad_fn=<NllLossBackward>)
[4492] Train Loss:  tensor(0.9392, device='cuda:0', grad_fn=<NllLossBackward>)
[4493] Train Loss:  tensor(0.9484, device='cuda:0', grad_fn=<NllLossBackward>)
[4494] Train Loss:  tensor(0.9466, device='cuda:0', grad_fn=<NllLossBackward>)
[4495] Train Loss:  tensor(0.9429, device='cuda:0', grad_fn=<NllLossBackward>)
[4496] Train Loss:  tensor(0.9463, device='cuda:0', grad_fn=<NllLossBackward>)
[4497] Train Loss:  tensor(0.9128, device='cuda:0', grad_fn=<NllLossBackward>)
[4498] Train Loss:  tensor(0.9118, device='cuda:0', grad_fn=<NllLossBackward>)
[4499] Train Loss:  tensor(0.9949, device='cuda:0', grad_fn=<NllLossBackward>)
[4500] Train Loss:  tensor(0.9480, device='cuda:0', grad_fn=<NllLossBackward>)
[4501] Train Loss:  tensor(0.9562, device='cuda:0', grad_fn=<NllLossBackward>)
[4502] Train Loss:  tensor(0.9549, device='cuda:0', grad_fn=<NllLossBackward>)
[4503] Train Loss:  tensor(0.9692, device='cuda:0', grad_fn=<NllLossBackward>)
[4504] Train Loss:  tensor(0.9514, device='cuda:0', grad_fn=<NllLossBackward>)
[4505] Train Loss:  tensor(0.9307, device='cuda:0', grad_fn=<NllLossBackward>)
[4506] Train Loss:  tensor(0.9684, device='cuda:0', grad_fn=<NllLossBackward>)
[4507] Train Loss:  tensor(0.9433, device='cuda:0', grad_fn=<NllLossBackward>)
[4508] Train Loss:  tensor(0.9488, device='cuda:0', grad_fn=<NllLossBackward>)
[4509] Train Loss:  tensor(0.9840, device='cuda:0', grad_fn=<NllLossBackward>)
[4510] Train Loss:  tensor(0.9617, device='cuda:0', grad_fn=<NllLossBackward>)
[4511] Train Loss:  tensor(0.9543, device='cuda:0', grad_fn=<NllLossBackward>)
Training Loss:  1.1194106106719015 Time:  2362.641988515854 s
Training Accuaracy:  66.7067399261691 %
[0] Valid Loss:  tensor(0.8637, device='cuda:0')
[1] Valid Loss:  tensor(1.0113, device='cuda:0')
[2] Valid Loss:  tensor(0.8432, device='cuda:0')
[3] Valid Loss:  tensor(0.7607, device='cuda:0')
[4] Valid Loss:  tensor(0.8756, device='cuda:0')
[5] Valid Loss:  tensor(0.8683, device='cuda:0')
[6] Valid Loss:  tensor(0.8889, device='cuda:0')
[7] Valid Loss:  tensor(0.9853, device='cuda:0')
[8] Valid Loss:  tensor(0.9288, device='cuda:0')
[9] Valid Loss:  tensor(0.9938, device='cuda:0')
[10] Valid Loss:  tensor(0.7898, device='cuda:0')
[11] Valid Loss:  tensor(0.8533, device='cuda:0')
[12] Valid Loss:  tensor(1.2511, device='cuda:0')
[13] Valid Loss:  tensor(1.1364, device='cuda:0')
[14] Valid Loss:  tensor(0.7474, device='cuda:0')
```

```
[15] Valid Loss:  tensor(0.9255, device='cuda:0')
[16] Valid Loss:  tensor(0.8013, device='cuda:0')
[17] Valid Loss:  tensor(0.9451, device='cuda:0')
[18] Valid Loss:  tensor(0.8543, device='cuda:0')
[19] Valid Loss:  tensor(0.7997, device='cuda:0')
[20] Valid Loss:  tensor(0.8385, device='cuda:0')
[21] Valid Loss:  tensor(0.7440, device='cuda:0')
[22] Valid Loss:  tensor(0.7468, device='cuda:0')
[23] Valid Loss:  tensor(0.9655, device='cuda:0')
[24] Valid Loss:  tensor(0.8775, device='cuda:0')
[25] Valid Loss:  tensor(1.0371, device='cuda:0')
[26] Valid Loss:  tensor(0.6692, device='cuda:0')
[27] Valid Loss:  tensor(1.0517, device='cuda:0')
[28] Valid Loss:  tensor(0.8330, device='cuda:0')
[29] Valid Loss:  tensor(0.8840, device='cuda:0')
[30] Valid Loss:  tensor(0.7594, device='cuda:0')
[31] Valid Loss:  tensor(0.9218, device='cuda:0')
[32] Valid Loss:  tensor(1.0016, device='cuda:0')
[33] Valid Loss:  tensor(0.9525, device='cuda:0')
[34] Valid Loss:  tensor(0.9175, device='cuda:0')
[35] Valid Loss:  tensor(1.0332, device='cuda:0')
[36] Valid Loss:  tensor(0.8774, device='cuda:0')
[37] Valid Loss:  tensor(1.0973, device='cuda:0')
[38] Valid Loss:  tensor(0.9999, device='cuda:0')
[39] Valid Loss:  tensor(0.8550, device='cuda:0')
[40] Valid Loss:  tensor(0.9347, device='cuda:0')
[41] Valid Loss:  tensor(0.9900, device='cuda:0')
[42] Valid Loss:  tensor(0.9008, device='cuda:0')
[43] Valid Loss:  tensor(0.6953, device='cuda:0')
[44] Valid Loss:  tensor(0.8426, device='cuda:0')
[45] Valid Loss:  tensor(0.7311, device='cuda:0')
[46] Valid Loss:  tensor(1.0267, device='cuda:0')
[47] Valid Loss:  tensor(0.9120, device='cuda:0')
[48] Valid Loss:  tensor(0.8350, device='cuda:0')
[49] Valid Loss:  tensor(0.9072, device='cuda:0')
[50] Valid Loss:  tensor(0.8715, device='cuda:0')
[51] Valid Loss:  tensor(0.8135, device='cuda:0')
[52] Valid Loss:  tensor(0.7872, device='cuda:0')
[53] Valid Loss:  tensor(1.0029, device='cuda:0')
[54] Valid Loss:  tensor(0.7701, device='cuda:0')
[55] Valid Loss:  tensor(0.7084, device='cuda:0')
[56] Valid Loss:  tensor(1.0600, device='cuda:0')
[57] Valid Loss:  tensor(0.9926, device='cuda:0')
[58] Valid Loss:  tensor(0.9485, device='cuda:0')
[59] Valid Loss:  tensor(0.8265, device='cuda:0')
[60] Valid Loss:  tensor(1.1731, device='cuda:0')
[61] Valid Loss:  tensor(0.8408, device='cuda:0')
[62] Valid Loss:  tensor(0.9876, device='cuda:0')
```

```
[63] Valid Loss:  tensor(0.8894, device='cuda:0')
[64] Valid Loss:  tensor(0.9049, device='cuda:0')
[65] Valid Loss:  tensor(0.9962, device='cuda:0')
[66] Valid Loss:  tensor(0.8537, device='cuda:0')
[67] Valid Loss:  tensor(0.8882, device='cuda:0')
[68] Valid Loss:  tensor(0.8459, device='cuda:0')
[69] Valid Loss:  tensor(0.7250, device='cuda:0')
[70] Valid Loss:  tensor(0.7648, device='cuda:0')
[71] Valid Loss:  tensor(0.9869, device='cuda:0')
[72] Valid Loss:  tensor(0.8996, device='cuda:0')
[73] Valid Loss:  tensor(0.9040, device='cuda:0')
[74] Valid Loss:  tensor(0.8775, device='cuda:0')
[75] Valid Loss:  tensor(0.9879, device='cuda:0')
[76] Valid Loss:  tensor(0.9229, device='cuda:0')
[77] Valid Loss:  tensor(0.8341, device='cuda:0')
[78] Valid Loss:  tensor(0.9099, device='cuda:0')
[79] Valid Loss:  tensor(0.8227, device='cuda:0')
[80] Valid Loss:  tensor(0.9034, device='cuda:0')
[81] Valid Loss:  tensor(0.7599, device='cuda:0')
[82] Valid Loss:  tensor(1.0213, device='cuda:0')
[83] Valid Loss:  tensor(0.9631, device='cuda:0')
[84] Valid Loss:  tensor(0.7804, device='cuda:0')
[85] Valid Loss:  tensor(0.8643, device='cuda:0')
[86] Valid Loss:  tensor(0.9154, device='cuda:0')
[87] Valid Loss:  tensor(0.9803, device='cuda:0')
[88] Valid Loss:  tensor(0.7391, device='cuda:0')
[89] Valid Loss:  tensor(0.9125, device='cuda:0')
[90] Valid Loss:  tensor(1.1029, device='cuda:0')
[91] Valid Loss:  tensor(0.8811, device='cuda:0')
[92] Valid Loss:  tensor(0.9582, device='cuda:0')
[93] Valid Loss:  tensor(0.9583, device='cuda:0')
[94] Valid Loss:  tensor(0.9671, device='cuda:0')
[95] Valid Loss:  tensor(0.8398, device='cuda:0')
[96] Valid Loss:  tensor(1.0178, device='cuda:0')
[97] Valid Loss:  tensor(0.9591, device='cuda:0')
[98] Valid Loss:  tensor(0.9341, device='cuda:0')
[99] Valid Loss:  tensor(0.7999, device='cuda:0')
[100] Valid Loss:  tensor(0.8791, device='cuda:0')
[101] Valid Loss:  tensor(0.9034, device='cuda:0')
[102] Valid Loss:  tensor(1.0374, device='cuda:0')
[103] Valid Loss:  tensor(1.0205, device='cuda:0')
[104] Valid Loss:  tensor(0.8350, device='cuda:0')
[105] Valid Loss:  tensor(0.8156, device='cuda:0')
[106] Valid Loss:  tensor(0.7397, device='cuda:0')
[107] Valid Loss:  tensor(0.9503, device='cuda:0')
[108] Valid Loss:  tensor(0.9883, device='cuda:0')
[109] Valid Loss:  tensor(0.8693, device='cuda:0')
[110] Valid Loss:  tensor(0.9539, device='cuda:0')
```

```
[111] Valid Loss:   tensor(0.9212, device='cuda:0')
[112] Valid Loss:   tensor(0.9443, device='cuda:0')
[113] Valid Loss:   tensor(0.9248, device='cuda:0')
[114] Valid Loss:   tensor(0.8675, device='cuda:0')
[115] Valid Loss:   tensor(0.9794, device='cuda:0')
[116] Valid Loss:   tensor(0.9729, device='cuda:0')
[117] Valid Loss:   tensor(0.9500, device='cuda:0')
[118] Valid Loss:   tensor(0.8488, device='cuda:0')
[119] Valid Loss:   tensor(0.7735, device='cuda:0')
[120] Valid Loss:   tensor(1.0721, device='cuda:0')
[121] Valid Loss:   tensor(0.9011, device='cuda:0')
[122] Valid Loss:   tensor(0.8728, device='cuda:0')
[123] Valid Loss:   tensor(0.8341, device='cuda:0')
[124] Valid Loss:   tensor(1.0526, device='cuda:0')
[125] Valid Loss:   tensor(1.1069, device='cuda:0')
[126] Valid Loss:   tensor(0.7292, device='cuda:0')
[127] Valid Loss:   tensor(0.8009, device='cuda:0')
[128] Valid Loss:   tensor(0.9688, device='cuda:0')
[129] Valid Loss:   tensor(0.8779, device='cuda:0')
[130] Valid Loss:   tensor(1.0077, device='cuda:0')
[131] Valid Loss:   tensor(0.8584, device='cuda:0')
[132] Valid Loss:   tensor(0.8772, device='cuda:0')
[133] Valid Loss:   tensor(0.8979, device='cuda:0')
[134] Valid Loss:   tensor(0.8280, device='cuda:0')
[135] Valid Loss:   tensor(0.7709, device='cuda:0')
[136] Valid Loss:   tensor(0.8185, device='cuda:0')
[137] Valid Loss:   tensor(1.0986, device='cuda:0')
[138] Valid Loss:   tensor(0.9958, device='cuda:0')
[139] Valid Loss:   tensor(1.1810, device='cuda:0')
[140] Valid Loss:   tensor(0.9296, device='cuda:0')
[141] Valid Loss:   tensor(0.8029, device='cuda:0')
[142] Valid Loss:   tensor(1.1441, device='cuda:0')
[143] Valid Loss:   tensor(0.7919, device='cuda:0')
[144] Valid Loss:   tensor(0.9166, device='cuda:0')
[145] Valid Loss:   tensor(1.0932, device='cuda:0')
[146] Valid Loss:   tensor(0.8645, device='cuda:0')
[147] Valid Loss:   tensor(0.9467, device='cuda:0')
[148] Valid Loss:   tensor(0.8763, device='cuda:0')
[149] Valid Loss:   tensor(0.9027, device='cuda:0')
[150] Valid Loss:   tensor(0.9446, device='cuda:0')
[151] Valid Loss:   tensor(0.9068, device='cuda:0')
[152] Valid Loss:   tensor(0.8431, device='cuda:0')
[153] Valid Loss:   tensor(0.8297, device='cuda:0')
[154] Valid Loss:   tensor(0.8347, device='cuda:0')
[155] Valid Loss:   tensor(0.9245, device='cuda:0')
[156] Valid Loss:   tensor(0.9353, device='cuda:0')
[157] Valid Loss:   tensor(0.7862, device='cuda:0')
[158] Valid Loss:   tensor(0.8475, device='cuda:0')
```

```
[159] Valid Loss:   tensor(0.8955, device='cuda:0')
[160] Valid Loss:   tensor(0.7662, device='cuda:0')
[161] Valid Loss:   tensor(0.9704, device='cuda:0')
[162] Valid Loss:   tensor(0.9290, device='cuda:0')
[163] Valid Loss:   tensor(0.8415, device='cuda:0')
[164] Valid Loss:   tensor(1.0882, device='cuda:0')
[165] Valid Loss:   tensor(0.8718, device='cuda:0')
[166] Valid Loss:   tensor(0.8936, device='cuda:0')
[167] Valid Loss:   tensor(0.7863, device='cuda:0')
[168] Valid Loss:   tensor(0.8407, device='cuda:0')
[169] Valid Loss:   tensor(0.8106, device='cuda:0')
[170] Valid Loss:   tensor(0.8029, device='cuda:0')
[171] Valid Loss:   tensor(0.7342, device='cuda:0')
[172] Valid Loss:   tensor(0.9406, device='cuda:0')
[173] Valid Loss:   tensor(0.8033, device='cuda:0')
[174] Valid Loss:   tensor(0.9226, device='cuda:0')
[175] Valid Loss:   tensor(0.7951, device='cuda:0')
[176] Valid Loss:   tensor(0.7558, device='cuda:0')
[177] Valid Loss:   tensor(0.8925, device='cuda:0')
[178] Valid Loss:   tensor(0.9593, device='cuda:0')
[179] Valid Loss:   tensor(0.8925, device='cuda:0')
[180] Valid Loss:   tensor(0.9526, device='cuda:0')
[181] Valid Loss:   tensor(0.8796, device='cuda:0')
[182] Valid Loss:   tensor(0.9339, device='cuda:0')
[183] Valid Loss:   tensor(0.9002, device='cuda:0')
[184] Valid Loss:   tensor(1.0547, device='cuda:0')
[185] Valid Loss:   tensor(0.9569, device='cuda:0')
[186] Valid Loss:   tensor(0.8729, device='cuda:0')
[187] Valid Loss:   tensor(0.9877, device='cuda:0')
[188] Valid Loss:   tensor(1.0629, device='cuda:0')
[189] Valid Loss:   tensor(0.9028, device='cuda:0')
[190] Valid Loss:   tensor(0.8875, device='cuda:0')
[191] Valid Loss:   tensor(0.7760, device='cuda:0')
[192] Valid Loss:   tensor(1.1589, device='cuda:0')
[193] Valid Loss:   tensor(0.9235, device='cuda:0')
[194] Valid Loss:   tensor(0.9425, device='cuda:0')
[195] Valid Loss:   tensor(0.9982, device='cuda:0')
[196] Valid Loss:   tensor(0.7635, device='cuda:0')
[197] Valid Loss:   tensor(0.9149, device='cuda:0')
[198] Valid Loss:   tensor(0.8647, device='cuda:0')
[199] Valid Loss:   tensor(0.9292, device='cuda:0')
[200] Valid Loss:   tensor(0.9154, device='cuda:0')
[201] Valid Loss:   tensor(0.8473, device='cuda:0')
[202] Valid Loss:   tensor(0.8244, device='cuda:0')
[203] Valid Loss:   tensor(1.0183, device='cuda:0')
[204] Valid Loss:   tensor(1.0241, device='cuda:0')
[205] Valid Loss:   tensor(0.9986, device='cuda:0')
[206] Valid Loss:   tensor(1.0724, device='cuda:0')
```

```
[207] Valid Loss:   tensor(0.9351, device='cuda:0')
[208] Valid Loss:   tensor(0.7878, device='cuda:0')
[209] Valid Loss:   tensor(0.8900, device='cuda:0')
[210] Valid Loss:   tensor(1.2735, device='cuda:0')
[211] Valid Loss:   tensor(0.8683, device='cuda:0')
[212] Valid Loss:   tensor(0.8247, device='cuda:0')
[213] Valid Loss:   tensor(0.8698, device='cuda:0')
[214] Valid Loss:   tensor(0.8483, device='cuda:0')
[215] Valid Loss:   tensor(0.9812, device='cuda:0')
[216] Valid Loss:   tensor(0.8879, device='cuda:0')
[217] Valid Loss:   tensor(0.9046, device='cuda:0')
[218] Valid Loss:   tensor(0.7231, device='cuda:0')
[219] Valid Loss:   tensor(0.8442, device='cuda:0')
[220] Valid Loss:   tensor(0.8304, device='cuda:0')
[221] Valid Loss:   tensor(0.9232, device='cuda:0')
[222] Valid Loss:   tensor(0.7344, device='cuda:0')
[223] Valid Loss:   tensor(0.8287, device='cuda:0')
[224] Valid Loss:   tensor(0.9860, device='cuda:0')
[225] Valid Loss:   tensor(0.7605, device='cuda:0')
[226] Valid Loss:   tensor(0.8871, device='cuda:0')
[227] Valid Loss:   tensor(0.7792, device='cuda:0')
[228] Valid Loss:   tensor(0.9417, device='cuda:0')
[229] Valid Loss:   tensor(0.8200, device='cuda:0')
[230] Valid Loss:   tensor(0.9211, device='cuda:0')
[231] Valid Loss:   tensor(0.8579, device='cuda:0')
[232] Valid Loss:   tensor(0.7255, device='cuda:0')
[233] Valid Loss:   tensor(0.8574, device='cuda:0')
[234] Valid Loss:   tensor(0.7647, device='cuda:0')
[235] Valid Loss:   tensor(0.7384, device='cuda:0')
[236] Valid Loss:   tensor(0.9162, device='cuda:0')
[237] Valid Loss:   tensor(0.9511, device='cuda:0')
[238] Valid Loss:   tensor(0.7287, device='cuda:0')
[239] Valid Loss:   tensor(0.9080, device='cuda:0')
[240] Valid Loss:   tensor(0.8748, device='cuda:0')
[241] Valid Loss:   tensor(0.8056, device='cuda:0')
[242] Valid Loss:   tensor(0.6994, device='cuda:0')
[243] Valid Loss:   tensor(1.2903, device='cuda:0')
[244] Valid Loss:   tensor(0.9695, device='cuda:0')
[245] Valid Loss:   tensor(0.8592, device='cuda:0')
[246] Valid Loss:   tensor(0.8348, device='cuda:0')
[247] Valid Loss:   tensor(0.8682, device='cuda:0')
[248] Valid Loss:   tensor(1.0669, device='cuda:0')
[249] Valid Loss:   tensor(0.8357, device='cuda:0')
[250] Valid Loss:   tensor(0.9275, device='cuda:0')
[251] Valid Loss:   tensor(0.9859, device='cuda:0')
[252] Valid Loss:   tensor(0.7989, device='cuda:0')
[253] Valid Loss:   tensor(0.7843, device='cuda:0')
[254] Valid Loss:   tensor(0.8511, device='cuda:0')
```

```
[255] Valid Loss:   tensor(0.9412, device='cuda:0')
[256] Valid Loss:   tensor(0.9635, device='cuda:0')
[257] Valid Loss:   tensor(0.7851, device='cuda:0')
[258] Valid Loss:   tensor(0.9902, device='cuda:0')
[259] Valid Loss:   tensor(0.8555, device='cuda:0')
[260] Valid Loss:   tensor(1.0249, device='cuda:0')
[261] Valid Loss:   tensor(0.8934, device='cuda:0')
[262] Valid Loss:   tensor(1.0536, device='cuda:0')
[263] Valid Loss:   tensor(0.9989, device='cuda:0')
[264] Valid Loss:   tensor(0.9872, device='cuda:0')
[265] Valid Loss:   tensor(0.9884, device='cuda:0')
[266] Valid Loss:   tensor(0.8278, device='cuda:0')
[267] Valid Loss:   tensor(0.8951, device='cuda:0')
[268] Valid Loss:   tensor(0.9039, device='cuda:0')
[269] Valid Loss:   tensor(0.8480, device='cuda:0')
[270] Valid Loss:   tensor(0.7696, device='cuda:0')
[271] Valid Loss:   tensor(1.1019, device='cuda:0')
[272] Valid Loss:   tensor(0.8973, device='cuda:0')
[273] Valid Loss:   tensor(1.0019, device='cuda:0')
[274] Valid Loss:   tensor(0.8964, device='cuda:0')
[275] Valid Loss:   tensor(0.9057, device='cuda:0')
[276] Valid Loss:   tensor(0.8092, device='cuda:0')
[277] Valid Loss:   tensor(0.7793, device='cuda:0')
[278] Valid Loss:   tensor(1.0045, device='cuda:0')
[279] Valid Loss:   tensor(0.9517, device='cuda:0')
[280] Valid Loss:   tensor(1.0176, device='cuda:0')
[281] Valid Loss:   tensor(0.7985, device='cuda:0')
[282] Valid Loss:   tensor(0.9477, device='cuda:0')
[283] Valid Loss:   tensor(1.0474, device='cuda:0')
[284] Valid Loss:   tensor(0.8039, device='cuda:0')
[285] Valid Loss:   tensor(0.7156, device='cuda:0')
[286] Valid Loss:   tensor(0.9456, device='cuda:0')
[287] Valid Loss:   tensor(1.0967, device='cuda:0')
[288] Valid Loss:   tensor(0.7435, device='cuda:0')
[289] Valid Loss:   tensor(0.9706, device='cuda:0')
[290] Valid Loss:   tensor(0.7488, device='cuda:0')
[291] Valid Loss:   tensor(0.8619, device='cuda:0')
[292] Valid Loss:   tensor(0.9990, device='cuda:0')
[293] Valid Loss:   tensor(0.8285, device='cuda:0')
[294] Valid Loss:   tensor(0.9748, device='cuda:0')
[295] Valid Loss:   tensor(0.7673, device='cuda:0')
[296] Valid Loss:   tensor(0.8995, device='cuda:0')
[297] Valid Loss:   tensor(0.9039, device='cuda:0')
[298] Valid Loss:   tensor(0.7961, device='cuda:0')
[299] Valid Loss:   tensor(0.9261, device='cuda:0')
[300] Valid Loss:   tensor(1.0742, device='cuda:0')
[301] Valid Loss:   tensor(0.7391, device='cuda:0')
[302] Valid Loss:   tensor(0.7176, device='cuda:0')
```

```
[303] Valid Loss:   tensor(0.8790, device='cuda:0')
[304] Valid Loss:   tensor(0.8163, device='cuda:0')
[305] Valid Loss:   tensor(0.8851, device='cuda:0')
[306] Valid Loss:   tensor(0.9357, device='cuda:0')
[307] Valid Loss:   tensor(0.7638, device='cuda:0')
[308] Valid Loss:   tensor(0.8122, device='cuda:0')
[309] Valid Loss:   tensor(0.8376, device='cuda:0')
[310] Valid Loss:   tensor(0.8697, device='cuda:0')
[311] Valid Loss:   tensor(0.8137, device='cuda:0')
[312] Valid Loss:   tensor(0.7849, device='cuda:0')
[313] Valid Loss:   tensor(0.8809, device='cuda:0')
[314] Valid Loss:   tensor(0.8564, device='cuda:0')
[315] Valid Loss:   tensor(1.1290, device='cuda:0')
[316] Valid Loss:   tensor(0.9749, device='cuda:0')
[317] Valid Loss:   tensor(0.8456, device='cuda:0')
[318] Valid Loss:   tensor(0.8580, device='cuda:0')
[319] Valid Loss:   tensor(0.7969, device='cuda:0')
[320] Valid Loss:   tensor(0.8422, device='cuda:0')
[321] Valid Loss:   tensor(0.8375, device='cuda:0')
[322] Valid Loss:   tensor(0.8124, device='cuda:0')
[323] Valid Loss:   tensor(0.9871, device='cuda:0')
[324] Valid Loss:   tensor(0.9261, device='cuda:0')
[325] Valid Loss:   tensor(0.7931, device='cuda:0')
[326] Valid Loss:   tensor(0.8805, device='cuda:0')
[327] Valid Loss:   tensor(0.6335, device='cuda:0')
[328] Valid Loss:   tensor(0.8543, device='cuda:0')
[329] Valid Loss:   tensor(0.9822, device='cuda:0')
[330] Valid Loss:   tensor(0.7506, device='cuda:0')
[331] Valid Loss:   tensor(1.0522, device='cuda:0')
[332] Valid Loss:   tensor(0.8034, device='cuda:0')
[333] Valid Loss:   tensor(0.7293, device='cuda:0')
[334] Valid Loss:   tensor(0.7450, device='cuda:0')
[335] Valid Loss:   tensor(0.8827, device='cuda:0')
[336] Valid Loss:   tensor(0.9579, device='cuda:0')
[337] Valid Loss:   tensor(1.0525, device='cuda:0')
[338] Valid Loss:   tensor(0.9355, device='cuda:0')
[339] Valid Loss:   tensor(0.8458, device='cuda:0')
[340] Valid Loss:   tensor(0.8931, device='cuda:0')
[341] Valid Loss:   tensor(0.9788, device='cuda:0')
[342] Valid Loss:   tensor(1.0438, device='cuda:0')
[343] Valid Loss:   tensor(0.9424, device='cuda:0')
[344] Valid Loss:   tensor(0.7995, device='cuda:0')
[345] Valid Loss:   tensor(0.7419, device='cuda:0')
[346] Valid Loss:   tensor(0.8299, device='cuda:0')
[347] Valid Loss:   tensor(0.9654, device='cuda:0')
[348] Valid Loss:   tensor(0.8678, device='cuda:0')
[349] Valid Loss:   tensor(0.8141, device='cuda:0')
[350] Valid Loss:   tensor(0.8758, device='cuda:0')
```

```
[351] Valid Loss:  tensor(1.0402, device='cuda:0')
[352] Valid Loss:  tensor(0.7855, device='cuda:0')
[353] Valid Loss:  tensor(0.8768, device='cuda:0')
[354] Valid Loss:  tensor(1.0193, device='cuda:0')
[355] Valid Loss:  tensor(0.7207, device='cuda:0')
[356] Valid Loss:  tensor(0.8112, device='cuda:0')
[357] Valid Loss:  tensor(0.9553, device='cuda:0')
[358] Valid Loss:  tensor(0.8540, device='cuda:0')
[359] Valid Loss:  tensor(0.9643, device='cuda:0')
[360] Valid Loss:  tensor(0.8776, device='cuda:0')
[361] Valid Loss:  tensor(0.8498, device='cuda:0')
[362] Valid Loss:  tensor(0.8700, device='cuda:0')
[363] Valid Loss:  tensor(0.9343, device='cuda:0')
[364] Valid Loss:  tensor(1.0450, device='cuda:0')
[365] Valid Loss:  tensor(0.8247, device='cuda:0')
[366] Valid Loss:  tensor(0.8508, device='cuda:0')
[367] Valid Loss:  tensor(1.0129, device='cuda:0')
[368] Valid Loss:  tensor(0.9886, device='cuda:0')
[369] Valid Loss:  tensor(0.8955, device='cuda:0')
[370] Valid Loss:  tensor(0.7801, device='cuda:0')
[371] Valid Loss:  tensor(0.7534, device='cuda:0')
[372] Valid Loss:  tensor(0.9766, device='cuda:0')
[373] Valid Loss:  tensor(0.7397, device='cuda:0')
[374] Valid Loss:  tensor(0.8362, device='cuda:0')
[375] Valid Loss:  tensor(0.9456, device='cuda:0')
[376] Valid Loss:  tensor(0.8042, device='cuda:0')
[377] Valid Loss:  tensor(1.0873, device='cuda:0')
[378] Valid Loss:  tensor(0.7867, device='cuda:0')
[379] Valid Loss:  tensor(0.9707, device='cuda:0')
[380] Valid Loss:  tensor(1.1413, device='cuda:0')
Testing Loss:  0.896486697078064
Testing Accuaracy:  72.36310154150874 %
Epoch Number:  1
[0] Train Loss:  tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[1] Train Loss:  tensor(0.9381, device='cuda:0', grad_fn=<NllLossBackward>)
[2] Train Loss:  tensor(0.9702, device='cuda:0', grad_fn=<NllLossBackward>)
[3] Train Loss:  tensor(0.9196, device='cuda:0', grad_fn=<NllLossBackward>)
[4] Train Loss:  tensor(0.9958, device='cuda:0', grad_fn=<NllLossBackward>)
[5] Train Loss:  tensor(0.9526, device='cuda:0', grad_fn=<NllLossBackward>)
[6] Train Loss:  tensor(0.9302, device='cuda:0', grad_fn=<NllLossBackward>)
[7] Train Loss:  tensor(0.9546, device='cuda:0', grad_fn=<NllLossBackward>)
[8] Train Loss:  tensor(0.9314, device='cuda:0', grad_fn=<NllLossBackward>)
[9] Train Loss:  tensor(0.9270, device='cuda:0', grad_fn=<NllLossBackward>)
[10] Train Loss:  tensor(0.9225, device='cuda:0', grad_fn=<NllLossBackward>)
[11] Train Loss:  tensor(0.9247, device='cuda:0', grad_fn=<NllLossBackward>)
[12] Train Loss:  tensor(0.9202, device='cuda:0', grad_fn=<NllLossBackward>)
[13] Train Loss:  tensor(0.9273, device='cuda:0', grad_fn=<NllLossBackward>)
[14] Train Loss:  tensor(0.9326, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[15] Train Loss:    tensor(0.9420, device='cuda:0', grad_fn=<NllLossBackward>)
[16] Train Loss:    tensor(0.9326, device='cuda:0', grad_fn=<NllLossBackward>)
[17] Train Loss:    tensor(0.9609, device='cuda:0', grad_fn=<NllLossBackward>)
[18] Train Loss:    tensor(0.9270, device='cuda:0', grad_fn=<NllLossBackward>)
[19] Train Loss:    tensor(0.9442, device='cuda:0', grad_fn=<NllLossBackward>)
[20] Train Loss:    tensor(0.9602, device='cuda:0', grad_fn=<NllLossBackward>)
[21] Train Loss:    tensor(0.9474, device='cuda:0', grad_fn=<NllLossBackward>)
[22] Train Loss:    tensor(0.9363, device='cuda:0', grad_fn=<NllLossBackward>)
[23] Train Loss:    tensor(0.9242, device='cuda:0', grad_fn=<NllLossBackward>)
[24] Train Loss:    tensor(0.9547, device='cuda:0', grad_fn=<NllLossBackward>)
[25] Train Loss:    tensor(0.9207, device='cuda:0', grad_fn=<NllLossBackward>)
[26] Train Loss:    tensor(0.9346, device='cuda:0', grad_fn=<NllLossBackward>)
[27] Train Loss:    tensor(0.9291, device='cuda:0', grad_fn=<NllLossBackward>)
[28] Train Loss:    tensor(0.9726, device='cuda:0', grad_fn=<NllLossBackward>)
[29] Train Loss:    tensor(0.9561, device='cuda:0', grad_fn=<NllLossBackward>)
[30] Train Loss:    tensor(0.9269, device='cuda:0', grad_fn=<NllLossBackward>)
[31] Train Loss:    tensor(0.9408, device='cuda:0', grad_fn=<NllLossBackward>)
[32] Train Loss:    tensor(0.9341, device='cuda:0', grad_fn=<NllLossBackward>)
[33] Train Loss:    tensor(0.9256, device='cuda:0', grad_fn=<NllLossBackward>)
[34] Train Loss:    tensor(0.9314, device='cuda:0', grad_fn=<NllLossBackward>)
[35] Train Loss:    tensor(0.9446, device='cuda:0', grad_fn=<NllLossBackward>)
[36] Train Loss:    tensor(0.9510, device='cuda:0', grad_fn=<NllLossBackward>)
[37] Train Loss:    tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[38] Train Loss:    tensor(0.8981, device='cuda:0', grad_fn=<NllLossBackward>)
[39] Train Loss:    tensor(0.9572, device='cuda:0', grad_fn=<NllLossBackward>)
[40] Train Loss:    tensor(0.9440, device='cuda:0', grad_fn=<NllLossBackward>)
[41] Train Loss:    tensor(0.9315, device='cuda:0', grad_fn=<NllLossBackward>)
[42] Train Loss:    tensor(0.9295, device='cuda:0', grad_fn=<NllLossBackward>)
[43] Train Loss:    tensor(0.9399, device='cuda:0', grad_fn=<NllLossBackward>)
[44] Train Loss:    tensor(0.9298, device='cuda:0', grad_fn=<NllLossBackward>)
[45] Train Loss:    tensor(0.9242, device='cuda:0', grad_fn=<NllLossBackward>)
[46] Train Loss:    tensor(0.9085, device='cuda:0', grad_fn=<NllLossBackward>)
[47] Train Loss:    tensor(0.9076, device='cuda:0', grad_fn=<NllLossBackward>)
[48] Train Loss:    tensor(0.9111, device='cuda:0', grad_fn=<NllLossBackward>)
[49] Train Loss:    tensor(0.9179, device='cuda:0', grad_fn=<NllLossBackward>)
[50] Train Loss:    tensor(0.9341, device='cuda:0', grad_fn=<NllLossBackward>)
[51] Train Loss:    tensor(0.9532, device='cuda:0', grad_fn=<NllLossBackward>)
[52] Train Loss:    tensor(0.9534, device='cuda:0', grad_fn=<NllLossBackward>)
[53] Train Loss:    tensor(0.9595, device='cuda:0', grad_fn=<NllLossBackward>)
[54] Train Loss:    tensor(0.9464, device='cuda:0', grad_fn=<NllLossBackward>)
[55] Train Loss:    tensor(0.9601, device='cuda:0', grad_fn=<NllLossBackward>)
[56] Train Loss:    tensor(0.9201, device='cuda:0', grad_fn=<NllLossBackward>)
[57] Train Loss:    tensor(0.9334, device='cuda:0', grad_fn=<NllLossBackward>)
[58] Train Loss:    tensor(0.9393, device='cuda:0', grad_fn=<NllLossBackward>)
[59] Train Loss:    tensor(0.9048, device='cuda:0', grad_fn=<NllLossBackward>)
[60] Train Loss:    tensor(0.9280, device='cuda:0', grad_fn=<NllLossBackward>)
[61] Train Loss:    tensor(0.9344, device='cuda:0', grad_fn=<NllLossBackward>)
[62] Train Loss:    tensor(0.9612, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[63] Train Loss:   tensor(0.9468, device='cuda:0', grad_fn=<NllLossBackward>)
[64] Train Loss:   tensor(0.9457, device='cuda:0', grad_fn=<NllLossBackward>)
[65] Train Loss:   tensor(0.9478, device='cuda:0', grad_fn=<NllLossBackward>)
[66] Train Loss:   tensor(0.9464, device='cuda:0', grad_fn=<NllLossBackward>)
[67] Train Loss:   tensor(0.9443, device='cuda:0', grad_fn=<NllLossBackward>)
[68] Train Loss:   tensor(0.9195, device='cuda:0', grad_fn=<NllLossBackward>)
[69] Train Loss:   tensor(0.9235, device='cuda:0', grad_fn=<NllLossBackward>)
[70] Train Loss:   tensor(0.9434, device='cuda:0', grad_fn=<NllLossBackward>)
[71] Train Loss:   tensor(0.9329, device='cuda:0', grad_fn=<NllLossBackward>)
[72] Train Loss:   tensor(0.9279, device='cuda:0', grad_fn=<NllLossBackward>)
[73] Train Loss:   tensor(0.8971, device='cuda:0', grad_fn=<NllLossBackward>)
[74] Train Loss:   tensor(0.9519, device='cuda:0', grad_fn=<NllLossBackward>)
[75] Train Loss:   tensor(0.9229, device='cuda:0', grad_fn=<NllLossBackward>)
[76] Train Loss:   tensor(0.9036, device='cuda:0', grad_fn=<NllLossBackward>)
[77] Train Loss:   tensor(0.9317, device='cuda:0', grad_fn=<NllLossBackward>)
[78] Train Loss:   tensor(0.8872, device='cuda:0', grad_fn=<NllLossBackward>)
[79] Train Loss:   tensor(0.9274, device='cuda:0', grad_fn=<NllLossBackward>)
[80] Train Loss:   tensor(0.9391, device='cuda:0', grad_fn=<NllLossBackward>)
[81] Train Loss:   tensor(0.9630, device='cuda:0', grad_fn=<NllLossBackward>)
[82] Train Loss:   tensor(0.9618, device='cuda:0', grad_fn=<NllLossBackward>)
[83] Train Loss:   tensor(0.9375, device='cuda:0', grad_fn=<NllLossBackward>)
[84] Train Loss:   tensor(0.9450, device='cuda:0', grad_fn=<NllLossBackward>)
[85] Train Loss:   tensor(0.9597, device='cuda:0', grad_fn=<NllLossBackward>)
[86] Train Loss:   tensor(0.9320, device='cuda:0', grad_fn=<NllLossBackward>)
[87] Train Loss:   tensor(0.9418, device='cuda:0', grad_fn=<NllLossBackward>)
[88] Train Loss:   tensor(0.9650, device='cuda:0', grad_fn=<NllLossBackward>)
[89] Train Loss:   tensor(0.9635, device='cuda:0', grad_fn=<NllLossBackward>)
[90] Train Loss:   tensor(0.9301, device='cuda:0', grad_fn=<NllLossBackward>)
[91] Train Loss:   tensor(0.9540, device='cuda:0', grad_fn=<NllLossBackward>)
[92] Train Loss:   tensor(0.9350, device='cuda:0', grad_fn=<NllLossBackward>)
[93] Train Loss:   tensor(0.9528, device='cuda:0', grad_fn=<NllLossBackward>)
[94] Train Loss:   tensor(0.9290, device='cuda:0', grad_fn=<NllLossBackward>)
[95] Train Loss:   tensor(0.9398, device='cuda:0', grad_fn=<NllLossBackward>)
[96] Train Loss:   tensor(0.9344, device='cuda:0', grad_fn=<NllLossBackward>)
[97] Train Loss:   tensor(0.9255, device='cuda:0', grad_fn=<NllLossBackward>)
[98] Train Loss:   tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[99] Train Loss:   tensor(0.9173, device='cuda:0', grad_fn=<NllLossBackward>)
[100] Train Loss:   tensor(0.9437, device='cuda:0', grad_fn=<NllLossBackward>)
[101] Train Loss:   tensor(0.8940, device='cuda:0', grad_fn=<NllLossBackward>)
[102] Train Loss:   tensor(0.9163, device='cuda:0', grad_fn=<NllLossBackward>)
[103] Train Loss:   tensor(0.9117, device='cuda:0', grad_fn=<NllLossBackward>)
[104] Train Loss:   tensor(0.9585, device='cuda:0', grad_fn=<NllLossBackward>)
[105] Train Loss:   tensor(0.8933, device='cuda:0', grad_fn=<NllLossBackward>)
[106] Train Loss:   tensor(0.9127, device='cuda:0', grad_fn=<NllLossBackward>)
[107] Train Loss:   tensor(0.8977, device='cuda:0', grad_fn=<NllLossBackward>)
[108] Train Loss:   tensor(0.9295, device='cuda:0', grad_fn=<NllLossBackward>)
[109] Train Loss:   tensor(0.9387, device='cuda:0', grad_fn=<NllLossBackward>)
[110] Train Loss:   tensor(0.9507, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[111] Train Loss:    tensor(0.9093, device='cuda:0', grad_fn=<NllLossBackward>)
[112] Train Loss:    tensor(0.9242, device='cuda:0', grad_fn=<NllLossBackward>)
[113] Train Loss:    tensor(0.9340, device='cuda:0', grad_fn=<NllLossBackward>)
[114] Train Loss:    tensor(0.9009, device='cuda:0', grad_fn=<NllLossBackward>)
[115] Train Loss:    tensor(0.9033, device='cuda:0', grad_fn=<NllLossBackward>)
[116] Train Loss:    tensor(0.9018, device='cuda:0', grad_fn=<NllLossBackward>)
[117] Train Loss:    tensor(0.9713, device='cuda:0', grad_fn=<NllLossBackward>)
[118] Train Loss:    tensor(0.9173, device='cuda:0', grad_fn=<NllLossBackward>)
[119] Train Loss:    tensor(0.9131, device='cuda:0', grad_fn=<NllLossBackward>)
[120] Train Loss:    tensor(0.9135, device='cuda:0', grad_fn=<NllLossBackward>)
[121] Train Loss:    tensor(0.9181, device='cuda:0', grad_fn=<NllLossBackward>)
[122] Train Loss:    tensor(0.9703, device='cuda:0', grad_fn=<NllLossBackward>)
[123] Train Loss:    tensor(0.9679, device='cuda:0', grad_fn=<NllLossBackward>)
[124] Train Loss:    tensor(0.9160, device='cuda:0', grad_fn=<NllLossBackward>)
[125] Train Loss:    tensor(0.9092, device='cuda:0', grad_fn=<NllLossBackward>)
[126] Train Loss:    tensor(0.9544, device='cuda:0', grad_fn=<NllLossBackward>)
[127] Train Loss:    tensor(0.8807, device='cuda:0', grad_fn=<NllLossBackward>)
[128] Train Loss:    tensor(0.9145, device='cuda:0', grad_fn=<NllLossBackward>)
[129] Train Loss:    tensor(0.9257, device='cuda:0', grad_fn=<NllLossBackward>)
[130] Train Loss:    tensor(0.9543, device='cuda:0', grad_fn=<NllLossBackward>)
[131] Train Loss:    tensor(0.9714, device='cuda:0', grad_fn=<NllLossBackward>)
[132] Train Loss:    tensor(0.8795, device='cuda:0', grad_fn=<NllLossBackward>)
[133] Train Loss:    tensor(0.9445, device='cuda:0', grad_fn=<NllLossBackward>)
[134] Train Loss:    tensor(0.8905, device='cuda:0', grad_fn=<NllLossBackward>)
[135] Train Loss:    tensor(0.9451, device='cuda:0', grad_fn=<NllLossBackward>)
[136] Train Loss:    tensor(0.9805, device='cuda:0', grad_fn=<NllLossBackward>)
[137] Train Loss:    tensor(0.9329, device='cuda:0', grad_fn=<NllLossBackward>)
[138] Train Loss:    tensor(0.9171, device='cuda:0', grad_fn=<NllLossBackward>)
[139] Train Loss:    tensor(0.8984, device='cuda:0', grad_fn=<NllLossBackward>)
[140] Train Loss:    tensor(0.9070, device='cuda:0', grad_fn=<NllLossBackward>)
[141] Train Loss:    tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[142] Train Loss:    tensor(0.9263, device='cuda:0', grad_fn=<NllLossBackward>)
[143] Train Loss:    tensor(0.9216, device='cuda:0', grad_fn=<NllLossBackward>)
[144] Train Loss:    tensor(0.9624, device='cuda:0', grad_fn=<NllLossBackward>)
[145] Train Loss:    tensor(0.9227, device='cuda:0', grad_fn=<NllLossBackward>)
[146] Train Loss:    tensor(0.8631, device='cuda:0', grad_fn=<NllLossBackward>)
[147] Train Loss:    tensor(0.9374, device='cuda:0', grad_fn=<NllLossBackward>)
[148] Train Loss:    tensor(0.9225, device='cuda:0', grad_fn=<NllLossBackward>)
[149] Train Loss:    tensor(0.9332, device='cuda:0', grad_fn=<NllLossBackward>)
[150] Train Loss:    tensor(0.9251, device='cuda:0', grad_fn=<NllLossBackward>)
[151] Train Loss:    tensor(0.9308, device='cuda:0', grad_fn=<NllLossBackward>)
[152] Train Loss:    tensor(0.8896, device='cuda:0', grad_fn=<NllLossBackward>)
[153] Train Loss:    tensor(0.9520, device='cuda:0', grad_fn=<NllLossBackward>)
[154] Train Loss:    tensor(0.9653, device='cuda:0', grad_fn=<NllLossBackward>)
[155] Train Loss:    tensor(0.8830, device='cuda:0', grad_fn=<NllLossBackward>)
[156] Train Loss:    tensor(0.9063, device='cuda:0', grad_fn=<NllLossBackward>)
[157] Train Loss:    tensor(0.8958, device='cuda:0', grad_fn=<NllLossBackward>)
[158] Train Loss:    tensor(0.9528, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[159] Train Loss:    tensor(0.9730, device='cuda:0', grad_fn=<NllLossBackward>)
[160] Train Loss:    tensor(0.9462, device='cuda:0', grad_fn=<NllLossBackward>)
[161] Train Loss:    tensor(0.9095, device='cuda:0', grad_fn=<NllLossBackward>)
[162] Train Loss:    tensor(0.9628, device='cuda:0', grad_fn=<NllLossBackward>)
[163] Train Loss:    tensor(0.9117, device='cuda:0', grad_fn=<NllLossBackward>)
[164] Train Loss:    tensor(0.9248, device='cuda:0', grad_fn=<NllLossBackward>)
[165] Train Loss:    tensor(0.8918, device='cuda:0', grad_fn=<NllLossBackward>)
[166] Train Loss:    tensor(0.9047, device='cuda:0', grad_fn=<NllLossBackward>)
[167] Train Loss:    tensor(0.9593, device='cuda:0', grad_fn=<NllLossBackward>)
[168] Train Loss:    tensor(0.9438, device='cuda:0', grad_fn=<NllLossBackward>)
[169] Train Loss:    tensor(0.9293, device='cuda:0', grad_fn=<NllLossBackward>)
[170] Train Loss:    tensor(0.9402, device='cuda:0', grad_fn=<NllLossBackward>)
[171] Train Loss:    tensor(0.9222, device='cuda:0', grad_fn=<NllLossBackward>)
[172] Train Loss:    tensor(0.9108, device='cuda:0', grad_fn=<NllLossBackward>)
[173] Train Loss:    tensor(0.9342, device='cuda:0', grad_fn=<NllLossBackward>)
[174] Train Loss:    tensor(0.9175, device='cuda:0', grad_fn=<NllLossBackward>)
[175] Train Loss:    tensor(0.9509, device='cuda:0', grad_fn=<NllLossBackward>)
[176] Train Loss:    tensor(0.9082, device='cuda:0', grad_fn=<NllLossBackward>)
[177] Train Loss:    tensor(0.9599, device='cuda:0', grad_fn=<NllLossBackward>)
[178] Train Loss:    tensor(0.9130, device='cuda:0', grad_fn=<NllLossBackward>)
[179] Train Loss:    tensor(0.9371, device='cuda:0', grad_fn=<NllLossBackward>)
[180] Train Loss:    tensor(0.9585, device='cuda:0', grad_fn=<NllLossBackward>)
[181] Train Loss:    tensor(0.9091, device='cuda:0', grad_fn=<NllLossBackward>)
[182] Train Loss:    tensor(0.9219, device='cuda:0', grad_fn=<NllLossBackward>)
[183] Train Loss:    tensor(0.9563, device='cuda:0', grad_fn=<NllLossBackward>)
[184] Train Loss:    tensor(0.9267, device='cuda:0', grad_fn=<NllLossBackward>)
[185] Train Loss:    tensor(0.9164, device='cuda:0', grad_fn=<NllLossBackward>)
[186] Train Loss:    tensor(0.9263, device='cuda:0', grad_fn=<NllLossBackward>)
[187] Train Loss:    tensor(0.9378, device='cuda:0', grad_fn=<NllLossBackward>)
[188] Train Loss:    tensor(0.9352, device='cuda:0', grad_fn=<NllLossBackward>)
[189] Train Loss:    tensor(0.9025, device='cuda:0', grad_fn=<NllLossBackward>)
[190] Train Loss:    tensor(0.9279, device='cuda:0', grad_fn=<NllLossBackward>)
[191] Train Loss:    tensor(0.9280, device='cuda:0', grad_fn=<NllLossBackward>)
[192] Train Loss:    tensor(0.9374, device='cuda:0', grad_fn=<NllLossBackward>)
[193] Train Loss:    tensor(0.8944, device='cuda:0', grad_fn=<NllLossBackward>)
[194] Train Loss:    tensor(0.9132, device='cuda:0', grad_fn=<NllLossBackward>)
[195] Train Loss:    tensor(0.9267, device='cuda:0', grad_fn=<NllLossBackward>)
[196] Train Loss:    tensor(0.9802, device='cuda:0', grad_fn=<NllLossBackward>)
[197] Train Loss:    tensor(0.9489, device='cuda:0', grad_fn=<NllLossBackward>)
[198] Train Loss:    tensor(0.9191, device='cuda:0', grad_fn=<NllLossBackward>)
[199] Train Loss:    tensor(0.9613, device='cuda:0', grad_fn=<NllLossBackward>)
[200] Train Loss:    tensor(0.9232, device='cuda:0', grad_fn=<NllLossBackward>)
[201] Train Loss:    tensor(0.9361, device='cuda:0', grad_fn=<NllLossBackward>)
[202] Train Loss:    tensor(0.9456, device='cuda:0', grad_fn=<NllLossBackward>)
[203] Train Loss:    tensor(0.9739, device='cuda:0', grad_fn=<NllLossBackward>)
[204] Train Loss:    tensor(0.9428, device='cuda:0', grad_fn=<NllLossBackward>)
[205] Train Loss:    tensor(0.9708, device='cuda:0', grad_fn=<NllLossBackward>)
[206] Train Loss:    tensor(0.9708, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[207] Train Loss:  tensor(0.9087, device='cuda:0', grad_fn=<NllLossBackward>)
[208] Train Loss:  tensor(0.8891, device='cuda:0', grad_fn=<NllLossBackward>)
[209] Train Loss:  tensor(0.8911, device='cuda:0', grad_fn=<NllLossBackward>)
[210] Train Loss:  tensor(0.9322, device='cuda:0', grad_fn=<NllLossBackward>)
[211] Train Loss:  tensor(0.9145, device='cuda:0', grad_fn=<NllLossBackward>)
[212] Train Loss:  tensor(0.9311, device='cuda:0', grad_fn=<NllLossBackward>)
[213] Train Loss:  tensor(0.9139, device='cuda:0', grad_fn=<NllLossBackward>)
[214] Train Loss:  tensor(0.9103, device='cuda:0', grad_fn=<NllLossBackward>)
[215] Train Loss:  tensor(0.9173, device='cuda:0', grad_fn=<NllLossBackward>)
[216] Train Loss:  tensor(0.9144, device='cuda:0', grad_fn=<NllLossBackward>)
[217] Train Loss:  tensor(0.9269, device='cuda:0', grad_fn=<NllLossBackward>)
[218] Train Loss:  tensor(0.9273, device='cuda:0', grad_fn=<NllLossBackward>)
[219] Train Loss:  tensor(0.8865, device='cuda:0', grad_fn=<NllLossBackward>)
[220] Train Loss:  tensor(0.9340, device='cuda:0', grad_fn=<NllLossBackward>)
[221] Train Loss:  tensor(0.9142, device='cuda:0', grad_fn=<NllLossBackward>)
[222] Train Loss:  tensor(0.8925, device='cuda:0', grad_fn=<NllLossBackward>)
[223] Train Loss:  tensor(0.9193, device='cuda:0', grad_fn=<NllLossBackward>)
[224] Train Loss:  tensor(0.9087, device='cuda:0', grad_fn=<NllLossBackward>)
[225] Train Loss:  tensor(0.9210, device='cuda:0', grad_fn=<NllLossBackward>)
[226] Train Loss:  tensor(0.9383, device='cuda:0', grad_fn=<NllLossBackward>)
[227] Train Loss:  tensor(0.9250, device='cuda:0', grad_fn=<NllLossBackward>)
[228] Train Loss:  tensor(0.9538, device='cuda:0', grad_fn=<NllLossBackward>)
[229] Train Loss:  tensor(0.9283, device='cuda:0', grad_fn=<NllLossBackward>)
[230] Train Loss:  tensor(0.9327, device='cuda:0', grad_fn=<NllLossBackward>)
[231] Train Loss:  tensor(0.9545, device='cuda:0', grad_fn=<NllLossBackward>)
[232] Train Loss:  tensor(0.9306, device='cuda:0', grad_fn=<NllLossBackward>)
[233] Train Loss:  tensor(0.9087, device='cuda:0', grad_fn=<NllLossBackward>)
[234] Train Loss:  tensor(0.9197, device='cuda:0', grad_fn=<NllLossBackward>)
[235] Train Loss:  tensor(0.9084, device='cuda:0', grad_fn=<NllLossBackward>)
[236] Train Loss:  tensor(0.9662, device='cuda:0', grad_fn=<NllLossBackward>)
[237] Train Loss:  tensor(0.9385, device='cuda:0', grad_fn=<NllLossBackward>)
[238] Train Loss:  tensor(0.9394, device='cuda:0', grad_fn=<NllLossBackward>)
[239] Train Loss:  tensor(0.9287, device='cuda:0', grad_fn=<NllLossBackward>)
[240] Train Loss:  tensor(0.9297, device='cuda:0', grad_fn=<NllLossBackward>)
[241] Train Loss:  tensor(0.9423, device='cuda:0', grad_fn=<NllLossBackward>)
[242] Train Loss:  tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[243] Train Loss:  tensor(0.9454, device='cuda:0', grad_fn=<NllLossBackward>)
[244] Train Loss:  tensor(0.9439, device='cuda:0', grad_fn=<NllLossBackward>)
[245] Train Loss:  tensor(0.9350, device='cuda:0', grad_fn=<NllLossBackward>)
[246] Train Loss:  tensor(0.9365, device='cuda:0', grad_fn=<NllLossBackward>)
[247] Train Loss:  tensor(0.9348, device='cuda:0', grad_fn=<NllLossBackward>)
[248] Train Loss:  tensor(0.9273, device='cuda:0', grad_fn=<NllLossBackward>)
[249] Train Loss:  tensor(0.9243, device='cuda:0', grad_fn=<NllLossBackward>)
[250] Train Loss:  tensor(0.9706, device='cuda:0', grad_fn=<NllLossBackward>)
[251] Train Loss:  tensor(0.9042, device='cuda:0', grad_fn=<NllLossBackward>)
[252] Train Loss:  tensor(0.9190, device='cuda:0', grad_fn=<NllLossBackward>)
[253] Train Loss:  tensor(0.9569, device='cuda:0', grad_fn=<NllLossBackward>)
[254] Train Loss:  tensor(0.9362, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[255] Train Loss:    tensor(0.9186, device='cuda:0', grad_fn=<NllLossBackward>)
[256] Train Loss:    tensor(0.9207, device='cuda:0', grad_fn=<NllLossBackward>)
[257] Train Loss:    tensor(0.9103, device='cuda:0', grad_fn=<NllLossBackward>)
[258] Train Loss:    tensor(0.9448, device='cuda:0', grad_fn=<NllLossBackward>)
[259] Train Loss:    tensor(0.9399, device='cuda:0', grad_fn=<NllLossBackward>)
[260] Train Loss:    tensor(0.8993, device='cuda:0', grad_fn=<NllLossBackward>)
[261] Train Loss:    tensor(0.9237, device='cuda:0', grad_fn=<NllLossBackward>)
[262] Train Loss:    tensor(0.9288, device='cuda:0', grad_fn=<NllLossBackward>)
[263] Train Loss:    tensor(0.9052, device='cuda:0', grad_fn=<NllLossBackward>)
[264] Train Loss:    tensor(0.9247, device='cuda:0', grad_fn=<NllLossBackward>)
[265] Train Loss:    tensor(0.9169, device='cuda:0', grad_fn=<NllLossBackward>)
[266] Train Loss:    tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[267] Train Loss:    tensor(0.9613, device='cuda:0', grad_fn=<NllLossBackward>)
[268] Train Loss:    tensor(0.9377, device='cuda:0', grad_fn=<NllLossBackward>)
[269] Train Loss:    tensor(0.9035, device='cuda:0', grad_fn=<NllLossBackward>)
[270] Train Loss:    tensor(0.9477, device='cuda:0', grad_fn=<NllLossBackward>)
[271] Train Loss:    tensor(0.9573, device='cuda:0', grad_fn=<NllLossBackward>)
[272] Train Loss:    tensor(0.9085, device='cuda:0', grad_fn=<NllLossBackward>)
[273] Train Loss:    tensor(0.9380, device='cuda:0', grad_fn=<NllLossBackward>)
[274] Train Loss:    tensor(0.9131, device='cuda:0', grad_fn=<NllLossBackward>)
[275] Train Loss:    tensor(0.9140, device='cuda:0', grad_fn=<NllLossBackward>)
[276] Train Loss:    tensor(0.9328, device='cuda:0', grad_fn=<NllLossBackward>)
[277] Train Loss:    tensor(0.9065, device='cuda:0', grad_fn=<NllLossBackward>)
[278] Train Loss:    tensor(0.9484, device='cuda:0', grad_fn=<NllLossBackward>)
[279] Train Loss:    tensor(0.9411, device='cuda:0', grad_fn=<NllLossBackward>)
[280] Train Loss:    tensor(0.9581, device='cuda:0', grad_fn=<NllLossBackward>)
[281] Train Loss:    tensor(0.9194, device='cuda:0', grad_fn=<NllLossBackward>)
[282] Train Loss:    tensor(0.9128, device='cuda:0', grad_fn=<NllLossBackward>)
[283] Train Loss:    tensor(0.8850, device='cuda:0', grad_fn=<NllLossBackward>)
[284] Train Loss:    tensor(0.9498, device='cuda:0', grad_fn=<NllLossBackward>)
[285] Train Loss:    tensor(0.9231, device='cuda:0', grad_fn=<NllLossBackward>)
[286] Train Loss:    tensor(0.9107, device='cuda:0', grad_fn=<NllLossBackward>)
[287] Train Loss:    tensor(0.9073, device='cuda:0', grad_fn=<NllLossBackward>)
[288] Train Loss:    tensor(0.9251, device='cuda:0', grad_fn=<NllLossBackward>)
[289] Train Loss:    tensor(0.9422, device='cuda:0', grad_fn=<NllLossBackward>)
[290] Train Loss:    tensor(0.9119, device='cuda:0', grad_fn=<NllLossBackward>)
[291] Train Loss:    tensor(0.9383, device='cuda:0', grad_fn=<NllLossBackward>)
[292] Train Loss:    tensor(0.8983, device='cuda:0', grad_fn=<NllLossBackward>)
[293] Train Loss:    tensor(0.9507, device='cuda:0', grad_fn=<NllLossBackward>)
[294] Train Loss:    tensor(0.9381, device='cuda:0', grad_fn=<NllLossBackward>)
[295] Train Loss:    tensor(0.9088, device='cuda:0', grad_fn=<NllLossBackward>)
[296] Train Loss:    tensor(0.9241, device='cuda:0', grad_fn=<NllLossBackward>)
[297] Train Loss:    tensor(0.9136, device='cuda:0', grad_fn=<NllLossBackward>)
[298] Train Loss:    tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[299] Train Loss:    tensor(0.9091, device='cuda:0', grad_fn=<NllLossBackward>)
[300] Train Loss:    tensor(0.9410, device='cuda:0', grad_fn=<NllLossBackward>)
[301] Train Loss:    tensor(0.8892, device='cuda:0', grad_fn=<NllLossBackward>)
[302] Train Loss:    tensor(0.9457, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[303] Train Loss:   tensor(0.8946, device='cuda:0', grad_fn=<NllLossBackward>)
[304] Train Loss:   tensor(0.9376, device='cuda:0', grad_fn=<NllLossBackward>)
[305] Train Loss:   tensor(0.9075, device='cuda:0', grad_fn=<NllLossBackward>)
[306] Train Loss:   tensor(0.9041, device='cuda:0', grad_fn=<NllLossBackward>)
[307] Train Loss:   tensor(0.9282, device='cuda:0', grad_fn=<NllLossBackward>)
[308] Train Loss:   tensor(0.9551, device='cuda:0', grad_fn=<NllLossBackward>)
[309] Train Loss:   tensor(0.9549, device='cuda:0', grad_fn=<NllLossBackward>)
[310] Train Loss:   tensor(0.9205, device='cuda:0', grad_fn=<NllLossBackward>)
[311] Train Loss:   tensor(0.9910, device='cuda:0', grad_fn=<NllLossBackward>)
[312] Train Loss:   tensor(0.9720, device='cuda:0', grad_fn=<NllLossBackward>)
[313] Train Loss:   tensor(0.9238, device='cuda:0', grad_fn=<NllLossBackward>)
[314] Train Loss:   tensor(0.8911, device='cuda:0', grad_fn=<NllLossBackward>)
[315] Train Loss:   tensor(0.9189, device='cuda:0', grad_fn=<NllLossBackward>)
[316] Train Loss:   tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[317] Train Loss:   tensor(0.9241, device='cuda:0', grad_fn=<NllLossBackward>)
[318] Train Loss:   tensor(0.9287, device='cuda:0', grad_fn=<NllLossBackward>)
[319] Train Loss:   tensor(0.9252, device='cuda:0', grad_fn=<NllLossBackward>)
[320] Train Loss:   tensor(0.9056, device='cuda:0', grad_fn=<NllLossBackward>)
[321] Train Loss:   tensor(0.9576, device='cuda:0', grad_fn=<NllLossBackward>)
[322] Train Loss:   tensor(0.9572, device='cuda:0', grad_fn=<NllLossBackward>)
[323] Train Loss:   tensor(0.9306, device='cuda:0', grad_fn=<NllLossBackward>)
[324] Train Loss:   tensor(0.9533, device='cuda:0', grad_fn=<NllLossBackward>)
[325] Train Loss:   tensor(0.9307, device='cuda:0', grad_fn=<NllLossBackward>)
[326] Train Loss:   tensor(0.9313, device='cuda:0', grad_fn=<NllLossBackward>)
[327] Train Loss:   tensor(0.8993, device='cuda:0', grad_fn=<NllLossBackward>)
[328] Train Loss:   tensor(0.9072, device='cuda:0', grad_fn=<NllLossBackward>)
[329] Train Loss:   tensor(0.9116, device='cuda:0', grad_fn=<NllLossBackward>)
[330] Train Loss:   tensor(0.9397, device='cuda:0', grad_fn=<NllLossBackward>)
[331] Train Loss:   tensor(0.8921, device='cuda:0', grad_fn=<NllLossBackward>)
[332] Train Loss:   tensor(0.8988, device='cuda:0', grad_fn=<NllLossBackward>)
[333] Train Loss:   tensor(0.9544, device='cuda:0', grad_fn=<NllLossBackward>)
[334] Train Loss:   tensor(0.9322, device='cuda:0', grad_fn=<NllLossBackward>)
[335] Train Loss:   tensor(0.9251, device='cuda:0', grad_fn=<NllLossBackward>)
[336] Train Loss:   tensor(0.9144, device='cuda:0', grad_fn=<NllLossBackward>)
[337] Train Loss:   tensor(0.9360, device='cuda:0', grad_fn=<NllLossBackward>)
[338] Train Loss:   tensor(0.9720, device='cuda:0', grad_fn=<NllLossBackward>)
[339] Train Loss:   tensor(0.8974, device='cuda:0', grad_fn=<NllLossBackward>)
[340] Train Loss:   tensor(0.9188, device='cuda:0', grad_fn=<NllLossBackward>)
[341] Train Loss:   tensor(0.8959, device='cuda:0', grad_fn=<NllLossBackward>)
[342] Train Loss:   tensor(0.9100, device='cuda:0', grad_fn=<NllLossBackward>)
[343] Train Loss:   tensor(0.9942, device='cuda:0', grad_fn=<NllLossBackward>)
[344] Train Loss:   tensor(0.9621, device='cuda:0', grad_fn=<NllLossBackward>)
[345] Train Loss:   tensor(0.9434, device='cuda:0', grad_fn=<NllLossBackward>)
[346] Train Loss:   tensor(0.9420, device='cuda:0', grad_fn=<NllLossBackward>)
[347] Train Loss:   tensor(0.9202, device='cuda:0', grad_fn=<NllLossBackward>)
[348] Train Loss:   tensor(0.9271, device='cuda:0', grad_fn=<NllLossBackward>)
[349] Train Loss:   tensor(0.9144, device='cuda:0', grad_fn=<NllLossBackward>)
[350] Train Loss:   tensor(0.9326, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[351] Train Loss:   tensor(0.9454, device='cuda:0', grad_fn=<NllLossBackward>)
[352] Train Loss:   tensor(0.8878, device='cuda:0', grad_fn=<NllLossBackward>)
[353] Train Loss:   tensor(0.9392, device='cuda:0', grad_fn=<NllLossBackward>)
[354] Train Loss:   tensor(0.9402, device='cuda:0', grad_fn=<NllLossBackward>)
[355] Train Loss:   tensor(0.8854, device='cuda:0', grad_fn=<NllLossBackward>)
[356] Train Loss:   tensor(0.9264, device='cuda:0', grad_fn=<NllLossBackward>)
[357] Train Loss:   tensor(0.9296, device='cuda:0', grad_fn=<NllLossBackward>)
[358] Train Loss:   tensor(0.9119, device='cuda:0', grad_fn=<NllLossBackward>)
[359] Train Loss:   tensor(0.9180, device='cuda:0', grad_fn=<NllLossBackward>)
[360] Train Loss:   tensor(0.9392, device='cuda:0', grad_fn=<NllLossBackward>)
[361] Train Loss:   tensor(0.9166, device='cuda:0', grad_fn=<NllLossBackward>)
[362] Train Loss:   tensor(0.9237, device='cuda:0', grad_fn=<NllLossBackward>)
[363] Train Loss:   tensor(0.9447, device='cuda:0', grad_fn=<NllLossBackward>)
[364] Train Loss:   tensor(0.8811, device='cuda:0', grad_fn=<NllLossBackward>)
[365] Train Loss:   tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[366] Train Loss:   tensor(0.9367, device='cuda:0', grad_fn=<NllLossBackward>)
[367] Train Loss:   tensor(0.9093, device='cuda:0', grad_fn=<NllLossBackward>)
[368] Train Loss:   tensor(0.9545, device='cuda:0', grad_fn=<NllLossBackward>)
[369] Train Loss:   tensor(0.9340, device='cuda:0', grad_fn=<NllLossBackward>)
[370] Train Loss:   tensor(0.9493, device='cuda:0', grad_fn=<NllLossBackward>)
[371] Train Loss:   tensor(0.9364, device='cuda:0', grad_fn=<NllLossBackward>)
[372] Train Loss:   tensor(0.9308, device='cuda:0', grad_fn=<NllLossBackward>)
[373] Train Loss:   tensor(0.9448, device='cuda:0', grad_fn=<NllLossBackward>)
[374] Train Loss:   tensor(0.8824, device='cuda:0', grad_fn=<NllLossBackward>)
[375] Train Loss:   tensor(0.9584, device='cuda:0', grad_fn=<NllLossBackward>)
[376] Train Loss:   tensor(0.9220, device='cuda:0', grad_fn=<NllLossBackward>)
[377] Train Loss:   tensor(0.9305, device='cuda:0', grad_fn=<NllLossBackward>)
[378] Train Loss:   tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[379] Train Loss:   tensor(0.9141, device='cuda:0', grad_fn=<NllLossBackward>)
[380] Train Loss:   tensor(0.9777, device='cuda:0', grad_fn=<NllLossBackward>)
[381] Train Loss:   tensor(0.9562, device='cuda:0', grad_fn=<NllLossBackward>)
[382] Train Loss:   tensor(0.9374, device='cuda:0', grad_fn=<NllLossBackward>)
[383] Train Loss:   tensor(0.9155, device='cuda:0', grad_fn=<NllLossBackward>)
[384] Train Loss:   tensor(0.9513, device='cuda:0', grad_fn=<NllLossBackward>)
[385] Train Loss:   tensor(0.9160, device='cuda:0', grad_fn=<NllLossBackward>)
[386] Train Loss:   tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[387] Train Loss:   tensor(0.9565, device='cuda:0', grad_fn=<NllLossBackward>)
[388] Train Loss:   tensor(0.9115, device='cuda:0', grad_fn=<NllLossBackward>)
[389] Train Loss:   tensor(0.9129, device='cuda:0', grad_fn=<NllLossBackward>)
[390] Train Loss:   tensor(0.8983, device='cuda:0', grad_fn=<NllLossBackward>)
[391] Train Loss:   tensor(0.9080, device='cuda:0', grad_fn=<NllLossBackward>)
[392] Train Loss:   tensor(0.9180, device='cuda:0', grad_fn=<NllLossBackward>)
[393] Train Loss:   tensor(0.9208, device='cuda:0', grad_fn=<NllLossBackward>)
[394] Train Loss:   tensor(0.9549, device='cuda:0', grad_fn=<NllLossBackward>)
[395] Train Loss:   tensor(0.9250, device='cuda:0', grad_fn=<NllLossBackward>)
[396] Train Loss:   tensor(0.9288, device='cuda:0', grad_fn=<NllLossBackward>)
[397] Train Loss:   tensor(0.8980, device='cuda:0', grad_fn=<NllLossBackward>)
[398] Train Loss:   tensor(0.9143, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[399] Train Loss:    tensor(0.9158, device='cuda:0', grad_fn=<NllLossBackward>)
[400] Train Loss:    tensor(0.9129, device='cuda:0', grad_fn=<NllLossBackward>)
[401] Train Loss:    tensor(0.9372, device='cuda:0', grad_fn=<NllLossBackward>)
[402] Train Loss:    tensor(0.9126, device='cuda:0', grad_fn=<NllLossBackward>)
[403] Train Loss:    tensor(0.9245, device='cuda:0', grad_fn=<NllLossBackward>)
[404] Train Loss:    tensor(0.9230, device='cuda:0', grad_fn=<NllLossBackward>)
[405] Train Loss:    tensor(0.9462, device='cuda:0', grad_fn=<NllLossBackward>)
[406] Train Loss:    tensor(0.9572, device='cuda:0', grad_fn=<NllLossBackward>)
[407] Train Loss:    tensor(0.9302, device='cuda:0', grad_fn=<NllLossBackward>)
[408] Train Loss:    tensor(0.9433, device='cuda:0', grad_fn=<NllLossBackward>)
[409] Train Loss:    tensor(0.8878, device='cuda:0', grad_fn=<NllLossBackward>)
[410] Train Loss:    tensor(0.9129, device='cuda:0', grad_fn=<NllLossBackward>)
[411] Train Loss:    tensor(0.9408, device='cuda:0', grad_fn=<NllLossBackward>)
[412] Train Loss:    tensor(0.9171, device='cuda:0', grad_fn=<NllLossBackward>)
[413] Train Loss:    tensor(0.9222, device='cuda:0', grad_fn=<NllLossBackward>)
[414] Train Loss:    tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[415] Train Loss:    tensor(0.9222, device='cuda:0', grad_fn=<NllLossBackward>)
[416] Train Loss:    tensor(0.9785, device='cuda:0', grad_fn=<NllLossBackward>)
[417] Train Loss:    tensor(0.9259, device='cuda:0', grad_fn=<NllLossBackward>)
[418] Train Loss:    tensor(0.9667, device='cuda:0', grad_fn=<NllLossBackward>)
[419] Train Loss:    tensor(0.8916, device='cuda:0', grad_fn=<NllLossBackward>)
[420] Train Loss:    tensor(0.9152, device='cuda:0', grad_fn=<NllLossBackward>)
[421] Train Loss:    tensor(0.9049, device='cuda:0', grad_fn=<NllLossBackward>)
[422] Train Loss:    tensor(0.8970, device='cuda:0', grad_fn=<NllLossBackward>)
[423] Train Loss:    tensor(1.0005, device='cuda:0', grad_fn=<NllLossBackward>)
[424] Train Loss:    tensor(0.9735, device='cuda:0', grad_fn=<NllLossBackward>)
[425] Train Loss:    tensor(0.9241, device='cuda:0', grad_fn=<NllLossBackward>)
[426] Train Loss:    tensor(0.9213, device='cuda:0', grad_fn=<NllLossBackward>)
[427] Train Loss:    tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[428] Train Loss:    tensor(0.9253, device='cuda:0', grad_fn=<NllLossBackward>)
[429] Train Loss:    tensor(0.9588, device='cuda:0', grad_fn=<NllLossBackward>)
[430] Train Loss:    tensor(0.9482, device='cuda:0', grad_fn=<NllLossBackward>)
[431] Train Loss:    tensor(0.9508, device='cuda:0', grad_fn=<NllLossBackward>)
[432] Train Loss:    tensor(0.9589, device='cuda:0', grad_fn=<NllLossBackward>)
[433] Train Loss:    tensor(0.9319, device='cuda:0', grad_fn=<NllLossBackward>)
[434] Train Loss:    tensor(0.8990, device='cuda:0', grad_fn=<NllLossBackward>)
[435] Train Loss:    tensor(0.9337, device='cuda:0', grad_fn=<NllLossBackward>)
[436] Train Loss:    tensor(0.9180, device='cuda:0', grad_fn=<NllLossBackward>)
[437] Train Loss:    tensor(0.9163, device='cuda:0', grad_fn=<NllLossBackward>)
[438] Train Loss:    tensor(0.9465, device='cuda:0', grad_fn=<NllLossBackward>)
[439] Train Loss:    tensor(0.9616, device='cuda:0', grad_fn=<NllLossBackward>)
[440] Train Loss:    tensor(0.9293, device='cuda:0', grad_fn=<NllLossBackward>)
[441] Train Loss:    tensor(0.9052, device='cuda:0', grad_fn=<NllLossBackward>)
[442] Train Loss:    tensor(0.9700, device='cuda:0', grad_fn=<NllLossBackward>)
[443] Train Loss:    tensor(0.9273, device='cuda:0', grad_fn=<NllLossBackward>)
[444] Train Loss:    tensor(0.9102, device='cuda:0', grad_fn=<NllLossBackward>)
[445] Train Loss:    tensor(0.9151, device='cuda:0', grad_fn=<NllLossBackward>)
[446] Train Loss:    tensor(0.9300, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[447] Train Loss:    tensor(0.9002, device='cuda:0', grad_fn=<NllLossBackward>)
[448] Train Loss:    tensor(0.9515, device='cuda:0', grad_fn=<NllLossBackward>)
[449] Train Loss:    tensor(0.9399, device='cuda:0', grad_fn=<NllLossBackward>)
[450] Train Loss:    tensor(0.9335, device='cuda:0', grad_fn=<NllLossBackward>)
[451] Train Loss:    tensor(0.9411, device='cuda:0', grad_fn=<NllLossBackward>)
[452] Train Loss:    tensor(0.9413, device='cuda:0', grad_fn=<NllLossBackward>)
[453] Train Loss:    tensor(0.9166, device='cuda:0', grad_fn=<NllLossBackward>)
[454] Train Loss:    tensor(0.9078, device='cuda:0', grad_fn=<NllLossBackward>)
[455] Train Loss:    tensor(0.9167, device='cuda:0', grad_fn=<NllLossBackward>)
[456] Train Loss:    tensor(0.9408, device='cuda:0', grad_fn=<NllLossBackward>)
[457] Train Loss:    tensor(0.9443, device='cuda:0', grad_fn=<NllLossBackward>)
[458] Train Loss:    tensor(0.9397, device='cuda:0', grad_fn=<NllLossBackward>)
[459] Train Loss:    tensor(0.9229, device='cuda:0', grad_fn=<NllLossBackward>)
[460] Train Loss:    tensor(0.9045, device='cuda:0', grad_fn=<NllLossBackward>)
[461] Train Loss:    tensor(0.9146, device='cuda:0', grad_fn=<NllLossBackward>)
[462] Train Loss:    tensor(0.9184, device='cuda:0', grad_fn=<NllLossBackward>)
[463] Train Loss:    tensor(0.9427, device='cuda:0', grad_fn=<NllLossBackward>)
[464] Train Loss:    tensor(0.9576, device='cuda:0', grad_fn=<NllLossBackward>)
[465] Train Loss:    tensor(0.9310, device='cuda:0', grad_fn=<NllLossBackward>)
[466] Train Loss:    tensor(0.9527, device='cuda:0', grad_fn=<NllLossBackward>)
[467] Train Loss:    tensor(0.9244, device='cuda:0', grad_fn=<NllLossBackward>)
[468] Train Loss:    tensor(0.9453, device='cuda:0', grad_fn=<NllLossBackward>)
[469] Train Loss:    tensor(0.8985, device='cuda:0', grad_fn=<NllLossBackward>)
[470] Train Loss:    tensor(0.9483, device='cuda:0', grad_fn=<NllLossBackward>)
[471] Train Loss:    tensor(0.9126, device='cuda:0', grad_fn=<NllLossBackward>)
[472] Train Loss:    tensor(0.9386, device='cuda:0', grad_fn=<NllLossBackward>)
[473] Train Loss:    tensor(0.9316, device='cuda:0', grad_fn=<NllLossBackward>)
[474] Train Loss:    tensor(0.9159, device='cuda:0', grad_fn=<NllLossBackward>)
[475] Train Loss:    tensor(0.9712, device='cuda:0', grad_fn=<NllLossBackward>)
[476] Train Loss:    tensor(0.9284, device='cuda:0', grad_fn=<NllLossBackward>)
[477] Train Loss:    tensor(0.9167, device='cuda:0', grad_fn=<NllLossBackward>)
[478] Train Loss:    tensor(0.9173, device='cuda:0', grad_fn=<NllLossBackward>)
[479] Train Loss:    tensor(0.9285, device='cuda:0', grad_fn=<NllLossBackward>)
[480] Train Loss:    tensor(0.9723, device='cuda:0', grad_fn=<NllLossBackward>)
[481] Train Loss:    tensor(0.9072, device='cuda:0', grad_fn=<NllLossBackward>)
[482] Train Loss:    tensor(0.9278, device='cuda:0', grad_fn=<NllLossBackward>)
[483] Train Loss:    tensor(0.9219, device='cuda:0', grad_fn=<NllLossBackward>)
[484] Train Loss:    tensor(0.9382, device='cuda:0', grad_fn=<NllLossBackward>)
[485] Train Loss:    tensor(0.9405, device='cuda:0', grad_fn=<NllLossBackward>)
[486] Train Loss:    tensor(0.9251, device='cuda:0', grad_fn=<NllLossBackward>)
[487] Train Loss:    tensor(0.9461, device='cuda:0', grad_fn=<NllLossBackward>)
[488] Train Loss:    tensor(0.9568, device='cuda:0', grad_fn=<NllLossBackward>)
[489] Train Loss:    tensor(0.9579, device='cuda:0', grad_fn=<NllLossBackward>)
[490] Train Loss:    tensor(0.9034, device='cuda:0', grad_fn=<NllLossBackward>)
[491] Train Loss:    tensor(0.8765, device='cuda:0', grad_fn=<NllLossBackward>)
[492] Train Loss:    tensor(0.9367, device='cuda:0', grad_fn=<NllLossBackward>)
[493] Train Loss:    tensor(0.9467, device='cuda:0', grad_fn=<NllLossBackward>)
[494] Train Loss:    tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[495] Train Loss:   tensor(0.9183, device='cuda:0', grad_fn=<NllLossBackward>)
[496] Train Loss:   tensor(0.9374, device='cuda:0', grad_fn=<NllLossBackward>)
[497] Train Loss:   tensor(0.8853, device='cuda:0', grad_fn=<NllLossBackward>)
[498] Train Loss:   tensor(0.9340, device='cuda:0', grad_fn=<NllLossBackward>)
[499] Train Loss:   tensor(0.9410, device='cuda:0', grad_fn=<NllLossBackward>)
[500] Train Loss:   tensor(0.9352, device='cuda:0', grad_fn=<NllLossBackward>)
[501] Train Loss:   tensor(0.9567, device='cuda:0', grad_fn=<NllLossBackward>)
[502] Train Loss:   tensor(0.8859, device='cuda:0', grad_fn=<NllLossBackward>)
[503] Train Loss:   tensor(0.9458, device='cuda:0', grad_fn=<NllLossBackward>)
[504] Train Loss:   tensor(0.9331, device='cuda:0', grad_fn=<NllLossBackward>)
[505] Train Loss:   tensor(0.9342, device='cuda:0', grad_fn=<NllLossBackward>)
[506] Train Loss:   tensor(0.9327, device='cuda:0', grad_fn=<NllLossBackward>)
[507] Train Loss:   tensor(0.9173, device='cuda:0', grad_fn=<NllLossBackward>)
[508] Train Loss:   tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[509] Train Loss:   tensor(0.9197, device='cuda:0', grad_fn=<NllLossBackward>)
[510] Train Loss:   tensor(0.8910, device='cuda:0', grad_fn=<NllLossBackward>)
[511] Train Loss:   tensor(0.9335, device='cuda:0', grad_fn=<NllLossBackward>)
[512] Train Loss:   tensor(0.9243, device='cuda:0', grad_fn=<NllLossBackward>)
[513] Train Loss:   tensor(0.9390, device='cuda:0', grad_fn=<NllLossBackward>)
[514] Train Loss:   tensor(0.9304, device='cuda:0', grad_fn=<NllLossBackward>)
[515] Train Loss:   tensor(0.9402, device='cuda:0', grad_fn=<NllLossBackward>)
[516] Train Loss:   tensor(0.9439, device='cuda:0', grad_fn=<NllLossBackward>)
[517] Train Loss:   tensor(0.9314, device='cuda:0', grad_fn=<NllLossBackward>)
[518] Train Loss:   tensor(0.9642, device='cuda:0', grad_fn=<NllLossBackward>)
[519] Train Loss:   tensor(0.9360, device='cuda:0', grad_fn=<NllLossBackward>)
[520] Train Loss:   tensor(0.9590, device='cuda:0', grad_fn=<NllLossBackward>)
[521] Train Loss:   tensor(0.9289, device='cuda:0', grad_fn=<NllLossBackward>)
[522] Train Loss:   tensor(0.9068, device='cuda:0', grad_fn=<NllLossBackward>)
[523] Train Loss:   tensor(0.9271, device='cuda:0', grad_fn=<NllLossBackward>)
[524] Train Loss:   tensor(0.9658, device='cuda:0', grad_fn=<NllLossBackward>)
[525] Train Loss:   tensor(0.9282, device='cuda:0', grad_fn=<NllLossBackward>)
[526] Train Loss:   tensor(0.9288, device='cuda:0', grad_fn=<NllLossBackward>)
[527] Train Loss:   tensor(0.9135, device='cuda:0', grad_fn=<NllLossBackward>)
[528] Train Loss:   tensor(0.9131, device='cuda:0', grad_fn=<NllLossBackward>)
[529] Train Loss:   tensor(0.9103, device='cuda:0', grad_fn=<NllLossBackward>)
[530] Train Loss:   tensor(0.9087, device='cuda:0', grad_fn=<NllLossBackward>)
[531] Train Loss:   tensor(0.9461, device='cuda:0', grad_fn=<NllLossBackward>)
[532] Train Loss:   tensor(0.9612, device='cuda:0', grad_fn=<NllLossBackward>)
[533] Train Loss:   tensor(0.9184, device='cuda:0', grad_fn=<NllLossBackward>)
[534] Train Loss:   tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[535] Train Loss:   tensor(0.9110, device='cuda:0', grad_fn=<NllLossBackward>)
[536] Train Loss:   tensor(0.9165, device='cuda:0', grad_fn=<NllLossBackward>)
[537] Train Loss:   tensor(0.9155, device='cuda:0', grad_fn=<NllLossBackward>)
[538] Train Loss:   tensor(0.9460, device='cuda:0', grad_fn=<NllLossBackward>)
[539] Train Loss:   tensor(0.9543, device='cuda:0', grad_fn=<NllLossBackward>)
[540] Train Loss:   tensor(0.9275, device='cuda:0', grad_fn=<NllLossBackward>)
[541] Train Loss:   tensor(0.9083, device='cuda:0', grad_fn=<NllLossBackward>)
[542] Train Loss:   tensor(0.9315, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[543] Train Loss:    tensor(0.9679, device='cuda:0', grad_fn=<NllLossBackward>)
[544] Train Loss:    tensor(0.9266, device='cuda:0', grad_fn=<NllLossBackward>)
[545] Train Loss:    tensor(0.9214, device='cuda:0', grad_fn=<NllLossBackward>)
[546] Train Loss:    tensor(0.9073, device='cuda:0', grad_fn=<NllLossBackward>)
[547] Train Loss:    tensor(0.9477, device='cuda:0', grad_fn=<NllLossBackward>)
[548] Train Loss:    tensor(0.9275, device='cuda:0', grad_fn=<NllLossBackward>)
[549] Train Loss:    tensor(0.9484, device='cuda:0', grad_fn=<NllLossBackward>)
[550] Train Loss:    tensor(0.9248, device='cuda:0', grad_fn=<NllLossBackward>)
[551] Train Loss:    tensor(0.9259, device='cuda:0', grad_fn=<NllLossBackward>)
[552] Train Loss:    tensor(0.8946, device='cuda:0', grad_fn=<NllLossBackward>)
[553] Train Loss:    tensor(0.9214, device='cuda:0', grad_fn=<NllLossBackward>)
[554] Train Loss:    tensor(0.8915, device='cuda:0', grad_fn=<NllLossBackward>)
[555] Train Loss:    tensor(0.9392, device='cuda:0', grad_fn=<NllLossBackward>)
[556] Train Loss:    tensor(0.9312, device='cuda:0', grad_fn=<NllLossBackward>)
[557] Train Loss:    tensor(0.9101, device='cuda:0', grad_fn=<NllLossBackward>)
[558] Train Loss:    tensor(0.9050, device='cuda:0', grad_fn=<NllLossBackward>)
[559] Train Loss:    tensor(0.9566, device='cuda:0', grad_fn=<NllLossBackward>)
[560] Train Loss:    tensor(0.9086, device='cuda:0', grad_fn=<NllLossBackward>)
[561] Train Loss:    tensor(0.9097, device='cuda:0', grad_fn=<NllLossBackward>)
[562] Train Loss:    tensor(0.9259, device='cuda:0', grad_fn=<NllLossBackward>)
[563] Train Loss:    tensor(0.9293, device='cuda:0', grad_fn=<NllLossBackward>)
[564] Train Loss:    tensor(0.9206, device='cuda:0', grad_fn=<NllLossBackward>)
[565] Train Loss:    tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[566] Train Loss:    tensor(0.9291, device='cuda:0', grad_fn=<NllLossBackward>)
[567] Train Loss:    tensor(0.9135, device='cuda:0', grad_fn=<NllLossBackward>)
[568] Train Loss:    tensor(0.8951, device='cuda:0', grad_fn=<NllLossBackward>)
[569] Train Loss:    tensor(0.9076, device='cuda:0', grad_fn=<NllLossBackward>)
[570] Train Loss:    tensor(0.9301, device='cuda:0', grad_fn=<NllLossBackward>)
[571] Train Loss:    tensor(0.9698, device='cuda:0', grad_fn=<NllLossBackward>)
[572] Train Loss:    tensor(0.8914, device='cuda:0', grad_fn=<NllLossBackward>)
[573] Train Loss:    tensor(0.8752, device='cuda:0', grad_fn=<NllLossBackward>)
[574] Train Loss:    tensor(0.8860, device='cuda:0', grad_fn=<NllLossBackward>)
[575] Train Loss:    tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[576] Train Loss:    tensor(0.9190, device='cuda:0', grad_fn=<NllLossBackward>)
[577] Train Loss:    tensor(0.9318, device='cuda:0', grad_fn=<NllLossBackward>)
[578] Train Loss:    tensor(0.9352, device='cuda:0', grad_fn=<NllLossBackward>)
[579] Train Loss:    tensor(0.9265, device='cuda:0', grad_fn=<NllLossBackward>)
[580] Train Loss:    tensor(0.9479, device='cuda:0', grad_fn=<NllLossBackward>)
[581] Train Loss:    tensor(0.8900, device='cuda:0', grad_fn=<NllLossBackward>)
[582] Train Loss:    tensor(0.9051, device='cuda:0', grad_fn=<NllLossBackward>)
[583] Train Loss:    tensor(0.9276, device='cuda:0', grad_fn=<NllLossBackward>)
[584] Train Loss:    tensor(0.9382, device='cuda:0', grad_fn=<NllLossBackward>)
[585] Train Loss:    tensor(0.9222, device='cuda:0', grad_fn=<NllLossBackward>)
[586] Train Loss:    tensor(0.9235, device='cuda:0', grad_fn=<NllLossBackward>)
[587] Train Loss:    tensor(0.9345, device='cuda:0', grad_fn=<NllLossBackward>)
[588] Train Loss:    tensor(0.9001, device='cuda:0', grad_fn=<NllLossBackward>)
[589] Train Loss:    tensor(0.9340, device='cuda:0', grad_fn=<NllLossBackward>)
[590] Train Loss:    tensor(0.9353, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[591] Train Loss:   tensor(0.9061, device='cuda:0', grad_fn=<NllLossBackward>)
[592] Train Loss:   tensor(0.9034, device='cuda:0', grad_fn=<NllLossBackward>)
[593] Train Loss:   tensor(0.9022, device='cuda:0', grad_fn=<NllLossBackward>)
[594] Train Loss:   tensor(0.9201, device='cuda:0', grad_fn=<NllLossBackward>)
[595] Train Loss:   tensor(0.9411, device='cuda:0', grad_fn=<NllLossBackward>)
[596] Train Loss:   tensor(0.9304, device='cuda:0', grad_fn=<NllLossBackward>)
[597] Train Loss:   tensor(0.9560, device='cuda:0', grad_fn=<NllLossBackward>)
[598] Train Loss:   tensor(0.9167, device='cuda:0', grad_fn=<NllLossBackward>)
[599] Train Loss:   tensor(0.9393, device='cuda:0', grad_fn=<NllLossBackward>)
[600] Train Loss:   tensor(0.9341, device='cuda:0', grad_fn=<NllLossBackward>)
[601] Train Loss:   tensor(0.9208, device='cuda:0', grad_fn=<NllLossBackward>)
[602] Train Loss:   tensor(0.9124, device='cuda:0', grad_fn=<NllLossBackward>)
[603] Train Loss:   tensor(0.9226, device='cuda:0', grad_fn=<NllLossBackward>)
[604] Train Loss:   tensor(0.9244, device='cuda:0', grad_fn=<NllLossBackward>)
[605] Train Loss:   tensor(0.9392, device='cuda:0', grad_fn=<NllLossBackward>)
[606] Train Loss:   tensor(0.9189, device='cuda:0', grad_fn=<NllLossBackward>)
[607] Train Loss:   tensor(0.9092, device='cuda:0', grad_fn=<NllLossBackward>)
[608] Train Loss:   tensor(0.9363, device='cuda:0', grad_fn=<NllLossBackward>)
[609] Train Loss:   tensor(0.9783, device='cuda:0', grad_fn=<NllLossBackward>)
[610] Train Loss:   tensor(0.9105, device='cuda:0', grad_fn=<NllLossBackward>)
[611] Train Loss:   tensor(0.9186, device='cuda:0', grad_fn=<NllLossBackward>)
[612] Train Loss:   tensor(0.9235, device='cuda:0', grad_fn=<NllLossBackward>)
[613] Train Loss:   tensor(0.9081, device='cuda:0', grad_fn=<NllLossBackward>)
[614] Train Loss:   tensor(0.9150, device='cuda:0', grad_fn=<NllLossBackward>)
[615] Train Loss:   tensor(0.9101, device='cuda:0', grad_fn=<NllLossBackward>)
[616] Train Loss:   tensor(0.9338, device='cuda:0', grad_fn=<NllLossBackward>)
[617] Train Loss:   tensor(0.9432, device='cuda:0', grad_fn=<NllLossBackward>)
[618] Train Loss:   tensor(0.8987, device='cuda:0', grad_fn=<NllLossBackward>)
[619] Train Loss:   tensor(0.9415, device='cuda:0', grad_fn=<NllLossBackward>)
[620] Train Loss:   tensor(0.9421, device='cuda:0', grad_fn=<NllLossBackward>)
[621] Train Loss:   tensor(0.9541, device='cuda:0', grad_fn=<NllLossBackward>)
[622] Train Loss:   tensor(0.9342, device='cuda:0', grad_fn=<NllLossBackward>)
[623] Train Loss:   tensor(0.8833, device='cuda:0', grad_fn=<NllLossBackward>)
[624] Train Loss:   tensor(0.9067, device='cuda:0', grad_fn=<NllLossBackward>)
[625] Train Loss:   tensor(0.9160, device='cuda:0', grad_fn=<NllLossBackward>)
[626] Train Loss:   tensor(0.9371, device='cuda:0', grad_fn=<NllLossBackward>)
[627] Train Loss:   tensor(0.9536, device='cuda:0', grad_fn=<NllLossBackward>)
[628] Train Loss:   tensor(0.9506, device='cuda:0', grad_fn=<NllLossBackward>)
[629] Train Loss:   tensor(0.9726, device='cuda:0', grad_fn=<NllLossBackward>)
[630] Train Loss:   tensor(0.9253, device='cuda:0', grad_fn=<NllLossBackward>)
[631] Train Loss:   tensor(0.9408, device='cuda:0', grad_fn=<NllLossBackward>)
[632] Train Loss:   tensor(0.9325, device='cuda:0', grad_fn=<NllLossBackward>)
[633] Train Loss:   tensor(0.9174, device='cuda:0', grad_fn=<NllLossBackward>)
[634] Train Loss:   tensor(0.8967, device='cuda:0', grad_fn=<NllLossBackward>)
[635] Train Loss:   tensor(0.9363, device='cuda:0', grad_fn=<NllLossBackward>)
[636] Train Loss:   tensor(0.9186, device='cuda:0', grad_fn=<NllLossBackward>)
[637] Train Loss:   tensor(0.9347, device='cuda:0', grad_fn=<NllLossBackward>)
[638] Train Loss:   tensor(0.9247, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[639] Train Loss:    tensor(0.9421, device='cuda:0', grad_fn=<NllLossBackward>)
[640] Train Loss:    tensor(0.9282, device='cuda:0', grad_fn=<NllLossBackward>)
[641] Train Loss:    tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[642] Train Loss:    tensor(0.9231, device='cuda:0', grad_fn=<NllLossBackward>)
[643] Train Loss:    tensor(0.9576, device='cuda:0', grad_fn=<NllLossBackward>)
[644] Train Loss:    tensor(0.9605, device='cuda:0', grad_fn=<NllLossBackward>)
[645] Train Loss:    tensor(0.9410, device='cuda:0', grad_fn=<NllLossBackward>)
[646] Train Loss:    tensor(0.9389, device='cuda:0', grad_fn=<NllLossBackward>)
[647] Train Loss:    tensor(0.8948, device='cuda:0', grad_fn=<NllLossBackward>)
[648] Train Loss:    tensor(0.9721, device='cuda:0', grad_fn=<NllLossBackward>)
[649] Train Loss:    tensor(0.9310, device='cuda:0', grad_fn=<NllLossBackward>)
[650] Train Loss:    tensor(0.8878, device='cuda:0', grad_fn=<NllLossBackward>)
[651] Train Loss:    tensor(0.8987, device='cuda:0', grad_fn=<NllLossBackward>)
[652] Train Loss:    tensor(0.9403, device='cuda:0', grad_fn=<NllLossBackward>)
[653] Train Loss:    tensor(0.9355, device='cuda:0', grad_fn=<NllLossBackward>)
[654] Train Loss:    tensor(0.9017, device='cuda:0', grad_fn=<NllLossBackward>)
[655] Train Loss:    tensor(0.9256, device='cuda:0', grad_fn=<NllLossBackward>)
[656] Train Loss:    tensor(0.9419, device='cuda:0', grad_fn=<NllLossBackward>)
[657] Train Loss:    tensor(0.9010, device='cuda:0', grad_fn=<NllLossBackward>)
[658] Train Loss:    tensor(0.8958, device='cuda:0', grad_fn=<NllLossBackward>)
[659] Train Loss:    tensor(0.9075, device='cuda:0', grad_fn=<NllLossBackward>)
[660] Train Loss:    tensor(0.9133, device='cuda:0', grad_fn=<NllLossBackward>)
[661] Train Loss:    tensor(0.9017, device='cuda:0', grad_fn=<NllLossBackward>)
[662] Train Loss:    tensor(0.8869, device='cuda:0', grad_fn=<NllLossBackward>)
[663] Train Loss:    tensor(0.9280, device='cuda:0', grad_fn=<NllLossBackward>)
[664] Train Loss:    tensor(0.9304, device='cuda:0', grad_fn=<NllLossBackward>)
[665] Train Loss:    tensor(0.9200, device='cuda:0', grad_fn=<NllLossBackward>)
[666] Train Loss:    tensor(0.9539, device='cuda:0', grad_fn=<NllLossBackward>)
[667] Train Loss:    tensor(0.9187, device='cuda:0', grad_fn=<NllLossBackward>)
[668] Train Loss:    tensor(0.9504, device='cuda:0', grad_fn=<NllLossBackward>)
[669] Train Loss:    tensor(0.9255, device='cuda:0', grad_fn=<NllLossBackward>)
[670] Train Loss:    tensor(0.9016, device='cuda:0', grad_fn=<NllLossBackward>)
[671] Train Loss:    tensor(0.9088, device='cuda:0', grad_fn=<NllLossBackward>)
[672] Train Loss:    tensor(0.9285, device='cuda:0', grad_fn=<NllLossBackward>)
[673] Train Loss:    tensor(0.9199, device='cuda:0', grad_fn=<NllLossBackward>)
[674] Train Loss:    tensor(0.9302, device='cuda:0', grad_fn=<NllLossBackward>)
[675] Train Loss:    tensor(0.9680, device='cuda:0', grad_fn=<NllLossBackward>)
[676] Train Loss:    tensor(0.9362, device='cuda:0', grad_fn=<NllLossBackward>)
[677] Train Loss:    tensor(0.9308, device='cuda:0', grad_fn=<NllLossBackward>)
[678] Train Loss:    tensor(0.9364, device='cuda:0', grad_fn=<NllLossBackward>)
[679] Train Loss:    tensor(0.9290, device='cuda:0', grad_fn=<NllLossBackward>)
[680] Train Loss:    tensor(0.9265, device='cuda:0', grad_fn=<NllLossBackward>)
[681] Train Loss:    tensor(0.9600, device='cuda:0', grad_fn=<NllLossBackward>)
[682] Train Loss:    tensor(0.9282, device='cuda:0', grad_fn=<NllLossBackward>)
[683] Train Loss:    tensor(0.9363, device='cuda:0', grad_fn=<NllLossBackward>)
[684] Train Loss:    tensor(0.9099, device='cuda:0', grad_fn=<NllLossBackward>)
[685] Train Loss:    tensor(0.9095, device='cuda:0', grad_fn=<NllLossBackward>)
[686] Train Loss:    tensor(0.9079, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[687] Train Loss:   tensor(0.9262, device='cuda:0', grad_fn=<NllLossBackward>)
[688] Train Loss:   tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[689] Train Loss:   tensor(0.9506, device='cuda:0', grad_fn=<NllLossBackward>)
[690] Train Loss:   tensor(0.8906, device='cuda:0', grad_fn=<NllLossBackward>)
[691] Train Loss:   tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[692] Train Loss:   tensor(0.9173, device='cuda:0', grad_fn=<NllLossBackward>)
[693] Train Loss:   tensor(0.9211, device='cuda:0', grad_fn=<NllLossBackward>)
[694] Train Loss:   tensor(0.9229, device='cuda:0', grad_fn=<NllLossBackward>)
[695] Train Loss:   tensor(0.9141, device='cuda:0', grad_fn=<NllLossBackward>)
[696] Train Loss:   tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[697] Train Loss:   tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[698] Train Loss:   tensor(0.9095, device='cuda:0', grad_fn=<NllLossBackward>)
[699] Train Loss:   tensor(0.9614, device='cuda:0', grad_fn=<NllLossBackward>)
[700] Train Loss:   tensor(0.9119, device='cuda:0', grad_fn=<NllLossBackward>)
[701] Train Loss:   tensor(0.9143, device='cuda:0', grad_fn=<NllLossBackward>)
[702] Train Loss:   tensor(0.9517, device='cuda:0', grad_fn=<NllLossBackward>)
[703] Train Loss:   tensor(0.9106, device='cuda:0', grad_fn=<NllLossBackward>)
[704] Train Loss:   tensor(0.9308, device='cuda:0', grad_fn=<NllLossBackward>)
[705] Train Loss:   tensor(0.9258, device='cuda:0', grad_fn=<NllLossBackward>)
[706] Train Loss:   tensor(0.9369, device='cuda:0', grad_fn=<NllLossBackward>)
[707] Train Loss:   tensor(0.9208, device='cuda:0', grad_fn=<NllLossBackward>)
[708] Train Loss:   tensor(0.9087, device='cuda:0', grad_fn=<NllLossBackward>)
[709] Train Loss:   tensor(0.9133, device='cuda:0', grad_fn=<NllLossBackward>)
[710] Train Loss:   tensor(0.9162, device='cuda:0', grad_fn=<NllLossBackward>)
[711] Train Loss:   tensor(0.8979, device='cuda:0', grad_fn=<NllLossBackward>)
[712] Train Loss:   tensor(0.9481, device='cuda:0', grad_fn=<NllLossBackward>)
[713] Train Loss:   tensor(0.8683, device='cuda:0', grad_fn=<NllLossBackward>)
[714] Train Loss:   tensor(0.9841, device='cuda:0', grad_fn=<NllLossBackward>)
[715] Train Loss:   tensor(0.9224, device='cuda:0', grad_fn=<NllLossBackward>)
[716] Train Loss:   tensor(0.9102, device='cuda:0', grad_fn=<NllLossBackward>)
[717] Train Loss:   tensor(0.8925, device='cuda:0', grad_fn=<NllLossBackward>)
[718] Train Loss:   tensor(0.9338, device='cuda:0', grad_fn=<NllLossBackward>)
[719] Train Loss:   tensor(0.9129, device='cuda:0', grad_fn=<NllLossBackward>)
[720] Train Loss:   tensor(0.9048, device='cuda:0', grad_fn=<NllLossBackward>)
[721] Train Loss:   tensor(0.9396, device='cuda:0', grad_fn=<NllLossBackward>)
[722] Train Loss:   tensor(0.9107, device='cuda:0', grad_fn=<NllLossBackward>)
[723] Train Loss:   tensor(0.8772, device='cuda:0', grad_fn=<NllLossBackward>)
[724] Train Loss:   tensor(0.9213, device='cuda:0', grad_fn=<NllLossBackward>)
[725] Train Loss:   tensor(0.9230, device='cuda:0', grad_fn=<NllLossBackward>)
[726] Train Loss:   tensor(0.9291, device='cuda:0', grad_fn=<NllLossBackward>)
[727] Train Loss:   tensor(0.9715, device='cuda:0', grad_fn=<NllLossBackward>)
[728] Train Loss:   tensor(0.9033, device='cuda:0', grad_fn=<NllLossBackward>)
[729] Train Loss:   tensor(0.9140, device='cuda:0', grad_fn=<NllLossBackward>)
[730] Train Loss:   tensor(0.9248, device='cuda:0', grad_fn=<NllLossBackward>)
[731] Train Loss:   tensor(0.9410, device='cuda:0', grad_fn=<NllLossBackward>)
[732] Train Loss:   tensor(0.9065, device='cuda:0', grad_fn=<NllLossBackward>)
[733] Train Loss:   tensor(0.9328, device='cuda:0', grad_fn=<NllLossBackward>)
[734] Train Loss:   tensor(0.9368, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[735] Train Loss:   tensor(0.9260, device='cuda:0', grad_fn=<NllLossBackward>)
[736] Train Loss:   tensor(0.9296, device='cuda:0', grad_fn=<NllLossBackward>)
[737] Train Loss:   tensor(0.9070, device='cuda:0', grad_fn=<NllLossBackward>)
[738] Train Loss:   tensor(0.9403, device='cuda:0', grad_fn=<NllLossBackward>)
[739] Train Loss:   tensor(0.9275, device='cuda:0', grad_fn=<NllLossBackward>)
[740] Train Loss:   tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[741] Train Loss:   tensor(0.9388, device='cuda:0', grad_fn=<NllLossBackward>)
[742] Train Loss:   tensor(0.9410, device='cuda:0', grad_fn=<NllLossBackward>)
[743] Train Loss:   tensor(0.9705, device='cuda:0', grad_fn=<NllLossBackward>)
[744] Train Loss:   tensor(0.9701, device='cuda:0', grad_fn=<NllLossBackward>)
[745] Train Loss:   tensor(0.9345, device='cuda:0', grad_fn=<NllLossBackward>)
[746] Train Loss:   tensor(0.9398, device='cuda:0', grad_fn=<NllLossBackward>)
[747] Train Loss:   tensor(0.9140, device='cuda:0', grad_fn=<NllLossBackward>)
[748] Train Loss:   tensor(0.9008, device='cuda:0', grad_fn=<NllLossBackward>)
[749] Train Loss:   tensor(0.9257, device='cuda:0', grad_fn=<NllLossBackward>)
[750] Train Loss:   tensor(0.9218, device='cuda:0', grad_fn=<NllLossBackward>)
[751] Train Loss:   tensor(0.8913, device='cuda:0', grad_fn=<NllLossBackward>)
[752] Train Loss:   tensor(0.8786, device='cuda:0', grad_fn=<NllLossBackward>)
[753] Train Loss:   tensor(0.9228, device='cuda:0', grad_fn=<NllLossBackward>)
[754] Train Loss:   tensor(0.9047, device='cuda:0', grad_fn=<NllLossBackward>)
[755] Train Loss:   tensor(0.9181, device='cuda:0', grad_fn=<NllLossBackward>)
[756] Train Loss:   tensor(0.9340, device='cuda:0', grad_fn=<NllLossBackward>)
[757] Train Loss:   tensor(0.9269, device='cuda:0', grad_fn=<NllLossBackward>)
[758] Train Loss:   tensor(0.9338, device='cuda:0', grad_fn=<NllLossBackward>)
[759] Train Loss:   tensor(0.9227, device='cuda:0', grad_fn=<NllLossBackward>)
[760] Train Loss:   tensor(0.9267, device='cuda:0', grad_fn=<NllLossBackward>)
[761] Train Loss:   tensor(0.9177, device='cuda:0', grad_fn=<NllLossBackward>)
[762] Train Loss:   tensor(0.9114, device='cuda:0', grad_fn=<NllLossBackward>)
[763] Train Loss:   tensor(0.9267, device='cuda:0', grad_fn=<NllLossBackward>)
[764] Train Loss:   tensor(0.9254, device='cuda:0', grad_fn=<NllLossBackward>)
[765] Train Loss:   tensor(0.9516, device='cuda:0', grad_fn=<NllLossBackward>)
[766] Train Loss:   tensor(0.9299, device='cuda:0', grad_fn=<NllLossBackward>)
[767] Train Loss:   tensor(0.9224, device='cuda:0', grad_fn=<NllLossBackward>)
[768] Train Loss:   tensor(0.9005, device='cuda:0', grad_fn=<NllLossBackward>)
[769] Train Loss:   tensor(0.9212, device='cuda:0', grad_fn=<NllLossBackward>)
[770] Train Loss:   tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[771] Train Loss:   tensor(0.9244, device='cuda:0', grad_fn=<NllLossBackward>)
[772] Train Loss:   tensor(0.9216, device='cuda:0', grad_fn=<NllLossBackward>)
[773] Train Loss:   tensor(0.9228, device='cuda:0', grad_fn=<NllLossBackward>)
[774] Train Loss:   tensor(0.9498, device='cuda:0', grad_fn=<NllLossBackward>)
[775] Train Loss:   tensor(0.9444, device='cuda:0', grad_fn=<NllLossBackward>)
[776] Train Loss:   tensor(0.9441, device='cuda:0', grad_fn=<NllLossBackward>)
[777] Train Loss:   tensor(0.9341, device='cuda:0', grad_fn=<NllLossBackward>)
[778] Train Loss:   tensor(0.9355, device='cuda:0', grad_fn=<NllLossBackward>)
[779] Train Loss:   tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[780] Train Loss:   tensor(0.9591, device='cuda:0', grad_fn=<NllLossBackward>)
[781] Train Loss:   tensor(0.9112, device='cuda:0', grad_fn=<NllLossBackward>)
[782] Train Loss:   tensor(0.9152, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[783] Train Loss:    tensor(0.9015, device='cuda:0', grad_fn=<NllLossBackward>)
[784] Train Loss:    tensor(0.9511, device='cuda:0', grad_fn=<NllLossBackward>)
[785] Train Loss:    tensor(0.8863, device='cuda:0', grad_fn=<NllLossBackward>)
[786] Train Loss:    tensor(0.9217, device='cuda:0', grad_fn=<NllLossBackward>)
[787] Train Loss:    tensor(0.9183, device='cuda:0', grad_fn=<NllLossBackward>)
[788] Train Loss:    tensor(0.9362, device='cuda:0', grad_fn=<NllLossBackward>)
[789] Train Loss:    tensor(0.9662, device='cuda:0', grad_fn=<NllLossBackward>)
[790] Train Loss:    tensor(0.9073, device='cuda:0', grad_fn=<NllLossBackward>)
[791] Train Loss:    tensor(0.9162, device='cuda:0', grad_fn=<NllLossBackward>)
[792] Train Loss:    tensor(0.9446, device='cuda:0', grad_fn=<NllLossBackward>)
[793] Train Loss:    tensor(0.9265, device='cuda:0', grad_fn=<NllLossBackward>)
[794] Train Loss:    tensor(0.9028, device='cuda:0', grad_fn=<NllLossBackward>)
[795] Train Loss:    tensor(0.9271, device='cuda:0', grad_fn=<NllLossBackward>)
[796] Train Loss:    tensor(0.9100, device='cuda:0', grad_fn=<NllLossBackward>)
[797] Train Loss:    tensor(0.9111, device='cuda:0', grad_fn=<NllLossBackward>)
[798] Train Loss:    tensor(0.9016, device='cuda:0', grad_fn=<NllLossBackward>)
[799] Train Loss:    tensor(0.9449, device='cuda:0', grad_fn=<NllLossBackward>)
[800] Train Loss:    tensor(0.8754, device='cuda:0', grad_fn=<NllLossBackward>)
[801] Train Loss:    tensor(0.9433, device='cuda:0', grad_fn=<NllLossBackward>)
[802] Train Loss:    tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[803] Train Loss:    tensor(0.9463, device='cuda:0', grad_fn=<NllLossBackward>)
[804] Train Loss:    tensor(0.8885, device='cuda:0', grad_fn=<NllLossBackward>)
[805] Train Loss:    tensor(0.9589, device='cuda:0', grad_fn=<NllLossBackward>)
[806] Train Loss:    tensor(0.9263, device='cuda:0', grad_fn=<NllLossBackward>)
[807] Train Loss:    tensor(0.8958, device='cuda:0', grad_fn=<NllLossBackward>)
[808] Train Loss:    tensor(0.9278, device='cuda:0', grad_fn=<NllLossBackward>)
[809] Train Loss:    tensor(0.9517, device='cuda:0', grad_fn=<NllLossBackward>)
[810] Train Loss:    tensor(0.9099, device='cuda:0', grad_fn=<NllLossBackward>)
[811] Train Loss:    tensor(0.9085, device='cuda:0', grad_fn=<NllLossBackward>)
[812] Train Loss:    tensor(0.9398, device='cuda:0', grad_fn=<NllLossBackward>)
[813] Train Loss:    tensor(0.8990, device='cuda:0', grad_fn=<NllLossBackward>)
[814] Train Loss:    tensor(0.8940, device='cuda:0', grad_fn=<NllLossBackward>)
[815] Train Loss:    tensor(0.9262, device='cuda:0', grad_fn=<NllLossBackward>)
[816] Train Loss:    tensor(0.9095, device='cuda:0', grad_fn=<NllLossBackward>)
[817] Train Loss:    tensor(0.9124, device='cuda:0', grad_fn=<NllLossBackward>)
[818] Train Loss:    tensor(0.9318, device='cuda:0', grad_fn=<NllLossBackward>)
[819] Train Loss:    tensor(0.9065, device='cuda:0', grad_fn=<NllLossBackward>)
[820] Train Loss:    tensor(0.9048, device='cuda:0', grad_fn=<NllLossBackward>)
[821] Train Loss:    tensor(0.9201, device='cuda:0', grad_fn=<NllLossBackward>)
[822] Train Loss:    tensor(0.9251, device='cuda:0', grad_fn=<NllLossBackward>)
[823] Train Loss:    tensor(0.9283, device='cuda:0', grad_fn=<NllLossBackward>)
[824] Train Loss:    tensor(0.8802, device='cuda:0', grad_fn=<NllLossBackward>)
[825] Train Loss:    tensor(0.9248, device='cuda:0', grad_fn=<NllLossBackward>)
[826] Train Loss:    tensor(0.8919, device='cuda:0', grad_fn=<NllLossBackward>)
[827] Train Loss:    tensor(0.9286, device='cuda:0', grad_fn=<NllLossBackward>)
[828] Train Loss:    tensor(0.9115, device='cuda:0', grad_fn=<NllLossBackward>)
[829] Train Loss:    tensor(0.9058, device='cuda:0', grad_fn=<NllLossBackward>)
[830] Train Loss:    tensor(0.9610, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[831] Train Loss:    tensor(0.9365, device='cuda:0', grad_fn=<NllLossBackward>)
[832] Train Loss:    tensor(0.9115, device='cuda:0', grad_fn=<NllLossBackward>)
[833] Train Loss:    tensor(0.8923, device='cuda:0', grad_fn=<NllLossBackward>)
[834] Train Loss:    tensor(0.9152, device='cuda:0', grad_fn=<NllLossBackward>)
[835] Train Loss:    tensor(0.8995, device='cuda:0', grad_fn=<NllLossBackward>)
[836] Train Loss:    tensor(0.9327, device='cuda:0', grad_fn=<NllLossBackward>)
[837] Train Loss:    tensor(0.9206, device='cuda:0', grad_fn=<NllLossBackward>)
[838] Train Loss:    tensor(0.8871, device='cuda:0', grad_fn=<NllLossBackward>)
[839] Train Loss:    tensor(0.8760, device='cuda:0', grad_fn=<NllLossBackward>)
[840] Train Loss:    tensor(0.9603, device='cuda:0', grad_fn=<NllLossBackward>)
[841] Train Loss:    tensor(0.9089, device='cuda:0', grad_fn=<NllLossBackward>)
[842] Train Loss:    tensor(0.9204, device='cuda:0', grad_fn=<NllLossBackward>)
[843] Train Loss:    tensor(0.9300, device='cuda:0', grad_fn=<NllLossBackward>)
[844] Train Loss:    tensor(0.9258, device='cuda:0', grad_fn=<NllLossBackward>)
[845] Train Loss:    tensor(0.9351, device='cuda:0', grad_fn=<NllLossBackward>)
[846] Train Loss:    tensor(0.9152, device='cuda:0', grad_fn=<NllLossBackward>)
[847] Train Loss:    tensor(0.9252, device='cuda:0', grad_fn=<NllLossBackward>)
[848] Train Loss:    tensor(0.9256, device='cuda:0', grad_fn=<NllLossBackward>)
[849] Train Loss:    tensor(0.9010, device='cuda:0', grad_fn=<NllLossBackward>)
[850] Train Loss:    tensor(0.8912, device='cuda:0', grad_fn=<NllLossBackward>)
[851] Train Loss:    tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[852] Train Loss:    tensor(0.9054, device='cuda:0', grad_fn=<NllLossBackward>)
[853] Train Loss:    tensor(0.9307, device='cuda:0', grad_fn=<NllLossBackward>)
[854] Train Loss:    tensor(0.9425, device='cuda:0', grad_fn=<NllLossBackward>)
[855] Train Loss:    tensor(0.9405, device='cuda:0', grad_fn=<NllLossBackward>)
[856] Train Loss:    tensor(0.9151, device='cuda:0', grad_fn=<NllLossBackward>)
[857] Train Loss:    tensor(0.9082, device='cuda:0', grad_fn=<NllLossBackward>)
[858] Train Loss:    tensor(0.9267, device='cuda:0', grad_fn=<NllLossBackward>)
[859] Train Loss:    tensor(0.9133, device='cuda:0', grad_fn=<NllLossBackward>)
[860] Train Loss:    tensor(0.9327, device='cuda:0', grad_fn=<NllLossBackward>)
[861] Train Loss:    tensor(0.9085, device='cuda:0', grad_fn=<NllLossBackward>)
[862] Train Loss:    tensor(0.8896, device='cuda:0', grad_fn=<NllLossBackward>)
[863] Train Loss:    tensor(0.9549, device='cuda:0', grad_fn=<NllLossBackward>)
[864] Train Loss:    tensor(0.9184, device='cuda:0', grad_fn=<NllLossBackward>)
[865] Train Loss:    tensor(0.9660, device='cuda:0', grad_fn=<NllLossBackward>)
[866] Train Loss:    tensor(0.9146, device='cuda:0', grad_fn=<NllLossBackward>)
[867] Train Loss:    tensor(0.9008, device='cuda:0', grad_fn=<NllLossBackward>)
[868] Train Loss:    tensor(0.9140, device='cuda:0', grad_fn=<NllLossBackward>)
[869] Train Loss:    tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[870] Train Loss:    tensor(0.9079, device='cuda:0', grad_fn=<NllLossBackward>)
[871] Train Loss:    tensor(0.9043, device='cuda:0', grad_fn=<NllLossBackward>)
[872] Train Loss:    tensor(0.9057, device='cuda:0', grad_fn=<NllLossBackward>)
[873] Train Loss:    tensor(0.9399, device='cuda:0', grad_fn=<NllLossBackward>)
[874] Train Loss:    tensor(0.9073, device='cuda:0', grad_fn=<NllLossBackward>)
[875] Train Loss:    tensor(0.9287, device='cuda:0', grad_fn=<NllLossBackward>)
[876] Train Loss:    tensor(0.9563, device='cuda:0', grad_fn=<NllLossBackward>)
[877] Train Loss:    tensor(0.8811, device='cuda:0', grad_fn=<NllLossBackward>)
[878] Train Loss:    tensor(0.9356, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[879] Train Loss:    tensor(0.8822, device='cuda:0', grad_fn=<NllLossBackward>)
[880] Train Loss:    tensor(0.9749, device='cuda:0', grad_fn=<NllLossBackward>)
[881] Train Loss:    tensor(0.9446, device='cuda:0', grad_fn=<NllLossBackward>)
[882] Train Loss:    tensor(0.9028, device='cuda:0', grad_fn=<NllLossBackward>)
[883] Train Loss:    tensor(0.9182, device='cuda:0', grad_fn=<NllLossBackward>)
[884] Train Loss:    tensor(0.9240, device='cuda:0', grad_fn=<NllLossBackward>)
[885] Train Loss:    tensor(0.9274, device='cuda:0', grad_fn=<NllLossBackward>)
[886] Train Loss:    tensor(0.9076, device='cuda:0', grad_fn=<NllLossBackward>)
[887] Train Loss:    tensor(0.9032, device='cuda:0', grad_fn=<NllLossBackward>)
[888] Train Loss:    tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[889] Train Loss:    tensor(0.8972, device='cuda:0', grad_fn=<NllLossBackward>)
[890] Train Loss:    tensor(0.9143, device='cuda:0', grad_fn=<NllLossBackward>)
[891] Train Loss:    tensor(0.9280, device='cuda:0', grad_fn=<NllLossBackward>)
[892] Train Loss:    tensor(0.9015, device='cuda:0', grad_fn=<NllLossBackward>)
[893] Train Loss:    tensor(0.9331, device='cuda:0', grad_fn=<NllLossBackward>)
[894] Train Loss:    tensor(0.9255, device='cuda:0', grad_fn=<NllLossBackward>)
[895] Train Loss:    tensor(0.9540, device='cuda:0', grad_fn=<NllLossBackward>)
[896] Train Loss:    tensor(0.9576, device='cuda:0', grad_fn=<NllLossBackward>)
[897] Train Loss:    tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[898] Train Loss:    tensor(0.9293, device='cuda:0', grad_fn=<NllLossBackward>)
[899] Train Loss:    tensor(0.9206, device='cuda:0', grad_fn=<NllLossBackward>)
[900] Train Loss:    tensor(0.8847, device='cuda:0', grad_fn=<NllLossBackward>)
[901] Train Loss:    tensor(0.9178, device='cuda:0', grad_fn=<NllLossBackward>)
[902] Train Loss:    tensor(0.9672, device='cuda:0', grad_fn=<NllLossBackward>)
[903] Train Loss:    tensor(0.9279, device='cuda:0', grad_fn=<NllLossBackward>)
[904] Train Loss:    tensor(0.9775, device='cuda:0', grad_fn=<NllLossBackward>)
[905] Train Loss:    tensor(0.9092, device='cuda:0', grad_fn=<NllLossBackward>)
[906] Train Loss:    tensor(0.9095, device='cuda:0', grad_fn=<NllLossBackward>)
[907] Train Loss:    tensor(0.9247, device='cuda:0', grad_fn=<NllLossBackward>)
[908] Train Loss:    tensor(0.9192, device='cuda:0', grad_fn=<NllLossBackward>)
[909] Train Loss:    tensor(0.9550, device='cuda:0', grad_fn=<NllLossBackward>)
[910] Train Loss:    tensor(0.9106, device='cuda:0', grad_fn=<NllLossBackward>)
[911] Train Loss:    tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[912] Train Loss:    tensor(0.9175, device='cuda:0', grad_fn=<NllLossBackward>)
[913] Train Loss:    tensor(0.9370, device='cuda:0', grad_fn=<NllLossBackward>)
[914] Train Loss:    tensor(0.9098, device='cuda:0', grad_fn=<NllLossBackward>)
[915] Train Loss:    tensor(0.9521, device='cuda:0', grad_fn=<NllLossBackward>)
[916] Train Loss:    tensor(0.9180, device='cuda:0', grad_fn=<NllLossBackward>)
[917] Train Loss:    tensor(0.9304, device='cuda:0', grad_fn=<NllLossBackward>)
[918] Train Loss:    tensor(0.9365, device='cuda:0', grad_fn=<NllLossBackward>)
[919] Train Loss:    tensor(0.8985, device='cuda:0', grad_fn=<NllLossBackward>)
[920] Train Loss:    tensor(0.8942, device='cuda:0', grad_fn=<NllLossBackward>)
[921] Train Loss:    tensor(0.8594, device='cuda:0', grad_fn=<NllLossBackward>)
[922] Train Loss:    tensor(0.9160, device='cuda:0', grad_fn=<NllLossBackward>)
[923] Train Loss:    tensor(0.9036, device='cuda:0', grad_fn=<NllLossBackward>)
[924] Train Loss:    tensor(0.9308, device='cuda:0', grad_fn=<NllLossBackward>)
[925] Train Loss:    tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[926] Train Loss:    tensor(0.9279, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[927] Train Loss:   tensor(0.9158, device='cuda:0', grad_fn=<NllLossBackward>)
[928] Train Loss:   tensor(0.9136, device='cuda:0', grad_fn=<NllLossBackward>)
[929] Train Loss:   tensor(0.9128, device='cuda:0', grad_fn=<NllLossBackward>)
[930] Train Loss:   tensor(0.9004, device='cuda:0', grad_fn=<NllLossBackward>)
[931] Train Loss:   tensor(0.9155, device='cuda:0', grad_fn=<NllLossBackward>)
[932] Train Loss:   tensor(0.9362, device='cuda:0', grad_fn=<NllLossBackward>)
[933] Train Loss:   tensor(0.9574, device='cuda:0', grad_fn=<NllLossBackward>)
[934] Train Loss:   tensor(0.9151, device='cuda:0', grad_fn=<NllLossBackward>)
[935] Train Loss:   tensor(0.8963, device='cuda:0', grad_fn=<NllLossBackward>)
[936] Train Loss:   tensor(0.9357, device='cuda:0', grad_fn=<NllLossBackward>)
[937] Train Loss:   tensor(0.8976, device='cuda:0', grad_fn=<NllLossBackward>)
[938] Train Loss:   tensor(0.9023, device='cuda:0', grad_fn=<NllLossBackward>)
[939] Train Loss:   tensor(0.8705, device='cuda:0', grad_fn=<NllLossBackward>)
[940] Train Loss:   tensor(0.9056, device='cuda:0', grad_fn=<NllLossBackward>)
[941] Train Loss:   tensor(0.9342, device='cuda:0', grad_fn=<NllLossBackward>)
[942] Train Loss:   tensor(0.9321, device='cuda:0', grad_fn=<NllLossBackward>)
[943] Train Loss:   tensor(0.9071, device='cuda:0', grad_fn=<NllLossBackward>)
[944] Train Loss:   tensor(0.8960, device='cuda:0', grad_fn=<NllLossBackward>)
[945] Train Loss:   tensor(0.9048, device='cuda:0', grad_fn=<NllLossBackward>)
[946] Train Loss:   tensor(0.9110, device='cuda:0', grad_fn=<NllLossBackward>)
[947] Train Loss:   tensor(0.9089, device='cuda:0', grad_fn=<NllLossBackward>)
[948] Train Loss:   tensor(0.9082, device='cuda:0', grad_fn=<NllLossBackward>)
[949] Train Loss:   tensor(0.8704, device='cuda:0', grad_fn=<NllLossBackward>)
[950] Train Loss:   tensor(0.9162, device='cuda:0', grad_fn=<NllLossBackward>)
[951] Train Loss:   tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[952] Train Loss:   tensor(0.9551, device='cuda:0', grad_fn=<NllLossBackward>)
[953] Train Loss:   tensor(0.9318, device='cuda:0', grad_fn=<NllLossBackward>)
[954] Train Loss:   tensor(0.8957, device='cuda:0', grad_fn=<NllLossBackward>)
[955] Train Loss:   tensor(0.9390, device='cuda:0', grad_fn=<NllLossBackward>)
[956] Train Loss:   tensor(0.9458, device='cuda:0', grad_fn=<NllLossBackward>)
[957] Train Loss:   tensor(0.9497, device='cuda:0', grad_fn=<NllLossBackward>)
[958] Train Loss:   tensor(0.9391, device='cuda:0', grad_fn=<NllLossBackward>)
[959] Train Loss:   tensor(0.9671, device='cuda:0', grad_fn=<NllLossBackward>)
[960] Train Loss:   tensor(0.8789, device='cuda:0', grad_fn=<NllLossBackward>)
[961] Train Loss:   tensor(0.9418, device='cuda:0', grad_fn=<NllLossBackward>)
[962] Train Loss:   tensor(0.9346, device='cuda:0', grad_fn=<NllLossBackward>)
[963] Train Loss:   tensor(0.9120, device='cuda:0', grad_fn=<NllLossBackward>)
[964] Train Loss:   tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[965] Train Loss:   tensor(0.9120, device='cuda:0', grad_fn=<NllLossBackward>)
[966] Train Loss:   tensor(0.9061, device='cuda:0', grad_fn=<NllLossBackward>)
[967] Train Loss:   tensor(0.9287, device='cuda:0', grad_fn=<NllLossBackward>)
[968] Train Loss:   tensor(0.9259, device='cuda:0', grad_fn=<NllLossBackward>)
[969] Train Loss:   tensor(0.9320, device='cuda:0', grad_fn=<NllLossBackward>)
[970] Train Loss:   tensor(0.8917, device='cuda:0', grad_fn=<NllLossBackward>)
[971] Train Loss:   tensor(0.9018, device='cuda:0', grad_fn=<NllLossBackward>)
[972] Train Loss:   tensor(0.9169, device='cuda:0', grad_fn=<NllLossBackward>)
[973] Train Loss:   tensor(0.8911, device='cuda:0', grad_fn=<NllLossBackward>)
[974] Train Loss:   tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[975] Train Loss:  tensor(0.8765, device='cuda:0', grad_fn=<NllLossBackward>)
[976] Train Loss:  tensor(0.9224, device='cuda:0', grad_fn=<NllLossBackward>)
[977] Train Loss:  tensor(0.9173, device='cuda:0', grad_fn=<NllLossBackward>)
[978] Train Loss:  tensor(0.9223, device='cuda:0', grad_fn=<NllLossBackward>)
[979] Train Loss:  tensor(0.9251, device='cuda:0', grad_fn=<NllLossBackward>)
[980] Train Loss:  tensor(0.8913, device='cuda:0', grad_fn=<NllLossBackward>)
[981] Train Loss:  tensor(0.8839, device='cuda:0', grad_fn=<NllLossBackward>)
[982] Train Loss:  tensor(0.9051, device='cuda:0', grad_fn=<NllLossBackward>)
[983] Train Loss:  tensor(0.9232, device='cuda:0', grad_fn=<NllLossBackward>)
[984] Train Loss:  tensor(0.9277, device='cuda:0', grad_fn=<NllLossBackward>)
[985] Train Loss:  tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[986] Train Loss:  tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[987] Train Loss:  tensor(0.9506, device='cuda:0', grad_fn=<NllLossBackward>)
[988] Train Loss:  tensor(0.8852, device='cuda:0', grad_fn=<NllLossBackward>)
[989] Train Loss:  tensor(0.8894, device='cuda:0', grad_fn=<NllLossBackward>)
[990] Train Loss:  tensor(0.9295, device='cuda:0', grad_fn=<NllLossBackward>)
[991] Train Loss:  tensor(0.9101, device='cuda:0', grad_fn=<NllLossBackward>)
[992] Train Loss:  tensor(0.9282, device='cuda:0', grad_fn=<NllLossBackward>)
[993] Train Loss:  tensor(0.9259, device='cuda:0', grad_fn=<NllLossBackward>)
[994] Train Loss:  tensor(0.9301, device='cuda:0', grad_fn=<NllLossBackward>)
[995] Train Loss:  tensor(0.9265, device='cuda:0', grad_fn=<NllLossBackward>)
[996] Train Loss:  tensor(0.9290, device='cuda:0', grad_fn=<NllLossBackward>)
[997] Train Loss:  tensor(0.9054, device='cuda:0', grad_fn=<NllLossBackward>)
[998] Train Loss:  tensor(0.9210, device='cuda:0', grad_fn=<NllLossBackward>)
[999] Train Loss:  tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[1000] Train Loss:  tensor(0.8970, device='cuda:0', grad_fn=<NllLossBackward>)
[1001] Train Loss:  tensor(0.9108, device='cuda:0', grad_fn=<NllLossBackward>)
[1002] Train Loss:  tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[1003] Train Loss:  tensor(0.9036, device='cuda:0', grad_fn=<NllLossBackward>)
[1004] Train Loss:  tensor(0.9204, device='cuda:0', grad_fn=<NllLossBackward>)
[1005] Train Loss:  tensor(0.8964, device='cuda:0', grad_fn=<NllLossBackward>)
[1006] Train Loss:  tensor(0.9161, device='cuda:0', grad_fn=<NllLossBackward>)
[1007] Train Loss:  tensor(0.9163, device='cuda:0', grad_fn=<NllLossBackward>)
[1008] Train Loss:  tensor(0.9185, device='cuda:0', grad_fn=<NllLossBackward>)
[1009] Train Loss:  tensor(0.8986, device='cuda:0', grad_fn=<NllLossBackward>)
[1010] Train Loss:  tensor(0.9519, device='cuda:0', grad_fn=<NllLossBackward>)
[1011] Train Loss:  tensor(0.9144, device='cuda:0', grad_fn=<NllLossBackward>)
[1012] Train Loss:  tensor(0.9114, device='cuda:0', grad_fn=<NllLossBackward>)
[1013] Train Loss:  tensor(0.9496, device='cuda:0', grad_fn=<NllLossBackward>)
[1014] Train Loss:  tensor(0.9360, device='cuda:0', grad_fn=<NllLossBackward>)
[1015] Train Loss:  tensor(0.9228, device='cuda:0', grad_fn=<NllLossBackward>)
[1016] Train Loss:  tensor(0.9099, device='cuda:0', grad_fn=<NllLossBackward>)
[1017] Train Loss:  tensor(0.8919, device='cuda:0', grad_fn=<NllLossBackward>)
[1018] Train Loss:  tensor(0.9168, device='cuda:0', grad_fn=<NllLossBackward>)
[1019] Train Loss:  tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[1020] Train Loss:  tensor(0.9331, device='cuda:0', grad_fn=<NllLossBackward>)
[1021] Train Loss:  tensor(0.9256, device='cuda:0', grad_fn=<NllLossBackward>)
[1022] Train Loss:  tensor(0.9117, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1023] Train Loss:    tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[1024] Train Loss:    tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[1025] Train Loss:    tensor(0.9095, device='cuda:0', grad_fn=<NllLossBackward>)
[1026] Train Loss:    tensor(0.8876, device='cuda:0', grad_fn=<NllLossBackward>)
[1027] Train Loss:    tensor(0.8980, device='cuda:0', grad_fn=<NllLossBackward>)
[1028] Train Loss:    tensor(0.9058, device='cuda:0', grad_fn=<NllLossBackward>)
[1029] Train Loss:    tensor(0.9276, device='cuda:0', grad_fn=<NllLossBackward>)
[1030] Train Loss:    tensor(0.8775, device='cuda:0', grad_fn=<NllLossBackward>)
[1031] Train Loss:    tensor(0.8910, device='cuda:0', grad_fn=<NllLossBackward>)
[1032] Train Loss:    tensor(0.9358, device='cuda:0', grad_fn=<NllLossBackward>)
[1033] Train Loss:    tensor(0.9108, device='cuda:0', grad_fn=<NllLossBackward>)
[1034] Train Loss:    tensor(0.9107, device='cuda:0', grad_fn=<NllLossBackward>)
[1035] Train Loss:    tensor(0.9455, device='cuda:0', grad_fn=<NllLossBackward>)
[1036] Train Loss:    tensor(0.9164, device='cuda:0', grad_fn=<NllLossBackward>)
[1037] Train Loss:    tensor(0.8894, device='cuda:0', grad_fn=<NllLossBackward>)
[1038] Train Loss:    tensor(0.8864, device='cuda:0', grad_fn=<NllLossBackward>)
[1039] Train Loss:    tensor(0.9391, device='cuda:0', grad_fn=<NllLossBackward>)
[1040] Train Loss:    tensor(0.8866, device='cuda:0', grad_fn=<NllLossBackward>)
[1041] Train Loss:    tensor(0.9176, device='cuda:0', grad_fn=<NllLossBackward>)
[1042] Train Loss:    tensor(0.8696, device='cuda:0', grad_fn=<NllLossBackward>)
[1043] Train Loss:    tensor(0.9592, device='cuda:0', grad_fn=<NllLossBackward>)
[1044] Train Loss:    tensor(0.8966, device='cuda:0', grad_fn=<NllLossBackward>)
[1045] Train Loss:    tensor(0.9179, device='cuda:0', grad_fn=<NllLossBackward>)
[1046] Train Loss:    tensor(0.8801, device='cuda:0', grad_fn=<NllLossBackward>)
[1047] Train Loss:    tensor(0.8914, device='cuda:0', grad_fn=<NllLossBackward>)
[1048] Train Loss:    tensor(0.9114, device='cuda:0', grad_fn=<NllLossBackward>)
[1049] Train Loss:    tensor(0.8956, device='cuda:0', grad_fn=<NllLossBackward>)
[1050] Train Loss:    tensor(0.9088, device='cuda:0', grad_fn=<NllLossBackward>)
[1051] Train Loss:    tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[1052] Train Loss:    tensor(0.9289, device='cuda:0', grad_fn=<NllLossBackward>)
[1053] Train Loss:    tensor(0.9328, device='cuda:0', grad_fn=<NllLossBackward>)
[1054] Train Loss:    tensor(0.9531, device='cuda:0', grad_fn=<NllLossBackward>)
[1055] Train Loss:    tensor(0.8916, device='cuda:0', grad_fn=<NllLossBackward>)
[1056] Train Loss:    tensor(0.8847, device='cuda:0', grad_fn=<NllLossBackward>)
[1057] Train Loss:    tensor(0.9235, device='cuda:0', grad_fn=<NllLossBackward>)
[1058] Train Loss:    tensor(0.8901, device='cuda:0', grad_fn=<NllLossBackward>)
[1059] Train Loss:    tensor(0.9132, device='cuda:0', grad_fn=<NllLossBackward>)
[1060] Train Loss:    tensor(0.9185, device='cuda:0', grad_fn=<NllLossBackward>)
[1061] Train Loss:    tensor(0.9086, device='cuda:0', grad_fn=<NllLossBackward>)
[1062] Train Loss:    tensor(0.9055, device='cuda:0', grad_fn=<NllLossBackward>)
[1063] Train Loss:    tensor(0.9088, device='cuda:0', grad_fn=<NllLossBackward>)
[1064] Train Loss:    tensor(0.9018, device='cuda:0', grad_fn=<NllLossBackward>)
[1065] Train Loss:    tensor(0.9004, device='cuda:0', grad_fn=<NllLossBackward>)
[1066] Train Loss:    tensor(0.9371, device='cuda:0', grad_fn=<NllLossBackward>)
[1067] Train Loss:    tensor(0.9276, device='cuda:0', grad_fn=<NllLossBackward>)
[1068] Train Loss:    tensor(0.9370, device='cuda:0', grad_fn=<NllLossBackward>)
[1069] Train Loss:    tensor(0.9301, device='cuda:0', grad_fn=<NllLossBackward>)
[1070] Train Loss:    tensor(0.9354, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1071] Train Loss:    tensor(0.8849, device='cuda:0', grad_fn=<NllLossBackward>)
[1072] Train Loss:    tensor(0.9037, device='cuda:0', grad_fn=<NllLossBackward>)
[1073] Train Loss:    tensor(0.9041, device='cuda:0', grad_fn=<NllLossBackward>)
[1074] Train Loss:    tensor(0.8760, device='cuda:0', grad_fn=<NllLossBackward>)
[1075] Train Loss:    tensor(0.9415, device='cuda:0', grad_fn=<NllLossBackward>)
[1076] Train Loss:    tensor(0.9323, device='cuda:0', grad_fn=<NllLossBackward>)
[1077] Train Loss:    tensor(0.8834, device='cuda:0', grad_fn=<NllLossBackward>)
[1078] Train Loss:    tensor(0.9169, device='cuda:0', grad_fn=<NllLossBackward>)
[1079] Train Loss:    tensor(0.9559, device='cuda:0', grad_fn=<NllLossBackward>)
[1080] Train Loss:    tensor(0.9114, device='cuda:0', grad_fn=<NllLossBackward>)
[1081] Train Loss:    tensor(0.9375, device='cuda:0', grad_fn=<NllLossBackward>)
[1082] Train Loss:    tensor(0.9146, device='cuda:0', grad_fn=<NllLossBackward>)
[1083] Train Loss:    tensor(0.9152, device='cuda:0', grad_fn=<NllLossBackward>)
[1084] Train Loss:    tensor(0.9401, device='cuda:0', grad_fn=<NllLossBackward>)
[1085] Train Loss:    tensor(0.9225, device='cuda:0', grad_fn=<NllLossBackward>)
[1086] Train Loss:    tensor(0.8812, device='cuda:0', grad_fn=<NllLossBackward>)
[1087] Train Loss:    tensor(0.9246, device='cuda:0', grad_fn=<NllLossBackward>)
[1088] Train Loss:    tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[1089] Train Loss:    tensor(0.9107, device='cuda:0', grad_fn=<NllLossBackward>)
[1090] Train Loss:    tensor(0.9198, device='cuda:0', grad_fn=<NllLossBackward>)
[1091] Train Loss:    tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[1092] Train Loss:    tensor(0.9038, device='cuda:0', grad_fn=<NllLossBackward>)
[1093] Train Loss:    tensor(0.9090, device='cuda:0', grad_fn=<NllLossBackward>)
[1094] Train Loss:    tensor(0.8984, device='cuda:0', grad_fn=<NllLossBackward>)
[1095] Train Loss:    tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[1096] Train Loss:    tensor(0.8867, device='cuda:0', grad_fn=<NllLossBackward>)
[1097] Train Loss:    tensor(0.9421, device='cuda:0', grad_fn=<NllLossBackward>)
[1098] Train Loss:    tensor(0.9147, device='cuda:0', grad_fn=<NllLossBackward>)
[1099] Train Loss:    tensor(0.8920, device='cuda:0', grad_fn=<NllLossBackward>)
[1100] Train Loss:    tensor(0.9044, device='cuda:0', grad_fn=<NllLossBackward>)
[1101] Train Loss:    tensor(0.8745, device='cuda:0', grad_fn=<NllLossBackward>)
[1102] Train Loss:    tensor(0.9225, device='cuda:0', grad_fn=<NllLossBackward>)
[1103] Train Loss:    tensor(0.8620, device='cuda:0', grad_fn=<NllLossBackward>)
[1104] Train Loss:    tensor(0.8989, device='cuda:0', grad_fn=<NllLossBackward>)
[1105] Train Loss:    tensor(0.8867, device='cuda:0', grad_fn=<NllLossBackward>)
[1106] Train Loss:    tensor(0.8970, device='cuda:0', grad_fn=<NllLossBackward>)
[1107] Train Loss:    tensor(0.8960, device='cuda:0', grad_fn=<NllLossBackward>)
[1108] Train Loss:    tensor(0.9316, device='cuda:0', grad_fn=<NllLossBackward>)
[1109] Train Loss:    tensor(0.9057, device='cuda:0', grad_fn=<NllLossBackward>)
[1110] Train Loss:    tensor(0.9005, device='cuda:0', grad_fn=<NllLossBackward>)
[1111] Train Loss:    tensor(0.8992, device='cuda:0', grad_fn=<NllLossBackward>)
[1112] Train Loss:    tensor(0.9223, device='cuda:0', grad_fn=<NllLossBackward>)
[1113] Train Loss:    tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[1114] Train Loss:    tensor(0.8985, device='cuda:0', grad_fn=<NllLossBackward>)
[1115] Train Loss:    tensor(0.8980, device='cuda:0', grad_fn=<NllLossBackward>)
[1116] Train Loss:    tensor(0.9140, device='cuda:0', grad_fn=<NllLossBackward>)
[1117] Train Loss:    tensor(0.9013, device='cuda:0', grad_fn=<NllLossBackward>)
[1118] Train Loss:    tensor(0.9029, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1119] Train Loss:   tensor(0.8943, device='cuda:0', grad_fn=<NllLossBackward>)
[1120] Train Loss:   tensor(0.9244, device='cuda:0', grad_fn=<NllLossBackward>)
[1121] Train Loss:   tensor(0.9452, device='cuda:0', grad_fn=<NllLossBackward>)
[1122] Train Loss:   tensor(0.9377, device='cuda:0', grad_fn=<NllLossBackward>)
[1123] Train Loss:   tensor(0.9098, device='cuda:0', grad_fn=<NllLossBackward>)
[1124] Train Loss:   tensor(0.9091, device='cuda:0', grad_fn=<NllLossBackward>)
[1125] Train Loss:   tensor(0.8910, device='cuda:0', grad_fn=<NllLossBackward>)
[1126] Train Loss:   tensor(0.9087, device='cuda:0', grad_fn=<NllLossBackward>)
[1127] Train Loss:   tensor(0.9445, device='cuda:0', grad_fn=<NllLossBackward>)
[1128] Train Loss:   tensor(0.9018, device='cuda:0', grad_fn=<NllLossBackward>)
[1129] Train Loss:   tensor(0.9106, device='cuda:0', grad_fn=<NllLossBackward>)
[1130] Train Loss:   tensor(0.9252, device='cuda:0', grad_fn=<NllLossBackward>)
[1131] Train Loss:   tensor(0.9290, device='cuda:0', grad_fn=<NllLossBackward>)
[1132] Train Loss:   tensor(0.8858, device='cuda:0', grad_fn=<NllLossBackward>)
[1133] Train Loss:   tensor(0.9201, device='cuda:0', grad_fn=<NllLossBackward>)
[1134] Train Loss:   tensor(0.9200, device='cuda:0', grad_fn=<NllLossBackward>)
[1135] Train Loss:   tensor(0.9190, device='cuda:0', grad_fn=<NllLossBackward>)
[1136] Train Loss:   tensor(0.9288, device='cuda:0', grad_fn=<NllLossBackward>)
[1137] Train Loss:   tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[1138] Train Loss:   tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[1139] Train Loss:   tensor(0.9089, device='cuda:0', grad_fn=<NllLossBackward>)
[1140] Train Loss:   tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[1141] Train Loss:   tensor(0.8963, device='cuda:0', grad_fn=<NllLossBackward>)
[1142] Train Loss:   tensor(0.9048, device='cuda:0', grad_fn=<NllLossBackward>)
[1143] Train Loss:   tensor(0.9316, device='cuda:0', grad_fn=<NllLossBackward>)
[1144] Train Loss:   tensor(0.9045, device='cuda:0', grad_fn=<NllLossBackward>)
[1145] Train Loss:   tensor(0.8970, device='cuda:0', grad_fn=<NllLossBackward>)
[1146] Train Loss:   tensor(0.8929, device='cuda:0', grad_fn=<NllLossBackward>)
[1147] Train Loss:   tensor(0.9002, device='cuda:0', grad_fn=<NllLossBackward>)
[1148] Train Loss:   tensor(0.9362, device='cuda:0', grad_fn=<NllLossBackward>)
[1149] Train Loss:   tensor(0.9479, device='cuda:0', grad_fn=<NllLossBackward>)
[1150] Train Loss:   tensor(0.9234, device='cuda:0', grad_fn=<NllLossBackward>)
[1151] Train Loss:   tensor(0.8919, device='cuda:0', grad_fn=<NllLossBackward>)
[1152] Train Loss:   tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[1153] Train Loss:   tensor(0.8962, device='cuda:0', grad_fn=<NllLossBackward>)
[1154] Train Loss:   tensor(0.8923, device='cuda:0', grad_fn=<NllLossBackward>)
[1155] Train Loss:   tensor(0.9115, device='cuda:0', grad_fn=<NllLossBackward>)
[1156] Train Loss:   tensor(0.9160, device='cuda:0', grad_fn=<NllLossBackward>)
[1157] Train Loss:   tensor(0.9242, device='cuda:0', grad_fn=<NllLossBackward>)
[1158] Train Loss:   tensor(0.9136, device='cuda:0', grad_fn=<NllLossBackward>)
[1159] Train Loss:   tensor(0.9272, device='cuda:0', grad_fn=<NllLossBackward>)
[1160] Train Loss:   tensor(0.9526, device='cuda:0', grad_fn=<NllLossBackward>)
[1161] Train Loss:   tensor(0.9037, device='cuda:0', grad_fn=<NllLossBackward>)
[1162] Train Loss:   tensor(0.9407, device='cuda:0', grad_fn=<NllLossBackward>)
[1163] Train Loss:   tensor(0.9171, device='cuda:0', grad_fn=<NllLossBackward>)
[1164] Train Loss:   tensor(0.8845, device='cuda:0', grad_fn=<NllLossBackward>)
[1165] Train Loss:   tensor(0.8831, device='cuda:0', grad_fn=<NllLossBackward>)
[1166] Train Loss:   tensor(0.9262, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1167] Train Loss:    tensor(0.8973, device='cuda:0', grad_fn=<NllLossBackward>)
[1168] Train Loss:    tensor(0.9238, device='cuda:0', grad_fn=<NllLossBackward>)
[1169] Train Loss:    tensor(0.9212, device='cuda:0', grad_fn=<NllLossBackward>)
[1170] Train Loss:    tensor(0.9386, device='cuda:0', grad_fn=<NllLossBackward>)
[1171] Train Loss:    tensor(0.9098, device='cuda:0', grad_fn=<NllLossBackward>)
[1172] Train Loss:    tensor(0.8800, device='cuda:0', grad_fn=<NllLossBackward>)
[1173] Train Loss:    tensor(0.9084, device='cuda:0', grad_fn=<NllLossBackward>)
[1174] Train Loss:    tensor(0.9194, device='cuda:0', grad_fn=<NllLossBackward>)
[1175] Train Loss:    tensor(0.8825, device='cuda:0', grad_fn=<NllLossBackward>)
[1176] Train Loss:    tensor(0.8900, device='cuda:0', grad_fn=<NllLossBackward>)
[1177] Train Loss:    tensor(0.8948, device='cuda:0', grad_fn=<NllLossBackward>)
[1178] Train Loss:    tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[1179] Train Loss:    tensor(0.9067, device='cuda:0', grad_fn=<NllLossBackward>)
[1180] Train Loss:    tensor(0.8719, device='cuda:0', grad_fn=<NllLossBackward>)
[1181] Train Loss:    tensor(0.9257, device='cuda:0', grad_fn=<NllLossBackward>)
[1182] Train Loss:    tensor(0.9283, device='cuda:0', grad_fn=<NllLossBackward>)
[1183] Train Loss:    tensor(0.8703, device='cuda:0', grad_fn=<NllLossBackward>)
[1184] Train Loss:    tensor(0.9107, device='cuda:0', grad_fn=<NllLossBackward>)
[1185] Train Loss:    tensor(0.9185, device='cuda:0', grad_fn=<NllLossBackward>)
[1186] Train Loss:    tensor(0.8886, device='cuda:0', grad_fn=<NllLossBackward>)
[1187] Train Loss:    tensor(0.9029, device='cuda:0', grad_fn=<NllLossBackward>)
[1188] Train Loss:    tensor(0.9435, device='cuda:0', grad_fn=<NllLossBackward>)
[1189] Train Loss:    tensor(0.9376, device='cuda:0', grad_fn=<NllLossBackward>)
[1190] Train Loss:    tensor(0.9072, device='cuda:0', grad_fn=<NllLossBackward>)
[1191] Train Loss:    tensor(0.9193, device='cuda:0', grad_fn=<NllLossBackward>)
[1192] Train Loss:    tensor(0.9346, device='cuda:0', grad_fn=<NllLossBackward>)
[1193] Train Loss:    tensor(0.9217, device='cuda:0', grad_fn=<NllLossBackward>)
[1194] Train Loss:    tensor(0.8928, device='cuda:0', grad_fn=<NllLossBackward>)
[1195] Train Loss:    tensor(0.8953, device='cuda:0', grad_fn=<NllLossBackward>)
[1196] Train Loss:    tensor(0.9456, device='cuda:0', grad_fn=<NllLossBackward>)
[1197] Train Loss:    tensor(0.9085, device='cuda:0', grad_fn=<NllLossBackward>)
[1198] Train Loss:    tensor(0.8972, device='cuda:0', grad_fn=<NllLossBackward>)
[1199] Train Loss:    tensor(0.8721, device='cuda:0', grad_fn=<NllLossBackward>)
[1200] Train Loss:    tensor(0.8957, device='cuda:0', grad_fn=<NllLossBackward>)
[1201] Train Loss:    tensor(0.8969, device='cuda:0', grad_fn=<NllLossBackward>)
[1202] Train Loss:    tensor(0.9226, device='cuda:0', grad_fn=<NllLossBackward>)
[1203] Train Loss:    tensor(0.9356, device='cuda:0', grad_fn=<NllLossBackward>)
[1204] Train Loss:    tensor(0.9371, device='cuda:0', grad_fn=<NllLossBackward>)
[1205] Train Loss:    tensor(0.8853, device='cuda:0', grad_fn=<NllLossBackward>)
[1206] Train Loss:    tensor(0.9218, device='cuda:0', grad_fn=<NllLossBackward>)
[1207] Train Loss:    tensor(0.9139, device='cuda:0', grad_fn=<NllLossBackward>)
[1208] Train Loss:    tensor(0.9120, device='cuda:0', grad_fn=<NllLossBackward>)
[1209] Train Loss:    tensor(0.9319, device='cuda:0', grad_fn=<NllLossBackward>)
[1210] Train Loss:    tensor(0.8974, device='cuda:0', grad_fn=<NllLossBackward>)
[1211] Train Loss:    tensor(0.9381, device='cuda:0', grad_fn=<NllLossBackward>)
[1212] Train Loss:    tensor(0.8776, device='cuda:0', grad_fn=<NllLossBackward>)
[1213] Train Loss:    tensor(0.9360, device='cuda:0', grad_fn=<NllLossBackward>)
[1214] Train Loss:    tensor(0.8917, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1215] Train Loss:    tensor(0.9024, device='cuda:0', grad_fn=<NllLossBackward>)
[1216] Train Loss:    tensor(0.9405, device='cuda:0', grad_fn=<NllLossBackward>)
[1217] Train Loss:    tensor(0.8735, device='cuda:0', grad_fn=<NllLossBackward>)
[1218] Train Loss:    tensor(0.8901, device='cuda:0', grad_fn=<NllLossBackward>)
[1219] Train Loss:    tensor(0.8876, device='cuda:0', grad_fn=<NllLossBackward>)
[1220] Train Loss:    tensor(0.9009, device='cuda:0', grad_fn=<NllLossBackward>)
[1221] Train Loss:    tensor(0.9159, device='cuda:0', grad_fn=<NllLossBackward>)
[1222] Train Loss:    tensor(0.9462, device='cuda:0', grad_fn=<NllLossBackward>)
[1223] Train Loss:    tensor(0.9347, device='cuda:0', grad_fn=<NllLossBackward>)
[1224] Train Loss:    tensor(0.9239, device='cuda:0', grad_fn=<NllLossBackward>)
[1225] Train Loss:    tensor(0.9385, device='cuda:0', grad_fn=<NllLossBackward>)
[1226] Train Loss:    tensor(0.9270, device='cuda:0', grad_fn=<NllLossBackward>)
[1227] Train Loss:    tensor(0.9305, device='cuda:0', grad_fn=<NllLossBackward>)
[1228] Train Loss:    tensor(0.9218, device='cuda:0', grad_fn=<NllLossBackward>)
[1229] Train Loss:    tensor(0.9007, device='cuda:0', grad_fn=<NllLossBackward>)
[1230] Train Loss:    tensor(0.9216, device='cuda:0', grad_fn=<NllLossBackward>)
[1231] Train Loss:    tensor(0.9024, device='cuda:0', grad_fn=<NllLossBackward>)
[1232] Train Loss:    tensor(0.9240, device='cuda:0', grad_fn=<NllLossBackward>)
[1233] Train Loss:    tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[1234] Train Loss:    tensor(0.9031, device='cuda:0', grad_fn=<NllLossBackward>)
[1235] Train Loss:    tensor(0.9223, device='cuda:0', grad_fn=<NllLossBackward>)
[1236] Train Loss:    tensor(0.9544, device='cuda:0', grad_fn=<NllLossBackward>)
[1237] Train Loss:    tensor(0.8949, device='cuda:0', grad_fn=<NllLossBackward>)
[1238] Train Loss:    tensor(0.9304, device='cuda:0', grad_fn=<NllLossBackward>)
[1239] Train Loss:    tensor(0.9346, device='cuda:0', grad_fn=<NllLossBackward>)
[1240] Train Loss:    tensor(0.8906, device='cuda:0', grad_fn=<NllLossBackward>)
[1241] Train Loss:    tensor(0.9132, device='cuda:0', grad_fn=<NllLossBackward>)
[1242] Train Loss:    tensor(0.9334, device='cuda:0', grad_fn=<NllLossBackward>)
[1243] Train Loss:    tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[1244] Train Loss:    tensor(0.9350, device='cuda:0', grad_fn=<NllLossBackward>)
[1245] Train Loss:    tensor(0.9161, device='cuda:0', grad_fn=<NllLossBackward>)
[1246] Train Loss:    tensor(0.9300, device='cuda:0', grad_fn=<NllLossBackward>)
[1247] Train Loss:    tensor(0.9420, device='cuda:0', grad_fn=<NllLossBackward>)
[1248] Train Loss:    tensor(0.9063, device='cuda:0', grad_fn=<NllLossBackward>)
[1249] Train Loss:    tensor(0.9268, device='cuda:0', grad_fn=<NllLossBackward>)
[1250] Train Loss:    tensor(0.9241, device='cuda:0', grad_fn=<NllLossBackward>)
[1251] Train Loss:    tensor(0.9018, device='cuda:0', grad_fn=<NllLossBackward>)
[1252] Train Loss:    tensor(0.9175, device='cuda:0', grad_fn=<NllLossBackward>)
[1253] Train Loss:    tensor(0.9038, device='cuda:0', grad_fn=<NllLossBackward>)
[1254] Train Loss:    tensor(0.8894, device='cuda:0', grad_fn=<NllLossBackward>)
[1255] Train Loss:    tensor(0.9295, device='cuda:0', grad_fn=<NllLossBackward>)
[1256] Train Loss:    tensor(0.9193, device='cuda:0', grad_fn=<NllLossBackward>)
[1257] Train Loss:    tensor(0.9427, device='cuda:0', grad_fn=<NllLossBackward>)
[1258] Train Loss:    tensor(0.8974, device='cuda:0', grad_fn=<NllLossBackward>)
[1259] Train Loss:    tensor(0.9325, device='cuda:0', grad_fn=<NllLossBackward>)
[1260] Train Loss:    tensor(0.9266, device='cuda:0', grad_fn=<NllLossBackward>)
[1261] Train Loss:    tensor(0.9352, device='cuda:0', grad_fn=<NllLossBackward>)
[1262] Train Loss:    tensor(0.9216, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1263] Train Loss:    tensor(0.9197, device='cuda:0', grad_fn=<NllLossBackward>)
[1264] Train Loss:    tensor(0.9054, device='cuda:0', grad_fn=<NllLossBackward>)
[1265] Train Loss:    tensor(0.8973, device='cuda:0', grad_fn=<NllLossBackward>)
[1266] Train Loss:    tensor(0.9306, device='cuda:0', grad_fn=<NllLossBackward>)
[1267] Train Loss:    tensor(0.9424, device='cuda:0', grad_fn=<NllLossBackward>)
[1268] Train Loss:    tensor(0.8841, device='cuda:0', grad_fn=<NllLossBackward>)
[1269] Train Loss:    tensor(0.9310, device='cuda:0', grad_fn=<NllLossBackward>)
[1270] Train Loss:    tensor(0.9070, device='cuda:0', grad_fn=<NllLossBackward>)
[1271] Train Loss:    tensor(0.8684, device='cuda:0', grad_fn=<NllLossBackward>)
[1272] Train Loss:    tensor(0.9177, device='cuda:0', grad_fn=<NllLossBackward>)
[1273] Train Loss:    tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[1274] Train Loss:    tensor(0.8848, device='cuda:0', grad_fn=<NllLossBackward>)
[1275] Train Loss:    tensor(0.8823, device='cuda:0', grad_fn=<NllLossBackward>)
[1276] Train Loss:    tensor(0.8952, device='cuda:0', grad_fn=<NllLossBackward>)
[1277] Train Loss:    tensor(0.8928, device='cuda:0', grad_fn=<NllLossBackward>)
[1278] Train Loss:    tensor(0.9037, device='cuda:0', grad_fn=<NllLossBackward>)
[1279] Train Loss:    tensor(0.9227, device='cuda:0', grad_fn=<NllLossBackward>)
[1280] Train Loss:    tensor(0.9104, device='cuda:0', grad_fn=<NllLossBackward>)
[1281] Train Loss:    tensor(0.8959, device='cuda:0', grad_fn=<NllLossBackward>)
[1282] Train Loss:    tensor(0.9039, device='cuda:0', grad_fn=<NllLossBackward>)
[1283] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[1284] Train Loss:    tensor(0.8873, device='cuda:0', grad_fn=<NllLossBackward>)
[1285] Train Loss:    tensor(0.9138, device='cuda:0', grad_fn=<NllLossBackward>)
[1286] Train Loss:    tensor(0.9342, device='cuda:0', grad_fn=<NllLossBackward>)
[1287] Train Loss:    tensor(0.9632, device='cuda:0', grad_fn=<NllLossBackward>)
[1288] Train Loss:    tensor(0.9407, device='cuda:0', grad_fn=<NllLossBackward>)
[1289] Train Loss:    tensor(0.9012, device='cuda:0', grad_fn=<NllLossBackward>)
[1290] Train Loss:    tensor(0.9378, device='cuda:0', grad_fn=<NllLossBackward>)
[1291] Train Loss:    tensor(0.9263, device='cuda:0', grad_fn=<NllLossBackward>)
[1292] Train Loss:    tensor(0.8946, device='cuda:0', grad_fn=<NllLossBackward>)
[1293] Train Loss:    tensor(0.9142, device='cuda:0', grad_fn=<NllLossBackward>)
[1294] Train Loss:    tensor(0.8726, device='cuda:0', grad_fn=<NllLossBackward>)
[1295] Train Loss:    tensor(0.9125, device='cuda:0', grad_fn=<NllLossBackward>)
[1296] Train Loss:    tensor(0.8934, device='cuda:0', grad_fn=<NllLossBackward>)
[1297] Train Loss:    tensor(0.8863, device='cuda:0', grad_fn=<NllLossBackward>)
[1298] Train Loss:    tensor(0.9184, device='cuda:0', grad_fn=<NllLossBackward>)
[1299] Train Loss:    tensor(0.8812, device='cuda:0', grad_fn=<NllLossBackward>)
[1300] Train Loss:    tensor(0.9111, device='cuda:0', grad_fn=<NllLossBackward>)
[1301] Train Loss:    tensor(0.9368, device='cuda:0', grad_fn=<NllLossBackward>)
[1302] Train Loss:    tensor(0.9379, device='cuda:0', grad_fn=<NllLossBackward>)
[1303] Train Loss:    tensor(0.8973, device='cuda:0', grad_fn=<NllLossBackward>)
[1304] Train Loss:    tensor(0.9237, device='cuda:0', grad_fn=<NllLossBackward>)
[1305] Train Loss:    tensor(0.9022, device='cuda:0', grad_fn=<NllLossBackward>)
[1306] Train Loss:    tensor(0.9329, device='cuda:0', grad_fn=<NllLossBackward>)
[1307] Train Loss:    tensor(0.9171, device='cuda:0', grad_fn=<NllLossBackward>)
[1308] Train Loss:    tensor(0.8934, device='cuda:0', grad_fn=<NllLossBackward>)
[1309] Train Loss:    tensor(0.9115, device='cuda:0', grad_fn=<NllLossBackward>)
[1310] Train Loss:    tensor(0.9082, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1311] Train Loss:    tensor(0.8786, device='cuda:0', grad_fn=<NllLossBackward>)
[1312] Train Loss:    tensor(0.9186, device='cuda:0', grad_fn=<NllLossBackward>)
[1313] Train Loss:    tensor(0.9064, device='cuda:0', grad_fn=<NllLossBackward>)
[1314] Train Loss:    tensor(0.9458, device='cuda:0', grad_fn=<NllLossBackward>)
[1315] Train Loss:    tensor(0.9809, device='cuda:0', grad_fn=<NllLossBackward>)
[1316] Train Loss:    tensor(0.9118, device='cuda:0', grad_fn=<NllLossBackward>)
[1317] Train Loss:    tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[1318] Train Loss:    tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[1319] Train Loss:    tensor(0.9013, device='cuda:0', grad_fn=<NllLossBackward>)
[1320] Train Loss:    tensor(0.9530, device='cuda:0', grad_fn=<NllLossBackward>)
[1321] Train Loss:    tensor(0.9025, device='cuda:0', grad_fn=<NllLossBackward>)
[1322] Train Loss:    tensor(0.8787, device='cuda:0', grad_fn=<NllLossBackward>)
[1323] Train Loss:    tensor(0.9069, device='cuda:0', grad_fn=<NllLossBackward>)
[1324] Train Loss:    tensor(0.9166, device='cuda:0', grad_fn=<NllLossBackward>)
[1325] Train Loss:    tensor(0.8942, device='cuda:0', grad_fn=<NllLossBackward>)
[1326] Train Loss:    tensor(0.9092, device='cuda:0', grad_fn=<NllLossBackward>)
[1327] Train Loss:    tensor(0.9035, device='cuda:0', grad_fn=<NllLossBackward>)
[1328] Train Loss:    tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[1329] Train Loss:    tensor(0.9408, device='cuda:0', grad_fn=<NllLossBackward>)
[1330] Train Loss:    tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[1331] Train Loss:    tensor(0.8839, device='cuda:0', grad_fn=<NllLossBackward>)
[1332] Train Loss:    tensor(0.9169, device='cuda:0', grad_fn=<NllLossBackward>)
[1333] Train Loss:    tensor(0.9091, device='cuda:0', grad_fn=<NllLossBackward>)
[1334] Train Loss:    tensor(0.8812, device='cuda:0', grad_fn=<NllLossBackward>)
[1335] Train Loss:    tensor(0.9060, device='cuda:0', grad_fn=<NllLossBackward>)
[1336] Train Loss:    tensor(0.9311, device='cuda:0', grad_fn=<NllLossBackward>)
[1337] Train Loss:    tensor(0.9240, device='cuda:0', grad_fn=<NllLossBackward>)
[1338] Train Loss:    tensor(0.9044, device='cuda:0', grad_fn=<NllLossBackward>)
[1339] Train Loss:    tensor(0.9200, device='cuda:0', grad_fn=<NllLossBackward>)
[1340] Train Loss:    tensor(0.9396, device='cuda:0', grad_fn=<NllLossBackward>)
[1341] Train Loss:    tensor(0.8968, device='cuda:0', grad_fn=<NllLossBackward>)
[1342] Train Loss:    tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[1343] Train Loss:    tensor(0.8949, device='cuda:0', grad_fn=<NllLossBackward>)
[1344] Train Loss:    tensor(0.9152, device='cuda:0', grad_fn=<NllLossBackward>)
[1345] Train Loss:    tensor(0.9138, device='cuda:0', grad_fn=<NllLossBackward>)
[1346] Train Loss:    tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[1347] Train Loss:    tensor(0.9392, device='cuda:0', grad_fn=<NllLossBackward>)
[1348] Train Loss:    tensor(0.8967, device='cuda:0', grad_fn=<NllLossBackward>)
[1349] Train Loss:    tensor(0.9243, device='cuda:0', grad_fn=<NllLossBackward>)
[1350] Train Loss:    tensor(0.9009, device='cuda:0', grad_fn=<NllLossBackward>)
[1351] Train Loss:    tensor(0.9400, device='cuda:0', grad_fn=<NllLossBackward>)
[1352] Train Loss:    tensor(0.8981, device='cuda:0', grad_fn=<NllLossBackward>)
[1353] Train Loss:    tensor(0.8810, device='cuda:0', grad_fn=<NllLossBackward>)
[1354] Train Loss:    tensor(0.9060, device='cuda:0', grad_fn=<NllLossBackward>)
[1355] Train Loss:    tensor(0.8863, device='cuda:0', grad_fn=<NllLossBackward>)
[1356] Train Loss:    tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[1357] Train Loss:    tensor(0.9244, device='cuda:0', grad_fn=<NllLossBackward>)
[1358] Train Loss:    tensor(0.8809, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1359] Train Loss:    tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[1360] Train Loss:    tensor(0.8982, device='cuda:0', grad_fn=<NllLossBackward>)
[1361] Train Loss:    tensor(0.9131, device='cuda:0', grad_fn=<NllLossBackward>)
[1362] Train Loss:    tensor(0.8932, device='cuda:0', grad_fn=<NllLossBackward>)
[1363] Train Loss:    tensor(0.9111, device='cuda:0', grad_fn=<NllLossBackward>)
[1364] Train Loss:    tensor(0.9328, device='cuda:0', grad_fn=<NllLossBackward>)
[1365] Train Loss:    tensor(0.9138, device='cuda:0', grad_fn=<NllLossBackward>)
[1366] Train Loss:    tensor(0.9178, device='cuda:0', grad_fn=<NllLossBackward>)
[1367] Train Loss:    tensor(0.9071, device='cuda:0', grad_fn=<NllLossBackward>)
[1368] Train Loss:    tensor(0.9303, device='cuda:0', grad_fn=<NllLossBackward>)
[1369] Train Loss:    tensor(0.9488, device='cuda:0', grad_fn=<NllLossBackward>)
[1370] Train Loss:    tensor(0.9157, device='cuda:0', grad_fn=<NllLossBackward>)
[1371] Train Loss:    tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[1372] Train Loss:    tensor(0.9095, device='cuda:0', grad_fn=<NllLossBackward>)
[1373] Train Loss:    tensor(0.9170, device='cuda:0', grad_fn=<NllLossBackward>)
[1374] Train Loss:    tensor(0.9142, device='cuda:0', grad_fn=<NllLossBackward>)
[1375] Train Loss:    tensor(0.9109, device='cuda:0', grad_fn=<NllLossBackward>)
[1376] Train Loss:    tensor(0.9055, device='cuda:0', grad_fn=<NllLossBackward>)
[1377] Train Loss:    tensor(0.8863, device='cuda:0', grad_fn=<NllLossBackward>)
[1378] Train Loss:    tensor(0.9097, device='cuda:0', grad_fn=<NllLossBackward>)
[1379] Train Loss:    tensor(0.9079, device='cuda:0', grad_fn=<NllLossBackward>)
[1380] Train Loss:    tensor(0.9089, device='cuda:0', grad_fn=<NllLossBackward>)
[1381] Train Loss:    tensor(0.8972, device='cuda:0', grad_fn=<NllLossBackward>)
[1382] Train Loss:    tensor(0.9336, device='cuda:0', grad_fn=<NllLossBackward>)
[1383] Train Loss:    tensor(0.9471, device='cuda:0', grad_fn=<NllLossBackward>)
[1384] Train Loss:    tensor(0.8800, device='cuda:0', grad_fn=<NllLossBackward>)
[1385] Train Loss:    tensor(0.9241, device='cuda:0', grad_fn=<NllLossBackward>)
[1386] Train Loss:    tensor(0.9661, device='cuda:0', grad_fn=<NllLossBackward>)
[1387] Train Loss:    tensor(0.9242, device='cuda:0', grad_fn=<NllLossBackward>)
[1388] Train Loss:    tensor(0.8607, device='cuda:0', grad_fn=<NllLossBackward>)
[1389] Train Loss:    tensor(0.9255, device='cuda:0', grad_fn=<NllLossBackward>)
[1390] Train Loss:    tensor(0.8929, device='cuda:0', grad_fn=<NllLossBackward>)
[1391] Train Loss:    tensor(0.9475, device='cuda:0', grad_fn=<NllLossBackward>)
[1392] Train Loss:    tensor(0.8910, device='cuda:0', grad_fn=<NllLossBackward>)
[1393] Train Loss:    tensor(0.9165, device='cuda:0', grad_fn=<NllLossBackward>)
[1394] Train Loss:    tensor(0.8919, device='cuda:0', grad_fn=<NllLossBackward>)
[1395] Train Loss:    tensor(0.9060, device='cuda:0', grad_fn=<NllLossBackward>)
[1396] Train Loss:    tensor(0.9408, device='cuda:0', grad_fn=<NllLossBackward>)
[1397] Train Loss:    tensor(0.9070, device='cuda:0', grad_fn=<NllLossBackward>)
[1398] Train Loss:    tensor(0.9265, device='cuda:0', grad_fn=<NllLossBackward>)
[1399] Train Loss:    tensor(0.9304, device='cuda:0', grad_fn=<NllLossBackward>)
[1400] Train Loss:    tensor(0.9567, device='cuda:0', grad_fn=<NllLossBackward>)
[1401] Train Loss:    tensor(0.9114, device='cuda:0', grad_fn=<NllLossBackward>)
[1402] Train Loss:    tensor(0.8865, device='cuda:0', grad_fn=<NllLossBackward>)
[1403] Train Loss:    tensor(0.9229, device='cuda:0', grad_fn=<NllLossBackward>)
[1404] Train Loss:    tensor(0.9178, device='cuda:0', grad_fn=<NllLossBackward>)
[1405] Train Loss:    tensor(0.9412, device='cuda:0', grad_fn=<NllLossBackward>)
[1406] Train Loss:    tensor(0.8918, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1407] Train Loss:    tensor(0.8676, device='cuda:0', grad_fn=<NllLossBackward>)
[1408] Train Loss:    tensor(0.8990, device='cuda:0', grad_fn=<NllLossBackward>)
[1409] Train Loss:    tensor(0.9120, device='cuda:0', grad_fn=<NllLossBackward>)
[1410] Train Loss:    tensor(0.9163, device='cuda:0', grad_fn=<NllLossBackward>)
[1411] Train Loss:    tensor(0.9087, device='cuda:0', grad_fn=<NllLossBackward>)
[1412] Train Loss:    tensor(0.8966, device='cuda:0', grad_fn=<NllLossBackward>)
[1413] Train Loss:    tensor(0.8818, device='cuda:0', grad_fn=<NllLossBackward>)
[1414] Train Loss:    tensor(0.8687, device='cuda:0', grad_fn=<NllLossBackward>)
[1415] Train Loss:    tensor(0.8964, device='cuda:0', grad_fn=<NllLossBackward>)
[1416] Train Loss:    tensor(0.9253, device='cuda:0', grad_fn=<NllLossBackward>)
[1417] Train Loss:    tensor(0.9282, device='cuda:0', grad_fn=<NllLossBackward>)
[1418] Train Loss:    tensor(0.8977, device='cuda:0', grad_fn=<NllLossBackward>)
[1419] Train Loss:    tensor(0.8892, device='cuda:0', grad_fn=<NllLossBackward>)
[1420] Train Loss:    tensor(0.9094, device='cuda:0', grad_fn=<NllLossBackward>)
[1421] Train Loss:    tensor(0.9129, device='cuda:0', grad_fn=<NllLossBackward>)
[1422] Train Loss:    tensor(0.8992, device='cuda:0', grad_fn=<NllLossBackward>)
[1423] Train Loss:    tensor(0.9100, device='cuda:0', grad_fn=<NllLossBackward>)
[1424] Train Loss:    tensor(0.8911, device='cuda:0', grad_fn=<NllLossBackward>)
[1425] Train Loss:    tensor(0.9348, device='cuda:0', grad_fn=<NllLossBackward>)
[1426] Train Loss:    tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[1427] Train Loss:    tensor(0.9195, device='cuda:0', grad_fn=<NllLossBackward>)
[1428] Train Loss:    tensor(0.9188, device='cuda:0', grad_fn=<NllLossBackward>)
[1429] Train Loss:    tensor(0.8936, device='cuda:0', grad_fn=<NllLossBackward>)
[1430] Train Loss:    tensor(0.8750, device='cuda:0', grad_fn=<NllLossBackward>)
[1431] Train Loss:    tensor(0.9414, device='cuda:0', grad_fn=<NllLossBackward>)
[1432] Train Loss:    tensor(0.9296, device='cuda:0', grad_fn=<NllLossBackward>)
[1433] Train Loss:    tensor(0.9161, device='cuda:0', grad_fn=<NllLossBackward>)
[1434] Train Loss:    tensor(0.8878, device='cuda:0', grad_fn=<NllLossBackward>)
[1435] Train Loss:    tensor(0.8694, device='cuda:0', grad_fn=<NllLossBackward>)
[1436] Train Loss:    tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[1437] Train Loss:    tensor(0.9428, device='cuda:0', grad_fn=<NllLossBackward>)
[1438] Train Loss:    tensor(0.9114, device='cuda:0', grad_fn=<NllLossBackward>)
[1439] Train Loss:    tensor(0.9178, device='cuda:0', grad_fn=<NllLossBackward>)
[1440] Train Loss:    tensor(0.9290, device='cuda:0', grad_fn=<NllLossBackward>)
[1441] Train Loss:    tensor(0.8905, device='cuda:0', grad_fn=<NllLossBackward>)
[1442] Train Loss:    tensor(0.9178, device='cuda:0', grad_fn=<NllLossBackward>)
[1443] Train Loss:    tensor(0.9227, device='cuda:0', grad_fn=<NllLossBackward>)
[1444] Train Loss:    tensor(0.8760, device='cuda:0', grad_fn=<NllLossBackward>)
[1445] Train Loss:    tensor(0.9017, device='cuda:0', grad_fn=<NllLossBackward>)
[1446] Train Loss:    tensor(0.9437, device='cuda:0', grad_fn=<NllLossBackward>)
[1447] Train Loss:    tensor(0.9409, device='cuda:0', grad_fn=<NllLossBackward>)
[1448] Train Loss:    tensor(0.9094, device='cuda:0', grad_fn=<NllLossBackward>)
[1449] Train Loss:    tensor(0.8735, device='cuda:0', grad_fn=<NllLossBackward>)
[1450] Train Loss:    tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[1451] Train Loss:    tensor(0.9237, device='cuda:0', grad_fn=<NllLossBackward>)
[1452] Train Loss:    tensor(0.9124, device='cuda:0', grad_fn=<NllLossBackward>)
[1453] Train Loss:    tensor(0.8787, device='cuda:0', grad_fn=<NllLossBackward>)
[1454] Train Loss:    tensor(0.8783, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1455] Train Loss:    tensor(0.9339, device='cuda:0', grad_fn=<NllLossBackward>)
[1456] Train Loss:    tensor(0.9083, device='cuda:0', grad_fn=<NllLossBackward>)
[1457] Train Loss:    tensor(0.9608, device='cuda:0', grad_fn=<NllLossBackward>)
[1458] Train Loss:    tensor(0.9233, device='cuda:0', grad_fn=<NllLossBackward>)
[1459] Train Loss:    tensor(0.9171, device='cuda:0', grad_fn=<NllLossBackward>)
[1460] Train Loss:    tensor(0.9065, device='cuda:0', grad_fn=<NllLossBackward>)
[1461] Train Loss:    tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[1462] Train Loss:    tensor(0.9249, device='cuda:0', grad_fn=<NllLossBackward>)
[1463] Train Loss:    tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[1464] Train Loss:    tensor(0.8789, device='cuda:0', grad_fn=<NllLossBackward>)
[1465] Train Loss:    tensor(0.9028, device='cuda:0', grad_fn=<NllLossBackward>)
[1466] Train Loss:    tensor(0.9013, device='cuda:0', grad_fn=<NllLossBackward>)
[1467] Train Loss:    tensor(0.8885, device='cuda:0', grad_fn=<NllLossBackward>)
[1468] Train Loss:    tensor(0.9162, device='cuda:0', grad_fn=<NllLossBackward>)
[1469] Train Loss:    tensor(0.8908, device='cuda:0', grad_fn=<NllLossBackward>)
[1470] Train Loss:    tensor(0.9137, device='cuda:0', grad_fn=<NllLossBackward>)
[1471] Train Loss:    tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[1472] Train Loss:    tensor(0.9206, device='cuda:0', grad_fn=<NllLossBackward>)
[1473] Train Loss:    tensor(0.8861, device='cuda:0', grad_fn=<NllLossBackward>)
[1474] Train Loss:    tensor(0.9131, device='cuda:0', grad_fn=<NllLossBackward>)
[1475] Train Loss:    tensor(0.9420, device='cuda:0', grad_fn=<NllLossBackward>)
[1476] Train Loss:    tensor(0.9020, device='cuda:0', grad_fn=<NllLossBackward>)
[1477] Train Loss:    tensor(0.9183, device='cuda:0', grad_fn=<NllLossBackward>)
[1478] Train Loss:    tensor(0.9211, device='cuda:0', grad_fn=<NllLossBackward>)
[1479] Train Loss:    tensor(0.9209, device='cuda:0', grad_fn=<NllLossBackward>)
[1480] Train Loss:    tensor(0.8960, device='cuda:0', grad_fn=<NllLossBackward>)
[1481] Train Loss:    tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[1482] Train Loss:    tensor(0.9314, device='cuda:0', grad_fn=<NllLossBackward>)
[1483] Train Loss:    tensor(0.9085, device='cuda:0', grad_fn=<NllLossBackward>)
[1484] Train Loss:    tensor(0.8458, device='cuda:0', grad_fn=<NllLossBackward>)
[1485] Train Loss:    tensor(0.8977, device='cuda:0', grad_fn=<NllLossBackward>)
[1486] Train Loss:    tensor(0.9264, device='cuda:0', grad_fn=<NllLossBackward>)
[1487] Train Loss:    tensor(0.9275, device='cuda:0', grad_fn=<NllLossBackward>)
[1488] Train Loss:    tensor(0.9046, device='cuda:0', grad_fn=<NllLossBackward>)
[1489] Train Loss:    tensor(0.9210, device='cuda:0', grad_fn=<NllLossBackward>)
[1490] Train Loss:    tensor(0.8847, device='cuda:0', grad_fn=<NllLossBackward>)
[1491] Train Loss:    tensor(0.8847, device='cuda:0', grad_fn=<NllLossBackward>)
[1492] Train Loss:    tensor(0.9060, device='cuda:0', grad_fn=<NllLossBackward>)
[1493] Train Loss:    tensor(0.8802, device='cuda:0', grad_fn=<NllLossBackward>)
[1494] Train Loss:    tensor(0.9001, device='cuda:0', grad_fn=<NllLossBackward>)
[1495] Train Loss:    tensor(0.8780, device='cuda:0', grad_fn=<NllLossBackward>)
[1496] Train Loss:    tensor(0.8880, device='cuda:0', grad_fn=<NllLossBackward>)
[1497] Train Loss:    tensor(0.9343, device='cuda:0', grad_fn=<NllLossBackward>)
[1498] Train Loss:    tensor(0.8621, device='cuda:0', grad_fn=<NllLossBackward>)
[1499] Train Loss:    tensor(0.9148, device='cuda:0', grad_fn=<NllLossBackward>)
[1500] Train Loss:    tensor(0.9388, device='cuda:0', grad_fn=<NllLossBackward>)
[1501] Train Loss:    tensor(0.8814, device='cuda:0', grad_fn=<NllLossBackward>)
[1502] Train Loss:    tensor(0.8540, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1503] Train Loss:    tensor(0.8884, device='cuda:0', grad_fn=<NllLossBackward>)
[1504] Train Loss:    tensor(0.8774, device='cuda:0', grad_fn=<NllLossBackward>)
[1505] Train Loss:    tensor(0.9242, device='cuda:0', grad_fn=<NllLossBackward>)
[1506] Train Loss:    tensor(0.9035, device='cuda:0', grad_fn=<NllLossBackward>)
[1507] Train Loss:    tensor(0.8077, device='cuda:0', grad_fn=<NllLossBackward>)
[1508] Train Loss:    tensor(0.8646, device='cuda:0', grad_fn=<NllLossBackward>)
[1509] Train Loss:    tensor(0.8751, device='cuda:0', grad_fn=<NllLossBackward>)
[1510] Train Loss:    tensor(0.9407, device='cuda:0', grad_fn=<NllLossBackward>)
[1511] Train Loss:    tensor(0.9119, device='cuda:0', grad_fn=<NllLossBackward>)
[1512] Train Loss:    tensor(0.8879, device='cuda:0', grad_fn=<NllLossBackward>)
[1513] Train Loss:    tensor(0.9139, device='cuda:0', grad_fn=<NllLossBackward>)
[1514] Train Loss:    tensor(0.9005, device='cuda:0', grad_fn=<NllLossBackward>)
[1515] Train Loss:    tensor(0.9218, device='cuda:0', grad_fn=<NllLossBackward>)
[1516] Train Loss:    tensor(0.9173, device='cuda:0', grad_fn=<NllLossBackward>)
[1517] Train Loss:    tensor(0.8840, device='cuda:0', grad_fn=<NllLossBackward>)
[1518] Train Loss:    tensor(0.8932, device='cuda:0', grad_fn=<NllLossBackward>)
[1519] Train Loss:    tensor(0.9011, device='cuda:0', grad_fn=<NllLossBackward>)
[1520] Train Loss:    tensor(0.8776, device='cuda:0', grad_fn=<NllLossBackward>)
[1521] Train Loss:    tensor(0.9069, device='cuda:0', grad_fn=<NllLossBackward>)
[1522] Train Loss:    tensor(0.9170, device='cuda:0', grad_fn=<NllLossBackward>)
[1523] Train Loss:    tensor(0.9085, device='cuda:0', grad_fn=<NllLossBackward>)
[1524] Train Loss:    tensor(0.8911, device='cuda:0', grad_fn=<NllLossBackward>)
[1525] Train Loss:    tensor(0.8856, device='cuda:0', grad_fn=<NllLossBackward>)
[1526] Train Loss:    tensor(0.9314, device='cuda:0', grad_fn=<NllLossBackward>)
[1527] Train Loss:    tensor(0.9044, device='cuda:0', grad_fn=<NllLossBackward>)
[1528] Train Loss:    tensor(0.8597, device='cuda:0', grad_fn=<NllLossBackward>)
[1529] Train Loss:    tensor(0.8966, device='cuda:0', grad_fn=<NllLossBackward>)
[1530] Train Loss:    tensor(0.9022, device='cuda:0', grad_fn=<NllLossBackward>)
[1531] Train Loss:    tensor(0.9380, device='cuda:0', grad_fn=<NllLossBackward>)
[1532] Train Loss:    tensor(0.8789, device='cuda:0', grad_fn=<NllLossBackward>)
[1533] Train Loss:    tensor(0.9108, device='cuda:0', grad_fn=<NllLossBackward>)
[1534] Train Loss:    tensor(0.8948, device='cuda:0', grad_fn=<NllLossBackward>)
[1535] Train Loss:    tensor(0.9168, device='cuda:0', grad_fn=<NllLossBackward>)
[1536] Train Loss:    tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[1537] Train Loss:    tensor(0.9215, device='cuda:0', grad_fn=<NllLossBackward>)
[1538] Train Loss:    tensor(0.9110, device='cuda:0', grad_fn=<NllLossBackward>)
[1539] Train Loss:    tensor(0.8860, device='cuda:0', grad_fn=<NllLossBackward>)
[1540] Train Loss:    tensor(0.8937, device='cuda:0', grad_fn=<NllLossBackward>)
[1541] Train Loss:    tensor(0.8755, device='cuda:0', grad_fn=<NllLossBackward>)
[1542] Train Loss:    tensor(0.8814, device='cuda:0', grad_fn=<NllLossBackward>)
[1543] Train Loss:    tensor(0.8922, device='cuda:0', grad_fn=<NllLossBackward>)
[1544] Train Loss:    tensor(0.8853, device='cuda:0', grad_fn=<NllLossBackward>)
[1545] Train Loss:    tensor(0.9342, device='cuda:0', grad_fn=<NllLossBackward>)
[1546] Train Loss:    tensor(0.8580, device='cuda:0', grad_fn=<NllLossBackward>)
[1547] Train Loss:    tensor(0.9193, device='cuda:0', grad_fn=<NllLossBackward>)
[1548] Train Loss:    tensor(0.9063, device='cuda:0', grad_fn=<NllLossBackward>)
[1549] Train Loss:    tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[1550] Train Loss:    tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1551] Train Loss:    tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[1552] Train Loss:    tensor(0.9020, device='cuda:0', grad_fn=<NllLossBackward>)
[1553] Train Loss:    tensor(0.9235, device='cuda:0', grad_fn=<NllLossBackward>)
[1554] Train Loss:    tensor(0.8999, device='cuda:0', grad_fn=<NllLossBackward>)
[1555] Train Loss:    tensor(0.9424, device='cuda:0', grad_fn=<NllLossBackward>)
[1556] Train Loss:    tensor(0.9453, device='cuda:0', grad_fn=<NllLossBackward>)
[1557] Train Loss:    tensor(0.9082, device='cuda:0', grad_fn=<NllLossBackward>)
[1558] Train Loss:    tensor(0.9677, device='cuda:0', grad_fn=<NllLossBackward>)
[1559] Train Loss:    tensor(0.8912, device='cuda:0', grad_fn=<NllLossBackward>)
[1560] Train Loss:    tensor(0.9190, device='cuda:0', grad_fn=<NllLossBackward>)
[1561] Train Loss:    tensor(0.9090, device='cuda:0', grad_fn=<NllLossBackward>)
[1562] Train Loss:    tensor(0.9113, device='cuda:0', grad_fn=<NllLossBackward>)
[1563] Train Loss:    tensor(0.9065, device='cuda:0', grad_fn=<NllLossBackward>)
[1564] Train Loss:    tensor(0.9193, device='cuda:0', grad_fn=<NllLossBackward>)
[1565] Train Loss:    tensor(0.8855, device='cuda:0', grad_fn=<NllLossBackward>)
[1566] Train Loss:    tensor(0.8962, device='cuda:0', grad_fn=<NllLossBackward>)
[1567] Train Loss:    tensor(0.9260, device='cuda:0', grad_fn=<NllLossBackward>)
[1568] Train Loss:    tensor(0.8805, device='cuda:0', grad_fn=<NllLossBackward>)
[1569] Train Loss:    tensor(0.8753, device='cuda:0', grad_fn=<NllLossBackward>)
[1570] Train Loss:    tensor(0.8935, device='cuda:0', grad_fn=<NllLossBackward>)
[1571] Train Loss:    tensor(0.9202, device='cuda:0', grad_fn=<NllLossBackward>)
[1572] Train Loss:    tensor(0.9282, device='cuda:0', grad_fn=<NllLossBackward>)
[1573] Train Loss:    tensor(0.8968, device='cuda:0', grad_fn=<NllLossBackward>)
[1574] Train Loss:    tensor(0.9409, device='cuda:0', grad_fn=<NllLossBackward>)
[1575] Train Loss:    tensor(0.8825, device='cuda:0', grad_fn=<NllLossBackward>)
[1576] Train Loss:    tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[1577] Train Loss:    tensor(0.9001, device='cuda:0', grad_fn=<NllLossBackward>)
[1578] Train Loss:    tensor(0.8710, device='cuda:0', grad_fn=<NllLossBackward>)
[1579] Train Loss:    tensor(0.9101, device='cuda:0', grad_fn=<NllLossBackward>)
[1580] Train Loss:    tensor(0.9508, device='cuda:0', grad_fn=<NllLossBackward>)
[1581] Train Loss:    tensor(0.9112, device='cuda:0', grad_fn=<NllLossBackward>)
[1582] Train Loss:    tensor(0.8801, device='cuda:0', grad_fn=<NllLossBackward>)
[1583] Train Loss:    tensor(0.9035, device='cuda:0', grad_fn=<NllLossBackward>)
[1584] Train Loss:    tensor(0.9054, device='cuda:0', grad_fn=<NllLossBackward>)
[1585] Train Loss:    tensor(0.9105, device='cuda:0', grad_fn=<NllLossBackward>)
[1586] Train Loss:    tensor(0.8906, device='cuda:0', grad_fn=<NllLossBackward>)
[1587] Train Loss:    tensor(0.9170, device='cuda:0', grad_fn=<NllLossBackward>)
[1588] Train Loss:    tensor(0.9158, device='cuda:0', grad_fn=<NllLossBackward>)
[1589] Train Loss:    tensor(0.8831, device='cuda:0', grad_fn=<NllLossBackward>)
[1590] Train Loss:    tensor(0.8901, device='cuda:0', grad_fn=<NllLossBackward>)
[1591] Train Loss:    tensor(0.9170, device='cuda:0', grad_fn=<NllLossBackward>)
[1592] Train Loss:    tensor(0.8953, device='cuda:0', grad_fn=<NllLossBackward>)
[1593] Train Loss:    tensor(0.9121, device='cuda:0', grad_fn=<NllLossBackward>)
[1594] Train Loss:    tensor(0.8792, device='cuda:0', grad_fn=<NllLossBackward>)
[1595] Train Loss:    tensor(0.8984, device='cuda:0', grad_fn=<NllLossBackward>)
[1596] Train Loss:    tensor(0.9142, device='cuda:0', grad_fn=<NllLossBackward>)
[1597] Train Loss:    tensor(0.9023, device='cuda:0', grad_fn=<NllLossBackward>)
[1598] Train Loss:    tensor(0.9436, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1599] Train Loss:   tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[1600] Train Loss:   tensor(0.9092, device='cuda:0', grad_fn=<NllLossBackward>)
[1601] Train Loss:   tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[1602] Train Loss:   tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[1603] Train Loss:   tensor(0.8987, device='cuda:0', grad_fn=<NllLossBackward>)
[1604] Train Loss:   tensor(0.8973, device='cuda:0', grad_fn=<NllLossBackward>)
[1605] Train Loss:   tensor(0.8829, device='cuda:0', grad_fn=<NllLossBackward>)
[1606] Train Loss:   tensor(0.9010, device='cuda:0', grad_fn=<NllLossBackward>)
[1607] Train Loss:   tensor(0.9195, device='cuda:0', grad_fn=<NllLossBackward>)
[1608] Train Loss:   tensor(0.8907, device='cuda:0', grad_fn=<NllLossBackward>)
[1609] Train Loss:   tensor(0.9485, device='cuda:0', grad_fn=<NllLossBackward>)
[1610] Train Loss:   tensor(0.9355, device='cuda:0', grad_fn=<NllLossBackward>)
[1611] Train Loss:   tensor(0.8871, device='cuda:0', grad_fn=<NllLossBackward>)
[1612] Train Loss:   tensor(0.8985, device='cuda:0', grad_fn=<NllLossBackward>)
[1613] Train Loss:   tensor(0.8928, device='cuda:0', grad_fn=<NllLossBackward>)
[1614] Train Loss:   tensor(0.8568, device='cuda:0', grad_fn=<NllLossBackward>)
[1615] Train Loss:   tensor(0.8626, device='cuda:0', grad_fn=<NllLossBackward>)
[1616] Train Loss:   tensor(0.8710, device='cuda:0', grad_fn=<NllLossBackward>)
[1617] Train Loss:   tensor(0.9069, device='cuda:0', grad_fn=<NllLossBackward>)
[1618] Train Loss:   tensor(0.8929, device='cuda:0', grad_fn=<NllLossBackward>)
[1619] Train Loss:   tensor(0.9143, device='cuda:0', grad_fn=<NllLossBackward>)
[1620] Train Loss:   tensor(0.8832, device='cuda:0', grad_fn=<NllLossBackward>)
[1621] Train Loss:   tensor(0.8921, device='cuda:0', grad_fn=<NllLossBackward>)
[1622] Train Loss:   tensor(0.9221, device='cuda:0', grad_fn=<NllLossBackward>)
[1623] Train Loss:   tensor(0.9170, device='cuda:0', grad_fn=<NllLossBackward>)
[1624] Train Loss:   tensor(0.9199, device='cuda:0', grad_fn=<NllLossBackward>)
[1625] Train Loss:   tensor(0.8825, device='cuda:0', grad_fn=<NllLossBackward>)
[1626] Train Loss:   tensor(0.8918, device='cuda:0', grad_fn=<NllLossBackward>)
[1627] Train Loss:   tensor(0.9160, device='cuda:0', grad_fn=<NllLossBackward>)
[1628] Train Loss:   tensor(0.8890, device='cuda:0', grad_fn=<NllLossBackward>)
[1629] Train Loss:   tensor(0.8854, device='cuda:0', grad_fn=<NllLossBackward>)
[1630] Train Loss:   tensor(0.9022, device='cuda:0', grad_fn=<NllLossBackward>)
[1631] Train Loss:   tensor(0.9219, device='cuda:0', grad_fn=<NllLossBackward>)
[1632] Train Loss:   tensor(0.9106, device='cuda:0', grad_fn=<NllLossBackward>)
[1633] Train Loss:   tensor(0.9274, device='cuda:0', grad_fn=<NllLossBackward>)
[1634] Train Loss:   tensor(0.8938, device='cuda:0', grad_fn=<NllLossBackward>)
[1635] Train Loss:   tensor(0.9072, device='cuda:0', grad_fn=<NllLossBackward>)
[1636] Train Loss:   tensor(0.9260, device='cuda:0', grad_fn=<NllLossBackward>)
[1637] Train Loss:   tensor(0.9128, device='cuda:0', grad_fn=<NllLossBackward>)
[1638] Train Loss:   tensor(0.8882, device='cuda:0', grad_fn=<NllLossBackward>)
[1639] Train Loss:   tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[1640] Train Loss:   tensor(0.9003, device='cuda:0', grad_fn=<NllLossBackward>)
[1641] Train Loss:   tensor(0.9107, device='cuda:0', grad_fn=<NllLossBackward>)
[1642] Train Loss:   tensor(0.9228, device='cuda:0', grad_fn=<NllLossBackward>)
[1643] Train Loss:   tensor(0.9165, device='cuda:0', grad_fn=<NllLossBackward>)
[1644] Train Loss:   tensor(0.9374, device='cuda:0', grad_fn=<NllLossBackward>)
[1645] Train Loss:   tensor(0.9052, device='cuda:0', grad_fn=<NllLossBackward>)
[1646] Train Loss:   tensor(0.9090, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1647] Train Loss:   tensor(0.8892, device='cuda:0', grad_fn=<NllLossBackward>)
[1648] Train Loss:   tensor(0.9114, device='cuda:0', grad_fn=<NllLossBackward>)
[1649] Train Loss:   tensor(0.9133, device='cuda:0', grad_fn=<NllLossBackward>)
[1650] Train Loss:   tensor(0.9217, device='cuda:0', grad_fn=<NllLossBackward>)
[1651] Train Loss:   tensor(0.9097, device='cuda:0', grad_fn=<NllLossBackward>)
[1652] Train Loss:   tensor(0.9148, device='cuda:0', grad_fn=<NllLossBackward>)
[1653] Train Loss:   tensor(0.9242, device='cuda:0', grad_fn=<NllLossBackward>)
[1654] Train Loss:   tensor(0.8997, device='cuda:0', grad_fn=<NllLossBackward>)
[1655] Train Loss:   tensor(0.9012, device='cuda:0', grad_fn=<NllLossBackward>)
[1656] Train Loss:   tensor(0.9276, device='cuda:0', grad_fn=<NllLossBackward>)
[1657] Train Loss:   tensor(0.9179, device='cuda:0', grad_fn=<NllLossBackward>)
[1658] Train Loss:   tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[1659] Train Loss:   tensor(0.9210, device='cuda:0', grad_fn=<NllLossBackward>)
[1660] Train Loss:   tensor(0.9161, device='cuda:0', grad_fn=<NllLossBackward>)
[1661] Train Loss:   tensor(0.8828, device='cuda:0', grad_fn=<NllLossBackward>)
[1662] Train Loss:   tensor(0.8926, device='cuda:0', grad_fn=<NllLossBackward>)
[1663] Train Loss:   tensor(0.9096, device='cuda:0', grad_fn=<NllLossBackward>)
[1664] Train Loss:   tensor(0.9022, device='cuda:0', grad_fn=<NllLossBackward>)
[1665] Train Loss:   tensor(0.9041, device='cuda:0', grad_fn=<NllLossBackward>)
[1666] Train Loss:   tensor(0.8705, device='cuda:0', grad_fn=<NllLossBackward>)
[1667] Train Loss:   tensor(0.9091, device='cuda:0', grad_fn=<NllLossBackward>)
[1668] Train Loss:   tensor(0.9009, device='cuda:0', grad_fn=<NllLossBackward>)
[1669] Train Loss:   tensor(0.8850, device='cuda:0', grad_fn=<NllLossBackward>)
[1670] Train Loss:   tensor(0.9180, device='cuda:0', grad_fn=<NllLossBackward>)
[1671] Train Loss:   tensor(0.9341, device='cuda:0', grad_fn=<NllLossBackward>)
[1672] Train Loss:   tensor(0.9172, device='cuda:0', grad_fn=<NllLossBackward>)
[1673] Train Loss:   tensor(0.9067, device='cuda:0', grad_fn=<NllLossBackward>)
[1674] Train Loss:   tensor(0.8866, device='cuda:0', grad_fn=<NllLossBackward>)
[1675] Train Loss:   tensor(0.8977, device='cuda:0', grad_fn=<NllLossBackward>)
[1676] Train Loss:   tensor(0.9001, device='cuda:0', grad_fn=<NllLossBackward>)
[1677] Train Loss:   tensor(0.8820, device='cuda:0', grad_fn=<NllLossBackward>)
[1678] Train Loss:   tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[1679] Train Loss:   tensor(0.9443, device='cuda:0', grad_fn=<NllLossBackward>)
[1680] Train Loss:   tensor(0.9178, device='cuda:0', grad_fn=<NllLossBackward>)
[1681] Train Loss:   tensor(0.9045, device='cuda:0', grad_fn=<NllLossBackward>)
[1682] Train Loss:   tensor(0.8973, device='cuda:0', grad_fn=<NllLossBackward>)
[1683] Train Loss:   tensor(0.9160, device='cuda:0', grad_fn=<NllLossBackward>)
[1684] Train Loss:   tensor(0.9256, device='cuda:0', grad_fn=<NllLossBackward>)
[1685] Train Loss:   tensor(0.9085, device='cuda:0', grad_fn=<NllLossBackward>)
[1686] Train Loss:   tensor(0.9170, device='cuda:0', grad_fn=<NllLossBackward>)
[1687] Train Loss:   tensor(0.8915, device='cuda:0', grad_fn=<NllLossBackward>)
[1688] Train Loss:   tensor(0.9229, device='cuda:0', grad_fn=<NllLossBackward>)
[1689] Train Loss:   tensor(0.9157, device='cuda:0', grad_fn=<NllLossBackward>)
[1690] Train Loss:   tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[1691] Train Loss:   tensor(0.9157, device='cuda:0', grad_fn=<NllLossBackward>)
[1692] Train Loss:   tensor(0.9056, device='cuda:0', grad_fn=<NllLossBackward>)
[1693] Train Loss:   tensor(0.9110, device='cuda:0', grad_fn=<NllLossBackward>)
[1694] Train Loss:   tensor(0.9038, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1695] Train Loss:    tensor(0.9585, device='cuda:0', grad_fn=<NllLossBackward>)
[1696] Train Loss:    tensor(0.8929, device='cuda:0', grad_fn=<NllLossBackward>)
[1697] Train Loss:    tensor(0.9268, device='cuda:0', grad_fn=<NllLossBackward>)
[1698] Train Loss:    tensor(0.9065, device='cuda:0', grad_fn=<NllLossBackward>)
[1699] Train Loss:    tensor(0.8787, device='cuda:0', grad_fn=<NllLossBackward>)
[1700] Train Loss:    tensor(0.9033, device='cuda:0', grad_fn=<NllLossBackward>)
[1701] Train Loss:    tensor(0.8891, device='cuda:0', grad_fn=<NllLossBackward>)
[1702] Train Loss:    tensor(0.9247, device='cuda:0', grad_fn=<NllLossBackward>)
[1703] Train Loss:    tensor(0.8910, device='cuda:0', grad_fn=<NllLossBackward>)
[1704] Train Loss:    tensor(0.8892, device='cuda:0', grad_fn=<NllLossBackward>)
[1705] Train Loss:    tensor(0.9292, device='cuda:0', grad_fn=<NllLossBackward>)
[1706] Train Loss:    tensor(0.9080, device='cuda:0', grad_fn=<NllLossBackward>)
[1707] Train Loss:    tensor(0.9149, device='cuda:0', grad_fn=<NllLossBackward>)
[1708] Train Loss:    tensor(0.9116, device='cuda:0', grad_fn=<NllLossBackward>)
[1709] Train Loss:    tensor(0.9347, device='cuda:0', grad_fn=<NllLossBackward>)
[1710] Train Loss:    tensor(0.8974, device='cuda:0', grad_fn=<NllLossBackward>)
[1711] Train Loss:    tensor(0.9201, device='cuda:0', grad_fn=<NllLossBackward>)
[1712] Train Loss:    tensor(0.8860, device='cuda:0', grad_fn=<NllLossBackward>)
[1713] Train Loss:    tensor(0.9530, device='cuda:0', grad_fn=<NllLossBackward>)
[1714] Train Loss:    tensor(0.8689, device='cuda:0', grad_fn=<NllLossBackward>)
[1715] Train Loss:    tensor(0.9322, device='cuda:0', grad_fn=<NllLossBackward>)
[1716] Train Loss:    tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
[1717] Train Loss:    tensor(0.9204, device='cuda:0', grad_fn=<NllLossBackward>)
[1718] Train Loss:    tensor(0.9301, device='cuda:0', grad_fn=<NllLossBackward>)
[1719] Train Loss:    tensor(0.9000, device='cuda:0', grad_fn=<NllLossBackward>)
[1720] Train Loss:    tensor(0.8990, device='cuda:0', grad_fn=<NllLossBackward>)
[1721] Train Loss:    tensor(0.8844, device='cuda:0', grad_fn=<NllLossBackward>)
[1722] Train Loss:    tensor(0.8907, device='cuda:0', grad_fn=<NllLossBackward>)
[1723] Train Loss:    tensor(0.8817, device='cuda:0', grad_fn=<NllLossBackward>)
[1724] Train Loss:    tensor(0.9281, device='cuda:0', grad_fn=<NllLossBackward>)
[1725] Train Loss:    tensor(0.8991, device='cuda:0', grad_fn=<NllLossBackward>)
[1726] Train Loss:    tensor(0.9247, device='cuda:0', grad_fn=<NllLossBackward>)
[1727] Train Loss:    tensor(0.9331, device='cuda:0', grad_fn=<NllLossBackward>)
[1728] Train Loss:    tensor(0.8785, device='cuda:0', grad_fn=<NllLossBackward>)
[1729] Train Loss:    tensor(0.9138, device='cuda:0', grad_fn=<NllLossBackward>)
[1730] Train Loss:    tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[1731] Train Loss:    tensor(0.9078, device='cuda:0', grad_fn=<NllLossBackward>)
[1732] Train Loss:    tensor(0.8904, device='cuda:0', grad_fn=<NllLossBackward>)
[1733] Train Loss:    tensor(0.8946, device='cuda:0', grad_fn=<NllLossBackward>)
[1734] Train Loss:    tensor(0.9256, device='cuda:0', grad_fn=<NllLossBackward>)
[1735] Train Loss:    tensor(0.9273, device='cuda:0', grad_fn=<NllLossBackward>)
[1736] Train Loss:    tensor(0.9158, device='cuda:0', grad_fn=<NllLossBackward>)
[1737] Train Loss:    tensor(0.8682, device='cuda:0', grad_fn=<NllLossBackward>)
[1738] Train Loss:    tensor(0.9177, device='cuda:0', grad_fn=<NllLossBackward>)
[1739] Train Loss:    tensor(0.8952, device='cuda:0', grad_fn=<NllLossBackward>)
[1740] Train Loss:    tensor(0.9142, device='cuda:0', grad_fn=<NllLossBackward>)
[1741] Train Loss:    tensor(0.9091, device='cuda:0', grad_fn=<NllLossBackward>)
[1742] Train Loss:    tensor(0.9567, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1743] Train Loss:  tensor(0.9078, device='cuda:0', grad_fn=<NllLossBackward>)
[1744] Train Loss:  tensor(0.9106, device='cuda:0', grad_fn=<NllLossBackward>)
[1745] Train Loss:  tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[1746] Train Loss:  tensor(0.9044, device='cuda:0', grad_fn=<NllLossBackward>)
[1747] Train Loss:  tensor(0.9089, device='cuda:0', grad_fn=<NllLossBackward>)
[1748] Train Loss:  tensor(0.9320, device='cuda:0', grad_fn=<NllLossBackward>)
[1749] Train Loss:  tensor(0.9161, device='cuda:0', grad_fn=<NllLossBackward>)
[1750] Train Loss:  tensor(0.8844, device='cuda:0', grad_fn=<NllLossBackward>)
[1751] Train Loss:  tensor(0.9051, device='cuda:0', grad_fn=<NllLossBackward>)
[1752] Train Loss:  tensor(0.8967, device='cuda:0', grad_fn=<NllLossBackward>)
[1753] Train Loss:  tensor(0.9308, device='cuda:0', grad_fn=<NllLossBackward>)
[1754] Train Loss:  tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[1755] Train Loss:  tensor(0.8969, device='cuda:0', grad_fn=<NllLossBackward>)
[1756] Train Loss:  tensor(0.9223, device='cuda:0', grad_fn=<NllLossBackward>)
[1757] Train Loss:  tensor(0.8821, device='cuda:0', grad_fn=<NllLossBackward>)
[1758] Train Loss:  tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[1759] Train Loss:  tensor(0.9009, device='cuda:0', grad_fn=<NllLossBackward>)
[1760] Train Loss:  tensor(0.9121, device='cuda:0', grad_fn=<NllLossBackward>)
[1761] Train Loss:  tensor(0.8920, device='cuda:0', grad_fn=<NllLossBackward>)
[1762] Train Loss:  tensor(0.9077, device='cuda:0', grad_fn=<NllLossBackward>)
[1763] Train Loss:  tensor(0.9126, device='cuda:0', grad_fn=<NllLossBackward>)
[1764] Train Loss:  tensor(0.9241, device='cuda:0', grad_fn=<NllLossBackward>)
[1765] Train Loss:  tensor(0.9611, device='cuda:0', grad_fn=<NllLossBackward>)
[1766] Train Loss:  tensor(0.8813, device='cuda:0', grad_fn=<NllLossBackward>)
[1767] Train Loss:  tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[1768] Train Loss:  tensor(0.9204, device='cuda:0', grad_fn=<NllLossBackward>)
[1769] Train Loss:  tensor(0.9165, device='cuda:0', grad_fn=<NllLossBackward>)
[1770] Train Loss:  tensor(0.9054, device='cuda:0', grad_fn=<NllLossBackward>)
[1771] Train Loss:  tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[1772] Train Loss:  tensor(0.8870, device='cuda:0', grad_fn=<NllLossBackward>)
[1773] Train Loss:  tensor(0.8942, device='cuda:0', grad_fn=<NllLossBackward>)
[1774] Train Loss:  tensor(0.8808, device='cuda:0', grad_fn=<NllLossBackward>)
[1775] Train Loss:  tensor(0.8898, device='cuda:0', grad_fn=<NllLossBackward>)
[1776] Train Loss:  tensor(0.8918, device='cuda:0', grad_fn=<NllLossBackward>)
[1777] Train Loss:  tensor(0.9309, device='cuda:0', grad_fn=<NllLossBackward>)
[1778] Train Loss:  tensor(0.9203, device='cuda:0', grad_fn=<NllLossBackward>)
[1779] Train Loss:  tensor(0.9332, device='cuda:0', grad_fn=<NllLossBackward>)
[1780] Train Loss:  tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
[1781] Train Loss:  tensor(0.8903, device='cuda:0', grad_fn=<NllLossBackward>)
[1782] Train Loss:  tensor(0.8618, device='cuda:0', grad_fn=<NllLossBackward>)
[1783] Train Loss:  tensor(0.9148, device='cuda:0', grad_fn=<NllLossBackward>)
[1784] Train Loss:  tensor(0.9040, device='cuda:0', grad_fn=<NllLossBackward>)
[1785] Train Loss:  tensor(0.9076, device='cuda:0', grad_fn=<NllLossBackward>)
[1786] Train Loss:  tensor(0.9196, device='cuda:0', grad_fn=<NllLossBackward>)
[1787] Train Loss:  tensor(0.9075, device='cuda:0', grad_fn=<NllLossBackward>)
[1788] Train Loss:  tensor(0.8755, device='cuda:0', grad_fn=<NllLossBackward>)
[1789] Train Loss:  tensor(0.9059, device='cuda:0', grad_fn=<NllLossBackward>)
[1790] Train Loss:  tensor(0.9004, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1791] Train Loss:    tensor(0.9457, device='cuda:0', grad_fn=<NllLossBackward>)
[1792] Train Loss:    tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[1793] Train Loss:    tensor(0.9199, device='cuda:0', grad_fn=<NllLossBackward>)
[1794] Train Loss:    tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[1795] Train Loss:    tensor(0.9137, device='cuda:0', grad_fn=<NllLossBackward>)
[1796] Train Loss:    tensor(0.9081, device='cuda:0', grad_fn=<NllLossBackward>)
[1797] Train Loss:    tensor(0.8464, device='cuda:0', grad_fn=<NllLossBackward>)
[1798] Train Loss:    tensor(0.8931, device='cuda:0', grad_fn=<NllLossBackward>)
[1799] Train Loss:    tensor(0.9274, device='cuda:0', grad_fn=<NllLossBackward>)
[1800] Train Loss:    tensor(0.8913, device='cuda:0', grad_fn=<NllLossBackward>)
[1801] Train Loss:    tensor(0.8987, device='cuda:0', grad_fn=<NllLossBackward>)
[1802] Train Loss:    tensor(0.9000, device='cuda:0', grad_fn=<NllLossBackward>)
[1803] Train Loss:    tensor(0.9016, device='cuda:0', grad_fn=<NllLossBackward>)
[1804] Train Loss:    tensor(0.9034, device='cuda:0', grad_fn=<NllLossBackward>)
[1805] Train Loss:    tensor(0.8909, device='cuda:0', grad_fn=<NllLossBackward>)
[1806] Train Loss:    tensor(0.9194, device='cuda:0', grad_fn=<NllLossBackward>)
[1807] Train Loss:    tensor(0.8927, device='cuda:0', grad_fn=<NllLossBackward>)
[1808] Train Loss:    tensor(0.8702, device='cuda:0', grad_fn=<NllLossBackward>)
[1809] Train Loss:    tensor(0.8727, device='cuda:0', grad_fn=<NllLossBackward>)
[1810] Train Loss:    tensor(0.9052, device='cuda:0', grad_fn=<NllLossBackward>)
[1811] Train Loss:    tensor(0.9245, device='cuda:0', grad_fn=<NllLossBackward>)
[1812] Train Loss:    tensor(0.9158, device='cuda:0', grad_fn=<NllLossBackward>)
[1813] Train Loss:    tensor(0.9207, device='cuda:0', grad_fn=<NllLossBackward>)
[1814] Train Loss:    tensor(0.9008, device='cuda:0', grad_fn=<NllLossBackward>)
[1815] Train Loss:    tensor(0.8726, device='cuda:0', grad_fn=<NllLossBackward>)
[1816] Train Loss:    tensor(0.9107, device='cuda:0', grad_fn=<NllLossBackward>)
[1817] Train Loss:    tensor(0.9100, device='cuda:0', grad_fn=<NllLossBackward>)
[1818] Train Loss:    tensor(0.9297, device='cuda:0', grad_fn=<NllLossBackward>)
[1819] Train Loss:    tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[1820] Train Loss:    tensor(0.9174, device='cuda:0', grad_fn=<NllLossBackward>)
[1821] Train Loss:    tensor(0.8821, device='cuda:0', grad_fn=<NllLossBackward>)
[1822] Train Loss:    tensor(0.9121, device='cuda:0', grad_fn=<NllLossBackward>)
[1823] Train Loss:    tensor(0.8978, device='cuda:0', grad_fn=<NllLossBackward>)
[1824] Train Loss:    tensor(0.8410, device='cuda:0', grad_fn=<NllLossBackward>)
[1825] Train Loss:    tensor(0.9236, device='cuda:0', grad_fn=<NllLossBackward>)
[1826] Train Loss:    tensor(0.9101, device='cuda:0', grad_fn=<NllLossBackward>)
[1827] Train Loss:    tensor(0.8903, device='cuda:0', grad_fn=<NllLossBackward>)
[1828] Train Loss:    tensor(0.9182, device='cuda:0', grad_fn=<NllLossBackward>)
[1829] Train Loss:    tensor(0.8808, device='cuda:0', grad_fn=<NllLossBackward>)
[1830] Train Loss:    tensor(0.8854, device='cuda:0', grad_fn=<NllLossBackward>)
[1831] Train Loss:    tensor(0.8963, device='cuda:0', grad_fn=<NllLossBackward>)
[1832] Train Loss:    tensor(0.8883, device='cuda:0', grad_fn=<NllLossBackward>)
[1833] Train Loss:    tensor(0.9110, device='cuda:0', grad_fn=<NllLossBackward>)
[1834] Train Loss:    tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[1835] Train Loss:    tensor(0.8916, device='cuda:0', grad_fn=<NllLossBackward>)
[1836] Train Loss:    tensor(0.8665, device='cuda:0', grad_fn=<NllLossBackward>)
[1837] Train Loss:    tensor(0.9080, device='cuda:0', grad_fn=<NllLossBackward>)
[1838] Train Loss:    tensor(0.9110, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1839] Train Loss:    tensor(0.9444, device='cuda:0', grad_fn=<NllLossBackward>)
[1840] Train Loss:    tensor(0.8848, device='cuda:0', grad_fn=<NllLossBackward>)
[1841] Train Loss:    tensor(0.9106, device='cuda:0', grad_fn=<NllLossBackward>)
[1842] Train Loss:    tensor(0.9070, device='cuda:0', grad_fn=<NllLossBackward>)
[1843] Train Loss:    tensor(0.9030, device='cuda:0', grad_fn=<NllLossBackward>)
[1844] Train Loss:    tensor(0.9179, device='cuda:0', grad_fn=<NllLossBackward>)
[1845] Train Loss:    tensor(0.8935, device='cuda:0', grad_fn=<NllLossBackward>)
[1846] Train Loss:    tensor(0.8852, device='cuda:0', grad_fn=<NllLossBackward>)
[1847] Train Loss:    tensor(0.9216, device='cuda:0', grad_fn=<NllLossBackward>)
[1848] Train Loss:    tensor(0.9403, device='cuda:0', grad_fn=<NllLossBackward>)
[1849] Train Loss:    tensor(0.8690, device='cuda:0', grad_fn=<NllLossBackward>)
[1850] Train Loss:    tensor(0.8905, device='cuda:0', grad_fn=<NllLossBackward>)
[1851] Train Loss:    tensor(0.9165, device='cuda:0', grad_fn=<NllLossBackward>)
[1852] Train Loss:    tensor(0.8886, device='cuda:0', grad_fn=<NllLossBackward>)
[1853] Train Loss:    tensor(0.8701, device='cuda:0', grad_fn=<NllLossBackward>)
[1854] Train Loss:    tensor(0.9110, device='cuda:0', grad_fn=<NllLossBackward>)
[1855] Train Loss:    tensor(0.9105, device='cuda:0', grad_fn=<NllLossBackward>)
[1856] Train Loss:    tensor(0.8861, device='cuda:0', grad_fn=<NllLossBackward>)
[1857] Train Loss:    tensor(0.8953, device='cuda:0', grad_fn=<NllLossBackward>)
[1858] Train Loss:    tensor(0.9263, device='cuda:0', grad_fn=<NllLossBackward>)
[1859] Train Loss:    tensor(0.9385, device='cuda:0', grad_fn=<NllLossBackward>)
[1860] Train Loss:    tensor(0.8541, device='cuda:0', grad_fn=<NllLossBackward>)
[1861] Train Loss:    tensor(0.8816, device='cuda:0', grad_fn=<NllLossBackward>)
[1862] Train Loss:    tensor(0.8926, device='cuda:0', grad_fn=<NllLossBackward>)
[1863] Train Loss:    tensor(0.8832, device='cuda:0', grad_fn=<NllLossBackward>)
[1864] Train Loss:    tensor(0.9044, device='cuda:0', grad_fn=<NllLossBackward>)
[1865] Train Loss:    tensor(0.8877, device='cuda:0', grad_fn=<NllLossBackward>)
[1866] Train Loss:    tensor(0.8677, device='cuda:0', grad_fn=<NllLossBackward>)
[1867] Train Loss:    tensor(0.8952, device='cuda:0', grad_fn=<NllLossBackward>)
[1868] Train Loss:    tensor(0.9430, device='cuda:0', grad_fn=<NllLossBackward>)
[1869] Train Loss:    tensor(0.8895, device='cuda:0', grad_fn=<NllLossBackward>)
[1870] Train Loss:    tensor(0.8836, device='cuda:0', grad_fn=<NllLossBackward>)
[1871] Train Loss:    tensor(0.9144, device='cuda:0', grad_fn=<NllLossBackward>)
[1872] Train Loss:    tensor(0.8958, device='cuda:0', grad_fn=<NllLossBackward>)
[1873] Train Loss:    tensor(0.8820, device='cuda:0', grad_fn=<NllLossBackward>)
[1874] Train Loss:    tensor(0.8882, device='cuda:0', grad_fn=<NllLossBackward>)
[1875] Train Loss:    tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[1876] Train Loss:    tensor(0.8946, device='cuda:0', grad_fn=<NllLossBackward>)
[1877] Train Loss:    tensor(0.8676, device='cuda:0', grad_fn=<NllLossBackward>)
[1878] Train Loss:    tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[1879] Train Loss:    tensor(0.8758, device='cuda:0', grad_fn=<NllLossBackward>)
[1880] Train Loss:    tensor(0.8652, device='cuda:0', grad_fn=<NllLossBackward>)
[1881] Train Loss:    tensor(0.8980, device='cuda:0', grad_fn=<NllLossBackward>)
[1882] Train Loss:    tensor(0.8649, device='cuda:0', grad_fn=<NllLossBackward>)
[1883] Train Loss:    tensor(0.8669, device='cuda:0', grad_fn=<NllLossBackward>)
[1884] Train Loss:    tensor(0.8959, device='cuda:0', grad_fn=<NllLossBackward>)
[1885] Train Loss:    tensor(0.9025, device='cuda:0', grad_fn=<NllLossBackward>)
[1886] Train Loss:    tensor(0.8910, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1887] Train Loss:   tensor(0.8875, device='cuda:0', grad_fn=<NllLossBackward>)
[1888] Train Loss:   tensor(0.9098, device='cuda:0', grad_fn=<NllLossBackward>)
[1889] Train Loss:   tensor(0.9032, device='cuda:0', grad_fn=<NllLossBackward>)
[1890] Train Loss:   tensor(0.9249, device='cuda:0', grad_fn=<NllLossBackward>)
[1891] Train Loss:   tensor(0.9169, device='cuda:0', grad_fn=<NllLossBackward>)
[1892] Train Loss:   tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[1893] Train Loss:   tensor(0.8918, device='cuda:0', grad_fn=<NllLossBackward>)
[1894] Train Loss:   tensor(0.8976, device='cuda:0', grad_fn=<NllLossBackward>)
[1895] Train Loss:   tensor(0.8809, device='cuda:0', grad_fn=<NllLossBackward>)
[1896] Train Loss:   tensor(0.8942, device='cuda:0', grad_fn=<NllLossBackward>)
[1897] Train Loss:   tensor(0.8715, device='cuda:0', grad_fn=<NllLossBackward>)
[1898] Train Loss:   tensor(0.9243, device='cuda:0', grad_fn=<NllLossBackward>)
[1899] Train Loss:   tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[1900] Train Loss:   tensor(0.9284, device='cuda:0', grad_fn=<NllLossBackward>)
[1901] Train Loss:   tensor(0.8927, device='cuda:0', grad_fn=<NllLossBackward>)
[1902] Train Loss:   tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
[1903] Train Loss:   tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[1904] Train Loss:   tensor(0.8979, device='cuda:0', grad_fn=<NllLossBackward>)
[1905] Train Loss:   tensor(0.9388, device='cuda:0', grad_fn=<NllLossBackward>)
[1906] Train Loss:   tensor(0.9030, device='cuda:0', grad_fn=<NllLossBackward>)
[1907] Train Loss:   tensor(0.9171, device='cuda:0', grad_fn=<NllLossBackward>)
[1908] Train Loss:   tensor(0.8844, device='cuda:0', grad_fn=<NllLossBackward>)
[1909] Train Loss:   tensor(0.8979, device='cuda:0', grad_fn=<NllLossBackward>)
[1910] Train Loss:   tensor(0.8767, device='cuda:0', grad_fn=<NllLossBackward>)
[1911] Train Loss:   tensor(0.8913, device='cuda:0', grad_fn=<NllLossBackward>)
[1912] Train Loss:   tensor(0.8873, device='cuda:0', grad_fn=<NllLossBackward>)
[1913] Train Loss:   tensor(0.9086, device='cuda:0', grad_fn=<NllLossBackward>)
[1914] Train Loss:   tensor(0.9198, device='cuda:0', grad_fn=<NllLossBackward>)
[1915] Train Loss:   tensor(0.8670, device='cuda:0', grad_fn=<NllLossBackward>)
[1916] Train Loss:   tensor(0.8789, device='cuda:0', grad_fn=<NllLossBackward>)
[1917] Train Loss:   tensor(0.8934, device='cuda:0', grad_fn=<NllLossBackward>)
[1918] Train Loss:   tensor(0.9138, device='cuda:0', grad_fn=<NllLossBackward>)
[1919] Train Loss:   tensor(0.9060, device='cuda:0', grad_fn=<NllLossBackward>)
[1920] Train Loss:   tensor(0.9055, device='cuda:0', grad_fn=<NllLossBackward>)
[1921] Train Loss:   tensor(0.8873, device='cuda:0', grad_fn=<NllLossBackward>)
[1922] Train Loss:   tensor(0.9328, device='cuda:0', grad_fn=<NllLossBackward>)
[1923] Train Loss:   tensor(0.9054, device='cuda:0', grad_fn=<NllLossBackward>)
[1924] Train Loss:   tensor(0.8850, device='cuda:0', grad_fn=<NllLossBackward>)
[1925] Train Loss:   tensor(0.8828, device='cuda:0', grad_fn=<NllLossBackward>)
[1926] Train Loss:   tensor(0.8975, device='cuda:0', grad_fn=<NllLossBackward>)
[1927] Train Loss:   tensor(0.9014, device='cuda:0', grad_fn=<NllLossBackward>)
[1928] Train Loss:   tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[1929] Train Loss:   tensor(0.8843, device='cuda:0', grad_fn=<NllLossBackward>)
[1930] Train Loss:   tensor(0.8915, device='cuda:0', grad_fn=<NllLossBackward>)
[1931] Train Loss:   tensor(0.8618, device='cuda:0', grad_fn=<NllLossBackward>)
[1932] Train Loss:   tensor(0.9249, device='cuda:0', grad_fn=<NllLossBackward>)
[1933] Train Loss:   tensor(0.8789, device='cuda:0', grad_fn=<NllLossBackward>)
[1934] Train Loss:   tensor(0.8935, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1935] Train Loss:    tensor(0.8797, device='cuda:0', grad_fn=<NllLossBackward>)
[1936] Train Loss:    tensor(0.9482, device='cuda:0', grad_fn=<NllLossBackward>)
[1937] Train Loss:    tensor(0.8921, device='cuda:0', grad_fn=<NllLossBackward>)
[1938] Train Loss:    tensor(0.8982, device='cuda:0', grad_fn=<NllLossBackward>)
[1939] Train Loss:    tensor(0.8890, device='cuda:0', grad_fn=<NllLossBackward>)
[1940] Train Loss:    tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[1941] Train Loss:    tensor(0.9175, device='cuda:0', grad_fn=<NllLossBackward>)
[1942] Train Loss:    tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[1943] Train Loss:    tensor(0.8906, device='cuda:0', grad_fn=<NllLossBackward>)
[1944] Train Loss:    tensor(0.9215, device='cuda:0', grad_fn=<NllLossBackward>)
[1945] Train Loss:    tensor(0.8987, device='cuda:0', grad_fn=<NllLossBackward>)
[1946] Train Loss:    tensor(0.9227, device='cuda:0', grad_fn=<NllLossBackward>)
[1947] Train Loss:    tensor(0.9219, device='cuda:0', grad_fn=<NllLossBackward>)
[1948] Train Loss:    tensor(0.9008, device='cuda:0', grad_fn=<NllLossBackward>)
[1949] Train Loss:    tensor(0.9096, device='cuda:0', grad_fn=<NllLossBackward>)
[1950] Train Loss:    tensor(0.8957, device='cuda:0', grad_fn=<NllLossBackward>)
[1951] Train Loss:    tensor(0.8672, device='cuda:0', grad_fn=<NllLossBackward>)
[1952] Train Loss:    tensor(0.9115, device='cuda:0', grad_fn=<NllLossBackward>)
[1953] Train Loss:    tensor(0.8871, device='cuda:0', grad_fn=<NllLossBackward>)
[1954] Train Loss:    tensor(0.9176, device='cuda:0', grad_fn=<NllLossBackward>)
[1955] Train Loss:    tensor(0.8722, device='cuda:0', grad_fn=<NllLossBackward>)
[1956] Train Loss:    tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[1957] Train Loss:    tensor(0.9388, device='cuda:0', grad_fn=<NllLossBackward>)
[1958] Train Loss:    tensor(0.9045, device='cuda:0', grad_fn=<NllLossBackward>)
[1959] Train Loss:    tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[1960] Train Loss:    tensor(0.8808, device='cuda:0', grad_fn=<NllLossBackward>)
[1961] Train Loss:    tensor(0.9040, device='cuda:0', grad_fn=<NllLossBackward>)
[1962] Train Loss:    tensor(0.8556, device='cuda:0', grad_fn=<NllLossBackward>)
[1963] Train Loss:    tensor(0.8771, device='cuda:0', grad_fn=<NllLossBackward>)
[1964] Train Loss:    tensor(0.9369, device='cuda:0', grad_fn=<NllLossBackward>)
[1965] Train Loss:    tensor(0.8879, device='cuda:0', grad_fn=<NllLossBackward>)
[1966] Train Loss:    tensor(0.9187, device='cuda:0', grad_fn=<NllLossBackward>)
[1967] Train Loss:    tensor(0.8945, device='cuda:0', grad_fn=<NllLossBackward>)
[1968] Train Loss:    tensor(0.9082, device='cuda:0', grad_fn=<NllLossBackward>)
[1969] Train Loss:    tensor(0.8878, device='cuda:0', grad_fn=<NllLossBackward>)
[1970] Train Loss:    tensor(0.9069, device='cuda:0', grad_fn=<NllLossBackward>)
[1971] Train Loss:    tensor(0.9244, device='cuda:0', grad_fn=<NllLossBackward>)
[1972] Train Loss:    tensor(0.9258, device='cuda:0', grad_fn=<NllLossBackward>)
[1973] Train Loss:    tensor(0.9094, device='cuda:0', grad_fn=<NllLossBackward>)
[1974] Train Loss:    tensor(0.8914, device='cuda:0', grad_fn=<NllLossBackward>)
[1975] Train Loss:    tensor(0.9323, device='cuda:0', grad_fn=<NllLossBackward>)
[1976] Train Loss:    tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[1977] Train Loss:    tensor(0.8839, device='cuda:0', grad_fn=<NllLossBackward>)
[1978] Train Loss:    tensor(0.9021, device='cuda:0', grad_fn=<NllLossBackward>)
[1979] Train Loss:    tensor(0.9234, device='cuda:0', grad_fn=<NllLossBackward>)
[1980] Train Loss:    tensor(0.8896, device='cuda:0', grad_fn=<NllLossBackward>)
[1981] Train Loss:    tensor(0.9155, device='cuda:0', grad_fn=<NllLossBackward>)
[1982] Train Loss:    tensor(0.8615, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1983] Train Loss:    tensor(0.9041, device='cuda:0', grad_fn=<NllLossBackward>)
[1984] Train Loss:    tensor(0.8926, device='cuda:0', grad_fn=<NllLossBackward>)
[1985] Train Loss:    tensor(0.8912, device='cuda:0', grad_fn=<NllLossBackward>)
[1986] Train Loss:    tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[1987] Train Loss:    tensor(0.8927, device='cuda:0', grad_fn=<NllLossBackward>)
[1988] Train Loss:    tensor(0.8952, device='cuda:0', grad_fn=<NllLossBackward>)
[1989] Train Loss:    tensor(0.9124, device='cuda:0', grad_fn=<NllLossBackward>)
[1990] Train Loss:    tensor(0.8724, device='cuda:0', grad_fn=<NllLossBackward>)
[1991] Train Loss:    tensor(0.9061, device='cuda:0', grad_fn=<NllLossBackward>)
[1992] Train Loss:    tensor(0.8832, device='cuda:0', grad_fn=<NllLossBackward>)
[1993] Train Loss:    tensor(0.8968, device='cuda:0', grad_fn=<NllLossBackward>)
[1994] Train Loss:    tensor(0.8785, device='cuda:0', grad_fn=<NllLossBackward>)
[1995] Train Loss:    tensor(0.8826, device='cuda:0', grad_fn=<NllLossBackward>)
[1996] Train Loss:    tensor(0.9196, device='cuda:0', grad_fn=<NllLossBackward>)
[1997] Train Loss:    tensor(0.8906, device='cuda:0', grad_fn=<NllLossBackward>)
[1998] Train Loss:    tensor(0.8891, device='cuda:0', grad_fn=<NllLossBackward>)
[1999] Train Loss:    tensor(0.9063, device='cuda:0', grad_fn=<NllLossBackward>)
[2000] Train Loss:    tensor(0.8965, device='cuda:0', grad_fn=<NllLossBackward>)
[2001] Train Loss:    tensor(0.8876, device='cuda:0', grad_fn=<NllLossBackward>)
[2002] Train Loss:    tensor(0.9103, device='cuda:0', grad_fn=<NllLossBackward>)
[2003] Train Loss:    tensor(0.9357, device='cuda:0', grad_fn=<NllLossBackward>)
[2004] Train Loss:    tensor(0.9022, device='cuda:0', grad_fn=<NllLossBackward>)
[2005] Train Loss:    tensor(0.9090, device='cuda:0', grad_fn=<NllLossBackward>)
[2006] Train Loss:    tensor(0.9025, device='cuda:0', grad_fn=<NllLossBackward>)
[2007] Train Loss:    tensor(0.8965, device='cuda:0', grad_fn=<NllLossBackward>)
[2008] Train Loss:    tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[2009] Train Loss:    tensor(0.9079, device='cuda:0', grad_fn=<NllLossBackward>)
[2010] Train Loss:    tensor(0.8945, device='cuda:0', grad_fn=<NllLossBackward>)
[2011] Train Loss:    tensor(0.9013, device='cuda:0', grad_fn=<NllLossBackward>)
[2012] Train Loss:    tensor(0.9136, device='cuda:0', grad_fn=<NllLossBackward>)
[2013] Train Loss:    tensor(0.8987, device='cuda:0', grad_fn=<NllLossBackward>)
[2014] Train Loss:    tensor(0.8799, device='cuda:0', grad_fn=<NllLossBackward>)
[2015] Train Loss:    tensor(0.8891, device='cuda:0', grad_fn=<NllLossBackward>)
[2016] Train Loss:    tensor(0.8744, device='cuda:0', grad_fn=<NllLossBackward>)
[2017] Train Loss:    tensor(0.8914, device='cuda:0', grad_fn=<NllLossBackward>)
[2018] Train Loss:    tensor(0.8873, device='cuda:0', grad_fn=<NllLossBackward>)
[2019] Train Loss:    tensor(0.8894, device='cuda:0', grad_fn=<NllLossBackward>)
[2020] Train Loss:    tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[2021] Train Loss:    tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[2022] Train Loss:    tensor(0.9132, device='cuda:0', grad_fn=<NllLossBackward>)
[2023] Train Loss:    tensor(0.8648, device='cuda:0', grad_fn=<NllLossBackward>)
[2024] Train Loss:    tensor(0.9004, device='cuda:0', grad_fn=<NllLossBackward>)
[2025] Train Loss:    tensor(0.9021, device='cuda:0', grad_fn=<NllLossBackward>)
[2026] Train Loss:    tensor(0.8734, device='cuda:0', grad_fn=<NllLossBackward>)
[2027] Train Loss:    tensor(0.8931, device='cuda:0', grad_fn=<NllLossBackward>)
[2028] Train Loss:    tensor(0.8860, device='cuda:0', grad_fn=<NllLossBackward>)
[2029] Train Loss:    tensor(0.9102, device='cuda:0', grad_fn=<NllLossBackward>)
[2030] Train Loss:    tensor(0.8970, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2031] Train Loss:   tensor(0.8929, device='cuda:0', grad_fn=<NllLossBackward>)
[2032] Train Loss:   tensor(0.9264, device='cuda:0', grad_fn=<NllLossBackward>)
[2033] Train Loss:   tensor(0.9028, device='cuda:0', grad_fn=<NllLossBackward>)
[2034] Train Loss:   tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
[2035] Train Loss:   tensor(0.9246, device='cuda:0', grad_fn=<NllLossBackward>)
[2036] Train Loss:   tensor(0.8997, device='cuda:0', grad_fn=<NllLossBackward>)
[2037] Train Loss:   tensor(0.9044, device='cuda:0', grad_fn=<NllLossBackward>)
[2038] Train Loss:   tensor(0.8828, device='cuda:0', grad_fn=<NllLossBackward>)
[2039] Train Loss:   tensor(0.8931, device='cuda:0', grad_fn=<NllLossBackward>)
[2040] Train Loss:   tensor(0.9075, device='cuda:0', grad_fn=<NllLossBackward>)
[2041] Train Loss:   tensor(0.8940, device='cuda:0', grad_fn=<NllLossBackward>)
[2042] Train Loss:   tensor(0.9176, device='cuda:0', grad_fn=<NllLossBackward>)
[2043] Train Loss:   tensor(0.8819, device='cuda:0', grad_fn=<NllLossBackward>)
[2044] Train Loss:   tensor(0.9097, device='cuda:0', grad_fn=<NllLossBackward>)
[2045] Train Loss:   tensor(0.8835, device='cuda:0', grad_fn=<NllLossBackward>)
[2046] Train Loss:   tensor(0.8557, device='cuda:0', grad_fn=<NllLossBackward>)
[2047] Train Loss:   tensor(0.8945, device='cuda:0', grad_fn=<NllLossBackward>)
[2048] Train Loss:   tensor(0.8753, device='cuda:0', grad_fn=<NllLossBackward>)
[2049] Train Loss:   tensor(0.9080, device='cuda:0', grad_fn=<NllLossBackward>)
[2050] Train Loss:   tensor(0.8701, device='cuda:0', grad_fn=<NllLossBackward>)
[2051] Train Loss:   tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[2052] Train Loss:   tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[2053] Train Loss:   tensor(0.9405, device='cuda:0', grad_fn=<NllLossBackward>)
[2054] Train Loss:   tensor(0.8982, device='cuda:0', grad_fn=<NllLossBackward>)
[2055] Train Loss:   tensor(0.9127, device='cuda:0', grad_fn=<NllLossBackward>)
[2056] Train Loss:   tensor(0.8983, device='cuda:0', grad_fn=<NllLossBackward>)
[2057] Train Loss:   tensor(0.8524, device='cuda:0', grad_fn=<NllLossBackward>)
[2058] Train Loss:   tensor(0.8972, device='cuda:0', grad_fn=<NllLossBackward>)
[2059] Train Loss:   tensor(0.8926, device='cuda:0', grad_fn=<NllLossBackward>)
[2060] Train Loss:   tensor(0.8659, device='cuda:0', grad_fn=<NllLossBackward>)
[2061] Train Loss:   tensor(0.8734, device='cuda:0', grad_fn=<NllLossBackward>)
[2062] Train Loss:   tensor(0.8734, device='cuda:0', grad_fn=<NllLossBackward>)
[2063] Train Loss:   tensor(0.9233, device='cuda:0', grad_fn=<NllLossBackward>)
[2064] Train Loss:   tensor(0.8872, device='cuda:0', grad_fn=<NllLossBackward>)
[2065] Train Loss:   tensor(0.8948, device='cuda:0', grad_fn=<NllLossBackward>)
[2066] Train Loss:   tensor(0.8821, device='cuda:0', grad_fn=<NllLossBackward>)
[2067] Train Loss:   tensor(0.8851, device='cuda:0', grad_fn=<NllLossBackward>)
[2068] Train Loss:   tensor(0.8808, device='cuda:0', grad_fn=<NllLossBackward>)
[2069] Train Loss:   tensor(0.8960, device='cuda:0', grad_fn=<NllLossBackward>)
[2070] Train Loss:   tensor(0.8771, device='cuda:0', grad_fn=<NllLossBackward>)
[2071] Train Loss:   tensor(0.9057, device='cuda:0', grad_fn=<NllLossBackward>)
[2072] Train Loss:   tensor(0.8880, device='cuda:0', grad_fn=<NllLossBackward>)
[2073] Train Loss:   tensor(0.8940, device='cuda:0', grad_fn=<NllLossBackward>)
[2074] Train Loss:   tensor(0.9012, device='cuda:0', grad_fn=<NllLossBackward>)
[2075] Train Loss:   tensor(0.8801, device='cuda:0', grad_fn=<NllLossBackward>)
[2076] Train Loss:   tensor(0.9035, device='cuda:0', grad_fn=<NllLossBackward>)
[2077] Train Loss:   tensor(0.8862, device='cuda:0', grad_fn=<NllLossBackward>)
[2078] Train Loss:   tensor(0.8966, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2079] Train Loss:    tensor(0.8712, device='cuda:0', grad_fn=<NllLossBackward>)
[2080] Train Loss:    tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[2081] Train Loss:    tensor(0.8972, device='cuda:0', grad_fn=<NllLossBackward>)
[2082] Train Loss:    tensor(0.9320, device='cuda:0', grad_fn=<NllLossBackward>)
[2083] Train Loss:    tensor(0.8878, device='cuda:0', grad_fn=<NllLossBackward>)
[2084] Train Loss:    tensor(0.9061, device='cuda:0', grad_fn=<NllLossBackward>)
[2085] Train Loss:    tensor(0.8967, device='cuda:0', grad_fn=<NllLossBackward>)
[2086] Train Loss:    tensor(0.8808, device='cuda:0', grad_fn=<NllLossBackward>)
[2087] Train Loss:    tensor(0.8733, device='cuda:0', grad_fn=<NllLossBackward>)
[2088] Train Loss:    tensor(0.8854, device='cuda:0', grad_fn=<NllLossBackward>)
[2089] Train Loss:    tensor(0.9145, device='cuda:0', grad_fn=<NllLossBackward>)
[2090] Train Loss:    tensor(0.8906, device='cuda:0', grad_fn=<NllLossBackward>)
[2091] Train Loss:    tensor(0.8884, device='cuda:0', grad_fn=<NllLossBackward>)
[2092] Train Loss:    tensor(0.9072, device='cuda:0', grad_fn=<NllLossBackward>)
[2093] Train Loss:    tensor(0.9249, device='cuda:0', grad_fn=<NllLossBackward>)
[2094] Train Loss:    tensor(0.9309, device='cuda:0', grad_fn=<NllLossBackward>)
[2095] Train Loss:    tensor(0.8841, device='cuda:0', grad_fn=<NllLossBackward>)
[2096] Train Loss:    tensor(0.8771, device='cuda:0', grad_fn=<NllLossBackward>)
[2097] Train Loss:    tensor(0.9031, device='cuda:0', grad_fn=<NllLossBackward>)
[2098] Train Loss:    tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[2099] Train Loss:    tensor(0.8871, device='cuda:0', grad_fn=<NllLossBackward>)
[2100] Train Loss:    tensor(0.8670, device='cuda:0', grad_fn=<NllLossBackward>)
[2101] Train Loss:    tensor(0.9158, device='cuda:0', grad_fn=<NllLossBackward>)
[2102] Train Loss:    tensor(0.8690, device='cuda:0', grad_fn=<NllLossBackward>)
[2103] Train Loss:    tensor(0.8880, device='cuda:0', grad_fn=<NllLossBackward>)
[2104] Train Loss:    tensor(0.8830, device='cuda:0', grad_fn=<NllLossBackward>)
[2105] Train Loss:    tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[2106] Train Loss:    tensor(0.9229, device='cuda:0', grad_fn=<NllLossBackward>)
[2107] Train Loss:    tensor(0.9379, device='cuda:0', grad_fn=<NllLossBackward>)
[2108] Train Loss:    tensor(0.8906, device='cuda:0', grad_fn=<NllLossBackward>)
[2109] Train Loss:    tensor(0.8989, device='cuda:0', grad_fn=<NllLossBackward>)
[2110] Train Loss:    tensor(0.8927, device='cuda:0', grad_fn=<NllLossBackward>)
[2111] Train Loss:    tensor(0.9204, device='cuda:0', grad_fn=<NllLossBackward>)
[2112] Train Loss:    tensor(0.8952, device='cuda:0', grad_fn=<NllLossBackward>)
[2113] Train Loss:    tensor(0.8750, device='cuda:0', grad_fn=<NllLossBackward>)
[2114] Train Loss:    tensor(0.8767, device='cuda:0', grad_fn=<NllLossBackward>)
[2115] Train Loss:    tensor(0.8879, device='cuda:0', grad_fn=<NllLossBackward>)
[2116] Train Loss:    tensor(0.8940, device='cuda:0', grad_fn=<NllLossBackward>)
[2117] Train Loss:    tensor(0.8755, device='cuda:0', grad_fn=<NllLossBackward>)
[2118] Train Loss:    tensor(0.8792, device='cuda:0', grad_fn=<NllLossBackward>)
[2119] Train Loss:    tensor(0.8780, device='cuda:0', grad_fn=<NllLossBackward>)
[2120] Train Loss:    tensor(0.9074, device='cuda:0', grad_fn=<NllLossBackward>)
[2121] Train Loss:    tensor(0.8937, device='cuda:0', grad_fn=<NllLossBackward>)
[2122] Train Loss:    tensor(0.8610, device='cuda:0', grad_fn=<NllLossBackward>)
[2123] Train Loss:    tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[2124] Train Loss:    tensor(0.8804, device='cuda:0', grad_fn=<NllLossBackward>)
[2125] Train Loss:    tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[2126] Train Loss:    tensor(0.9021, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2127] Train Loss:    tensor(0.8959, device='cuda:0', grad_fn=<NllLossBackward>)
[2128] Train Loss:    tensor(0.8571, device='cuda:0', grad_fn=<NllLossBackward>)
[2129] Train Loss:    tensor(0.9231, device='cuda:0', grad_fn=<NllLossBackward>)
[2130] Train Loss:    tensor(0.9087, device='cuda:0', grad_fn=<NllLossBackward>)
[2131] Train Loss:    tensor(0.8902, device='cuda:0', grad_fn=<NllLossBackward>)
[2132] Train Loss:    tensor(0.9340, device='cuda:0', grad_fn=<NllLossBackward>)
[2133] Train Loss:    tensor(0.8899, device='cuda:0', grad_fn=<NllLossBackward>)
[2134] Train Loss:    tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[2135] Train Loss:    tensor(0.8849, device='cuda:0', grad_fn=<NllLossBackward>)
[2136] Train Loss:    tensor(0.8920, device='cuda:0', grad_fn=<NllLossBackward>)
[2137] Train Loss:    tensor(0.8878, device='cuda:0', grad_fn=<NllLossBackward>)
[2138] Train Loss:    tensor(0.8538, device='cuda:0', grad_fn=<NllLossBackward>)
[2139] Train Loss:    tensor(0.8856, device='cuda:0', grad_fn=<NllLossBackward>)
[2140] Train Loss:    tensor(0.9191, device='cuda:0', grad_fn=<NllLossBackward>)
[2141] Train Loss:    tensor(0.8799, device='cuda:0', grad_fn=<NllLossBackward>)
[2142] Train Loss:    tensor(0.8835, device='cuda:0', grad_fn=<NllLossBackward>)
[2143] Train Loss:    tensor(0.9146, device='cuda:0', grad_fn=<NllLossBackward>)
[2144] Train Loss:    tensor(0.8738, device='cuda:0', grad_fn=<NllLossBackward>)
[2145] Train Loss:    tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[2146] Train Loss:    tensor(0.9464, device='cuda:0', grad_fn=<NllLossBackward>)
[2147] Train Loss:    tensor(0.8926, device='cuda:0', grad_fn=<NllLossBackward>)
[2148] Train Loss:    tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[2149] Train Loss:    tensor(0.9029, device='cuda:0', grad_fn=<NllLossBackward>)
[2150] Train Loss:    tensor(0.8545, device='cuda:0', grad_fn=<NllLossBackward>)
[2151] Train Loss:    tensor(0.8912, device='cuda:0', grad_fn=<NllLossBackward>)
[2152] Train Loss:    tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[2153] Train Loss:    tensor(0.9203, device='cuda:0', grad_fn=<NllLossBackward>)
[2154] Train Loss:    tensor(0.8796, device='cuda:0', grad_fn=<NllLossBackward>)
[2155] Train Loss:    tensor(0.8899, device='cuda:0', grad_fn=<NllLossBackward>)
[2156] Train Loss:    tensor(0.8916, device='cuda:0', grad_fn=<NllLossBackward>)
[2157] Train Loss:    tensor(0.9479, device='cuda:0', grad_fn=<NllLossBackward>)
[2158] Train Loss:    tensor(0.8906, device='cuda:0', grad_fn=<NllLossBackward>)
[2159] Train Loss:    tensor(0.9095, device='cuda:0', grad_fn=<NllLossBackward>)
[2160] Train Loss:    tensor(0.8903, device='cuda:0', grad_fn=<NllLossBackward>)
[2161] Train Loss:    tensor(0.8978, device='cuda:0', grad_fn=<NllLossBackward>)
[2162] Train Loss:    tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[2163] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[2164] Train Loss:    tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[2165] Train Loss:    tensor(0.9051, device='cuda:0', grad_fn=<NllLossBackward>)
[2166] Train Loss:    tensor(0.9368, device='cuda:0', grad_fn=<NllLossBackward>)
[2167] Train Loss:    tensor(0.9340, device='cuda:0', grad_fn=<NllLossBackward>)
[2168] Train Loss:    tensor(0.9392, device='cuda:0', grad_fn=<NllLossBackward>)
[2169] Train Loss:    tensor(0.9162, device='cuda:0', grad_fn=<NllLossBackward>)
[2170] Train Loss:    tensor(0.9125, device='cuda:0', grad_fn=<NllLossBackward>)
[2171] Train Loss:    tensor(0.8805, device='cuda:0', grad_fn=<NllLossBackward>)
[2172] Train Loss:    tensor(0.8835, device='cuda:0', grad_fn=<NllLossBackward>)
[2173] Train Loss:    tensor(0.8840, device='cuda:0', grad_fn=<NllLossBackward>)
[2174] Train Loss:    tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2175] Train Loss:  tensor(0.9177, device='cuda:0', grad_fn=<NllLossBackward>)
[2176] Train Loss:  tensor(0.8766, device='cuda:0', grad_fn=<NllLossBackward>)
[2177] Train Loss:  tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[2178] Train Loss:  tensor(0.8875, device='cuda:0', grad_fn=<NllLossBackward>)
[2179] Train Loss:  tensor(0.8783, device='cuda:0', grad_fn=<NllLossBackward>)
[2180] Train Loss:  tensor(0.9114, device='cuda:0', grad_fn=<NllLossBackward>)
[2181] Train Loss:  tensor(0.9067, device='cuda:0', grad_fn=<NllLossBackward>)
[2182] Train Loss:  tensor(0.9150, device='cuda:0', grad_fn=<NllLossBackward>)
[2183] Train Loss:  tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[2184] Train Loss:  tensor(0.9031, device='cuda:0', grad_fn=<NllLossBackward>)
[2185] Train Loss:  tensor(0.8917, device='cuda:0', grad_fn=<NllLossBackward>)
[2186] Train Loss:  tensor(0.8823, device='cuda:0', grad_fn=<NllLossBackward>)
[2187] Train Loss:  tensor(0.8831, device='cuda:0', grad_fn=<NllLossBackward>)
[2188] Train Loss:  tensor(0.8981, device='cuda:0', grad_fn=<NllLossBackward>)
[2189] Train Loss:  tensor(0.8918, device='cuda:0', grad_fn=<NllLossBackward>)
[2190] Train Loss:  tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[2191] Train Loss:  tensor(0.9280, device='cuda:0', grad_fn=<NllLossBackward>)
[2192] Train Loss:  tensor(0.8728, device='cuda:0', grad_fn=<NllLossBackward>)
[2193] Train Loss:  tensor(0.8845, device='cuda:0', grad_fn=<NllLossBackward>)
[2194] Train Loss:  tensor(0.9362, device='cuda:0', grad_fn=<NllLossBackward>)
[2195] Train Loss:  tensor(0.8753, device='cuda:0', grad_fn=<NllLossBackward>)
[2196] Train Loss:  tensor(0.8781, device='cuda:0', grad_fn=<NllLossBackward>)
[2197] Train Loss:  tensor(0.9074, device='cuda:0', grad_fn=<NllLossBackward>)
[2198] Train Loss:  tensor(0.8923, device='cuda:0', grad_fn=<NllLossBackward>)
[2199] Train Loss:  tensor(0.9075, device='cuda:0', grad_fn=<NllLossBackward>)
[2200] Train Loss:  tensor(0.8843, device='cuda:0', grad_fn=<NllLossBackward>)
[2201] Train Loss:  tensor(0.8646, device='cuda:0', grad_fn=<NllLossBackward>)
[2202] Train Loss:  tensor(0.9305, device='cuda:0', grad_fn=<NllLossBackward>)
[2203] Train Loss:  tensor(0.8898, device='cuda:0', grad_fn=<NllLossBackward>)
[2204] Train Loss:  tensor(0.8988, device='cuda:0', grad_fn=<NllLossBackward>)
[2205] Train Loss:  tensor(0.8882, device='cuda:0', grad_fn=<NllLossBackward>)
[2206] Train Loss:  tensor(0.8640, device='cuda:0', grad_fn=<NllLossBackward>)
[2207] Train Loss:  tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[2208] Train Loss:  tensor(0.8734, device='cuda:0', grad_fn=<NllLossBackward>)
[2209] Train Loss:  tensor(0.8535, device='cuda:0', grad_fn=<NllLossBackward>)
[2210] Train Loss:  tensor(0.9057, device='cuda:0', grad_fn=<NllLossBackward>)
[2211] Train Loss:  tensor(0.8686, device='cuda:0', grad_fn=<NllLossBackward>)
[2212] Train Loss:  tensor(0.8869, device='cuda:0', grad_fn=<NllLossBackward>)
[2213] Train Loss:  tensor(0.8831, device='cuda:0', grad_fn=<NllLossBackward>)
[2214] Train Loss:  tensor(0.8939, device='cuda:0', grad_fn=<NllLossBackward>)
[2215] Train Loss:  tensor(0.9040, device='cuda:0', grad_fn=<NllLossBackward>)
[2216] Train Loss:  tensor(0.9099, device='cuda:0', grad_fn=<NllLossBackward>)
[2217] Train Loss:  tensor(0.9039, device='cuda:0', grad_fn=<NllLossBackward>)
[2218] Train Loss:  tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[2219] Train Loss:  tensor(0.8705, device='cuda:0', grad_fn=<NllLossBackward>)
[2220] Train Loss:  tensor(0.8822, device='cuda:0', grad_fn=<NllLossBackward>)
[2221] Train Loss:  tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
[2222] Train Loss:  tensor(0.9033, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2223] Train Loss:   tensor(0.9007, device='cuda:0', grad_fn=<NllLossBackward>)
[2224] Train Loss:   tensor(0.8873, device='cuda:0', grad_fn=<NllLossBackward>)
[2225] Train Loss:   tensor(0.9005, device='cuda:0', grad_fn=<NllLossBackward>)
[2226] Train Loss:   tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[2227] Train Loss:   tensor(0.9251, device='cuda:0', grad_fn=<NllLossBackward>)
[2228] Train Loss:   tensor(0.8348, device='cuda:0', grad_fn=<NllLossBackward>)
[2229] Train Loss:   tensor(0.8792, device='cuda:0', grad_fn=<NllLossBackward>)
[2230] Train Loss:   tensor(0.8928, device='cuda:0', grad_fn=<NllLossBackward>)
[2231] Train Loss:   tensor(0.8932, device='cuda:0', grad_fn=<NllLossBackward>)
[2232] Train Loss:   tensor(0.8818, device='cuda:0', grad_fn=<NllLossBackward>)
[2233] Train Loss:   tensor(0.8841, device='cuda:0', grad_fn=<NllLossBackward>)
[2234] Train Loss:   tensor(0.9214, device='cuda:0', grad_fn=<NllLossBackward>)
[2235] Train Loss:   tensor(0.9045, device='cuda:0', grad_fn=<NllLossBackward>)
[2236] Train Loss:   tensor(0.8949, device='cuda:0', grad_fn=<NllLossBackward>)
[2237] Train Loss:   tensor(0.8702, device='cuda:0', grad_fn=<NllLossBackward>)
[2238] Train Loss:   tensor(0.8670, device='cuda:0', grad_fn=<NllLossBackward>)
[2239] Train Loss:   tensor(0.8552, device='cuda:0', grad_fn=<NllLossBackward>)
[2240] Train Loss:   tensor(0.8792, device='cuda:0', grad_fn=<NllLossBackward>)
[2241] Train Loss:   tensor(0.9277, device='cuda:0', grad_fn=<NllLossBackward>)
[2242] Train Loss:   tensor(0.9302, device='cuda:0', grad_fn=<NllLossBackward>)
[2243] Train Loss:   tensor(0.9366, device='cuda:0', grad_fn=<NllLossBackward>)
[2244] Train Loss:   tensor(0.8912, device='cuda:0', grad_fn=<NllLossBackward>)
[2245] Train Loss:   tensor(0.8932, device='cuda:0', grad_fn=<NllLossBackward>)
[2246] Train Loss:   tensor(0.8818, device='cuda:0', grad_fn=<NllLossBackward>)
[2247] Train Loss:   tensor(0.8786, device='cuda:0', grad_fn=<NllLossBackward>)
[2248] Train Loss:   tensor(0.8863, device='cuda:0', grad_fn=<NllLossBackward>)
[2249] Train Loss:   tensor(0.8881, device='cuda:0', grad_fn=<NllLossBackward>)
[2250] Train Loss:   tensor(0.8881, device='cuda:0', grad_fn=<NllLossBackward>)
[2251] Train Loss:   tensor(0.9038, device='cuda:0', grad_fn=<NllLossBackward>)
[2252] Train Loss:   tensor(0.9109, device='cuda:0', grad_fn=<NllLossBackward>)
[2253] Train Loss:   tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
[2254] Train Loss:   tensor(0.9439, device='cuda:0', grad_fn=<NllLossBackward>)
[2255] Train Loss:   tensor(0.8543, device='cuda:0', grad_fn=<NllLossBackward>)
[2256] Train Loss:   tensor(0.8515, device='cuda:0', grad_fn=<NllLossBackward>)
[2257] Train Loss:   tensor(0.8778, device='cuda:0', grad_fn=<NllLossBackward>)
[2258] Train Loss:   tensor(0.8963, device='cuda:0', grad_fn=<NllLossBackward>)
[2259] Train Loss:   tensor(0.8619, device='cuda:0', grad_fn=<NllLossBackward>)
[2260] Train Loss:   tensor(0.8894, device='cuda:0', grad_fn=<NllLossBackward>)
[2261] Train Loss:   tensor(0.9299, device='cuda:0', grad_fn=<NllLossBackward>)
[2262] Train Loss:   tensor(0.8950, device='cuda:0', grad_fn=<NllLossBackward>)
[2263] Train Loss:   tensor(0.9074, device='cuda:0', grad_fn=<NllLossBackward>)
[2264] Train Loss:   tensor(0.8819, device='cuda:0', grad_fn=<NllLossBackward>)
[2265] Train Loss:   tensor(0.9177, device='cuda:0', grad_fn=<NllLossBackward>)
[2266] Train Loss:   tensor(0.9121, device='cuda:0', grad_fn=<NllLossBackward>)
[2267] Train Loss:   tensor(0.8860, device='cuda:0', grad_fn=<NllLossBackward>)
[2268] Train Loss:   tensor(0.8800, device='cuda:0', grad_fn=<NllLossBackward>)
[2269] Train Loss:   tensor(0.8876, device='cuda:0', grad_fn=<NllLossBackward>)
[2270] Train Loss:   tensor(0.9140, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2271] Train Loss:    tensor(0.8850, device='cuda:0', grad_fn=<NllLossBackward>)
[2272] Train Loss:    tensor(0.8902, device='cuda:0', grad_fn=<NllLossBackward>)
[2273] Train Loss:    tensor(0.8940, device='cuda:0', grad_fn=<NllLossBackward>)
[2274] Train Loss:    tensor(0.8864, device='cuda:0', grad_fn=<NllLossBackward>)
[2275] Train Loss:    tensor(0.8562, device='cuda:0', grad_fn=<NllLossBackward>)
[2276] Train Loss:    tensor(0.9346, device='cuda:0', grad_fn=<NllLossBackward>)
[2277] Train Loss:    tensor(0.8574, device='cuda:0', grad_fn=<NllLossBackward>)
[2278] Train Loss:    tensor(0.9369, device='cuda:0', grad_fn=<NllLossBackward>)
[2279] Train Loss:    tensor(0.8886, device='cuda:0', grad_fn=<NllLossBackward>)
[2280] Train Loss:    tensor(0.9047, device='cuda:0', grad_fn=<NllLossBackward>)
[2281] Train Loss:    tensor(0.9141, device='cuda:0', grad_fn=<NllLossBackward>)
[2282] Train Loss:    tensor(0.9050, device='cuda:0', grad_fn=<NllLossBackward>)
[2283] Train Loss:    tensor(0.8937, device='cuda:0', grad_fn=<NllLossBackward>)
[2284] Train Loss:    tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[2285] Train Loss:    tensor(0.8930, device='cuda:0', grad_fn=<NllLossBackward>)
[2286] Train Loss:    tensor(0.9165, device='cuda:0', grad_fn=<NllLossBackward>)
[2287] Train Loss:    tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[2288] Train Loss:    tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[2289] Train Loss:    tensor(0.8561, device='cuda:0', grad_fn=<NllLossBackward>)
[2290] Train Loss:    tensor(0.8956, device='cuda:0', grad_fn=<NllLossBackward>)
[2291] Train Loss:    tensor(0.8664, device='cuda:0', grad_fn=<NllLossBackward>)
[2292] Train Loss:    tensor(0.8887, device='cuda:0', grad_fn=<NllLossBackward>)
[2293] Train Loss:    tensor(0.8912, device='cuda:0', grad_fn=<NllLossBackward>)
[2294] Train Loss:    tensor(0.8930, device='cuda:0', grad_fn=<NllLossBackward>)
[2295] Train Loss:    tensor(0.9019, device='cuda:0', grad_fn=<NllLossBackward>)
[2296] Train Loss:    tensor(0.8865, device='cuda:0', grad_fn=<NllLossBackward>)
[2297] Train Loss:    tensor(0.8900, device='cuda:0', grad_fn=<NllLossBackward>)
[2298] Train Loss:    tensor(0.8402, device='cuda:0', grad_fn=<NllLossBackward>)
[2299] Train Loss:    tensor(0.8639, device='cuda:0', grad_fn=<NllLossBackward>)
[2300] Train Loss:    tensor(0.8554, device='cuda:0', grad_fn=<NllLossBackward>)
[2301] Train Loss:    tensor(0.8648, device='cuda:0', grad_fn=<NllLossBackward>)
[2302] Train Loss:    tensor(0.8868, device='cuda:0', grad_fn=<NllLossBackward>)
[2303] Train Loss:    tensor(0.8410, device='cuda:0', grad_fn=<NllLossBackward>)
[2304] Train Loss:    tensor(0.8858, device='cuda:0', grad_fn=<NllLossBackward>)
[2305] Train Loss:    tensor(0.8934, device='cuda:0', grad_fn=<NllLossBackward>)
[2306] Train Loss:    tensor(0.8831, device='cuda:0', grad_fn=<NllLossBackward>)
[2307] Train Loss:    tensor(0.8726, device='cuda:0', grad_fn=<NllLossBackward>)
[2308] Train Loss:    tensor(0.9194, device='cuda:0', grad_fn=<NllLossBackward>)
[2309] Train Loss:    tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[2310] Train Loss:    tensor(0.8967, device='cuda:0', grad_fn=<NllLossBackward>)
[2311] Train Loss:    tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[2312] Train Loss:    tensor(0.8741, device='cuda:0', grad_fn=<NllLossBackward>)
[2313] Train Loss:    tensor(0.8643, device='cuda:0', grad_fn=<NllLossBackward>)
[2314] Train Loss:    tensor(0.8580, device='cuda:0', grad_fn=<NllLossBackward>)
[2315] Train Loss:    tensor(0.8882, device='cuda:0', grad_fn=<NllLossBackward>)
[2316] Train Loss:    tensor(0.9114, device='cuda:0', grad_fn=<NllLossBackward>)
[2317] Train Loss:    tensor(0.9037, device='cuda:0', grad_fn=<NllLossBackward>)
[2318] Train Loss:    tensor(0.9016, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2319] Train Loss:    tensor(0.9213, device='cuda:0', grad_fn=<NllLossBackward>)
[2320] Train Loss:    tensor(0.9190, device='cuda:0', grad_fn=<NllLossBackward>)
[2321] Train Loss:    tensor(0.8807, device='cuda:0', grad_fn=<NllLossBackward>)
[2322] Train Loss:    tensor(0.9423, device='cuda:0', grad_fn=<NllLossBackward>)
[2323] Train Loss:    tensor(0.8959, device='cuda:0', grad_fn=<NllLossBackward>)
[2324] Train Loss:    tensor(0.8511, device='cuda:0', grad_fn=<NllLossBackward>)
[2325] Train Loss:    tensor(0.8753, device='cuda:0', grad_fn=<NllLossBackward>)
[2326] Train Loss:    tensor(0.9286, device='cuda:0', grad_fn=<NllLossBackward>)
[2327] Train Loss:    tensor(0.9079, device='cuda:0', grad_fn=<NllLossBackward>)
[2328] Train Loss:    tensor(0.8904, device='cuda:0', grad_fn=<NllLossBackward>)
[2329] Train Loss:    tensor(0.8900, device='cuda:0', grad_fn=<NllLossBackward>)
[2330] Train Loss:    tensor(0.8756, device='cuda:0', grad_fn=<NllLossBackward>)
[2331] Train Loss:    tensor(0.8807, device='cuda:0', grad_fn=<NllLossBackward>)
[2332] Train Loss:    tensor(0.8681, device='cuda:0', grad_fn=<NllLossBackward>)
[2333] Train Loss:    tensor(0.9179, device='cuda:0', grad_fn=<NllLossBackward>)
[2334] Train Loss:    tensor(0.8997, device='cuda:0', grad_fn=<NllLossBackward>)
[2335] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[2336] Train Loss:    tensor(0.9198, device='cuda:0', grad_fn=<NllLossBackward>)
[2337] Train Loss:    tensor(0.8697, device='cuda:0', grad_fn=<NllLossBackward>)
[2338] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[2339] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[2340] Train Loss:    tensor(0.8970, device='cuda:0', grad_fn=<NllLossBackward>)
[2341] Train Loss:    tensor(0.8701, device='cuda:0', grad_fn=<NllLossBackward>)
[2342] Train Loss:    tensor(0.8709, device='cuda:0', grad_fn=<NllLossBackward>)
[2343] Train Loss:    tensor(0.8782, device='cuda:0', grad_fn=<NllLossBackward>)
[2344] Train Loss:    tensor(0.8739, device='cuda:0', grad_fn=<NllLossBackward>)
[2345] Train Loss:    tensor(0.8830, device='cuda:0', grad_fn=<NllLossBackward>)
[2346] Train Loss:    tensor(0.9116, device='cuda:0', grad_fn=<NllLossBackward>)
[2347] Train Loss:    tensor(0.8604, device='cuda:0', grad_fn=<NllLossBackward>)
[2348] Train Loss:    tensor(0.9130, device='cuda:0', grad_fn=<NllLossBackward>)
[2349] Train Loss:    tensor(0.8739, device='cuda:0', grad_fn=<NllLossBackward>)
[2350] Train Loss:    tensor(0.9215, device='cuda:0', grad_fn=<NllLossBackward>)
[2351] Train Loss:    tensor(0.8718, device='cuda:0', grad_fn=<NllLossBackward>)
[2352] Train Loss:    tensor(0.9119, device='cuda:0', grad_fn=<NllLossBackward>)
[2353] Train Loss:    tensor(0.8853, device='cuda:0', grad_fn=<NllLossBackward>)
[2354] Train Loss:    tensor(0.8747, device='cuda:0', grad_fn=<NllLossBackward>)
[2355] Train Loss:    tensor(0.8782, device='cuda:0', grad_fn=<NllLossBackward>)
[2356] Train Loss:    tensor(0.8696, device='cuda:0', grad_fn=<NllLossBackward>)
[2357] Train Loss:    tensor(0.8579, device='cuda:0', grad_fn=<NllLossBackward>)
[2358] Train Loss:    tensor(0.8903, device='cuda:0', grad_fn=<NllLossBackward>)
[2359] Train Loss:    tensor(0.8937, device='cuda:0', grad_fn=<NllLossBackward>)
[2360] Train Loss:    tensor(0.8901, device='cuda:0', grad_fn=<NllLossBackward>)
[2361] Train Loss:    tensor(0.9184, device='cuda:0', grad_fn=<NllLossBackward>)
[2362] Train Loss:    tensor(0.8823, device='cuda:0', grad_fn=<NllLossBackward>)
[2363] Train Loss:    tensor(0.9332, device='cuda:0', grad_fn=<NllLossBackward>)
[2364] Train Loss:    tensor(0.8914, device='cuda:0', grad_fn=<NllLossBackward>)
[2365] Train Loss:    tensor(0.8445, device='cuda:0', grad_fn=<NllLossBackward>)
[2366] Train Loss:    tensor(0.9300, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2367] Train Loss:    tensor(0.8750, device='cuda:0', grad_fn=<NllLossBackward>)
[2368] Train Loss:    tensor(0.8806, device='cuda:0', grad_fn=<NllLossBackward>)
[2369] Train Loss:    tensor(0.9067, device='cuda:0', grad_fn=<NllLossBackward>)
[2370] Train Loss:    tensor(0.8705, device='cuda:0', grad_fn=<NllLossBackward>)
[2371] Train Loss:    tensor(0.8749, device='cuda:0', grad_fn=<NllLossBackward>)
[2372] Train Loss:    tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[2373] Train Loss:    tensor(0.9018, device='cuda:0', grad_fn=<NllLossBackward>)
[2374] Train Loss:    tensor(0.8814, device='cuda:0', grad_fn=<NllLossBackward>)
[2375] Train Loss:    tensor(0.8641, device='cuda:0', grad_fn=<NllLossBackward>)
[2376] Train Loss:    tensor(0.8859, device='cuda:0', grad_fn=<NllLossBackward>)
[2377] Train Loss:    tensor(0.9196, device='cuda:0', grad_fn=<NllLossBackward>)
[2378] Train Loss:    tensor(0.8760, device='cuda:0', grad_fn=<NllLossBackward>)
[2379] Train Loss:    tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[2380] Train Loss:    tensor(0.8550, device='cuda:0', grad_fn=<NllLossBackward>)
[2381] Train Loss:    tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[2382] Train Loss:    tensor(0.8960, device='cuda:0', grad_fn=<NllLossBackward>)
[2383] Train Loss:    tensor(0.9116, device='cuda:0', grad_fn=<NllLossBackward>)
[2384] Train Loss:    tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[2385] Train Loss:    tensor(0.8847, device='cuda:0', grad_fn=<NllLossBackward>)
[2386] Train Loss:    tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[2387] Train Loss:    tensor(0.8750, device='cuda:0', grad_fn=<NllLossBackward>)
[2388] Train Loss:    tensor(0.9165, device='cuda:0', grad_fn=<NllLossBackward>)
[2389] Train Loss:    tensor(0.8899, device='cuda:0', grad_fn=<NllLossBackward>)
[2390] Train Loss:    tensor(0.9019, device='cuda:0', grad_fn=<NllLossBackward>)
[2391] Train Loss:    tensor(0.9242, device='cuda:0', grad_fn=<NllLossBackward>)
[2392] Train Loss:    tensor(0.8750, device='cuda:0', grad_fn=<NllLossBackward>)
[2393] Train Loss:    tensor(0.8905, device='cuda:0', grad_fn=<NllLossBackward>)
[2394] Train Loss:    tensor(0.9671, device='cuda:0', grad_fn=<NllLossBackward>)
[2395] Train Loss:    tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
[2396] Train Loss:    tensor(0.8932, device='cuda:0', grad_fn=<NllLossBackward>)
[2397] Train Loss:    tensor(0.8866, device='cuda:0', grad_fn=<NllLossBackward>)
[2398] Train Loss:    tensor(0.8803, device='cuda:0', grad_fn=<NllLossBackward>)
[2399] Train Loss:    tensor(0.8903, device='cuda:0', grad_fn=<NllLossBackward>)
[2400] Train Loss:    tensor(0.8752, device='cuda:0', grad_fn=<NllLossBackward>)
[2401] Train Loss:    tensor(0.8819, device='cuda:0', grad_fn=<NllLossBackward>)
[2402] Train Loss:    tensor(0.8909, device='cuda:0', grad_fn=<NllLossBackward>)
[2403] Train Loss:    tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[2404] Train Loss:    tensor(0.8748, device='cuda:0', grad_fn=<NllLossBackward>)
[2405] Train Loss:    tensor(0.9054, device='cuda:0', grad_fn=<NllLossBackward>)
[2406] Train Loss:    tensor(0.8653, device='cuda:0', grad_fn=<NllLossBackward>)
[2407] Train Loss:    tensor(0.9282, device='cuda:0', grad_fn=<NllLossBackward>)
[2408] Train Loss:    tensor(0.8481, device='cuda:0', grad_fn=<NllLossBackward>)
[2409] Train Loss:    tensor(0.9117, device='cuda:0', grad_fn=<NllLossBackward>)
[2410] Train Loss:    tensor(0.9000, device='cuda:0', grad_fn=<NllLossBackward>)
[2411] Train Loss:    tensor(0.8866, device='cuda:0', grad_fn=<NllLossBackward>)
[2412] Train Loss:    tensor(0.8860, device='cuda:0', grad_fn=<NllLossBackward>)
[2413] Train Loss:    tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[2414] Train Loss:    tensor(0.8973, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2415] Train Loss:    tensor(0.8930, device='cuda:0', grad_fn=<NllLossBackward>)
[2416] Train Loss:    tensor(0.8981, device='cuda:0', grad_fn=<NllLossBackward>)
[2417] Train Loss:    tensor(0.8896, device='cuda:0', grad_fn=<NllLossBackward>)
[2418] Train Loss:    tensor(0.8934, device='cuda:0', grad_fn=<NllLossBackward>)
[2419] Train Loss:    tensor(0.8829, device='cuda:0', grad_fn=<NllLossBackward>)
[2420] Train Loss:    tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
[2421] Train Loss:    tensor(0.8799, device='cuda:0', grad_fn=<NllLossBackward>)
[2422] Train Loss:    tensor(0.8776, device='cuda:0', grad_fn=<NllLossBackward>)
[2423] Train Loss:    tensor(0.8995, device='cuda:0', grad_fn=<NllLossBackward>)
[2424] Train Loss:    tensor(0.9058, device='cuda:0', grad_fn=<NllLossBackward>)
[2425] Train Loss:    tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[2426] Train Loss:    tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[2427] Train Loss:    tensor(0.8986, device='cuda:0', grad_fn=<NllLossBackward>)
[2428] Train Loss:    tensor(0.9129, device='cuda:0', grad_fn=<NllLossBackward>)
[2429] Train Loss:    tensor(0.9172, device='cuda:0', grad_fn=<NllLossBackward>)
[2430] Train Loss:    tensor(0.8858, device='cuda:0', grad_fn=<NllLossBackward>)
[2431] Train Loss:    tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[2432] Train Loss:    tensor(0.8926, device='cuda:0', grad_fn=<NllLossBackward>)
[2433] Train Loss:    tensor(0.8756, device='cuda:0', grad_fn=<NllLossBackward>)
[2434] Train Loss:    tensor(0.8768, device='cuda:0', grad_fn=<NllLossBackward>)
[2435] Train Loss:    tensor(0.8930, device='cuda:0', grad_fn=<NllLossBackward>)
[2436] Train Loss:    tensor(0.8669, device='cuda:0', grad_fn=<NllLossBackward>)
[2437] Train Loss:    tensor(0.9051, device='cuda:0', grad_fn=<NllLossBackward>)
[2438] Train Loss:    tensor(0.8923, device='cuda:0', grad_fn=<NllLossBackward>)
[2439] Train Loss:    tensor(0.8522, device='cuda:0', grad_fn=<NllLossBackward>)
[2440] Train Loss:    tensor(0.8963, device='cuda:0', grad_fn=<NllLossBackward>)
[2441] Train Loss:    tensor(0.8724, device='cuda:0', grad_fn=<NllLossBackward>)
[2442] Train Loss:    tensor(0.8615, device='cuda:0', grad_fn=<NllLossBackward>)
[2443] Train Loss:    tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[2444] Train Loss:    tensor(0.9384, device='cuda:0', grad_fn=<NllLossBackward>)
[2445] Train Loss:    tensor(0.9116, device='cuda:0', grad_fn=<NllLossBackward>)
[2446] Train Loss:    tensor(0.8901, device='cuda:0', grad_fn=<NllLossBackward>)
[2447] Train Loss:    tensor(0.8653, device='cuda:0', grad_fn=<NllLossBackward>)
[2448] Train Loss:    tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[2449] Train Loss:    tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[2450] Train Loss:    tensor(0.8717, device='cuda:0', grad_fn=<NllLossBackward>)
[2451] Train Loss:    tensor(0.8501, device='cuda:0', grad_fn=<NllLossBackward>)
[2452] Train Loss:    tensor(0.9091, device='cuda:0', grad_fn=<NllLossBackward>)
[2453] Train Loss:    tensor(0.8813, device='cuda:0', grad_fn=<NllLossBackward>)
[2454] Train Loss:    tensor(0.9117, device='cuda:0', grad_fn=<NllLossBackward>)
[2455] Train Loss:    tensor(0.8678, device='cuda:0', grad_fn=<NllLossBackward>)
[2456] Train Loss:    tensor(0.8678, device='cuda:0', grad_fn=<NllLossBackward>)
[2457] Train Loss:    tensor(0.9106, device='cuda:0', grad_fn=<NllLossBackward>)
[2458] Train Loss:    tensor(0.8647, device='cuda:0', grad_fn=<NllLossBackward>)
[2459] Train Loss:    tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[2460] Train Loss:    tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[2461] Train Loss:    tensor(0.8612, device='cuda:0', grad_fn=<NllLossBackward>)
[2462] Train Loss:    tensor(0.8357, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2463] Train Loss:   tensor(0.9158, device='cuda:0', grad_fn=<NllLossBackward>)
[2464] Train Loss:   tensor(0.8849, device='cuda:0', grad_fn=<NllLossBackward>)
[2465] Train Loss:   tensor(0.8978, device='cuda:0', grad_fn=<NllLossBackward>)
[2466] Train Loss:   tensor(0.8900, device='cuda:0', grad_fn=<NllLossBackward>)
[2467] Train Loss:   tensor(0.8703, device='cuda:0', grad_fn=<NllLossBackward>)
[2468] Train Loss:   tensor(0.8762, device='cuda:0', grad_fn=<NllLossBackward>)
[2469] Train Loss:   tensor(0.8943, device='cuda:0', grad_fn=<NllLossBackward>)
[2470] Train Loss:   tensor(0.8740, device='cuda:0', grad_fn=<NllLossBackward>)
[2471] Train Loss:   tensor(0.9297, device='cuda:0', grad_fn=<NllLossBackward>)
[2472] Train Loss:   tensor(0.8804, device='cuda:0', grad_fn=<NllLossBackward>)
[2473] Train Loss:   tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[2474] Train Loss:   tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[2475] Train Loss:   tensor(0.9314, device='cuda:0', grad_fn=<NllLossBackward>)
[2476] Train Loss:   tensor(0.9222, device='cuda:0', grad_fn=<NllLossBackward>)
[2477] Train Loss:   tensor(0.8868, device='cuda:0', grad_fn=<NllLossBackward>)
[2478] Train Loss:   tensor(0.8671, device='cuda:0', grad_fn=<NllLossBackward>)
[2479] Train Loss:   tensor(0.8693, device='cuda:0', grad_fn=<NllLossBackward>)
[2480] Train Loss:   tensor(0.8738, device='cuda:0', grad_fn=<NllLossBackward>)
[2481] Train Loss:   tensor(0.8862, device='cuda:0', grad_fn=<NllLossBackward>)
[2482] Train Loss:   tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[2483] Train Loss:   tensor(0.8805, device='cuda:0', grad_fn=<NllLossBackward>)
[2484] Train Loss:   tensor(0.8871, device='cuda:0', grad_fn=<NllLossBackward>)
[2485] Train Loss:   tensor(0.9067, device='cuda:0', grad_fn=<NllLossBackward>)
[2486] Train Loss:   tensor(0.8851, device='cuda:0', grad_fn=<NllLossBackward>)
[2487] Train Loss:   tensor(0.9012, device='cuda:0', grad_fn=<NllLossBackward>)
[2488] Train Loss:   tensor(0.8579, device='cuda:0', grad_fn=<NllLossBackward>)
[2489] Train Loss:   tensor(0.8718, device='cuda:0', grad_fn=<NllLossBackward>)
[2490] Train Loss:   tensor(0.9106, device='cuda:0', grad_fn=<NllLossBackward>)
[2491] Train Loss:   tensor(0.8563, device='cuda:0', grad_fn=<NllLossBackward>)
[2492] Train Loss:   tensor(0.8842, device='cuda:0', grad_fn=<NllLossBackward>)
[2493] Train Loss:   tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[2494] Train Loss:   tensor(0.9293, device='cuda:0', grad_fn=<NllLossBackward>)
[2495] Train Loss:   tensor(0.8872, device='cuda:0', grad_fn=<NllLossBackward>)
[2496] Train Loss:   tensor(0.8500, device='cuda:0', grad_fn=<NllLossBackward>)
[2497] Train Loss:   tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[2498] Train Loss:   tensor(0.8799, device='cuda:0', grad_fn=<NllLossBackward>)
[2499] Train Loss:   tensor(0.9099, device='cuda:0', grad_fn=<NllLossBackward>)
[2500] Train Loss:   tensor(0.8910, device='cuda:0', grad_fn=<NllLossBackward>)
[2501] Train Loss:   tensor(0.8953, device='cuda:0', grad_fn=<NllLossBackward>)
[2502] Train Loss:   tensor(0.8927, device='cuda:0', grad_fn=<NllLossBackward>)
[2503] Train Loss:   tensor(0.8889, device='cuda:0', grad_fn=<NllLossBackward>)
[2504] Train Loss:   tensor(0.8809, device='cuda:0', grad_fn=<NllLossBackward>)
[2505] Train Loss:   tensor(0.8670, device='cuda:0', grad_fn=<NllLossBackward>)
[2506] Train Loss:   tensor(0.9258, device='cuda:0', grad_fn=<NllLossBackward>)
[2507] Train Loss:   tensor(0.9007, device='cuda:0', grad_fn=<NllLossBackward>)
[2508] Train Loss:   tensor(0.8968, device='cuda:0', grad_fn=<NllLossBackward>)
[2509] Train Loss:   tensor(0.8944, device='cuda:0', grad_fn=<NllLossBackward>)
[2510] Train Loss:   tensor(0.9281, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2511] Train Loss:    tensor(0.8829, device='cuda:0', grad_fn=<NllLossBackward>)
[2512] Train Loss:    tensor(0.8719, device='cuda:0', grad_fn=<NllLossBackward>)
[2513] Train Loss:    tensor(0.8902, device='cuda:0', grad_fn=<NllLossBackward>)
[2514] Train Loss:    tensor(0.8890, device='cuda:0', grad_fn=<NllLossBackward>)
[2515] Train Loss:    tensor(0.8987, device='cuda:0', grad_fn=<NllLossBackward>)
[2516] Train Loss:    tensor(0.9042, device='cuda:0', grad_fn=<NllLossBackward>)
[2517] Train Loss:    tensor(0.8735, device='cuda:0', grad_fn=<NllLossBackward>)
[2518] Train Loss:    tensor(0.8688, device='cuda:0', grad_fn=<NllLossBackward>)
[2519] Train Loss:    tensor(0.9038, device='cuda:0', grad_fn=<NllLossBackward>)
[2520] Train Loss:    tensor(0.8405, device='cuda:0', grad_fn=<NllLossBackward>)
[2521] Train Loss:    tensor(0.9127, device='cuda:0', grad_fn=<NllLossBackward>)
[2522] Train Loss:    tensor(0.8961, device='cuda:0', grad_fn=<NllLossBackward>)
[2523] Train Loss:    tensor(0.8783, device='cuda:0', grad_fn=<NllLossBackward>)
[2524] Train Loss:    tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[2525] Train Loss:    tensor(0.9157, device='cuda:0', grad_fn=<NllLossBackward>)
[2526] Train Loss:    tensor(0.8956, device='cuda:0', grad_fn=<NllLossBackward>)
[2527] Train Loss:    tensor(0.8811, device='cuda:0', grad_fn=<NllLossBackward>)
[2528] Train Loss:    tensor(0.9086, device='cuda:0', grad_fn=<NllLossBackward>)
[2529] Train Loss:    tensor(0.8990, device='cuda:0', grad_fn=<NllLossBackward>)
[2530] Train Loss:    tensor(0.8949, device='cuda:0', grad_fn=<NllLossBackward>)
[2531] Train Loss:    tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
[2532] Train Loss:    tensor(0.8811, device='cuda:0', grad_fn=<NllLossBackward>)
[2533] Train Loss:    tensor(0.8798, device='cuda:0', grad_fn=<NllLossBackward>)
[2534] Train Loss:    tensor(0.9001, device='cuda:0', grad_fn=<NllLossBackward>)
[2535] Train Loss:    tensor(0.9133, device='cuda:0', grad_fn=<NllLossBackward>)
[2536] Train Loss:    tensor(0.8856, device='cuda:0', grad_fn=<NllLossBackward>)
[2537] Train Loss:    tensor(0.9231, device='cuda:0', grad_fn=<NllLossBackward>)
[2538] Train Loss:    tensor(0.8652, device='cuda:0', grad_fn=<NllLossBackward>)
[2539] Train Loss:    tensor(0.8887, device='cuda:0', grad_fn=<NllLossBackward>)
[2540] Train Loss:    tensor(0.9120, device='cuda:0', grad_fn=<NllLossBackward>)
[2541] Train Loss:    tensor(0.8817, device='cuda:0', grad_fn=<NllLossBackward>)
[2542] Train Loss:    tensor(0.9067, device='cuda:0', grad_fn=<NllLossBackward>)
[2543] Train Loss:    tensor(0.9086, device='cuda:0', grad_fn=<NllLossBackward>)
[2544] Train Loss:    tensor(0.8641, device='cuda:0', grad_fn=<NllLossBackward>)
[2545] Train Loss:    tensor(0.8734, device='cuda:0', grad_fn=<NllLossBackward>)
[2546] Train Loss:    tensor(0.8853, device='cuda:0', grad_fn=<NllLossBackward>)
[2547] Train Loss:    tensor(0.9117, device='cuda:0', grad_fn=<NllLossBackward>)
[2548] Train Loss:    tensor(0.8807, device='cuda:0', grad_fn=<NllLossBackward>)
[2549] Train Loss:    tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[2550] Train Loss:    tensor(0.9277, device='cuda:0', grad_fn=<NllLossBackward>)
[2551] Train Loss:    tensor(0.8778, device='cuda:0', grad_fn=<NllLossBackward>)
[2552] Train Loss:    tensor(0.8752, device='cuda:0', grad_fn=<NllLossBackward>)
[2553] Train Loss:    tensor(0.9205, device='cuda:0', grad_fn=<NllLossBackward>)
[2554] Train Loss:    tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[2555] Train Loss:    tensor(0.8989, device='cuda:0', grad_fn=<NllLossBackward>)
[2556] Train Loss:    tensor(0.8926, device='cuda:0', grad_fn=<NllLossBackward>)
[2557] Train Loss:    tensor(0.8923, device='cuda:0', grad_fn=<NllLossBackward>)
[2558] Train Loss:    tensor(0.8672, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2559] Train Loss:    tensor(0.8950, device='cuda:0', grad_fn=<NllLossBackward>)
[2560] Train Loss:    tensor(0.8700, device='cuda:0', grad_fn=<NllLossBackward>)
[2561] Train Loss:    tensor(0.9089, device='cuda:0', grad_fn=<NllLossBackward>)
[2562] Train Loss:    tensor(0.8845, device='cuda:0', grad_fn=<NllLossBackward>)
[2563] Train Loss:    tensor(0.8966, device='cuda:0', grad_fn=<NllLossBackward>)
[2564] Train Loss:    tensor(0.8918, device='cuda:0', grad_fn=<NllLossBackward>)
[2565] Train Loss:    tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[2566] Train Loss:    tensor(0.8913, device='cuda:0', grad_fn=<NllLossBackward>)
[2567] Train Loss:    tensor(0.9328, device='cuda:0', grad_fn=<NllLossBackward>)
[2568] Train Loss:    tensor(0.9044, device='cuda:0', grad_fn=<NllLossBackward>)
[2569] Train Loss:    tensor(0.9036, device='cuda:0', grad_fn=<NllLossBackward>)
[2570] Train Loss:    tensor(0.9014, device='cuda:0', grad_fn=<NllLossBackward>)
[2571] Train Loss:    tensor(0.9130, device='cuda:0', grad_fn=<NllLossBackward>)
[2572] Train Loss:    tensor(0.9024, device='cuda:0', grad_fn=<NllLossBackward>)
[2573] Train Loss:    tensor(0.8939, device='cuda:0', grad_fn=<NllLossBackward>)
[2574] Train Loss:    tensor(0.8970, device='cuda:0', grad_fn=<NllLossBackward>)
[2575] Train Loss:    tensor(0.9020, device='cuda:0', grad_fn=<NllLossBackward>)
[2576] Train Loss:    tensor(0.8784, device='cuda:0', grad_fn=<NllLossBackward>)
[2577] Train Loss:    tensor(0.8711, device='cuda:0', grad_fn=<NllLossBackward>)
[2578] Train Loss:    tensor(0.8549, device='cuda:0', grad_fn=<NllLossBackward>)
[2579] Train Loss:    tensor(0.8635, device='cuda:0', grad_fn=<NllLossBackward>)
[2580] Train Loss:    tensor(0.8561, device='cuda:0', grad_fn=<NllLossBackward>)
[2581] Train Loss:    tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[2582] Train Loss:    tensor(0.8776, device='cuda:0', grad_fn=<NllLossBackward>)
[2583] Train Loss:    tensor(0.8454, device='cuda:0', grad_fn=<NllLossBackward>)
[2584] Train Loss:    tensor(0.9360, device='cuda:0', grad_fn=<NllLossBackward>)
[2585] Train Loss:    tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[2586] Train Loss:    tensor(0.9405, device='cuda:0', grad_fn=<NllLossBackward>)
[2587] Train Loss:    tensor(0.8636, device='cuda:0', grad_fn=<NllLossBackward>)
[2588] Train Loss:    tensor(0.8541, device='cuda:0', grad_fn=<NllLossBackward>)
[2589] Train Loss:    tensor(0.8510, device='cuda:0', grad_fn=<NllLossBackward>)
[2590] Train Loss:    tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[2591] Train Loss:    tensor(0.9153, device='cuda:0', grad_fn=<NllLossBackward>)
[2592] Train Loss:    tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[2593] Train Loss:    tensor(0.8902, device='cuda:0', grad_fn=<NllLossBackward>)
[2594] Train Loss:    tensor(0.8875, device='cuda:0', grad_fn=<NllLossBackward>)
[2595] Train Loss:    tensor(0.9147, device='cuda:0', grad_fn=<NllLossBackward>)
[2596] Train Loss:    tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[2597] Train Loss:    tensor(0.8774, device='cuda:0', grad_fn=<NllLossBackward>)
[2598] Train Loss:    tensor(0.8965, device='cuda:0', grad_fn=<NllLossBackward>)
[2599] Train Loss:    tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[2600] Train Loss:    tensor(0.8771, device='cuda:0', grad_fn=<NllLossBackward>)
[2601] Train Loss:    tensor(0.8813, device='cuda:0', grad_fn=<NllLossBackward>)
[2602] Train Loss:    tensor(0.8839, device='cuda:0', grad_fn=<NllLossBackward>)
[2603] Train Loss:    tensor(0.9080, device='cuda:0', grad_fn=<NllLossBackward>)
[2604] Train Loss:    tensor(0.8928, device='cuda:0', grad_fn=<NllLossBackward>)
[2605] Train Loss:    tensor(0.8657, device='cuda:0', grad_fn=<NllLossBackward>)
[2606] Train Loss:    tensor(0.8989, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2607] Train Loss:    tensor(0.8589, device='cuda:0', grad_fn=<NllLossBackward>)
[2608] Train Loss:    tensor(0.8803, device='cuda:0', grad_fn=<NllLossBackward>)
[2609] Train Loss:    tensor(0.8846, device='cuda:0', grad_fn=<NllLossBackward>)
[2610] Train Loss:    tensor(0.9301, device='cuda:0', grad_fn=<NllLossBackward>)
[2611] Train Loss:    tensor(0.8717, device='cuda:0', grad_fn=<NllLossBackward>)
[2612] Train Loss:    tensor(0.8889, device='cuda:0', grad_fn=<NllLossBackward>)
[2613] Train Loss:    tensor(0.8820, device='cuda:0', grad_fn=<NllLossBackward>)
[2614] Train Loss:    tensor(0.9135, device='cuda:0', grad_fn=<NllLossBackward>)
[2615] Train Loss:    tensor(0.9281, device='cuda:0', grad_fn=<NllLossBackward>)
[2616] Train Loss:    tensor(0.8813, device='cuda:0', grad_fn=<NllLossBackward>)
[2617] Train Loss:    tensor(0.8995, device='cuda:0', grad_fn=<NllLossBackward>)
[2618] Train Loss:    tensor(0.8719, device='cuda:0', grad_fn=<NllLossBackward>)
[2619] Train Loss:    tensor(0.8856, device='cuda:0', grad_fn=<NllLossBackward>)
[2620] Train Loss:    tensor(0.8772, device='cuda:0', grad_fn=<NllLossBackward>)
[2621] Train Loss:    tensor(0.8975, device='cuda:0', grad_fn=<NllLossBackward>)
[2622] Train Loss:    tensor(0.8783, device='cuda:0', grad_fn=<NllLossBackward>)
[2623] Train Loss:    tensor(0.8848, device='cuda:0', grad_fn=<NllLossBackward>)
[2624] Train Loss:    tensor(0.8818, device='cuda:0', grad_fn=<NllLossBackward>)
[2625] Train Loss:    tensor(0.8856, device='cuda:0', grad_fn=<NllLossBackward>)
[2626] Train Loss:    tensor(0.8652, device='cuda:0', grad_fn=<NllLossBackward>)
[2627] Train Loss:    tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[2628] Train Loss:    tensor(0.8999, device='cuda:0', grad_fn=<NllLossBackward>)
[2629] Train Loss:    tensor(0.9123, device='cuda:0', grad_fn=<NllLossBackward>)
[2630] Train Loss:    tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[2631] Train Loss:    tensor(0.8891, device='cuda:0', grad_fn=<NllLossBackward>)
[2632] Train Loss:    tensor(0.9059, device='cuda:0', grad_fn=<NllLossBackward>)
[2633] Train Loss:    tensor(0.8904, device='cuda:0', grad_fn=<NllLossBackward>)
[2634] Train Loss:    tensor(0.9200, device='cuda:0', grad_fn=<NllLossBackward>)
[2635] Train Loss:    tensor(0.9113, device='cuda:0', grad_fn=<NllLossBackward>)
[2636] Train Loss:    tensor(0.8784, device='cuda:0', grad_fn=<NllLossBackward>)
[2637] Train Loss:    tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
[2638] Train Loss:    tensor(0.8863, device='cuda:0', grad_fn=<NllLossBackward>)
[2639] Train Loss:    tensor(0.8913, device='cuda:0', grad_fn=<NllLossBackward>)
[2640] Train Loss:    tensor(0.8409, device='cuda:0', grad_fn=<NllLossBackward>)
[2641] Train Loss:    tensor(0.9271, device='cuda:0', grad_fn=<NllLossBackward>)
[2642] Train Loss:    tensor(0.9205, device='cuda:0', grad_fn=<NllLossBackward>)
[2643] Train Loss:    tensor(0.9110, device='cuda:0', grad_fn=<NllLossBackward>)
[2644] Train Loss:    tensor(0.8640, device='cuda:0', grad_fn=<NllLossBackward>)
[2645] Train Loss:    tensor(0.8753, device='cuda:0', grad_fn=<NllLossBackward>)
[2646] Train Loss:    tensor(0.8797, device='cuda:0', grad_fn=<NllLossBackward>)
[2647] Train Loss:    tensor(0.8715, device='cuda:0', grad_fn=<NllLossBackward>)
[2648] Train Loss:    tensor(0.8778, device='cuda:0', grad_fn=<NllLossBackward>)
[2649] Train Loss:    tensor(0.8995, device='cuda:0', grad_fn=<NllLossBackward>)
[2650] Train Loss:    tensor(0.9077, device='cuda:0', grad_fn=<NllLossBackward>)
[2651] Train Loss:    tensor(0.9207, device='cuda:0', grad_fn=<NllLossBackward>)
[2652] Train Loss:    tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[2653] Train Loss:    tensor(0.8892, device='cuda:0', grad_fn=<NllLossBackward>)
[2654] Train Loss:    tensor(0.8537, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2655] Train Loss:    tensor(0.8755, device='cuda:0', grad_fn=<NllLossBackward>)
[2656] Train Loss:    tensor(0.8849, device='cuda:0', grad_fn=<NllLossBackward>)
[2657] Train Loss:    tensor(0.9136, device='cuda:0', grad_fn=<NllLossBackward>)
[2658] Train Loss:    tensor(0.8850, device='cuda:0', grad_fn=<NllLossBackward>)
[2659] Train Loss:    tensor(0.9041, device='cuda:0', grad_fn=<NllLossBackward>)
[2660] Train Loss:    tensor(0.9198, device='cuda:0', grad_fn=<NllLossBackward>)
[2661] Train Loss:    tensor(0.8601, device='cuda:0', grad_fn=<NllLossBackward>)
[2662] Train Loss:    tensor(0.8665, device='cuda:0', grad_fn=<NllLossBackward>)
[2663] Train Loss:    tensor(0.8962, device='cuda:0', grad_fn=<NllLossBackward>)
[2664] Train Loss:    tensor(0.8970, device='cuda:0', grad_fn=<NllLossBackward>)
[2665] Train Loss:    tensor(0.9015, device='cuda:0', grad_fn=<NllLossBackward>)
[2666] Train Loss:    tensor(0.8733, device='cuda:0', grad_fn=<NllLossBackward>)
[2667] Train Loss:    tensor(0.8978, device='cuda:0', grad_fn=<NllLossBackward>)
[2668] Train Loss:    tensor(0.8848, device='cuda:0', grad_fn=<NllLossBackward>)
[2669] Train Loss:    tensor(0.9047, device='cuda:0', grad_fn=<NllLossBackward>)
[2670] Train Loss:    tensor(0.8731, device='cuda:0', grad_fn=<NllLossBackward>)
[2671] Train Loss:    tensor(0.8961, device='cuda:0', grad_fn=<NllLossBackward>)
[2672] Train Loss:    tensor(0.9075, device='cuda:0', grad_fn=<NllLossBackward>)
[2673] Train Loss:    tensor(0.8974, device='cuda:0', grad_fn=<NllLossBackward>)
[2674] Train Loss:    tensor(0.9100, device='cuda:0', grad_fn=<NllLossBackward>)
[2675] Train Loss:    tensor(0.8517, device='cuda:0', grad_fn=<NllLossBackward>)
[2676] Train Loss:    tensor(0.9411, device='cuda:0', grad_fn=<NllLossBackward>)
[2677] Train Loss:    tensor(0.8635, device='cuda:0', grad_fn=<NllLossBackward>)
[2678] Train Loss:    tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[2679] Train Loss:    tensor(0.8838, device='cuda:0', grad_fn=<NllLossBackward>)
[2680] Train Loss:    tensor(0.8612, device='cuda:0', grad_fn=<NllLossBackward>)
[2681] Train Loss:    tensor(0.9388, device='cuda:0', grad_fn=<NllLossBackward>)
[2682] Train Loss:    tensor(0.8684, device='cuda:0', grad_fn=<NllLossBackward>)
[2683] Train Loss:    tensor(0.8552, device='cuda:0', grad_fn=<NllLossBackward>)
[2684] Train Loss:    tensor(0.8859, device='cuda:0', grad_fn=<NllLossBackward>)
[2685] Train Loss:    tensor(0.8645, device='cuda:0', grad_fn=<NllLossBackward>)
[2686] Train Loss:    tensor(0.8821, device='cuda:0', grad_fn=<NllLossBackward>)
[2687] Train Loss:    tensor(0.8809, device='cuda:0', grad_fn=<NllLossBackward>)
[2688] Train Loss:    tensor(0.8874, device='cuda:0', grad_fn=<NllLossBackward>)
[2689] Train Loss:    tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[2690] Train Loss:    tensor(0.9256, device='cuda:0', grad_fn=<NllLossBackward>)
[2691] Train Loss:    tensor(0.8471, device='cuda:0', grad_fn=<NllLossBackward>)
[2692] Train Loss:    tensor(0.8942, device='cuda:0', grad_fn=<NllLossBackward>)
[2693] Train Loss:    tensor(0.8587, device='cuda:0', grad_fn=<NllLossBackward>)
[2694] Train Loss:    tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[2695] Train Loss:    tensor(0.8461, device='cuda:0', grad_fn=<NllLossBackward>)
[2696] Train Loss:    tensor(0.8769, device='cuda:0', grad_fn=<NllLossBackward>)
[2697] Train Loss:    tensor(0.9090, device='cuda:0', grad_fn=<NllLossBackward>)
[2698] Train Loss:    tensor(0.8511, device='cuda:0', grad_fn=<NllLossBackward>)
[2699] Train Loss:    tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[2700] Train Loss:    tensor(0.8593, device='cuda:0', grad_fn=<NllLossBackward>)
[2701] Train Loss:    tensor(0.8852, device='cuda:0', grad_fn=<NllLossBackward>)
[2702] Train Loss:    tensor(0.8565, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2703] Train Loss:    tensor(0.8690, device='cuda:0', grad_fn=<NllLossBackward>)
[2704] Train Loss:    tensor(0.8611, device='cuda:0', grad_fn=<NllLossBackward>)
[2705] Train Loss:    tensor(0.8786, device='cuda:0', grad_fn=<NllLossBackward>)
[2706] Train Loss:    tensor(0.9085, device='cuda:0', grad_fn=<NllLossBackward>)
[2707] Train Loss:    tensor(0.8866, device='cuda:0', grad_fn=<NllLossBackward>)
[2708] Train Loss:    tensor(0.8836, device='cuda:0', grad_fn=<NllLossBackward>)
[2709] Train Loss:    tensor(0.9313, device='cuda:0', grad_fn=<NllLossBackward>)
[2710] Train Loss:    tensor(0.9146, device='cuda:0', grad_fn=<NllLossBackward>)
[2711] Train Loss:    tensor(0.8520, device='cuda:0', grad_fn=<NllLossBackward>)
[2712] Train Loss:    tensor(0.9229, device='cuda:0', grad_fn=<NllLossBackward>)
[2713] Train Loss:    tensor(0.8977, device='cuda:0', grad_fn=<NllLossBackward>)
[2714] Train Loss:    tensor(0.8564, device='cuda:0', grad_fn=<NllLossBackward>)
[2715] Train Loss:    tensor(0.8611, device='cuda:0', grad_fn=<NllLossBackward>)
[2716] Train Loss:    tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[2717] Train Loss:    tensor(0.8969, device='cuda:0', grad_fn=<NllLossBackward>)
[2718] Train Loss:    tensor(0.8717, device='cuda:0', grad_fn=<NllLossBackward>)
[2719] Train Loss:    tensor(0.8843, device='cuda:0', grad_fn=<NllLossBackward>)
[2720] Train Loss:    tensor(0.8975, device='cuda:0', grad_fn=<NllLossBackward>)
[2721] Train Loss:    tensor(0.8822, device='cuda:0', grad_fn=<NllLossBackward>)
[2722] Train Loss:    tensor(0.8938, device='cuda:0', grad_fn=<NllLossBackward>)
[2723] Train Loss:    tensor(0.8754, device='cuda:0', grad_fn=<NllLossBackward>)
[2724] Train Loss:    tensor(0.8824, device='cuda:0', grad_fn=<NllLossBackward>)
[2725] Train Loss:    tensor(0.9165, device='cuda:0', grad_fn=<NllLossBackward>)
[2726] Train Loss:    tensor(0.8861, device='cuda:0', grad_fn=<NllLossBackward>)
[2727] Train Loss:    tensor(0.8787, device='cuda:0', grad_fn=<NllLossBackward>)
[2728] Train Loss:    tensor(0.8865, device='cuda:0', grad_fn=<NllLossBackward>)
[2729] Train Loss:    tensor(0.8522, device='cuda:0', grad_fn=<NllLossBackward>)
[2730] Train Loss:    tensor(0.9177, device='cuda:0', grad_fn=<NllLossBackward>)
[2731] Train Loss:    tensor(0.8952, device='cuda:0', grad_fn=<NllLossBackward>)
[2732] Train Loss:    tensor(0.8684, device='cuda:0', grad_fn=<NllLossBackward>)
[2733] Train Loss:    tensor(0.8530, device='cuda:0', grad_fn=<NllLossBackward>)
[2734] Train Loss:    tensor(0.8705, device='cuda:0', grad_fn=<NllLossBackward>)
[2735] Train Loss:    tensor(0.8962, device='cuda:0', grad_fn=<NllLossBackward>)
[2736] Train Loss:    tensor(0.8807, device='cuda:0', grad_fn=<NllLossBackward>)
[2737] Train Loss:    tensor(0.8577, device='cuda:0', grad_fn=<NllLossBackward>)
[2738] Train Loss:    tensor(0.8938, device='cuda:0', grad_fn=<NllLossBackward>)
[2739] Train Loss:    tensor(0.8547, device='cuda:0', grad_fn=<NllLossBackward>)
[2740] Train Loss:    tensor(0.8792, device='cuda:0', grad_fn=<NllLossBackward>)
[2741] Train Loss:    tensor(0.9150, device='cuda:0', grad_fn=<NllLossBackward>)
[2742] Train Loss:    tensor(0.8485, device='cuda:0', grad_fn=<NllLossBackward>)
[2743] Train Loss:    tensor(0.9121, device='cuda:0', grad_fn=<NllLossBackward>)
[2744] Train Loss:    tensor(0.8868, device='cuda:0', grad_fn=<NllLossBackward>)
[2745] Train Loss:    tensor(0.8872, device='cuda:0', grad_fn=<NllLossBackward>)
[2746] Train Loss:    tensor(0.8889, device='cuda:0', grad_fn=<NllLossBackward>)
[2747] Train Loss:    tensor(0.8881, device='cuda:0', grad_fn=<NllLossBackward>)
[2748] Train Loss:    tensor(0.8739, device='cuda:0', grad_fn=<NllLossBackward>)
[2749] Train Loss:    tensor(0.9011, device='cuda:0', grad_fn=<NllLossBackward>)
[2750] Train Loss:    tensor(0.8849, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2751] Train Loss:   tensor(0.8690, device='cuda:0', grad_fn=<NllLossBackward>)
[2752] Train Loss:   tensor(0.8636, device='cuda:0', grad_fn=<NllLossBackward>)
[2753] Train Loss:   tensor(0.8578, device='cuda:0', grad_fn=<NllLossBackward>)
[2754] Train Loss:   tensor(0.8686, device='cuda:0', grad_fn=<NllLossBackward>)
[2755] Train Loss:   tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[2756] Train Loss:   tensor(0.8810, device='cuda:0', grad_fn=<NllLossBackward>)
[2757] Train Loss:   tensor(0.9060, device='cuda:0', grad_fn=<NllLossBackward>)
[2758] Train Loss:   tensor(0.9005, device='cuda:0', grad_fn=<NllLossBackward>)
[2759] Train Loss:   tensor(0.9283, device='cuda:0', grad_fn=<NllLossBackward>)
[2760] Train Loss:   tensor(0.9014, device='cuda:0', grad_fn=<NllLossBackward>)
[2761] Train Loss:   tensor(0.9150, device='cuda:0', grad_fn=<NllLossBackward>)
[2762] Train Loss:   tensor(0.8987, device='cuda:0', grad_fn=<NllLossBackward>)
[2763] Train Loss:   tensor(0.8649, device='cuda:0', grad_fn=<NllLossBackward>)
[2764] Train Loss:   tensor(0.9020, device='cuda:0', grad_fn=<NllLossBackward>)
[2765] Train Loss:   tensor(0.8783, device='cuda:0', grad_fn=<NllLossBackward>)
[2766] Train Loss:   tensor(0.8665, device='cuda:0', grad_fn=<NllLossBackward>)
[2767] Train Loss:   tensor(0.9016, device='cuda:0', grad_fn=<NllLossBackward>)
[2768] Train Loss:   tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[2769] Train Loss:   tensor(0.8710, device='cuda:0', grad_fn=<NllLossBackward>)
[2770] Train Loss:   tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[2771] Train Loss:   tensor(0.8446, device='cuda:0', grad_fn=<NllLossBackward>)
[2772] Train Loss:   tensor(0.9184, device='cuda:0', grad_fn=<NllLossBackward>)
[2773] Train Loss:   tensor(0.8623, device='cuda:0', grad_fn=<NllLossBackward>)
[2774] Train Loss:   tensor(0.8693, device='cuda:0', grad_fn=<NllLossBackward>)
[2775] Train Loss:   tensor(0.8927, device='cuda:0', grad_fn=<NllLossBackward>)
[2776] Train Loss:   tensor(0.8957, device='cuda:0', grad_fn=<NllLossBackward>)
[2777] Train Loss:   tensor(0.9176, device='cuda:0', grad_fn=<NllLossBackward>)
[2778] Train Loss:   tensor(0.9085, device='cuda:0', grad_fn=<NllLossBackward>)
[2779] Train Loss:   tensor(0.8629, device='cuda:0', grad_fn=<NllLossBackward>)
[2780] Train Loss:   tensor(0.9124, device='cuda:0', grad_fn=<NllLossBackward>)
[2781] Train Loss:   tensor(0.8901, device='cuda:0', grad_fn=<NllLossBackward>)
[2782] Train Loss:   tensor(0.9100, device='cuda:0', grad_fn=<NllLossBackward>)
[2783] Train Loss:   tensor(0.9033, device='cuda:0', grad_fn=<NllLossBackward>)
[2784] Train Loss:   tensor(0.8997, device='cuda:0', grad_fn=<NllLossBackward>)
[2785] Train Loss:   tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[2786] Train Loss:   tensor(0.8628, device='cuda:0', grad_fn=<NllLossBackward>)
[2787] Train Loss:   tensor(0.9033, device='cuda:0', grad_fn=<NllLossBackward>)
[2788] Train Loss:   tensor(0.8751, device='cuda:0', grad_fn=<NllLossBackward>)
[2789] Train Loss:   tensor(0.8922, device='cuda:0', grad_fn=<NllLossBackward>)
[2790] Train Loss:   tensor(0.9018, device='cuda:0', grad_fn=<NllLossBackward>)
[2791] Train Loss:   tensor(0.8624, device='cuda:0', grad_fn=<NllLossBackward>)
[2792] Train Loss:   tensor(0.8569, device='cuda:0', grad_fn=<NllLossBackward>)
[2793] Train Loss:   tensor(0.8714, device='cuda:0', grad_fn=<NllLossBackward>)
[2794] Train Loss:   tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[2795] Train Loss:   tensor(0.8519, device='cuda:0', grad_fn=<NllLossBackward>)
[2796] Train Loss:   tensor(0.9273, device='cuda:0', grad_fn=<NllLossBackward>)
[2797] Train Loss:   tensor(0.9159, device='cuda:0', grad_fn=<NllLossBackward>)
[2798] Train Loss:   tensor(0.9079, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2799] Train Loss:    tensor(0.9242, device='cuda:0', grad_fn=<NllLossBackward>)
[2800] Train Loss:    tensor(0.8825, device='cuda:0', grad_fn=<NllLossBackward>)
[2801] Train Loss:    tensor(0.8620, device='cuda:0', grad_fn=<NllLossBackward>)
[2802] Train Loss:    tensor(0.8923, device='cuda:0', grad_fn=<NllLossBackward>)
[2803] Train Loss:    tensor(0.8697, device='cuda:0', grad_fn=<NllLossBackward>)
[2804] Train Loss:    tensor(0.8574, device='cuda:0', grad_fn=<NllLossBackward>)
[2805] Train Loss:    tensor(0.8799, device='cuda:0', grad_fn=<NllLossBackward>)
[2806] Train Loss:    tensor(0.8928, device='cuda:0', grad_fn=<NllLossBackward>)
[2807] Train Loss:    tensor(0.8795, device='cuda:0', grad_fn=<NllLossBackward>)
[2808] Train Loss:    tensor(0.8847, device='cuda:0', grad_fn=<NllLossBackward>)
[2809] Train Loss:    tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
[2810] Train Loss:    tensor(0.8733, device='cuda:0', grad_fn=<NllLossBackward>)
[2811] Train Loss:    tensor(0.8918, device='cuda:0', grad_fn=<NllLossBackward>)
[2812] Train Loss:    tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[2813] Train Loss:    tensor(0.8931, device='cuda:0', grad_fn=<NllLossBackward>)
[2814] Train Loss:    tensor(0.8824, device='cuda:0', grad_fn=<NllLossBackward>)
[2815] Train Loss:    tensor(0.9050, device='cuda:0', grad_fn=<NllLossBackward>)
[2816] Train Loss:    tensor(0.8926, device='cuda:0', grad_fn=<NllLossBackward>)
[2817] Train Loss:    tensor(0.8710, device='cuda:0', grad_fn=<NllLossBackward>)
[2818] Train Loss:    tensor(0.8757, device='cuda:0', grad_fn=<NllLossBackward>)
[2819] Train Loss:    tensor(0.8800, device='cuda:0', grad_fn=<NllLossBackward>)
[2820] Train Loss:    tensor(0.8791, device='cuda:0', grad_fn=<NllLossBackward>)
[2821] Train Loss:    tensor(0.8859, device='cuda:0', grad_fn=<NllLossBackward>)
[2822] Train Loss:    tensor(0.8851, device='cuda:0', grad_fn=<NllLossBackward>)
[2823] Train Loss:    tensor(0.8466, device='cuda:0', grad_fn=<NllLossBackward>)
[2824] Train Loss:    tensor(0.8466, device='cuda:0', grad_fn=<NllLossBackward>)
[2825] Train Loss:    tensor(0.9142, device='cuda:0', grad_fn=<NllLossBackward>)
[2826] Train Loss:    tensor(0.8859, device='cuda:0', grad_fn=<NllLossBackward>)
[2827] Train Loss:    tensor(0.9062, device='cuda:0', grad_fn=<NllLossBackward>)
[2828] Train Loss:    tensor(0.8864, device='cuda:0', grad_fn=<NllLossBackward>)
[2829] Train Loss:    tensor(0.8807, device='cuda:0', grad_fn=<NllLossBackward>)
[2830] Train Loss:    tensor(0.8855, device='cuda:0', grad_fn=<NllLossBackward>)
[2831] Train Loss:    tensor(0.8629, device='cuda:0', grad_fn=<NllLossBackward>)
[2832] Train Loss:    tensor(0.8821, device='cuda:0', grad_fn=<NllLossBackward>)
[2833] Train Loss:    tensor(0.8722, device='cuda:0', grad_fn=<NllLossBackward>)
[2834] Train Loss:    tensor(0.8751, device='cuda:0', grad_fn=<NllLossBackward>)
[2835] Train Loss:    tensor(0.8802, device='cuda:0', grad_fn=<NllLossBackward>)
[2836] Train Loss:    tensor(0.8699, device='cuda:0', grad_fn=<NllLossBackward>)
[2837] Train Loss:    tensor(0.8854, device='cuda:0', grad_fn=<NllLossBackward>)
[2838] Train Loss:    tensor(0.8839, device='cuda:0', grad_fn=<NllLossBackward>)
[2839] Train Loss:    tensor(0.9100, device='cuda:0', grad_fn=<NllLossBackward>)
[2840] Train Loss:    tensor(0.9074, device='cuda:0', grad_fn=<NllLossBackward>)
[2841] Train Loss:    tensor(0.9041, device='cuda:0', grad_fn=<NllLossBackward>)
[2842] Train Loss:    tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[2843] Train Loss:    tensor(0.8789, device='cuda:0', grad_fn=<NllLossBackward>)
[2844] Train Loss:    tensor(0.8428, device='cuda:0', grad_fn=<NllLossBackward>)
[2845] Train Loss:    tensor(0.9063, device='cuda:0', grad_fn=<NllLossBackward>)
[2846] Train Loss:    tensor(0.8979, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2847] Train Loss:    tensor(0.8804, device='cuda:0', grad_fn=<NllLossBackward>)
[2848] Train Loss:    tensor(0.9064, device='cuda:0', grad_fn=<NllLossBackward>)
[2849] Train Loss:    tensor(0.8543, device='cuda:0', grad_fn=<NllLossBackward>)
[2850] Train Loss:    tensor(0.8685, device='cuda:0', grad_fn=<NllLossBackward>)
[2851] Train Loss:    tensor(0.9197, device='cuda:0', grad_fn=<NllLossBackward>)
[2852] Train Loss:    tensor(0.9159, device='cuda:0', grad_fn=<NllLossBackward>)
[2853] Train Loss:    tensor(0.9286, device='cuda:0', grad_fn=<NllLossBackward>)
[2854] Train Loss:    tensor(0.8550, device='cuda:0', grad_fn=<NllLossBackward>)
[2855] Train Loss:    tensor(0.9141, device='cuda:0', grad_fn=<NllLossBackward>)
[2856] Train Loss:    tensor(0.8690, device='cuda:0', grad_fn=<NllLossBackward>)
[2857] Train Loss:    tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[2858] Train Loss:    tensor(0.8887, device='cuda:0', grad_fn=<NllLossBackward>)
[2859] Train Loss:    tensor(0.8619, device='cuda:0', grad_fn=<NllLossBackward>)
[2860] Train Loss:    tensor(0.8635, device='cuda:0', grad_fn=<NllLossBackward>)
[2861] Train Loss:    tensor(0.9040, device='cuda:0', grad_fn=<NllLossBackward>)
[2862] Train Loss:    tensor(0.8361, device='cuda:0', grad_fn=<NllLossBackward>)
[2863] Train Loss:    tensor(0.9040, device='cuda:0', grad_fn=<NllLossBackward>)
[2864] Train Loss:    tensor(0.8813, device='cuda:0', grad_fn=<NllLossBackward>)
[2865] Train Loss:    tensor(0.9187, device='cuda:0', grad_fn=<NllLossBackward>)
[2866] Train Loss:    tensor(0.8760, device='cuda:0', grad_fn=<NllLossBackward>)
[2867] Train Loss:    tensor(0.8769, device='cuda:0', grad_fn=<NllLossBackward>)
[2868] Train Loss:    tensor(0.8890, device='cuda:0', grad_fn=<NllLossBackward>)
[2869] Train Loss:    tensor(0.8599, device='cuda:0', grad_fn=<NllLossBackward>)
[2870] Train Loss:    tensor(0.8496, device='cuda:0', grad_fn=<NllLossBackward>)
[2871] Train Loss:    tensor(0.8577, device='cuda:0', grad_fn=<NllLossBackward>)
[2872] Train Loss:    tensor(0.8831, device='cuda:0', grad_fn=<NllLossBackward>)
[2873] Train Loss:    tensor(0.8660, device='cuda:0', grad_fn=<NllLossBackward>)
[2874] Train Loss:    tensor(0.8961, device='cuda:0', grad_fn=<NllLossBackward>)
[2875] Train Loss:    tensor(0.9020, device='cuda:0', grad_fn=<NllLossBackward>)
[2876] Train Loss:    tensor(0.8897, device='cuda:0', grad_fn=<NllLossBackward>)
[2877] Train Loss:    tensor(0.9036, device='cuda:0', grad_fn=<NllLossBackward>)
[2878] Train Loss:    tensor(0.8554, device='cuda:0', grad_fn=<NllLossBackward>)
[2879] Train Loss:    tensor(0.8822, device='cuda:0', grad_fn=<NllLossBackward>)
[2880] Train Loss:    tensor(0.8736, device='cuda:0', grad_fn=<NllLossBackward>)
[2881] Train Loss:    tensor(0.9040, device='cuda:0', grad_fn=<NllLossBackward>)
[2882] Train Loss:    tensor(0.8511, device='cuda:0', grad_fn=<NllLossBackward>)
[2883] Train Loss:    tensor(0.8841, device='cuda:0', grad_fn=<NllLossBackward>)
[2884] Train Loss:    tensor(0.8806, device='cuda:0', grad_fn=<NllLossBackward>)
[2885] Train Loss:    tensor(0.8665, device='cuda:0', grad_fn=<NllLossBackward>)
[2886] Train Loss:    tensor(0.8786, device='cuda:0', grad_fn=<NllLossBackward>)
[2887] Train Loss:    tensor(0.8648, device='cuda:0', grad_fn=<NllLossBackward>)
[2888] Train Loss:    tensor(0.8846, device='cuda:0', grad_fn=<NllLossBackward>)
[2889] Train Loss:    tensor(0.8303, device='cuda:0', grad_fn=<NllLossBackward>)
[2890] Train Loss:    tensor(0.9120, device='cuda:0', grad_fn=<NllLossBackward>)
[2891] Train Loss:    tensor(0.8655, device='cuda:0', grad_fn=<NllLossBackward>)
[2892] Train Loss:    tensor(0.8973, device='cuda:0', grad_fn=<NllLossBackward>)
[2893] Train Loss:    tensor(0.9128, device='cuda:0', grad_fn=<NllLossBackward>)
[2894] Train Loss:    tensor(0.8594, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2895] Train Loss:    tensor(0.8785, device='cuda:0', grad_fn=<NllLossBackward>)
[2896] Train Loss:    tensor(0.8896, device='cuda:0', grad_fn=<NllLossBackward>)
[2897] Train Loss:    tensor(0.8944, device='cuda:0', grad_fn=<NllLossBackward>)
[2898] Train Loss:    tensor(0.8696, device='cuda:0', grad_fn=<NllLossBackward>)
[2899] Train Loss:    tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[2900] Train Loss:    tensor(0.8712, device='cuda:0', grad_fn=<NllLossBackward>)
[2901] Train Loss:    tensor(0.8290, device='cuda:0', grad_fn=<NllLossBackward>)
[2902] Train Loss:    tensor(0.8838, device='cuda:0', grad_fn=<NllLossBackward>)
[2903] Train Loss:    tensor(0.9194, device='cuda:0', grad_fn=<NllLossBackward>)
[2904] Train Loss:    tensor(0.9072, device='cuda:0', grad_fn=<NllLossBackward>)
[2905] Train Loss:    tensor(0.8812, device='cuda:0', grad_fn=<NllLossBackward>)
[2906] Train Loss:    tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[2907] Train Loss:    tensor(0.9114, device='cuda:0', grad_fn=<NllLossBackward>)
[2908] Train Loss:    tensor(0.8607, device='cuda:0', grad_fn=<NllLossBackward>)
[2909] Train Loss:    tensor(0.9285, device='cuda:0', grad_fn=<NllLossBackward>)
[2910] Train Loss:    tensor(0.8956, device='cuda:0', grad_fn=<NllLossBackward>)
[2911] Train Loss:    tensor(0.8881, device='cuda:0', grad_fn=<NllLossBackward>)
[2912] Train Loss:    tensor(0.8782, device='cuda:0', grad_fn=<NllLossBackward>)
[2913] Train Loss:    tensor(0.9041, device='cuda:0', grad_fn=<NllLossBackward>)
[2914] Train Loss:    tensor(0.8520, device='cuda:0', grad_fn=<NllLossBackward>)
[2915] Train Loss:    tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[2916] Train Loss:    tensor(0.8795, device='cuda:0', grad_fn=<NllLossBackward>)
[2917] Train Loss:    tensor(0.8300, device='cuda:0', grad_fn=<NllLossBackward>)
[2918] Train Loss:    tensor(0.8816, device='cuda:0', grad_fn=<NllLossBackward>)
[2919] Train Loss:    tensor(0.8889, device='cuda:0', grad_fn=<NllLossBackward>)
[2920] Train Loss:    tensor(0.8817, device='cuda:0', grad_fn=<NllLossBackward>)
[2921] Train Loss:    tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
[2922] Train Loss:    tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[2923] Train Loss:    tensor(0.8818, device='cuda:0', grad_fn=<NllLossBackward>)
[2924] Train Loss:    tensor(0.8810, device='cuda:0', grad_fn=<NllLossBackward>)
[2925] Train Loss:    tensor(0.8901, device='cuda:0', grad_fn=<NllLossBackward>)
[2926] Train Loss:    tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[2927] Train Loss:    tensor(0.8917, device='cuda:0', grad_fn=<NllLossBackward>)
[2928] Train Loss:    tensor(0.8700, device='cuda:0', grad_fn=<NllLossBackward>)
[2929] Train Loss:    tensor(0.8881, device='cuda:0', grad_fn=<NllLossBackward>)
[2930] Train Loss:    tensor(0.8820, device='cuda:0', grad_fn=<NllLossBackward>)
[2931] Train Loss:    tensor(0.8963, device='cuda:0', grad_fn=<NllLossBackward>)
[2932] Train Loss:    tensor(0.8717, device='cuda:0', grad_fn=<NllLossBackward>)
[2933] Train Loss:    tensor(0.9027, device='cuda:0', grad_fn=<NllLossBackward>)
[2934] Train Loss:    tensor(0.8530, device='cuda:0', grad_fn=<NllLossBackward>)
[2935] Train Loss:    tensor(0.9084, device='cuda:0', grad_fn=<NllLossBackward>)
[2936] Train Loss:    tensor(0.8598, device='cuda:0', grad_fn=<NllLossBackward>)
[2937] Train Loss:    tensor(0.9367, device='cuda:0', grad_fn=<NllLossBackward>)
[2938] Train Loss:    tensor(0.8825, device='cuda:0', grad_fn=<NllLossBackward>)
[2939] Train Loss:    tensor(0.8721, device='cuda:0', grad_fn=<NllLossBackward>)
[2940] Train Loss:    tensor(0.8690, device='cuda:0', grad_fn=<NllLossBackward>)
[2941] Train Loss:    tensor(0.9086, device='cuda:0', grad_fn=<NllLossBackward>)
[2942] Train Loss:    tensor(0.8832, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2943] Train Loss:    tensor(0.8684, device='cuda:0', grad_fn=<NllLossBackward>)
[2944] Train Loss:    tensor(0.8945, device='cuda:0', grad_fn=<NllLossBackward>)
[2945] Train Loss:    tensor(0.8966, device='cuda:0', grad_fn=<NllLossBackward>)
[2946] Train Loss:    tensor(0.8926, device='cuda:0', grad_fn=<NllLossBackward>)
[2947] Train Loss:    tensor(0.8745, device='cuda:0', grad_fn=<NllLossBackward>)
[2948] Train Loss:    tensor(0.8903, device='cuda:0', grad_fn=<NllLossBackward>)
[2949] Train Loss:    tensor(0.8451, device='cuda:0', grad_fn=<NllLossBackward>)
[2950] Train Loss:    tensor(0.8845, device='cuda:0', grad_fn=<NllLossBackward>)
[2951] Train Loss:    tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[2952] Train Loss:    tensor(0.8987, device='cuda:0', grad_fn=<NllLossBackward>)
[2953] Train Loss:    tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
[2954] Train Loss:    tensor(0.9199, device='cuda:0', grad_fn=<NllLossBackward>)
[2955] Train Loss:    tensor(0.8535, device='cuda:0', grad_fn=<NllLossBackward>)
[2956] Train Loss:    tensor(0.8877, device='cuda:0', grad_fn=<NllLossBackward>)
[2957] Train Loss:    tensor(0.9021, device='cuda:0', grad_fn=<NllLossBackward>)
[2958] Train Loss:    tensor(0.8590, device='cuda:0', grad_fn=<NllLossBackward>)
[2959] Train Loss:    tensor(0.9116, device='cuda:0', grad_fn=<NllLossBackward>)
[2960] Train Loss:    tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[2961] Train Loss:    tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[2962] Train Loss:    tensor(0.8937, device='cuda:0', grad_fn=<NllLossBackward>)
[2963] Train Loss:    tensor(0.8852, device='cuda:0', grad_fn=<NllLossBackward>)
[2964] Train Loss:    tensor(0.8645, device='cuda:0', grad_fn=<NllLossBackward>)
[2965] Train Loss:    tensor(0.9052, device='cuda:0', grad_fn=<NllLossBackward>)
[2966] Train Loss:    tensor(0.8700, device='cuda:0', grad_fn=<NllLossBackward>)
[2967] Train Loss:    tensor(0.9045, device='cuda:0', grad_fn=<NllLossBackward>)
[2968] Train Loss:    tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[2969] Train Loss:    tensor(0.8960, device='cuda:0', grad_fn=<NllLossBackward>)
[2970] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[2971] Train Loss:    tensor(0.8880, device='cuda:0', grad_fn=<NllLossBackward>)
[2972] Train Loss:    tensor(0.8631, device='cuda:0', grad_fn=<NllLossBackward>)
[2973] Train Loss:    tensor(0.8681, device='cuda:0', grad_fn=<NllLossBackward>)
[2974] Train Loss:    tensor(0.8630, device='cuda:0', grad_fn=<NllLossBackward>)
[2975] Train Loss:    tensor(0.8741, device='cuda:0', grad_fn=<NllLossBackward>)
[2976] Train Loss:    tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[2977] Train Loss:    tensor(0.8907, device='cuda:0', grad_fn=<NllLossBackward>)
[2978] Train Loss:    tensor(0.8672, device='cuda:0', grad_fn=<NllLossBackward>)
[2979] Train Loss:    tensor(0.8640, device='cuda:0', grad_fn=<NllLossBackward>)
[2980] Train Loss:    tensor(0.8889, device='cuda:0', grad_fn=<NllLossBackward>)
[2981] Train Loss:    tensor(0.8645, device='cuda:0', grad_fn=<NllLossBackward>)
[2982] Train Loss:    tensor(0.8549, device='cuda:0', grad_fn=<NllLossBackward>)
[2983] Train Loss:    tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[2984] Train Loss:    tensor(0.8738, device='cuda:0', grad_fn=<NllLossBackward>)
[2985] Train Loss:    tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[2986] Train Loss:    tensor(0.8904, device='cuda:0', grad_fn=<NllLossBackward>)
[2987] Train Loss:    tensor(0.8540, device='cuda:0', grad_fn=<NllLossBackward>)
[2988] Train Loss:    tensor(0.8707, device='cuda:0', grad_fn=<NllLossBackward>)
[2989] Train Loss:    tensor(0.8788, device='cuda:0', grad_fn=<NllLossBackward>)
[2990] Train Loss:    tensor(0.8910, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2991] Train Loss:    tensor(0.8765, device='cuda:0', grad_fn=<NllLossBackward>)
[2992] Train Loss:    tensor(0.8736, device='cuda:0', grad_fn=<NllLossBackward>)
[2993] Train Loss:    tensor(0.8880, device='cuda:0', grad_fn=<NllLossBackward>)
[2994] Train Loss:    tensor(0.8828, device='cuda:0', grad_fn=<NllLossBackward>)
[2995] Train Loss:    tensor(0.8633, device='cuda:0', grad_fn=<NllLossBackward>)
[2996] Train Loss:    tensor(0.8565, device='cuda:0', grad_fn=<NllLossBackward>)
[2997] Train Loss:    tensor(0.9032, device='cuda:0', grad_fn=<NllLossBackward>)
[2998] Train Loss:    tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[2999] Train Loss:    tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[3000] Train Loss:    tensor(0.8847, device='cuda:0', grad_fn=<NllLossBackward>)
[3001] Train Loss:    tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[3002] Train Loss:    tensor(0.8881, device='cuda:0', grad_fn=<NllLossBackward>)
[3003] Train Loss:    tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[3004] Train Loss:    tensor(0.8866, device='cuda:0', grad_fn=<NllLossBackward>)
[3005] Train Loss:    tensor(0.8765, device='cuda:0', grad_fn=<NllLossBackward>)
[3006] Train Loss:    tensor(0.8859, device='cuda:0', grad_fn=<NllLossBackward>)
[3007] Train Loss:    tensor(0.8816, device='cuda:0', grad_fn=<NllLossBackward>)
[3008] Train Loss:    tensor(0.8893, device='cuda:0', grad_fn=<NllLossBackward>)
[3009] Train Loss:    tensor(0.8468, device='cuda:0', grad_fn=<NllLossBackward>)
[3010] Train Loss:    tensor(0.8699, device='cuda:0', grad_fn=<NllLossBackward>)
[3011] Train Loss:    tensor(0.8895, device='cuda:0', grad_fn=<NllLossBackward>)
[3012] Train Loss:    tensor(0.8539, device='cuda:0', grad_fn=<NllLossBackward>)
[3013] Train Loss:    tensor(0.8630, device='cuda:0', grad_fn=<NllLossBackward>)
[3014] Train Loss:    tensor(0.8657, device='cuda:0', grad_fn=<NllLossBackward>)
[3015] Train Loss:    tensor(0.8832, device='cuda:0', grad_fn=<NllLossBackward>)
[3016] Train Loss:    tensor(0.9037, device='cuda:0', grad_fn=<NllLossBackward>)
[3017] Train Loss:    tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[3018] Train Loss:    tensor(0.8841, device='cuda:0', grad_fn=<NllLossBackward>)
[3019] Train Loss:    tensor(0.8993, device='cuda:0', grad_fn=<NllLossBackward>)
[3020] Train Loss:    tensor(0.9083, device='cuda:0', grad_fn=<NllLossBackward>)
[3021] Train Loss:    tensor(0.8514, device='cuda:0', grad_fn=<NllLossBackward>)
[3022] Train Loss:    tensor(0.8683, device='cuda:0', grad_fn=<NllLossBackward>)
[3023] Train Loss:    tensor(0.8683, device='cuda:0', grad_fn=<NllLossBackward>)
[3024] Train Loss:    tensor(0.8545, device='cuda:0', grad_fn=<NllLossBackward>)
[3025] Train Loss:    tensor(0.8846, device='cuda:0', grad_fn=<NllLossBackward>)
[3026] Train Loss:    tensor(0.8612, device='cuda:0', grad_fn=<NllLossBackward>)
[3027] Train Loss:    tensor(0.8597, device='cuda:0', grad_fn=<NllLossBackward>)
[3028] Train Loss:    tensor(0.8573, device='cuda:0', grad_fn=<NllLossBackward>)
[3029] Train Loss:    tensor(0.8975, device='cuda:0', grad_fn=<NllLossBackward>)
[3030] Train Loss:    tensor(0.9018, device='cuda:0', grad_fn=<NllLossBackward>)
[3031] Train Loss:    tensor(0.9063, device='cuda:0', grad_fn=<NllLossBackward>)
[3032] Train Loss:    tensor(0.8782, device='cuda:0', grad_fn=<NllLossBackward>)
[3033] Train Loss:    tensor(0.8822, device='cuda:0', grad_fn=<NllLossBackward>)
[3034] Train Loss:    tensor(0.8555, device='cuda:0', grad_fn=<NllLossBackward>)
[3035] Train Loss:    tensor(0.8702, device='cuda:0', grad_fn=<NllLossBackward>)
[3036] Train Loss:    tensor(0.8691, device='cuda:0', grad_fn=<NllLossBackward>)
[3037] Train Loss:    tensor(0.8961, device='cuda:0', grad_fn=<NllLossBackward>)
[3038] Train Loss:    tensor(0.9062, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3039] Train Loss:   tensor(0.9174, device='cuda:0', grad_fn=<NllLossBackward>)
[3040] Train Loss:   tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[3041] Train Loss:   tensor(0.8889, device='cuda:0', grad_fn=<NllLossBackward>)
[3042] Train Loss:   tensor(0.8839, device='cuda:0', grad_fn=<NllLossBackward>)
[3043] Train Loss:   tensor(0.9013, device='cuda:0', grad_fn=<NllLossBackward>)
[3044] Train Loss:   tensor(0.8748, device='cuda:0', grad_fn=<NllLossBackward>)
[3045] Train Loss:   tensor(0.9012, device='cuda:0', grad_fn=<NllLossBackward>)
[3046] Train Loss:   tensor(0.8929, device='cuda:0', grad_fn=<NllLossBackward>)
[3047] Train Loss:   tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[3048] Train Loss:   tensor(0.8993, device='cuda:0', grad_fn=<NllLossBackward>)
[3049] Train Loss:   tensor(0.8552, device='cuda:0', grad_fn=<NllLossBackward>)
[3050] Train Loss:   tensor(0.8689, device='cuda:0', grad_fn=<NllLossBackward>)
[3051] Train Loss:   tensor(0.8776, device='cuda:0', grad_fn=<NllLossBackward>)
[3052] Train Loss:   tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[3053] Train Loss:   tensor(0.8841, device='cuda:0', grad_fn=<NllLossBackward>)
[3054] Train Loss:   tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[3055] Train Loss:   tensor(0.8494, device='cuda:0', grad_fn=<NllLossBackward>)
[3056] Train Loss:   tensor(0.9172, device='cuda:0', grad_fn=<NllLossBackward>)
[3057] Train Loss:   tensor(0.8775, device='cuda:0', grad_fn=<NllLossBackward>)
[3058] Train Loss:   tensor(0.8931, device='cuda:0', grad_fn=<NllLossBackward>)
[3059] Train Loss:   tensor(0.8800, device='cuda:0', grad_fn=<NllLossBackward>)
[3060] Train Loss:   tensor(0.8879, device='cuda:0', grad_fn=<NllLossBackward>)
[3061] Train Loss:   tensor(0.8830, device='cuda:0', grad_fn=<NllLossBackward>)
[3062] Train Loss:   tensor(0.8843, device='cuda:0', grad_fn=<NllLossBackward>)
[3063] Train Loss:   tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[3064] Train Loss:   tensor(0.8830, device='cuda:0', grad_fn=<NllLossBackward>)
[3065] Train Loss:   tensor(0.8818, device='cuda:0', grad_fn=<NllLossBackward>)
[3066] Train Loss:   tensor(0.8944, device='cuda:0', grad_fn=<NllLossBackward>)
[3067] Train Loss:   tensor(0.8831, device='cuda:0', grad_fn=<NllLossBackward>)
[3068] Train Loss:   tensor(0.8938, device='cuda:0', grad_fn=<NllLossBackward>)
[3069] Train Loss:   tensor(0.8704, device='cuda:0', grad_fn=<NllLossBackward>)
[3070] Train Loss:   tensor(0.9088, device='cuda:0', grad_fn=<NllLossBackward>)
[3071] Train Loss:   tensor(0.8925, device='cuda:0', grad_fn=<NllLossBackward>)
[3072] Train Loss:   tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[3073] Train Loss:   tensor(0.8925, device='cuda:0', grad_fn=<NllLossBackward>)
[3074] Train Loss:   tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[3075] Train Loss:   tensor(0.9056, device='cuda:0', grad_fn=<NllLossBackward>)
[3076] Train Loss:   tensor(0.8866, device='cuda:0', grad_fn=<NllLossBackward>)
[3077] Train Loss:   tensor(0.8914, device='cuda:0', grad_fn=<NllLossBackward>)
[3078] Train Loss:   tensor(0.8439, device='cuda:0', grad_fn=<NllLossBackward>)
[3079] Train Loss:   tensor(0.8845, device='cuda:0', grad_fn=<NllLossBackward>)
[3080] Train Loss:   tensor(0.8808, device='cuda:0', grad_fn=<NllLossBackward>)
[3081] Train Loss:   tensor(0.8598, device='cuda:0', grad_fn=<NllLossBackward>)
[3082] Train Loss:   tensor(0.8988, device='cuda:0', grad_fn=<NllLossBackward>)
[3083] Train Loss:   tensor(0.8738, device='cuda:0', grad_fn=<NllLossBackward>)
[3084] Train Loss:   tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[3085] Train Loss:   tensor(0.9147, device='cuda:0', grad_fn=<NllLossBackward>)
[3086] Train Loss:   tensor(0.8641, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3087] Train Loss:    tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[3088] Train Loss:    tensor(0.8611, device='cuda:0', grad_fn=<NllLossBackward>)
[3089] Train Loss:    tensor(0.8446, device='cuda:0', grad_fn=<NllLossBackward>)
[3090] Train Loss:    tensor(0.8876, device='cuda:0', grad_fn=<NllLossBackward>)
[3091] Train Loss:    tensor(0.8551, device='cuda:0', grad_fn=<NllLossBackward>)
[3092] Train Loss:    tensor(0.8568, device='cuda:0', grad_fn=<NllLossBackward>)
[3093] Train Loss:    tensor(0.9120, device='cuda:0', grad_fn=<NllLossBackward>)
[3094] Train Loss:    tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
[3095] Train Loss:    tensor(0.8753, device='cuda:0', grad_fn=<NllLossBackward>)
[3096] Train Loss:    tensor(0.8836, device='cuda:0', grad_fn=<NllLossBackward>)
[3097] Train Loss:    tensor(0.8900, device='cuda:0', grad_fn=<NllLossBackward>)
[3098] Train Loss:    tensor(0.8652, device='cuda:0', grad_fn=<NllLossBackward>)
[3099] Train Loss:    tensor(0.8890, device='cuda:0', grad_fn=<NllLossBackward>)
[3100] Train Loss:    tensor(0.8611, device='cuda:0', grad_fn=<NllLossBackward>)
[3101] Train Loss:    tensor(0.8790, device='cuda:0', grad_fn=<NllLossBackward>)
[3102] Train Loss:    tensor(0.8727, device='cuda:0', grad_fn=<NllLossBackward>)
[3103] Train Loss:    tensor(0.8914, device='cuda:0', grad_fn=<NllLossBackward>)
[3104] Train Loss:    tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[3105] Train Loss:    tensor(0.8972, device='cuda:0', grad_fn=<NllLossBackward>)
[3106] Train Loss:    tensor(0.9086, device='cuda:0', grad_fn=<NllLossBackward>)
[3107] Train Loss:    tensor(0.8831, device='cuda:0', grad_fn=<NllLossBackward>)
[3108] Train Loss:    tensor(0.9028, device='cuda:0', grad_fn=<NllLossBackward>)
[3109] Train Loss:    tensor(0.8569, device='cuda:0', grad_fn=<NllLossBackward>)
[3110] Train Loss:    tensor(0.8756, device='cuda:0', grad_fn=<NllLossBackward>)
[3111] Train Loss:    tensor(0.9018, device='cuda:0', grad_fn=<NllLossBackward>)
[3112] Train Loss:    tensor(0.8601, device='cuda:0', grad_fn=<NllLossBackward>)
[3113] Train Loss:    tensor(0.8700, device='cuda:0', grad_fn=<NllLossBackward>)
[3114] Train Loss:    tensor(0.8605, device='cuda:0', grad_fn=<NllLossBackward>)
[3115] Train Loss:    tensor(0.8816, device='cuda:0', grad_fn=<NllLossBackward>)
[3116] Train Loss:    tensor(0.8886, device='cuda:0', grad_fn=<NllLossBackward>)
[3117] Train Loss:    tensor(0.8824, device='cuda:0', grad_fn=<NllLossBackward>)
[3118] Train Loss:    tensor(0.8739, device='cuda:0', grad_fn=<NllLossBackward>)
[3119] Train Loss:    tensor(0.9144, device='cuda:0', grad_fn=<NllLossBackward>)
[3120] Train Loss:    tensor(0.9214, device='cuda:0', grad_fn=<NllLossBackward>)
[3121] Train Loss:    tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[3122] Train Loss:    tensor(0.8860, device='cuda:0', grad_fn=<NllLossBackward>)
[3123] Train Loss:    tensor(0.8862, device='cuda:0', grad_fn=<NllLossBackward>)
[3124] Train Loss:    tensor(0.8530, device='cuda:0', grad_fn=<NllLossBackward>)
[3125] Train Loss:    tensor(0.8794, device='cuda:0', grad_fn=<NllLossBackward>)
[3126] Train Loss:    tensor(0.8290, device='cuda:0', grad_fn=<NllLossBackward>)
[3127] Train Loss:    tensor(0.8900, device='cuda:0', grad_fn=<NllLossBackward>)
[3128] Train Loss:    tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[3129] Train Loss:    tensor(0.9146, device='cuda:0', grad_fn=<NllLossBackward>)
[3130] Train Loss:    tensor(0.8865, device='cuda:0', grad_fn=<NllLossBackward>)
[3131] Train Loss:    tensor(0.8957, device='cuda:0', grad_fn=<NllLossBackward>)
[3132] Train Loss:    tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[3133] Train Loss:    tensor(0.8955, device='cuda:0', grad_fn=<NllLossBackward>)
[3134] Train Loss:    tensor(0.8691, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3135] Train Loss:    tensor(0.8884, device='cuda:0', grad_fn=<NllLossBackward>)
[3136] Train Loss:    tensor(0.8718, device='cuda:0', grad_fn=<NllLossBackward>)
[3137] Train Loss:    tensor(0.8933, device='cuda:0', grad_fn=<NllLossBackward>)
[3138] Train Loss:    tensor(0.8756, device='cuda:0', grad_fn=<NllLossBackward>)
[3139] Train Loss:    tensor(0.8955, device='cuda:0', grad_fn=<NllLossBackward>)
[3140] Train Loss:    tensor(0.8518, device='cuda:0', grad_fn=<NllLossBackward>)
[3141] Train Loss:    tensor(0.9021, device='cuda:0', grad_fn=<NllLossBackward>)
[3142] Train Loss:    tensor(0.8570, device='cuda:0', grad_fn=<NllLossBackward>)
[3143] Train Loss:    tensor(0.8712, device='cuda:0', grad_fn=<NllLossBackward>)
[3144] Train Loss:    tensor(0.8547, device='cuda:0', grad_fn=<NllLossBackward>)
[3145] Train Loss:    tensor(0.8752, device='cuda:0', grad_fn=<NllLossBackward>)
[3146] Train Loss:    tensor(0.8846, device='cuda:0', grad_fn=<NllLossBackward>)
[3147] Train Loss:    tensor(0.9077, device='cuda:0', grad_fn=<NllLossBackward>)
[3148] Train Loss:    tensor(0.8959, device='cuda:0', grad_fn=<NllLossBackward>)
[3149] Train Loss:    tensor(0.8504, device='cuda:0', grad_fn=<NllLossBackward>)
[3150] Train Loss:    tensor(0.8685, device='cuda:0', grad_fn=<NllLossBackward>)
[3151] Train Loss:    tensor(0.8713, device='cuda:0', grad_fn=<NllLossBackward>)
[3152] Train Loss:    tensor(0.8821, device='cuda:0', grad_fn=<NllLossBackward>)
[3153] Train Loss:    tensor(0.9049, device='cuda:0', grad_fn=<NllLossBackward>)
[3154] Train Loss:    tensor(0.9044, device='cuda:0', grad_fn=<NllLossBackward>)
[3155] Train Loss:    tensor(0.8983, device='cuda:0', grad_fn=<NllLossBackward>)
[3156] Train Loss:    tensor(0.8786, device='cuda:0', grad_fn=<NllLossBackward>)
[3157] Train Loss:    tensor(0.9099, device='cuda:0', grad_fn=<NllLossBackward>)
[3158] Train Loss:    tensor(0.8626, device='cuda:0', grad_fn=<NllLossBackward>)
[3159] Train Loss:    tensor(0.8954, device='cuda:0', grad_fn=<NllLossBackward>)
[3160] Train Loss:    tensor(0.8707, device='cuda:0', grad_fn=<NllLossBackward>)
[3161] Train Loss:    tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
[3162] Train Loss:    tensor(0.8986, device='cuda:0', grad_fn=<NllLossBackward>)
[3163] Train Loss:    tensor(0.8689, device='cuda:0', grad_fn=<NllLossBackward>)
[3164] Train Loss:    tensor(0.9083, device='cuda:0', grad_fn=<NllLossBackward>)
[3165] Train Loss:    tensor(0.8760, device='cuda:0', grad_fn=<NllLossBackward>)
[3166] Train Loss:    tensor(0.9068, device='cuda:0', grad_fn=<NllLossBackward>)
[3167] Train Loss:    tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[3168] Train Loss:    tensor(0.8630, device='cuda:0', grad_fn=<NllLossBackward>)
[3169] Train Loss:    tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
[3170] Train Loss:    tensor(0.8775, device='cuda:0', grad_fn=<NllLossBackward>)
[3171] Train Loss:    tensor(0.8607, device='cuda:0', grad_fn=<NllLossBackward>)
[3172] Train Loss:    tensor(0.8868, device='cuda:0', grad_fn=<NllLossBackward>)
[3173] Train Loss:    tensor(0.9038, device='cuda:0', grad_fn=<NllLossBackward>)
[3174] Train Loss:    tensor(0.8719, device='cuda:0', grad_fn=<NllLossBackward>)
[3175] Train Loss:    tensor(0.8636, device='cuda:0', grad_fn=<NllLossBackward>)
[3176] Train Loss:    tensor(0.8879, device='cuda:0', grad_fn=<NllLossBackward>)
[3177] Train Loss:    tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[3178] Train Loss:    tensor(0.8839, device='cuda:0', grad_fn=<NllLossBackward>)
[3179] Train Loss:    tensor(0.9273, device='cuda:0', grad_fn=<NllLossBackward>)
[3180] Train Loss:    tensor(0.8712, device='cuda:0', grad_fn=<NllLossBackward>)
[3181] Train Loss:    tensor(0.8794, device='cuda:0', grad_fn=<NllLossBackward>)
[3182] Train Loss:    tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3183] Train Loss:    tensor(0.8514, device='cuda:0', grad_fn=<NllLossBackward>)
[3184] Train Loss:    tensor(0.9313, device='cuda:0', grad_fn=<NllLossBackward>)
[3185] Train Loss:    tensor(0.8733, device='cuda:0', grad_fn=<NllLossBackward>)
[3186] Train Loss:    tensor(0.8685, device='cuda:0', grad_fn=<NllLossBackward>)
[3187] Train Loss:    tensor(0.8677, device='cuda:0', grad_fn=<NllLossBackward>)
[3188] Train Loss:    tensor(0.8783, device='cuda:0', grad_fn=<NllLossBackward>)
[3189] Train Loss:    tensor(0.9030, device='cuda:0', grad_fn=<NllLossBackward>)
[3190] Train Loss:    tensor(0.8723, device='cuda:0', grad_fn=<NllLossBackward>)
[3191] Train Loss:    tensor(0.9378, device='cuda:0', grad_fn=<NllLossBackward>)
[3192] Train Loss:    tensor(0.8454, device='cuda:0', grad_fn=<NllLossBackward>)
[3193] Train Loss:    tensor(0.8758, device='cuda:0', grad_fn=<NllLossBackward>)
[3194] Train Loss:    tensor(0.8693, device='cuda:0', grad_fn=<NllLossBackward>)
[3195] Train Loss:    tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[3196] Train Loss:    tensor(0.8817, device='cuda:0', grad_fn=<NllLossBackward>)
[3197] Train Loss:    tensor(0.8840, device='cuda:0', grad_fn=<NllLossBackward>)
[3198] Train Loss:    tensor(0.8663, device='cuda:0', grad_fn=<NllLossBackward>)
[3199] Train Loss:    tensor(0.8842, device='cuda:0', grad_fn=<NllLossBackward>)
[3200] Train Loss:    tensor(0.8800, device='cuda:0', grad_fn=<NllLossBackward>)
[3201] Train Loss:    tensor(0.9167, device='cuda:0', grad_fn=<NllLossBackward>)
[3202] Train Loss:    tensor(0.9234, device='cuda:0', grad_fn=<NllLossBackward>)
[3203] Train Loss:    tensor(0.8646, device='cuda:0', grad_fn=<NllLossBackward>)
[3204] Train Loss:    tensor(0.8801, device='cuda:0', grad_fn=<NllLossBackward>)
[3205] Train Loss:    tensor(0.9021, device='cuda:0', grad_fn=<NllLossBackward>)
[3206] Train Loss:    tensor(0.8617, device='cuda:0', grad_fn=<NllLossBackward>)
[3207] Train Loss:    tensor(0.8703, device='cuda:0', grad_fn=<NllLossBackward>)
[3208] Train Loss:    tensor(0.9029, device='cuda:0', grad_fn=<NllLossBackward>)
[3209] Train Loss:    tensor(0.8639, device='cuda:0', grad_fn=<NllLossBackward>)
[3210] Train Loss:    tensor(0.8952, device='cuda:0', grad_fn=<NllLossBackward>)
[3211] Train Loss:    tensor(0.8932, device='cuda:0', grad_fn=<NllLossBackward>)
[3212] Train Loss:    tensor(0.9007, device='cuda:0', grad_fn=<NllLossBackward>)
[3213] Train Loss:    tensor(0.8672, device='cuda:0', grad_fn=<NllLossBackward>)
[3214] Train Loss:    tensor(0.8557, device='cuda:0', grad_fn=<NllLossBackward>)
[3215] Train Loss:    tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[3216] Train Loss:    tensor(0.8738, device='cuda:0', grad_fn=<NllLossBackward>)
[3217] Train Loss:    tensor(0.8796, device='cuda:0', grad_fn=<NllLossBackward>)
[3218] Train Loss:    tensor(0.8695, device='cuda:0', grad_fn=<NllLossBackward>)
[3219] Train Loss:    tensor(0.8753, device='cuda:0', grad_fn=<NllLossBackward>)
[3220] Train Loss:    tensor(0.8822, device='cuda:0', grad_fn=<NllLossBackward>)
[3221] Train Loss:    tensor(0.8784, device='cuda:0', grad_fn=<NllLossBackward>)
[3222] Train Loss:    tensor(0.8831, device='cuda:0', grad_fn=<NllLossBackward>)
[3223] Train Loss:    tensor(0.8739, device='cuda:0', grad_fn=<NllLossBackward>)
[3224] Train Loss:    tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[3225] Train Loss:    tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[3226] Train Loss:    tensor(0.8707, device='cuda:0', grad_fn=<NllLossBackward>)
[3227] Train Loss:    tensor(0.8847, device='cuda:0', grad_fn=<NllLossBackward>)
[3228] Train Loss:    tensor(0.8710, device='cuda:0', grad_fn=<NllLossBackward>)
[3229] Train Loss:    tensor(0.8756, device='cuda:0', grad_fn=<NllLossBackward>)
[3230] Train Loss:    tensor(0.8644, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3231] Train Loss:    tensor(0.8909, device='cuda:0', grad_fn=<NllLossBackward>)
[3232] Train Loss:    tensor(0.8708, device='cuda:0', grad_fn=<NllLossBackward>)
[3233] Train Loss:    tensor(0.8798, device='cuda:0', grad_fn=<NllLossBackward>)
[3234] Train Loss:    tensor(0.8623, device='cuda:0', grad_fn=<NllLossBackward>)
[3235] Train Loss:    tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[3236] Train Loss:    tensor(0.8898, device='cuda:0', grad_fn=<NllLossBackward>)
[3237] Train Loss:    tensor(0.8659, device='cuda:0', grad_fn=<NllLossBackward>)
[3238] Train Loss:    tensor(0.9069, device='cuda:0', grad_fn=<NllLossBackward>)
[3239] Train Loss:    tensor(0.8640, device='cuda:0', grad_fn=<NllLossBackward>)
[3240] Train Loss:    tensor(0.8985, device='cuda:0', grad_fn=<NllLossBackward>)
[3241] Train Loss:    tensor(0.8930, device='cuda:0', grad_fn=<NllLossBackward>)
[3242] Train Loss:    tensor(0.8823, device='cuda:0', grad_fn=<NllLossBackward>)
[3243] Train Loss:    tensor(0.8666, device='cuda:0', grad_fn=<NllLossBackward>)
[3244] Train Loss:    tensor(0.8788, device='cuda:0', grad_fn=<NllLossBackward>)
[3245] Train Loss:    tensor(0.9208, device='cuda:0', grad_fn=<NllLossBackward>)
[3246] Train Loss:    tensor(0.8730, device='cuda:0', grad_fn=<NllLossBackward>)
[3247] Train Loss:    tensor(0.8903, device='cuda:0', grad_fn=<NllLossBackward>)
[3248] Train Loss:    tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
[3249] Train Loss:    tensor(0.8890, device='cuda:0', grad_fn=<NllLossBackward>)
[3250] Train Loss:    tensor(0.8862, device='cuda:0', grad_fn=<NllLossBackward>)
[3251] Train Loss:    tensor(0.8838, device='cuda:0', grad_fn=<NllLossBackward>)
[3252] Train Loss:    tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[3253] Train Loss:    tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[3254] Train Loss:    tensor(0.8857, device='cuda:0', grad_fn=<NllLossBackward>)
[3255] Train Loss:    tensor(0.8962, device='cuda:0', grad_fn=<NllLossBackward>)
[3256] Train Loss:    tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[3257] Train Loss:    tensor(0.8717, device='cuda:0', grad_fn=<NllLossBackward>)
[3258] Train Loss:    tensor(0.8674, device='cuda:0', grad_fn=<NllLossBackward>)
[3259] Train Loss:    tensor(0.8474, device='cuda:0', grad_fn=<NllLossBackward>)
[3260] Train Loss:    tensor(0.8807, device='cuda:0', grad_fn=<NllLossBackward>)
[3261] Train Loss:    tensor(0.8745, device='cuda:0', grad_fn=<NllLossBackward>)
[3262] Train Loss:    tensor(0.8985, device='cuda:0', grad_fn=<NllLossBackward>)
[3263] Train Loss:    tensor(0.8829, device='cuda:0', grad_fn=<NllLossBackward>)
[3264] Train Loss:    tensor(0.9065, device='cuda:0', grad_fn=<NllLossBackward>)
[3265] Train Loss:    tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[3266] Train Loss:    tensor(0.8886, device='cuda:0', grad_fn=<NllLossBackward>)
[3267] Train Loss:    tensor(0.8566, device='cuda:0', grad_fn=<NllLossBackward>)
[3268] Train Loss:    tensor(0.9184, device='cuda:0', grad_fn=<NllLossBackward>)
[3269] Train Loss:    tensor(0.8772, device='cuda:0', grad_fn=<NllLossBackward>)
[3270] Train Loss:    tensor(0.8772, device='cuda:0', grad_fn=<NllLossBackward>)
[3271] Train Loss:    tensor(0.9195, device='cuda:0', grad_fn=<NllLossBackward>)
[3272] Train Loss:    tensor(0.8700, device='cuda:0', grad_fn=<NllLossBackward>)
[3273] Train Loss:    tensor(0.8912, device='cuda:0', grad_fn=<NllLossBackward>)
[3274] Train Loss:    tensor(0.8893, device='cuda:0', grad_fn=<NllLossBackward>)
[3275] Train Loss:    tensor(0.8388, device='cuda:0', grad_fn=<NllLossBackward>)
[3276] Train Loss:    tensor(0.9187, device='cuda:0', grad_fn=<NllLossBackward>)
[3277] Train Loss:    tensor(0.8539, device='cuda:0', grad_fn=<NllLossBackward>)
[3278] Train Loss:    tensor(0.8663, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3279] Train Loss:    tensor(0.8856, device='cuda:0', grad_fn=<NllLossBackward>)
[3280] Train Loss:    tensor(0.8939, device='cuda:0', grad_fn=<NllLossBackward>)
[3281] Train Loss:    tensor(0.8466, device='cuda:0', grad_fn=<NllLossBackward>)
[3282] Train Loss:    tensor(0.8845, device='cuda:0', grad_fn=<NllLossBackward>)
[3283] Train Loss:    tensor(0.8851, device='cuda:0', grad_fn=<NllLossBackward>)
[3284] Train Loss:    tensor(0.9124, device='cuda:0', grad_fn=<NllLossBackward>)
[3285] Train Loss:    tensor(0.8612, device='cuda:0', grad_fn=<NllLossBackward>)
[3286] Train Loss:    tensor(0.8791, device='cuda:0', grad_fn=<NllLossBackward>)
[3287] Train Loss:    tensor(0.9081, device='cuda:0', grad_fn=<NllLossBackward>)
[3288] Train Loss:    tensor(0.8869, device='cuda:0', grad_fn=<NllLossBackward>)
[3289] Train Loss:    tensor(0.8652, device='cuda:0', grad_fn=<NllLossBackward>)
[3290] Train Loss:    tensor(0.8711, device='cuda:0', grad_fn=<NllLossBackward>)
[3291] Train Loss:    tensor(0.9345, device='cuda:0', grad_fn=<NllLossBackward>)
[3292] Train Loss:    tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[3293] Train Loss:    tensor(0.8683, device='cuda:0', grad_fn=<NllLossBackward>)
[3294] Train Loss:    tensor(0.8735, device='cuda:0', grad_fn=<NllLossBackward>)
[3295] Train Loss:    tensor(0.8611, device='cuda:0', grad_fn=<NllLossBackward>)
[3296] Train Loss:    tensor(0.8867, device='cuda:0', grad_fn=<NllLossBackward>)
[3297] Train Loss:    tensor(0.8612, device='cuda:0', grad_fn=<NllLossBackward>)
[3298] Train Loss:    tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[3299] Train Loss:    tensor(0.8685, device='cuda:0', grad_fn=<NllLossBackward>)
[3300] Train Loss:    tensor(0.8906, device='cuda:0', grad_fn=<NllLossBackward>)
[3301] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[3302] Train Loss:    tensor(0.8807, device='cuda:0', grad_fn=<NllLossBackward>)
[3303] Train Loss:    tensor(0.8934, device='cuda:0', grad_fn=<NllLossBackward>)
[3304] Train Loss:    tensor(0.8590, device='cuda:0', grad_fn=<NllLossBackward>)
[3305] Train Loss:    tensor(0.8596, device='cuda:0', grad_fn=<NllLossBackward>)
[3306] Train Loss:    tensor(0.9223, device='cuda:0', grad_fn=<NllLossBackward>)
[3307] Train Loss:    tensor(0.8854, device='cuda:0', grad_fn=<NllLossBackward>)
[3308] Train Loss:    tensor(0.8980, device='cuda:0', grad_fn=<NllLossBackward>)
[3309] Train Loss:    tensor(0.9136, device='cuda:0', grad_fn=<NllLossBackward>)
[3310] Train Loss:    tensor(0.8584, device='cuda:0', grad_fn=<NllLossBackward>)
[3311] Train Loss:    tensor(0.9197, device='cuda:0', grad_fn=<NllLossBackward>)
[3312] Train Loss:    tensor(0.8623, device='cuda:0', grad_fn=<NllLossBackward>)
[3313] Train Loss:    tensor(0.9015, device='cuda:0', grad_fn=<NllLossBackward>)
[3314] Train Loss:    tensor(0.8745, device='cuda:0', grad_fn=<NllLossBackward>)
[3315] Train Loss:    tensor(0.8758, device='cuda:0', grad_fn=<NllLossBackward>)
[3316] Train Loss:    tensor(0.8841, device='cuda:0', grad_fn=<NllLossBackward>)
[3317] Train Loss:    tensor(0.8644, device='cuda:0', grad_fn=<NllLossBackward>)
[3318] Train Loss:    tensor(0.9015, device='cuda:0', grad_fn=<NllLossBackward>)
[3319] Train Loss:    tensor(0.8474, device='cuda:0', grad_fn=<NllLossBackward>)
[3320] Train Loss:    tensor(0.8612, device='cuda:0', grad_fn=<NllLossBackward>)
[3321] Train Loss:    tensor(0.8843, device='cuda:0', grad_fn=<NllLossBackward>)
[3322] Train Loss:    tensor(0.8544, device='cuda:0', grad_fn=<NllLossBackward>)
[3323] Train Loss:    tensor(0.8782, device='cuda:0', grad_fn=<NllLossBackward>)
[3324] Train Loss:    tensor(0.9083, device='cuda:0', grad_fn=<NllLossBackward>)
[3325] Train Loss:    tensor(0.8449, device='cuda:0', grad_fn=<NllLossBackward>)
[3326] Train Loss:    tensor(0.8587, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3327] Train Loss:    tensor(0.8734, device='cuda:0', grad_fn=<NllLossBackward>)
[3328] Train Loss:    tensor(0.8809, device='cuda:0', grad_fn=<NllLossBackward>)
[3329] Train Loss:    tensor(0.8734, device='cuda:0', grad_fn=<NllLossBackward>)
[3330] Train Loss:    tensor(0.8883, device='cuda:0', grad_fn=<NllLossBackward>)
[3331] Train Loss:    tensor(0.8603, device='cuda:0', grad_fn=<NllLossBackward>)
[3332] Train Loss:    tensor(0.9058, device='cuda:0', grad_fn=<NllLossBackward>)
[3333] Train Loss:    tensor(0.8682, device='cuda:0', grad_fn=<NllLossBackward>)
[3334] Train Loss:    tensor(0.8608, device='cuda:0', grad_fn=<NllLossBackward>)
[3335] Train Loss:    tensor(0.9134, device='cuda:0', grad_fn=<NllLossBackward>)
[3336] Train Loss:    tensor(0.8925, device='cuda:0', grad_fn=<NllLossBackward>)
[3337] Train Loss:    tensor(0.8591, device='cuda:0', grad_fn=<NllLossBackward>)
[3338] Train Loss:    tensor(0.8537, device='cuda:0', grad_fn=<NllLossBackward>)
[3339] Train Loss:    tensor(0.8942, device='cuda:0', grad_fn=<NllLossBackward>)
[3340] Train Loss:    tensor(0.8719, device='cuda:0', grad_fn=<NllLossBackward>)
[3341] Train Loss:    tensor(0.8939, device='cuda:0', grad_fn=<NllLossBackward>)
[3342] Train Loss:    tensor(0.8759, device='cuda:0', grad_fn=<NllLossBackward>)
[3343] Train Loss:    tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[3344] Train Loss:    tensor(0.8762, device='cuda:0', grad_fn=<NllLossBackward>)
[3345] Train Loss:    tensor(0.8722, device='cuda:0', grad_fn=<NllLossBackward>)
[3346] Train Loss:    tensor(0.9044, device='cuda:0', grad_fn=<NllLossBackward>)
[3347] Train Loss:    tensor(0.8579, device='cuda:0', grad_fn=<NllLossBackward>)
[3348] Train Loss:    tensor(0.8505, device='cuda:0', grad_fn=<NllLossBackward>)
[3349] Train Loss:    tensor(0.8635, device='cuda:0', grad_fn=<NllLossBackward>)
[3350] Train Loss:    tensor(0.8799, device='cuda:0', grad_fn=<NllLossBackward>)
[3351] Train Loss:    tensor(0.8758, device='cuda:0', grad_fn=<NllLossBackward>)
[3352] Train Loss:    tensor(0.8753, device='cuda:0', grad_fn=<NllLossBackward>)
[3353] Train Loss:    tensor(0.9065, device='cuda:0', grad_fn=<NllLossBackward>)
[3354] Train Loss:    tensor(0.8916, device='cuda:0', grad_fn=<NllLossBackward>)
[3355] Train Loss:    tensor(0.8896, device='cuda:0', grad_fn=<NllLossBackward>)
[3356] Train Loss:    tensor(0.8969, device='cuda:0', grad_fn=<NllLossBackward>)
[3357] Train Loss:    tensor(0.8713, device='cuda:0', grad_fn=<NllLossBackward>)
[3358] Train Loss:    tensor(0.8629, device='cuda:0', grad_fn=<NllLossBackward>)
[3359] Train Loss:    tensor(0.8833, device='cuda:0', grad_fn=<NllLossBackward>)
[3360] Train Loss:    tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[3361] Train Loss:    tensor(0.8834, device='cuda:0', grad_fn=<NllLossBackward>)
[3362] Train Loss:    tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[3363] Train Loss:    tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[3364] Train Loss:    tensor(0.8889, device='cuda:0', grad_fn=<NllLossBackward>)
[3365] Train Loss:    tensor(0.8557, device='cuda:0', grad_fn=<NllLossBackward>)
[3366] Train Loss:    tensor(0.8899, device='cuda:0', grad_fn=<NllLossBackward>)
[3367] Train Loss:    tensor(0.8883, device='cuda:0', grad_fn=<NllLossBackward>)
[3368] Train Loss:    tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[3369] Train Loss:    tensor(0.9039, device='cuda:0', grad_fn=<NllLossBackward>)
[3370] Train Loss:    tensor(0.8540, device='cuda:0', grad_fn=<NllLossBackward>)
[3371] Train Loss:    tensor(0.8490, device='cuda:0', grad_fn=<NllLossBackward>)
[3372] Train Loss:    tensor(0.8624, device='cuda:0', grad_fn=<NllLossBackward>)
[3373] Train Loss:    tensor(0.8753, device='cuda:0', grad_fn=<NllLossBackward>)
[3374] Train Loss:    tensor(0.8481, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3375] Train Loss:    tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[3376] Train Loss:    tensor(0.8682, device='cuda:0', grad_fn=<NllLossBackward>)
[3377] Train Loss:    tensor(0.9011, device='cuda:0', grad_fn=<NllLossBackward>)
[3378] Train Loss:    tensor(0.8639, device='cuda:0', grad_fn=<NllLossBackward>)
[3379] Train Loss:    tensor(0.8855, device='cuda:0', grad_fn=<NllLossBackward>)
[3380] Train Loss:    tensor(0.8908, device='cuda:0', grad_fn=<NllLossBackward>)
[3381] Train Loss:    tensor(0.9005, device='cuda:0', grad_fn=<NllLossBackward>)
[3382] Train Loss:    tensor(0.9036, device='cuda:0', grad_fn=<NllLossBackward>)
[3383] Train Loss:    tensor(0.8798, device='cuda:0', grad_fn=<NllLossBackward>)
[3384] Train Loss:    tensor(0.8974, device='cuda:0', grad_fn=<NllLossBackward>)
[3385] Train Loss:    tensor(0.8903, device='cuda:0', grad_fn=<NllLossBackward>)
[3386] Train Loss:    tensor(0.8658, device='cuda:0', grad_fn=<NllLossBackward>)
[3387] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[3388] Train Loss:    tensor(0.8822, device='cuda:0', grad_fn=<NllLossBackward>)
[3389] Train Loss:    tensor(0.8615, device='cuda:0', grad_fn=<NllLossBackward>)
[3390] Train Loss:    tensor(0.9072, device='cuda:0', grad_fn=<NllLossBackward>)
[3391] Train Loss:    tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[3392] Train Loss:    tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[3393] Train Loss:    tensor(0.8397, device='cuda:0', grad_fn=<NllLossBackward>)
[3394] Train Loss:    tensor(0.8754, device='cuda:0', grad_fn=<NllLossBackward>)
[3395] Train Loss:    tensor(0.8813, device='cuda:0', grad_fn=<NllLossBackward>)
[3396] Train Loss:    tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[3397] Train Loss:    tensor(0.9096, device='cuda:0', grad_fn=<NllLossBackward>)
[3398] Train Loss:    tensor(0.8735, device='cuda:0', grad_fn=<NllLossBackward>)
[3399] Train Loss:    tensor(0.8755, device='cuda:0', grad_fn=<NllLossBackward>)
[3400] Train Loss:    tensor(0.8821, device='cuda:0', grad_fn=<NllLossBackward>)
[3401] Train Loss:    tensor(0.8699, device='cuda:0', grad_fn=<NllLossBackward>)
[3402] Train Loss:    tensor(0.8591, device='cuda:0', grad_fn=<NllLossBackward>)
[3403] Train Loss:    tensor(0.8478, device='cuda:0', grad_fn=<NllLossBackward>)
[3404] Train Loss:    tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[3405] Train Loss:    tensor(0.9011, device='cuda:0', grad_fn=<NllLossBackward>)
[3406] Train Loss:    tensor(0.8609, device='cuda:0', grad_fn=<NllLossBackward>)
[3407] Train Loss:    tensor(0.8818, device='cuda:0', grad_fn=<NllLossBackward>)
[3408] Train Loss:    tensor(0.8836, device='cuda:0', grad_fn=<NllLossBackward>)
[3409] Train Loss:    tensor(0.8562, device='cuda:0', grad_fn=<NllLossBackward>)
[3410] Train Loss:    tensor(0.8834, device='cuda:0', grad_fn=<NllLossBackward>)
[3411] Train Loss:    tensor(0.8721, device='cuda:0', grad_fn=<NllLossBackward>)
[3412] Train Loss:    tensor(0.8862, device='cuda:0', grad_fn=<NllLossBackward>)
[3413] Train Loss:    tensor(0.8578, device='cuda:0', grad_fn=<NllLossBackward>)
[3414] Train Loss:    tensor(0.8898, device='cuda:0', grad_fn=<NllLossBackward>)
[3415] Train Loss:    tensor(0.8781, device='cuda:0', grad_fn=<NllLossBackward>)
[3416] Train Loss:    tensor(0.9031, device='cuda:0', grad_fn=<NllLossBackward>)
[3417] Train Loss:    tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[3418] Train Loss:    tensor(0.8898, device='cuda:0', grad_fn=<NllLossBackward>)
[3419] Train Loss:    tensor(0.8698, device='cuda:0', grad_fn=<NllLossBackward>)
[3420] Train Loss:    tensor(0.8640, device='cuda:0', grad_fn=<NllLossBackward>)
[3421] Train Loss:    tensor(0.8473, device='cuda:0', grad_fn=<NllLossBackward>)
[3422] Train Loss:    tensor(0.9320, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3423] Train Loss:    tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[3424] Train Loss:    tensor(0.8772, device='cuda:0', grad_fn=<NllLossBackward>)
[3425] Train Loss:    tensor(0.9024, device='cuda:0', grad_fn=<NllLossBackward>)
[3426] Train Loss:    tensor(0.8784, device='cuda:0', grad_fn=<NllLossBackward>)
[3427] Train Loss:    tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[3428] Train Loss:    tensor(0.9001, device='cuda:0', grad_fn=<NllLossBackward>)
[3429] Train Loss:    tensor(0.8437, device='cuda:0', grad_fn=<NllLossBackward>)
[3430] Train Loss:    tensor(0.8654, device='cuda:0', grad_fn=<NllLossBackward>)
[3431] Train Loss:    tensor(0.8738, device='cuda:0', grad_fn=<NllLossBackward>)
[3432] Train Loss:    tensor(0.9005, device='cuda:0', grad_fn=<NllLossBackward>)
[3433] Train Loss:    tensor(0.8848, device='cuda:0', grad_fn=<NllLossBackward>)
[3434] Train Loss:    tensor(0.8671, device='cuda:0', grad_fn=<NllLossBackward>)
[3435] Train Loss:    tensor(0.8771, device='cuda:0', grad_fn=<NllLossBackward>)
[3436] Train Loss:    tensor(0.8838, device='cuda:0', grad_fn=<NllLossBackward>)
[3437] Train Loss:    tensor(0.8419, device='cuda:0', grad_fn=<NllLossBackward>)
[3438] Train Loss:    tensor(0.8687, device='cuda:0', grad_fn=<NllLossBackward>)
[3439] Train Loss:    tensor(0.8502, device='cuda:0', grad_fn=<NllLossBackward>)
[3440] Train Loss:    tensor(0.8728, device='cuda:0', grad_fn=<NllLossBackward>)
[3441] Train Loss:    tensor(0.8903, device='cuda:0', grad_fn=<NllLossBackward>)
[3442] Train Loss:    tensor(0.8617, device='cuda:0', grad_fn=<NllLossBackward>)
[3443] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[3444] Train Loss:    tensor(0.8985, device='cuda:0', grad_fn=<NllLossBackward>)
[3445] Train Loss:    tensor(0.8774, device='cuda:0', grad_fn=<NllLossBackward>)
[3446] Train Loss:    tensor(0.8740, device='cuda:0', grad_fn=<NllLossBackward>)
[3447] Train Loss:    tensor(0.9290, device='cuda:0', grad_fn=<NllLossBackward>)
[3448] Train Loss:    tensor(0.8778, device='cuda:0', grad_fn=<NllLossBackward>)
[3449] Train Loss:    tensor(0.8684, device='cuda:0', grad_fn=<NllLossBackward>)
[3450] Train Loss:    tensor(0.8676, device='cuda:0', grad_fn=<NllLossBackward>)
[3451] Train Loss:    tensor(0.8792, device='cuda:0', grad_fn=<NllLossBackward>)
[3452] Train Loss:    tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[3453] Train Loss:    tensor(0.9234, device='cuda:0', grad_fn=<NllLossBackward>)
[3454] Train Loss:    tensor(0.8596, device='cuda:0', grad_fn=<NllLossBackward>)
[3455] Train Loss:    tensor(0.8617, device='cuda:0', grad_fn=<NllLossBackward>)
[3456] Train Loss:    tensor(0.8688, device='cuda:0', grad_fn=<NllLossBackward>)
[3457] Train Loss:    tensor(0.8368, device='cuda:0', grad_fn=<NllLossBackward>)
[3458] Train Loss:    tensor(0.8918, device='cuda:0', grad_fn=<NllLossBackward>)
[3459] Train Loss:    tensor(0.8984, device='cuda:0', grad_fn=<NllLossBackward>)
[3460] Train Loss:    tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[3461] Train Loss:    tensor(0.8577, device='cuda:0', grad_fn=<NllLossBackward>)
[3462] Train Loss:    tensor(0.8413, device='cuda:0', grad_fn=<NllLossBackward>)
[3463] Train Loss:    tensor(0.8938, device='cuda:0', grad_fn=<NllLossBackward>)
[3464] Train Loss:    tensor(0.9056, device='cuda:0', grad_fn=<NllLossBackward>)
[3465] Train Loss:    tensor(0.8650, device='cuda:0', grad_fn=<NllLossBackward>)
[3466] Train Loss:    tensor(0.8461, device='cuda:0', grad_fn=<NllLossBackward>)
[3467] Train Loss:    tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[3468] Train Loss:    tensor(0.8883, device='cuda:0', grad_fn=<NllLossBackward>)
[3469] Train Loss:    tensor(0.8949, device='cuda:0', grad_fn=<NllLossBackward>)
[3470] Train Loss:    tensor(0.8613, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3471] Train Loss:    tensor(0.8882, device='cuda:0', grad_fn=<NllLossBackward>)
[3472] Train Loss:    tensor(0.8713, device='cuda:0', grad_fn=<NllLossBackward>)
[3473] Train Loss:    tensor(0.8614, device='cuda:0', grad_fn=<NllLossBackward>)
[3474] Train Loss:    tensor(0.8740, device='cuda:0', grad_fn=<NllLossBackward>)
[3475] Train Loss:    tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[3476] Train Loss:    tensor(0.8603, device='cuda:0', grad_fn=<NllLossBackward>)
[3477] Train Loss:    tensor(0.8632, device='cuda:0', grad_fn=<NllLossBackward>)
[3478] Train Loss:    tensor(0.8843, device='cuda:0', grad_fn=<NllLossBackward>)
[3479] Train Loss:    tensor(0.8244, device='cuda:0', grad_fn=<NllLossBackward>)
[3480] Train Loss:    tensor(0.8381, device='cuda:0', grad_fn=<NllLossBackward>)
[3481] Train Loss:    tensor(0.8746, device='cuda:0', grad_fn=<NllLossBackward>)
[3482] Train Loss:    tensor(0.9064, device='cuda:0', grad_fn=<NllLossBackward>)
[3483] Train Loss:    tensor(0.8920, device='cuda:0', grad_fn=<NllLossBackward>)
[3484] Train Loss:    tensor(0.9279, device='cuda:0', grad_fn=<NllLossBackward>)
[3485] Train Loss:    tensor(0.8742, device='cuda:0', grad_fn=<NllLossBackward>)
[3486] Train Loss:    tensor(0.8993, device='cuda:0', grad_fn=<NllLossBackward>)
[3487] Train Loss:    tensor(0.8793, device='cuda:0', grad_fn=<NllLossBackward>)
[3488] Train Loss:    tensor(0.8845, device='cuda:0', grad_fn=<NllLossBackward>)
[3489] Train Loss:    tensor(0.8645, device='cuda:0', grad_fn=<NllLossBackward>)
[3490] Train Loss:    tensor(0.8518, device='cuda:0', grad_fn=<NllLossBackward>)
[3491] Train Loss:    tensor(0.8522, device='cuda:0', grad_fn=<NllLossBackward>)
[3492] Train Loss:    tensor(0.8654, device='cuda:0', grad_fn=<NllLossBackward>)
[3493] Train Loss:    tensor(0.8741, device='cuda:0', grad_fn=<NllLossBackward>)
[3494] Train Loss:    tensor(0.9073, device='cuda:0', grad_fn=<NllLossBackward>)
[3495] Train Loss:    tensor(0.8721, device='cuda:0', grad_fn=<NllLossBackward>)
[3496] Train Loss:    tensor(0.8852, device='cuda:0', grad_fn=<NllLossBackward>)
[3497] Train Loss:    tensor(0.8821, device='cuda:0', grad_fn=<NllLossBackward>)
[3498] Train Loss:    tensor(0.8501, device='cuda:0', grad_fn=<NllLossBackward>)
[3499] Train Loss:    tensor(0.8451, device='cuda:0', grad_fn=<NllLossBackward>)
[3500] Train Loss:    tensor(0.8434, device='cuda:0', grad_fn=<NllLossBackward>)
[3501] Train Loss:    tensor(0.8909, device='cuda:0', grad_fn=<NllLossBackward>)
[3502] Train Loss:    tensor(0.8802, device='cuda:0', grad_fn=<NllLossBackward>)
[3503] Train Loss:    tensor(0.8624, device='cuda:0', grad_fn=<NllLossBackward>)
[3504] Train Loss:    tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[3505] Train Loss:    tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[3506] Train Loss:    tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[3507] Train Loss:    tensor(0.8703, device='cuda:0', grad_fn=<NllLossBackward>)
[3508] Train Loss:    tensor(0.8615, device='cuda:0', grad_fn=<NllLossBackward>)
[3509] Train Loss:    tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[3510] Train Loss:    tensor(0.8808, device='cuda:0', grad_fn=<NllLossBackward>)
[3511] Train Loss:    tensor(0.8691, device='cuda:0', grad_fn=<NllLossBackward>)
[3512] Train Loss:    tensor(0.8685, device='cuda:0', grad_fn=<NllLossBackward>)
[3513] Train Loss:    tensor(0.8701, device='cuda:0', grad_fn=<NllLossBackward>)
[3514] Train Loss:    tensor(0.8885, device='cuda:0', grad_fn=<NllLossBackward>)
[3515] Train Loss:    tensor(0.8618, device='cuda:0', grad_fn=<NllLossBackward>)
[3516] Train Loss:    tensor(0.8806, device='cuda:0', grad_fn=<NllLossBackward>)
[3517] Train Loss:    tensor(0.8666, device='cuda:0', grad_fn=<NllLossBackward>)
[3518] Train Loss:    tensor(0.8660, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3519] Train Loss:    tensor(0.8515, device='cuda:0', grad_fn=<NllLossBackward>)
[3520] Train Loss:    tensor(0.8563, device='cuda:0', grad_fn=<NllLossBackward>)
[3521] Train Loss:    tensor(0.8365, device='cuda:0', grad_fn=<NllLossBackward>)
[3522] Train Loss:    tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[3523] Train Loss:    tensor(0.8511, device='cuda:0', grad_fn=<NllLossBackward>)
[3524] Train Loss:    tensor(0.8836, device='cuda:0', grad_fn=<NllLossBackward>)
[3525] Train Loss:    tensor(0.8795, device='cuda:0', grad_fn=<NllLossBackward>)
[3526] Train Loss:    tensor(0.9076, device='cuda:0', grad_fn=<NllLossBackward>)
[3527] Train Loss:    tensor(0.8631, device='cuda:0', grad_fn=<NllLossBackward>)
[3528] Train Loss:    tensor(0.8671, device='cuda:0', grad_fn=<NllLossBackward>)
[3529] Train Loss:    tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
[3530] Train Loss:    tensor(0.8594, device='cuda:0', grad_fn=<NllLossBackward>)
[3531] Train Loss:    tensor(0.8758, device='cuda:0', grad_fn=<NllLossBackward>)
[3532] Train Loss:    tensor(0.8587, device='cuda:0', grad_fn=<NllLossBackward>)
[3533] Train Loss:    tensor(0.9054, device='cuda:0', grad_fn=<NllLossBackward>)
[3534] Train Loss:    tensor(0.8788, device='cuda:0', grad_fn=<NllLossBackward>)
[3535] Train Loss:    tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[3536] Train Loss:    tensor(0.8674, device='cuda:0', grad_fn=<NllLossBackward>)
[3537] Train Loss:    tensor(0.8849, device='cuda:0', grad_fn=<NllLossBackward>)
[3538] Train Loss:    tensor(0.9009, device='cuda:0', grad_fn=<NllLossBackward>)
[3539] Train Loss:    tensor(0.8532, device='cuda:0', grad_fn=<NllLossBackward>)
[3540] Train Loss:    tensor(0.8954, device='cuda:0', grad_fn=<NllLossBackward>)
[3541] Train Loss:    tensor(0.8393, device='cuda:0', grad_fn=<NllLossBackward>)
[3542] Train Loss:    tensor(0.9255, device='cuda:0', grad_fn=<NllLossBackward>)
[3543] Train Loss:    tensor(0.8739, device='cuda:0', grad_fn=<NllLossBackward>)
[3544] Train Loss:    tensor(0.9334, device='cuda:0', grad_fn=<NllLossBackward>)
[3545] Train Loss:    tensor(0.8841, device='cuda:0', grad_fn=<NllLossBackward>)
[3546] Train Loss:    tensor(0.8556, device='cuda:0', grad_fn=<NllLossBackward>)
[3547] Train Loss:    tensor(0.9002, device='cuda:0', grad_fn=<NllLossBackward>)
[3548] Train Loss:    tensor(0.8904, device='cuda:0', grad_fn=<NllLossBackward>)
[3549] Train Loss:    tensor(0.8839, device='cuda:0', grad_fn=<NllLossBackward>)
[3550] Train Loss:    tensor(0.8645, device='cuda:0', grad_fn=<NllLossBackward>)
[3551] Train Loss:    tensor(0.8945, device='cuda:0', grad_fn=<NllLossBackward>)
[3552] Train Loss:    tensor(0.8940, device='cuda:0', grad_fn=<NllLossBackward>)
[3553] Train Loss:    tensor(0.8383, device='cuda:0', grad_fn=<NllLossBackward>)
[3554] Train Loss:    tensor(0.8786, device='cuda:0', grad_fn=<NllLossBackward>)
[3555] Train Loss:    tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[3556] Train Loss:    tensor(0.8469, device='cuda:0', grad_fn=<NllLossBackward>)
[3557] Train Loss:    tensor(0.8709, device='cuda:0', grad_fn=<NllLossBackward>)
[3558] Train Loss:    tensor(0.8599, device='cuda:0', grad_fn=<NllLossBackward>)
[3559] Train Loss:    tensor(0.8586, device='cuda:0', grad_fn=<NllLossBackward>)
[3560] Train Loss:    tensor(0.8551, device='cuda:0', grad_fn=<NllLossBackward>)
[3561] Train Loss:    tensor(0.8762, device='cuda:0', grad_fn=<NllLossBackward>)
[3562] Train Loss:    tensor(0.8713, device='cuda:0', grad_fn=<NllLossBackward>)
[3563] Train Loss:    tensor(0.8765, device='cuda:0', grad_fn=<NllLossBackward>)
[3564] Train Loss:    tensor(0.8561, device='cuda:0', grad_fn=<NllLossBackward>)
[3565] Train Loss:    tensor(0.8697, device='cuda:0', grad_fn=<NllLossBackward>)
[3566] Train Loss:    tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3567] Train Loss:    tensor(0.8615, device='cuda:0', grad_fn=<NllLossBackward>)
[3568] Train Loss:    tensor(0.8799, device='cuda:0', grad_fn=<NllLossBackward>)
[3569] Train Loss:    tensor(0.8728, device='cuda:0', grad_fn=<NllLossBackward>)
[3570] Train Loss:    tensor(0.8692, device='cuda:0', grad_fn=<NllLossBackward>)
[3571] Train Loss:    tensor(0.8705, device='cuda:0', grad_fn=<NllLossBackward>)
[3572] Train Loss:    tensor(0.8915, device='cuda:0', grad_fn=<NllLossBackward>)
[3573] Train Loss:    tensor(0.8691, device='cuda:0', grad_fn=<NllLossBackward>)
[3574] Train Loss:    tensor(0.8553, device='cuda:0', grad_fn=<NllLossBackward>)
[3575] Train Loss:    tensor(0.8404, device='cuda:0', grad_fn=<NllLossBackward>)
[3576] Train Loss:    tensor(0.8775, device='cuda:0', grad_fn=<NllLossBackward>)
[3577] Train Loss:    tensor(0.8822, device='cuda:0', grad_fn=<NllLossBackward>)
[3578] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[3579] Train Loss:    tensor(0.8916, device='cuda:0', grad_fn=<NllLossBackward>)
[3580] Train Loss:    tensor(0.9125, device='cuda:0', grad_fn=<NllLossBackward>)
[3581] Train Loss:    tensor(0.8843, device='cuda:0', grad_fn=<NllLossBackward>)
[3582] Train Loss:    tensor(0.9000, device='cuda:0', grad_fn=<NllLossBackward>)
[3583] Train Loss:    tensor(0.8994, device='cuda:0', grad_fn=<NllLossBackward>)
[3584] Train Loss:    tensor(0.8864, device='cuda:0', grad_fn=<NllLossBackward>)
[3585] Train Loss:    tensor(0.8518, device='cuda:0', grad_fn=<NllLossBackward>)
[3586] Train Loss:    tensor(0.8558, device='cuda:0', grad_fn=<NllLossBackward>)
[3587] Train Loss:    tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[3588] Train Loss:    tensor(0.8515, device='cuda:0', grad_fn=<NllLossBackward>)
[3589] Train Loss:    tensor(0.8446, device='cuda:0', grad_fn=<NllLossBackward>)
[3590] Train Loss:    tensor(0.8876, device='cuda:0', grad_fn=<NllLossBackward>)
[3591] Train Loss:    tensor(0.8614, device='cuda:0', grad_fn=<NllLossBackward>)
[3592] Train Loss:    tensor(0.8752, device='cuda:0', grad_fn=<NllLossBackward>)
[3593] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[3594] Train Loss:    tensor(0.8606, device='cuda:0', grad_fn=<NllLossBackward>)
[3595] Train Loss:    tensor(0.9141, device='cuda:0', grad_fn=<NllLossBackward>)
[3596] Train Loss:    tensor(0.8963, device='cuda:0', grad_fn=<NllLossBackward>)
[3597] Train Loss:    tensor(0.8921, device='cuda:0', grad_fn=<NllLossBackward>)
[3598] Train Loss:    tensor(0.8839, device='cuda:0', grad_fn=<NllLossBackward>)
[3599] Train Loss:    tensor(0.8671, device='cuda:0', grad_fn=<NllLossBackward>)
[3600] Train Loss:    tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[3601] Train Loss:    tensor(0.8521, device='cuda:0', grad_fn=<NllLossBackward>)
[3602] Train Loss:    tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[3603] Train Loss:    tensor(0.8880, device='cuda:0', grad_fn=<NllLossBackward>)
[3604] Train Loss:    tensor(0.8830, device='cuda:0', grad_fn=<NllLossBackward>)
[3605] Train Loss:    tensor(0.8617, device='cuda:0', grad_fn=<NllLossBackward>)
[3606] Train Loss:    tensor(0.9021, device='cuda:0', grad_fn=<NllLossBackward>)
[3607] Train Loss:    tensor(0.8543, device='cuda:0', grad_fn=<NllLossBackward>)
[3608] Train Loss:    tensor(0.8947, device='cuda:0', grad_fn=<NllLossBackward>)
[3609] Train Loss:    tensor(0.8924, device='cuda:0', grad_fn=<NllLossBackward>)
[3610] Train Loss:    tensor(0.8618, device='cuda:0', grad_fn=<NllLossBackward>)
[3611] Train Loss:    tensor(0.8674, device='cuda:0', grad_fn=<NllLossBackward>)
[3612] Train Loss:    tensor(0.8595, device='cuda:0', grad_fn=<NllLossBackward>)
[3613] Train Loss:    tensor(0.8630, device='cuda:0', grad_fn=<NllLossBackward>)
[3614] Train Loss:    tensor(0.8810, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3615] Train Loss:    tensor(0.8396, device='cuda:0', grad_fn=<NllLossBackward>)
[3616] Train Loss:    tensor(0.8715, device='cuda:0', grad_fn=<NllLossBackward>)
[3617] Train Loss:    tensor(0.8912, device='cuda:0', grad_fn=<NllLossBackward>)
[3618] Train Loss:    tensor(0.8813, device='cuda:0', grad_fn=<NllLossBackward>)
[3619] Train Loss:    tensor(0.8645, device='cuda:0', grad_fn=<NllLossBackward>)
[3620] Train Loss:    tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[3621] Train Loss:    tensor(0.8505, device='cuda:0', grad_fn=<NllLossBackward>)
[3622] Train Loss:    tensor(0.8622, device='cuda:0', grad_fn=<NllLossBackward>)
[3623] Train Loss:    tensor(0.8740, device='cuda:0', grad_fn=<NllLossBackward>)
[3624] Train Loss:    tensor(0.8768, device='cuda:0', grad_fn=<NllLossBackward>)
[3625] Train Loss:    tensor(0.8919, device='cuda:0', grad_fn=<NllLossBackward>)
[3626] Train Loss:    tensor(0.8820, device='cuda:0', grad_fn=<NllLossBackward>)
[3627] Train Loss:    tensor(0.9125, device='cuda:0', grad_fn=<NllLossBackward>)
[3628] Train Loss:    tensor(0.8854, device='cuda:0', grad_fn=<NllLossBackward>)
[3629] Train Loss:    tensor(0.8638, device='cuda:0', grad_fn=<NllLossBackward>)
[3630] Train Loss:    tensor(0.8660, device='cuda:0', grad_fn=<NllLossBackward>)
[3631] Train Loss:    tensor(0.8453, device='cuda:0', grad_fn=<NllLossBackward>)
[3632] Train Loss:    tensor(0.8449, device='cuda:0', grad_fn=<NllLossBackward>)
[3633] Train Loss:    tensor(0.8728, device='cuda:0', grad_fn=<NllLossBackward>)
[3634] Train Loss:    tensor(0.8649, device='cuda:0', grad_fn=<NllLossBackward>)
[3635] Train Loss:    tensor(0.8730, device='cuda:0', grad_fn=<NllLossBackward>)
[3636] Train Loss:    tensor(0.8665, device='cuda:0', grad_fn=<NllLossBackward>)
[3637] Train Loss:    tensor(0.8731, device='cuda:0', grad_fn=<NllLossBackward>)
[3638] Train Loss:    tensor(0.8579, device='cuda:0', grad_fn=<NllLossBackward>)
[3639] Train Loss:    tensor(0.8494, device='cuda:0', grad_fn=<NllLossBackward>)
[3640] Train Loss:    tensor(0.8805, device='cuda:0', grad_fn=<NllLossBackward>)
[3641] Train Loss:    tensor(0.8641, device='cuda:0', grad_fn=<NllLossBackward>)
[3642] Train Loss:    tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[3643] Train Loss:    tensor(0.8771, device='cuda:0', grad_fn=<NllLossBackward>)
[3644] Train Loss:    tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[3645] Train Loss:    tensor(0.8991, device='cuda:0', grad_fn=<NllLossBackward>)
[3646] Train Loss:    tensor(0.8785, device='cuda:0', grad_fn=<NllLossBackward>)
[3647] Train Loss:    tensor(0.8800, device='cuda:0', grad_fn=<NllLossBackward>)
[3648] Train Loss:    tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[3649] Train Loss:    tensor(0.8694, device='cuda:0', grad_fn=<NllLossBackward>)
[3650] Train Loss:    tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[3651] Train Loss:    tensor(0.8701, device='cuda:0', grad_fn=<NllLossBackward>)
[3652] Train Loss:    tensor(0.8739, device='cuda:0', grad_fn=<NllLossBackward>)
[3653] Train Loss:    tensor(0.8442, device='cuda:0', grad_fn=<NllLossBackward>)
[3654] Train Loss:    tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[3655] Train Loss:    tensor(0.8623, device='cuda:0', grad_fn=<NllLossBackward>)
[3656] Train Loss:    tensor(0.8835, device='cuda:0', grad_fn=<NllLossBackward>)
[3657] Train Loss:    tensor(0.9014, device='cuda:0', grad_fn=<NllLossBackward>)
[3658] Train Loss:    tensor(0.8602, device='cuda:0', grad_fn=<NllLossBackward>)
[3659] Train Loss:    tensor(0.8960, device='cuda:0', grad_fn=<NllLossBackward>)
[3660] Train Loss:    tensor(0.8671, device='cuda:0', grad_fn=<NllLossBackward>)
[3661] Train Loss:    tensor(0.8304, device='cuda:0', grad_fn=<NllLossBackward>)
[3662] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3663] Train Loss:    tensor(0.8930, device='cuda:0', grad_fn=<NllLossBackward>)
[3664] Train Loss:    tensor(0.8653, device='cuda:0', grad_fn=<NllLossBackward>)
[3665] Train Loss:    tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[3666] Train Loss:    tensor(0.8602, device='cuda:0', grad_fn=<NllLossBackward>)
[3667] Train Loss:    tensor(0.8426, device='cuda:0', grad_fn=<NllLossBackward>)
[3668] Train Loss:    tensor(0.8554, device='cuda:0', grad_fn=<NllLossBackward>)
[3669] Train Loss:    tensor(0.9088, device='cuda:0', grad_fn=<NllLossBackward>)
[3670] Train Loss:    tensor(0.8537, device='cuda:0', grad_fn=<NllLossBackward>)
[3671] Train Loss:    tensor(0.8623, device='cuda:0', grad_fn=<NllLossBackward>)
[3672] Train Loss:    tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[3673] Train Loss:    tensor(0.8496, device='cuda:0', grad_fn=<NllLossBackward>)
[3674] Train Loss:    tensor(0.8677, device='cuda:0', grad_fn=<NllLossBackward>)
[3675] Train Loss:    tensor(0.8910, device='cuda:0', grad_fn=<NllLossBackward>)
[3676] Train Loss:    tensor(0.8649, device='cuda:0', grad_fn=<NllLossBackward>)
[3677] Train Loss:    tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[3678] Train Loss:    tensor(0.8795, device='cuda:0', grad_fn=<NllLossBackward>)
[3679] Train Loss:    tensor(0.8753, device='cuda:0', grad_fn=<NllLossBackward>)
[3680] Train Loss:    tensor(0.8747, device='cuda:0', grad_fn=<NllLossBackward>)
[3681] Train Loss:    tensor(0.8402, device='cuda:0', grad_fn=<NllLossBackward>)
[3682] Train Loss:    tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[3683] Train Loss:    tensor(0.8784, device='cuda:0', grad_fn=<NllLossBackward>)
[3684] Train Loss:    tensor(0.8798, device='cuda:0', grad_fn=<NllLossBackward>)
[3685] Train Loss:    tensor(0.8944, device='cuda:0', grad_fn=<NllLossBackward>)
[3686] Train Loss:    tensor(0.9084, device='cuda:0', grad_fn=<NllLossBackward>)
[3687] Train Loss:    tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[3688] Train Loss:    tensor(0.8712, device='cuda:0', grad_fn=<NllLossBackward>)
[3689] Train Loss:    tensor(0.8703, device='cuda:0', grad_fn=<NllLossBackward>)
[3690] Train Loss:    tensor(0.8653, device='cuda:0', grad_fn=<NllLossBackward>)
[3691] Train Loss:    tensor(0.8775, device='cuda:0', grad_fn=<NllLossBackward>)
[3692] Train Loss:    tensor(0.8797, device='cuda:0', grad_fn=<NllLossBackward>)
[3693] Train Loss:    tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[3694] Train Loss:    tensor(0.8701, device='cuda:0', grad_fn=<NllLossBackward>)
[3695] Train Loss:    tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[3696] Train Loss:    tensor(0.8762, device='cuda:0', grad_fn=<NllLossBackward>)
[3697] Train Loss:    tensor(0.8620, device='cuda:0', grad_fn=<NllLossBackward>)
[3698] Train Loss:    tensor(0.8767, device='cuda:0', grad_fn=<NllLossBackward>)
[3699] Train Loss:    tensor(0.8815, device='cuda:0', grad_fn=<NllLossBackward>)
[3700] Train Loss:    tensor(0.8266, device='cuda:0', grad_fn=<NllLossBackward>)
[3701] Train Loss:    tensor(0.8954, device='cuda:0', grad_fn=<NllLossBackward>)
[3702] Train Loss:    tensor(0.8629, device='cuda:0', grad_fn=<NllLossBackward>)
[3703] Train Loss:    tensor(0.8428, device='cuda:0', grad_fn=<NllLossBackward>)
[3704] Train Loss:    tensor(0.8690, device='cuda:0', grad_fn=<NllLossBackward>)
[3705] Train Loss:    tensor(0.8741, device='cuda:0', grad_fn=<NllLossBackward>)
[3706] Train Loss:    tensor(0.8486, device='cuda:0', grad_fn=<NllLossBackward>)
[3707] Train Loss:    tensor(0.8517, device='cuda:0', grad_fn=<NllLossBackward>)
[3708] Train Loss:    tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[3709] Train Loss:    tensor(0.8708, device='cuda:0', grad_fn=<NllLossBackward>)
[3710] Train Loss:    tensor(0.8464, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3711] Train Loss:    tensor(0.8832, device='cuda:0', grad_fn=<NllLossBackward>)
[3712] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[3713] Train Loss:    tensor(0.8567, device='cuda:0', grad_fn=<NllLossBackward>)
[3714] Train Loss:    tensor(0.8747, device='cuda:0', grad_fn=<NllLossBackward>)
[3715] Train Loss:    tensor(0.9034, device='cuda:0', grad_fn=<NllLossBackward>)
[3716] Train Loss:    tensor(0.8696, device='cuda:0', grad_fn=<NllLossBackward>)
[3717] Train Loss:    tensor(0.8864, device='cuda:0', grad_fn=<NllLossBackward>)
[3718] Train Loss:    tensor(0.8793, device='cuda:0', grad_fn=<NllLossBackward>)
[3719] Train Loss:    tensor(0.8758, device='cuda:0', grad_fn=<NllLossBackward>)
[3720] Train Loss:    tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[3721] Train Loss:    tensor(0.8743, device='cuda:0', grad_fn=<NllLossBackward>)
[3722] Train Loss:    tensor(0.8524, device='cuda:0', grad_fn=<NllLossBackward>)
[3723] Train Loss:    tensor(0.8954, device='cuda:0', grad_fn=<NllLossBackward>)
[3724] Train Loss:    tensor(0.8635, device='cuda:0', grad_fn=<NllLossBackward>)
[3725] Train Loss:    tensor(0.8685, device='cuda:0', grad_fn=<NllLossBackward>)
[3726] Train Loss:    tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[3727] Train Loss:    tensor(0.8811, device='cuda:0', grad_fn=<NllLossBackward>)
[3728] Train Loss:    tensor(0.8954, device='cuda:0', grad_fn=<NllLossBackward>)
[3729] Train Loss:    tensor(0.9234, device='cuda:0', grad_fn=<NllLossBackward>)
[3730] Train Loss:    tensor(0.8195, device='cuda:0', grad_fn=<NllLossBackward>)
[3731] Train Loss:    tensor(0.8806, device='cuda:0', grad_fn=<NllLossBackward>)
[3732] Train Loss:    tensor(0.8750, device='cuda:0', grad_fn=<NllLossBackward>)
[3733] Train Loss:    tensor(0.8892, device='cuda:0', grad_fn=<NllLossBackward>)
[3734] Train Loss:    tensor(0.9021, device='cuda:0', grad_fn=<NllLossBackward>)
[3735] Train Loss:    tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[3736] Train Loss:    tensor(0.8805, device='cuda:0', grad_fn=<NllLossBackward>)
[3737] Train Loss:    tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[3738] Train Loss:    tensor(0.8887, device='cuda:0', grad_fn=<NllLossBackward>)
[3739] Train Loss:    tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[3740] Train Loss:    tensor(0.8787, device='cuda:0', grad_fn=<NllLossBackward>)
[3741] Train Loss:    tensor(0.8620, device='cuda:0', grad_fn=<NllLossBackward>)
[3742] Train Loss:    tensor(0.8569, device='cuda:0', grad_fn=<NllLossBackward>)
[3743] Train Loss:    tensor(0.8630, device='cuda:0', grad_fn=<NllLossBackward>)
[3744] Train Loss:    tensor(0.8819, device='cuda:0', grad_fn=<NllLossBackward>)
[3745] Train Loss:    tensor(0.8462, device='cuda:0', grad_fn=<NllLossBackward>)
[3746] Train Loss:    tensor(0.8759, device='cuda:0', grad_fn=<NllLossBackward>)
[3747] Train Loss:    tensor(0.8478, device='cuda:0', grad_fn=<NllLossBackward>)
[3748] Train Loss:    tensor(0.8688, device='cuda:0', grad_fn=<NllLossBackward>)
[3749] Train Loss:    tensor(0.9019, device='cuda:0', grad_fn=<NllLossBackward>)
[3750] Train Loss:    tensor(0.8816, device='cuda:0', grad_fn=<NllLossBackward>)
[3751] Train Loss:    tensor(0.8364, device='cuda:0', grad_fn=<NllLossBackward>)
[3752] Train Loss:    tensor(0.8651, device='cuda:0', grad_fn=<NllLossBackward>)
[3753] Train Loss:    tensor(0.8862, device='cuda:0', grad_fn=<NllLossBackward>)
[3754] Train Loss:    tensor(0.8735, device='cuda:0', grad_fn=<NllLossBackward>)
[3755] Train Loss:    tensor(0.8792, device='cuda:0', grad_fn=<NllLossBackward>)
[3756] Train Loss:    tensor(0.8493, device='cuda:0', grad_fn=<NllLossBackward>)
[3757] Train Loss:    tensor(0.8974, device='cuda:0', grad_fn=<NllLossBackward>)
[3758] Train Loss:    tensor(0.8418, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3759] Train Loss:    tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[3760] Train Loss:    tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[3761] Train Loss:    tensor(0.8746, device='cuda:0', grad_fn=<NllLossBackward>)
[3762] Train Loss:    tensor(0.8814, device='cuda:0', grad_fn=<NllLossBackward>)
[3763] Train Loss:    tensor(0.8905, device='cuda:0', grad_fn=<NllLossBackward>)
[3764] Train Loss:    tensor(0.8928, device='cuda:0', grad_fn=<NllLossBackward>)
[3765] Train Loss:    tensor(0.8765, device='cuda:0', grad_fn=<NllLossBackward>)
[3766] Train Loss:    tensor(0.8681, device='cuda:0', grad_fn=<NllLossBackward>)
[3767] Train Loss:    tensor(0.8887, device='cuda:0', grad_fn=<NllLossBackward>)
[3768] Train Loss:    tensor(0.8633, device='cuda:0', grad_fn=<NllLossBackward>)
[3769] Train Loss:    tensor(0.8761, device='cuda:0', grad_fn=<NllLossBackward>)
[3770] Train Loss:    tensor(0.8635, device='cuda:0', grad_fn=<NllLossBackward>)
[3771] Train Loss:    tensor(0.8838, device='cuda:0', grad_fn=<NllLossBackward>)
[3772] Train Loss:    tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[3773] Train Loss:    tensor(0.8795, device='cuda:0', grad_fn=<NllLossBackward>)
[3774] Train Loss:    tensor(0.8829, device='cuda:0', grad_fn=<NllLossBackward>)
[3775] Train Loss:    tensor(0.8692, device='cuda:0', grad_fn=<NllLossBackward>)
[3776] Train Loss:    tensor(0.8496, device='cuda:0', grad_fn=<NllLossBackward>)
[3777] Train Loss:    tensor(0.8643, device='cuda:0', grad_fn=<NllLossBackward>)
[3778] Train Loss:    tensor(0.8688, device='cuda:0', grad_fn=<NllLossBackward>)
[3779] Train Loss:    tensor(0.8458, device='cuda:0', grad_fn=<NllLossBackward>)
[3780] Train Loss:    tensor(0.8968, device='cuda:0', grad_fn=<NllLossBackward>)
[3781] Train Loss:    tensor(0.8525, device='cuda:0', grad_fn=<NllLossBackward>)
[3782] Train Loss:    tensor(0.8862, device='cuda:0', grad_fn=<NllLossBackward>)
[3783] Train Loss:    tensor(0.8838, device='cuda:0', grad_fn=<NllLossBackward>)
[3784] Train Loss:    tensor(0.8715, device='cuda:0', grad_fn=<NllLossBackward>)
[3785] Train Loss:    tensor(0.8478, device='cuda:0', grad_fn=<NllLossBackward>)
[3786] Train Loss:    tensor(0.8719, device='cuda:0', grad_fn=<NllLossBackward>)
[3787] Train Loss:    tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[3788] Train Loss:    tensor(0.8511, device='cuda:0', grad_fn=<NllLossBackward>)
[3789] Train Loss:    tensor(0.8544, device='cuda:0', grad_fn=<NllLossBackward>)
[3790] Train Loss:    tensor(0.8650, device='cuda:0', grad_fn=<NllLossBackward>)
[3791] Train Loss:    tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[3792] Train Loss:    tensor(0.8669, device='cuda:0', grad_fn=<NllLossBackward>)
[3793] Train Loss:    tensor(0.8639, device='cuda:0', grad_fn=<NllLossBackward>)
[3794] Train Loss:    tensor(0.8659, device='cuda:0', grad_fn=<NllLossBackward>)
[3795] Train Loss:    tensor(0.8812, device='cuda:0', grad_fn=<NllLossBackward>)
[3796] Train Loss:    tensor(0.8759, device='cuda:0', grad_fn=<NllLossBackward>)
[3797] Train Loss:    tensor(0.9177, device='cuda:0', grad_fn=<NllLossBackward>)
[3798] Train Loss:    tensor(0.8623, device='cuda:0', grad_fn=<NllLossBackward>)
[3799] Train Loss:    tensor(0.8758, device='cuda:0', grad_fn=<NllLossBackward>)
[3800] Train Loss:    tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[3801] Train Loss:    tensor(0.8521, device='cuda:0', grad_fn=<NllLossBackward>)
[3802] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[3803] Train Loss:    tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[3804] Train Loss:    tensor(0.8663, device='cuda:0', grad_fn=<NllLossBackward>)
[3805] Train Loss:    tensor(0.8718, device='cuda:0', grad_fn=<NllLossBackward>)
[3806] Train Loss:    tensor(0.8660, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3807] Train Loss:   tensor(0.8514, device='cuda:0', grad_fn=<NllLossBackward>)
[3808] Train Loss:   tensor(0.8869, device='cuda:0', grad_fn=<NllLossBackward>)
[3809] Train Loss:   tensor(0.8674, device='cuda:0', grad_fn=<NllLossBackward>)
[3810] Train Loss:   tensor(0.8709, device='cuda:0', grad_fn=<NllLossBackward>)
[3811] Train Loss:   tensor(0.8940, device='cuda:0', grad_fn=<NllLossBackward>)
[3812] Train Loss:   tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[3813] Train Loss:   tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[3814] Train Loss:   tensor(0.9022, device='cuda:0', grad_fn=<NllLossBackward>)
[3815] Train Loss:   tensor(0.8701, device='cuda:0', grad_fn=<NllLossBackward>)
[3816] Train Loss:   tensor(0.9048, device='cuda:0', grad_fn=<NllLossBackward>)
[3817] Train Loss:   tensor(0.8574, device='cuda:0', grad_fn=<NllLossBackward>)
[3818] Train Loss:   tensor(0.8859, device='cuda:0', grad_fn=<NllLossBackward>)
[3819] Train Loss:   tensor(0.8774, device='cuda:0', grad_fn=<NllLossBackward>)
[3820] Train Loss:   tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[3821] Train Loss:   tensor(0.8704, device='cuda:0', grad_fn=<NllLossBackward>)
[3822] Train Loss:   tensor(0.8910, device='cuda:0', grad_fn=<NllLossBackward>)
[3823] Train Loss:   tensor(0.8919, device='cuda:0', grad_fn=<NllLossBackward>)
[3824] Train Loss:   tensor(0.8556, device='cuda:0', grad_fn=<NllLossBackward>)
[3825] Train Loss:   tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[3826] Train Loss:   tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[3827] Train Loss:   tensor(0.9022, device='cuda:0', grad_fn=<NllLossBackward>)
[3828] Train Loss:   tensor(0.8762, device='cuda:0', grad_fn=<NllLossBackward>)
[3829] Train Loss:   tensor(0.8861, device='cuda:0', grad_fn=<NllLossBackward>)
[3830] Train Loss:   tensor(0.8649, device='cuda:0', grad_fn=<NllLossBackward>)
[3831] Train Loss:   tensor(0.8437, device='cuda:0', grad_fn=<NllLossBackward>)
[3832] Train Loss:   tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[3833] Train Loss:   tensor(0.8571, device='cuda:0', grad_fn=<NllLossBackward>)
[3834] Train Loss:   tensor(0.8552, device='cuda:0', grad_fn=<NllLossBackward>)
[3835] Train Loss:   tensor(0.8735, device='cuda:0', grad_fn=<NllLossBackward>)
[3836] Train Loss:   tensor(0.8598, device='cuda:0', grad_fn=<NllLossBackward>)
[3837] Train Loss:   tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[3838] Train Loss:   tensor(0.9146, device='cuda:0', grad_fn=<NllLossBackward>)
[3839] Train Loss:   tensor(0.9025, device='cuda:0', grad_fn=<NllLossBackward>)
[3840] Train Loss:   tensor(0.8987, device='cuda:0', grad_fn=<NllLossBackward>)
[3841] Train Loss:   tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[3842] Train Loss:   tensor(0.8610, device='cuda:0', grad_fn=<NllLossBackward>)
[3843] Train Loss:   tensor(0.8635, device='cuda:0', grad_fn=<NllLossBackward>)
[3844] Train Loss:   tensor(0.8684, device='cuda:0', grad_fn=<NllLossBackward>)
[3845] Train Loss:   tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[3846] Train Loss:   tensor(0.8665, device='cuda:0', grad_fn=<NllLossBackward>)
[3847] Train Loss:   tensor(0.8712, device='cuda:0', grad_fn=<NllLossBackward>)
[3848] Train Loss:   tensor(0.8838, device='cuda:0', grad_fn=<NllLossBackward>)
[3849] Train Loss:   tensor(0.8898, device='cuda:0', grad_fn=<NllLossBackward>)
[3850] Train Loss:   tensor(0.8644, device='cuda:0', grad_fn=<NllLossBackward>)
[3851] Train Loss:   tensor(0.8501, device='cuda:0', grad_fn=<NllLossBackward>)
[3852] Train Loss:   tensor(0.8857, device='cuda:0', grad_fn=<NllLossBackward>)
[3853] Train Loss:   tensor(0.8506, device='cuda:0', grad_fn=<NllLossBackward>)
[3854] Train Loss:   tensor(0.8768, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3855] Train Loss:   tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[3856] Train Loss:   tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[3857] Train Loss:   tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[3858] Train Loss:   tensor(0.8360, device='cuda:0', grad_fn=<NllLossBackward>)
[3859] Train Loss:   tensor(0.8762, device='cuda:0', grad_fn=<NllLossBackward>)
[3860] Train Loss:   tensor(0.8862, device='cuda:0', grad_fn=<NllLossBackward>)
[3861] Train Loss:   tensor(0.8809, device='cuda:0', grad_fn=<NllLossBackward>)
[3862] Train Loss:   tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[3863] Train Loss:   tensor(0.8933, device='cuda:0', grad_fn=<NllLossBackward>)
[3864] Train Loss:   tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[3865] Train Loss:   tensor(0.8899, device='cuda:0', grad_fn=<NllLossBackward>)
[3866] Train Loss:   tensor(0.8579, device='cuda:0', grad_fn=<NllLossBackward>)
[3867] Train Loss:   tensor(0.8439, device='cuda:0', grad_fn=<NllLossBackward>)
[3868] Train Loss:   tensor(0.8418, device='cuda:0', grad_fn=<NllLossBackward>)
[3869] Train Loss:   tensor(0.8497, device='cuda:0', grad_fn=<NllLossBackward>)
[3870] Train Loss:   tensor(0.8631, device='cuda:0', grad_fn=<NllLossBackward>)
[3871] Train Loss:   tensor(0.8447, device='cuda:0', grad_fn=<NllLossBackward>)
[3872] Train Loss:   tensor(0.8662, device='cuda:0', grad_fn=<NllLossBackward>)
[3873] Train Loss:   tensor(0.8347, device='cuda:0', grad_fn=<NllLossBackward>)
[3874] Train Loss:   tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[3875] Train Loss:   tensor(0.8738, device='cuda:0', grad_fn=<NllLossBackward>)
[3876] Train Loss:   tensor(0.9155, device='cuda:0', grad_fn=<NllLossBackward>)
[3877] Train Loss:   tensor(0.8439, device='cuda:0', grad_fn=<NllLossBackward>)
[3878] Train Loss:   tensor(0.8772, device='cuda:0', grad_fn=<NllLossBackward>)
[3879] Train Loss:   tensor(0.8524, device='cuda:0', grad_fn=<NllLossBackward>)
[3880] Train Loss:   tensor(0.8783, device='cuda:0', grad_fn=<NllLossBackward>)
[3881] Train Loss:   tensor(0.8715, device='cuda:0', grad_fn=<NllLossBackward>)
[3882] Train Loss:   tensor(0.8562, device='cuda:0', grad_fn=<NllLossBackward>)
[3883] Train Loss:   tensor(0.8609, device='cuda:0', grad_fn=<NllLossBackward>)
[3884] Train Loss:   tensor(0.8743, device='cuda:0', grad_fn=<NllLossBackward>)
[3885] Train Loss:   tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[3886] Train Loss:   tensor(0.8523, device='cuda:0', grad_fn=<NllLossBackward>)
[3887] Train Loss:   tensor(0.8607, device='cuda:0', grad_fn=<NllLossBackward>)
[3888] Train Loss:   tensor(0.8669, device='cuda:0', grad_fn=<NllLossBackward>)
[3889] Train Loss:   tensor(0.8653, device='cuda:0', grad_fn=<NllLossBackward>)
[3890] Train Loss:   tensor(0.8872, device='cuda:0', grad_fn=<NllLossBackward>)
[3891] Train Loss:   tensor(0.8514, device='cuda:0', grad_fn=<NllLossBackward>)
[3892] Train Loss:   tensor(0.8473, device='cuda:0', grad_fn=<NllLossBackward>)
[3893] Train Loss:   tensor(0.8651, device='cuda:0', grad_fn=<NllLossBackward>)
[3894] Train Loss:   tensor(0.8657, device='cuda:0', grad_fn=<NllLossBackward>)
[3895] Train Loss:   tensor(0.8621, device='cuda:0', grad_fn=<NllLossBackward>)
[3896] Train Loss:   tensor(0.8768, device='cuda:0', grad_fn=<NllLossBackward>)
[3897] Train Loss:   tensor(0.8794, device='cuda:0', grad_fn=<NllLossBackward>)
[3898] Train Loss:   tensor(0.8632, device='cuda:0', grad_fn=<NllLossBackward>)
[3899] Train Loss:   tensor(0.8622, device='cuda:0', grad_fn=<NllLossBackward>)
[3900] Train Loss:   tensor(0.9124, device='cuda:0', grad_fn=<NllLossBackward>)
[3901] Train Loss:   tensor(0.8576, device='cuda:0', grad_fn=<NllLossBackward>)
[3902] Train Loss:   tensor(0.8760, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3903] Train Loss:    tensor(0.8931, device='cuda:0', grad_fn=<NllLossBackward>)
[3904] Train Loss:    tensor(0.8698, device='cuda:0', grad_fn=<NllLossBackward>)
[3905] Train Loss:    tensor(0.8707, device='cuda:0', grad_fn=<NllLossBackward>)
[3906] Train Loss:    tensor(0.8573, device='cuda:0', grad_fn=<NllLossBackward>)
[3907] Train Loss:    tensor(0.9154, device='cuda:0', grad_fn=<NllLossBackward>)
[3908] Train Loss:    tensor(0.8539, device='cuda:0', grad_fn=<NllLossBackward>)
[3909] Train Loss:    tensor(0.8440, device='cuda:0', grad_fn=<NllLossBackward>)
[3910] Train Loss:    tensor(0.8758, device='cuda:0', grad_fn=<NllLossBackward>)
[3911] Train Loss:    tensor(0.8586, device='cuda:0', grad_fn=<NllLossBackward>)
[3912] Train Loss:    tensor(0.8741, device='cuda:0', grad_fn=<NllLossBackward>)
[3913] Train Loss:    tensor(0.8541, device='cuda:0', grad_fn=<NllLossBackward>)
[3914] Train Loss:    tensor(0.8551, device='cuda:0', grad_fn=<NllLossBackward>)
[3915] Train Loss:    tensor(0.8736, device='cuda:0', grad_fn=<NllLossBackward>)
[3916] Train Loss:    tensor(0.8842, device='cuda:0', grad_fn=<NllLossBackward>)
[3917] Train Loss:    tensor(0.8769, device='cuda:0', grad_fn=<NllLossBackward>)
[3918] Train Loss:    tensor(0.8629, device='cuda:0', grad_fn=<NllLossBackward>)
[3919] Train Loss:    tensor(0.8962, device='cuda:0', grad_fn=<NllLossBackward>)
[3920] Train Loss:    tensor(0.8632, device='cuda:0', grad_fn=<NllLossBackward>)
[3921] Train Loss:    tensor(0.9135, device='cuda:0', grad_fn=<NllLossBackward>)
[3922] Train Loss:    tensor(0.9222, device='cuda:0', grad_fn=<NllLossBackward>)
[3923] Train Loss:    tensor(0.8508, device='cuda:0', grad_fn=<NllLossBackward>)
[3924] Train Loss:    tensor(0.8452, device='cuda:0', grad_fn=<NllLossBackward>)
[3925] Train Loss:    tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[3926] Train Loss:    tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[3927] Train Loss:    tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[3928] Train Loss:    tensor(0.8346, device='cuda:0', grad_fn=<NllLossBackward>)
[3929] Train Loss:    tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[3930] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[3931] Train Loss:    tensor(0.8551, device='cuda:0', grad_fn=<NllLossBackward>)
[3932] Train Loss:    tensor(0.8941, device='cuda:0', grad_fn=<NllLossBackward>)
[3933] Train Loss:    tensor(0.9071, device='cuda:0', grad_fn=<NllLossBackward>)
[3934] Train Loss:    tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[3935] Train Loss:    tensor(0.8666, device='cuda:0', grad_fn=<NllLossBackward>)
[3936] Train Loss:    tensor(0.8343, device='cuda:0', grad_fn=<NllLossBackward>)
[3937] Train Loss:    tensor(0.8787, device='cuda:0', grad_fn=<NllLossBackward>)
[3938] Train Loss:    tensor(0.8747, device='cuda:0', grad_fn=<NllLossBackward>)
[3939] Train Loss:    tensor(0.8628, device='cuda:0', grad_fn=<NllLossBackward>)
[3940] Train Loss:    tensor(0.8713, device='cuda:0', grad_fn=<NllLossBackward>)
[3941] Train Loss:    tensor(0.8893, device='cuda:0', grad_fn=<NllLossBackward>)
[3942] Train Loss:    tensor(0.8646, device='cuda:0', grad_fn=<NllLossBackward>)
[3943] Train Loss:    tensor(0.8343, device='cuda:0', grad_fn=<NllLossBackward>)
[3944] Train Loss:    tensor(0.8747, device='cuda:0', grad_fn=<NllLossBackward>)
[3945] Train Loss:    tensor(0.8689, device='cuda:0', grad_fn=<NllLossBackward>)
[3946] Train Loss:    tensor(0.8681, device='cuda:0', grad_fn=<NllLossBackward>)
[3947] Train Loss:    tensor(0.8922, device='cuda:0', grad_fn=<NllLossBackward>)
[3948] Train Loss:    tensor(0.8508, device='cuda:0', grad_fn=<NllLossBackward>)
[3949] Train Loss:    tensor(0.8553, device='cuda:0', grad_fn=<NllLossBackward>)
[3950] Train Loss:    tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3951] Train Loss:   tensor(0.8981, device='cuda:0', grad_fn=<NllLossBackward>)
[3952] Train Loss:   tensor(0.8388, device='cuda:0', grad_fn=<NllLossBackward>)
[3953] Train Loss:   tensor(0.8692, device='cuda:0', grad_fn=<NllLossBackward>)
[3954] Train Loss:   tensor(0.8453, device='cuda:0', grad_fn=<NllLossBackward>)
[3955] Train Loss:   tensor(0.8819, device='cuda:0', grad_fn=<NllLossBackward>)
[3956] Train Loss:   tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[3957] Train Loss:   tensor(0.9233, device='cuda:0', grad_fn=<NllLossBackward>)
[3958] Train Loss:   tensor(0.8816, device='cuda:0', grad_fn=<NllLossBackward>)
[3959] Train Loss:   tensor(0.8643, device='cuda:0', grad_fn=<NllLossBackward>)
[3960] Train Loss:   tensor(0.8436, device='cuda:0', grad_fn=<NllLossBackward>)
[3961] Train Loss:   tensor(0.8722, device='cuda:0', grad_fn=<NllLossBackward>)
[3962] Train Loss:   tensor(0.8712, device='cuda:0', grad_fn=<NllLossBackward>)
[3963] Train Loss:   tensor(0.8886, device='cuda:0', grad_fn=<NllLossBackward>)
[3964] Train Loss:   tensor(0.8621, device='cuda:0', grad_fn=<NllLossBackward>)
[3965] Train Loss:   tensor(0.8526, device='cuda:0', grad_fn=<NllLossBackward>)
[3966] Train Loss:   tensor(0.8418, device='cuda:0', grad_fn=<NllLossBackward>)
[3967] Train Loss:   tensor(0.8897, device='cuda:0', grad_fn=<NllLossBackward>)
[3968] Train Loss:   tensor(0.8817, device='cuda:0', grad_fn=<NllLossBackward>)
[3969] Train Loss:   tensor(0.8686, device='cuda:0', grad_fn=<NllLossBackward>)
[3970] Train Loss:   tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
[3971] Train Loss:   tensor(0.8718, device='cuda:0', grad_fn=<NllLossBackward>)
[3972] Train Loss:   tensor(0.8714, device='cuda:0', grad_fn=<NllLossBackward>)
[3973] Train Loss:   tensor(0.8754, device='cuda:0', grad_fn=<NllLossBackward>)
[3974] Train Loss:   tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[3975] Train Loss:   tensor(0.8779, device='cuda:0', grad_fn=<NllLossBackward>)
[3976] Train Loss:   tensor(0.8755, device='cuda:0', grad_fn=<NllLossBackward>)
[3977] Train Loss:   tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[3978] Train Loss:   tensor(0.8666, device='cuda:0', grad_fn=<NllLossBackward>)
[3979] Train Loss:   tensor(0.8959, device='cuda:0', grad_fn=<NllLossBackward>)
[3980] Train Loss:   tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[3981] Train Loss:   tensor(0.8635, device='cuda:0', grad_fn=<NllLossBackward>)
[3982] Train Loss:   tensor(0.8906, device='cuda:0', grad_fn=<NllLossBackward>)
[3983] Train Loss:   tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[3984] Train Loss:   tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[3985] Train Loss:   tensor(0.8756, device='cuda:0', grad_fn=<NllLossBackward>)
[3986] Train Loss:   tensor(0.8574, device='cuda:0', grad_fn=<NllLossBackward>)
[3987] Train Loss:   tensor(0.9180, device='cuda:0', grad_fn=<NllLossBackward>)
[3988] Train Loss:   tensor(0.8730, device='cuda:0', grad_fn=<NllLossBackward>)
[3989] Train Loss:   tensor(0.8575, device='cuda:0', grad_fn=<NllLossBackward>)
[3990] Train Loss:   tensor(0.8589, device='cuda:0', grad_fn=<NllLossBackward>)
[3991] Train Loss:   tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
[3992] Train Loss:   tensor(0.9021, device='cuda:0', grad_fn=<NllLossBackward>)
[3993] Train Loss:   tensor(0.8374, device='cuda:0', grad_fn=<NllLossBackward>)
[3994] Train Loss:   tensor(0.8644, device='cuda:0', grad_fn=<NllLossBackward>)
[3995] Train Loss:   tensor(0.8636, device='cuda:0', grad_fn=<NllLossBackward>)
[3996] Train Loss:   tensor(0.8687, device='cuda:0', grad_fn=<NllLossBackward>)
[3997] Train Loss:   tensor(0.8658, device='cuda:0', grad_fn=<NllLossBackward>)
[3998] Train Loss:   tensor(0.8458, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3999] Train Loss:   tensor(0.8948, device='cuda:0', grad_fn=<NllLossBackward>)
[4000] Train Loss:   tensor(0.8780, device='cuda:0', grad_fn=<NllLossBackward>)
[4001] Train Loss:   tensor(0.8804, device='cuda:0', grad_fn=<NllLossBackward>)
[4002] Train Loss:   tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[4003] Train Loss:   tensor(0.8374, device='cuda:0', grad_fn=<NllLossBackward>)
[4004] Train Loss:   tensor(0.8546, device='cuda:0', grad_fn=<NllLossBackward>)
[4005] Train Loss:   tensor(0.8512, device='cuda:0', grad_fn=<NllLossBackward>)
[4006] Train Loss:   tensor(0.8908, device='cuda:0', grad_fn=<NllLossBackward>)
[4007] Train Loss:   tensor(0.8459, device='cuda:0', grad_fn=<NllLossBackward>)
[4008] Train Loss:   tensor(0.8315, device='cuda:0', grad_fn=<NllLossBackward>)
[4009] Train Loss:   tensor(0.8748, device='cuda:0', grad_fn=<NllLossBackward>)
[4010] Train Loss:   tensor(0.8531, device='cuda:0', grad_fn=<NllLossBackward>)
[4011] Train Loss:   tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[4012] Train Loss:   tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[4013] Train Loss:   tensor(0.8862, device='cuda:0', grad_fn=<NllLossBackward>)
[4014] Train Loss:   tensor(0.8485, device='cuda:0', grad_fn=<NllLossBackward>)
[4015] Train Loss:   tensor(0.8421, device='cuda:0', grad_fn=<NllLossBackward>)
[4016] Train Loss:   tensor(0.8786, device='cuda:0', grad_fn=<NllLossBackward>)
[4017] Train Loss:   tensor(0.8669, device='cuda:0', grad_fn=<NllLossBackward>)
[4018] Train Loss:   tensor(0.8584, device='cuda:0', grad_fn=<NllLossBackward>)
[4019] Train Loss:   tensor(0.8320, device='cuda:0', grad_fn=<NllLossBackward>)
[4020] Train Loss:   tensor(0.8838, device='cuda:0', grad_fn=<NllLossBackward>)
[4021] Train Loss:   tensor(0.8085, device='cuda:0', grad_fn=<NllLossBackward>)
[4022] Train Loss:   tensor(0.8454, device='cuda:0', grad_fn=<NllLossBackward>)
[4023] Train Loss:   tensor(0.8874, device='cuda:0', grad_fn=<NllLossBackward>)
[4024] Train Loss:   tensor(0.8451, device='cuda:0', grad_fn=<NllLossBackward>)
[4025] Train Loss:   tensor(0.8499, device='cuda:0', grad_fn=<NllLossBackward>)
[4026] Train Loss:   tensor(0.8792, device='cuda:0', grad_fn=<NllLossBackward>)
[4027] Train Loss:   tensor(0.8683, device='cuda:0', grad_fn=<NllLossBackward>)
[4028] Train Loss:   tensor(0.8334, device='cuda:0', grad_fn=<NllLossBackward>)
[4029] Train Loss:   tensor(0.9141, device='cuda:0', grad_fn=<NllLossBackward>)
[4030] Train Loss:   tensor(0.8782, device='cuda:0', grad_fn=<NllLossBackward>)
[4031] Train Loss:   tensor(0.9068, device='cuda:0', grad_fn=<NllLossBackward>)
[4032] Train Loss:   tensor(0.8979, device='cuda:0', grad_fn=<NllLossBackward>)
[4033] Train Loss:   tensor(0.8387, device='cuda:0', grad_fn=<NllLossBackward>)
[4034] Train Loss:   tensor(0.8422, device='cuda:0', grad_fn=<NllLossBackward>)
[4035] Train Loss:   tensor(0.8630, device='cuda:0', grad_fn=<NllLossBackward>)
[4036] Train Loss:   tensor(0.8575, device='cuda:0', grad_fn=<NllLossBackward>)
[4037] Train Loss:   tensor(0.8356, device='cuda:0', grad_fn=<NllLossBackward>)
[4038] Train Loss:   tensor(0.8628, device='cuda:0', grad_fn=<NllLossBackward>)
[4039] Train Loss:   tensor(0.8814, device='cuda:0', grad_fn=<NllLossBackward>)
[4040] Train Loss:   tensor(0.8626, device='cuda:0', grad_fn=<NllLossBackward>)
[4041] Train Loss:   tensor(0.8584, device='cuda:0', grad_fn=<NllLossBackward>)
[4042] Train Loss:   tensor(0.8960, device='cuda:0', grad_fn=<NllLossBackward>)
[4043] Train Loss:   tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[4044] Train Loss:   tensor(0.8641, device='cuda:0', grad_fn=<NllLossBackward>)
[4045] Train Loss:   tensor(0.8556, device='cuda:0', grad_fn=<NllLossBackward>)
[4046] Train Loss:   tensor(0.9007, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4047] Train Loss:    tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[4048] Train Loss:    tensor(0.8711, device='cuda:0', grad_fn=<NllLossBackward>)
[4049] Train Loss:    tensor(0.8984, device='cuda:0', grad_fn=<NllLossBackward>)
[4050] Train Loss:    tensor(0.8687, device='cuda:0', grad_fn=<NllLossBackward>)
[4051] Train Loss:    tensor(0.8641, device='cuda:0', grad_fn=<NllLossBackward>)
[4052] Train Loss:    tensor(0.8553, device='cuda:0', grad_fn=<NllLossBackward>)
[4053] Train Loss:    tensor(0.8407, device='cuda:0', grad_fn=<NllLossBackward>)
[4054] Train Loss:    tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[4055] Train Loss:    tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[4056] Train Loss:    tensor(0.8465, device='cuda:0', grad_fn=<NllLossBackward>)
[4057] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[4058] Train Loss:    tensor(0.8780, device='cuda:0', grad_fn=<NllLossBackward>)
[4059] Train Loss:    tensor(0.8968, device='cuda:0', grad_fn=<NllLossBackward>)
[4060] Train Loss:    tensor(0.8607, device='cuda:0', grad_fn=<NllLossBackward>)
[4061] Train Loss:    tensor(0.8473, device='cuda:0', grad_fn=<NllLossBackward>)
[4062] Train Loss:    tensor(0.8390, device='cuda:0', grad_fn=<NllLossBackward>)
[4063] Train Loss:    tensor(0.8518, device='cuda:0', grad_fn=<NllLossBackward>)
[4064] Train Loss:    tensor(0.8815, device='cuda:0', grad_fn=<NllLossBackward>)
[4065] Train Loss:    tensor(0.8801, device='cuda:0', grad_fn=<NllLossBackward>)
[4066] Train Loss:    tensor(0.8666, device='cuda:0', grad_fn=<NllLossBackward>)
[4067] Train Loss:    tensor(0.8330, device='cuda:0', grad_fn=<NllLossBackward>)
[4068] Train Loss:    tensor(0.9052, device='cuda:0', grad_fn=<NllLossBackward>)
[4069] Train Loss:    tensor(0.8998, device='cuda:0', grad_fn=<NllLossBackward>)
[4070] Train Loss:    tensor(0.8584, device='cuda:0', grad_fn=<NllLossBackward>)
[4071] Train Loss:    tensor(0.8232, device='cuda:0', grad_fn=<NllLossBackward>)
[4072] Train Loss:    tensor(0.8745, device='cuda:0', grad_fn=<NllLossBackward>)
[4073] Train Loss:    tensor(0.8698, device='cuda:0', grad_fn=<NllLossBackward>)
[4074] Train Loss:    tensor(0.8666, device='cuda:0', grad_fn=<NllLossBackward>)
[4075] Train Loss:    tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[4076] Train Loss:    tensor(0.8825, device='cuda:0', grad_fn=<NllLossBackward>)
[4077] Train Loss:    tensor(0.8263, device='cuda:0', grad_fn=<NllLossBackward>)
[4078] Train Loss:    tensor(0.8408, device='cuda:0', grad_fn=<NllLossBackward>)
[4079] Train Loss:    tensor(0.8606, device='cuda:0', grad_fn=<NllLossBackward>)
[4080] Train Loss:    tensor(0.8820, device='cuda:0', grad_fn=<NllLossBackward>)
[4081] Train Loss:    tensor(0.8499, device='cuda:0', grad_fn=<NllLossBackward>)
[4082] Train Loss:    tensor(0.8439, device='cuda:0', grad_fn=<NllLossBackward>)
[4083] Train Loss:    tensor(0.8954, device='cuda:0', grad_fn=<NllLossBackward>)
[4084] Train Loss:    tensor(0.8709, device='cuda:0', grad_fn=<NllLossBackward>)
[4085] Train Loss:    tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[4086] Train Loss:    tensor(0.8819, device='cuda:0', grad_fn=<NllLossBackward>)
[4087] Train Loss:    tensor(0.8903, device='cuda:0', grad_fn=<NllLossBackward>)
[4088] Train Loss:    tensor(0.8760, device='cuda:0', grad_fn=<NllLossBackward>)
[4089] Train Loss:    tensor(0.8551, device='cuda:0', grad_fn=<NllLossBackward>)
[4090] Train Loss:    tensor(0.8643, device='cuda:0', grad_fn=<NllLossBackward>)
[4091] Train Loss:    tensor(0.8865, device='cuda:0', grad_fn=<NllLossBackward>)
[4092] Train Loss:    tensor(0.8574, device='cuda:0', grad_fn=<NllLossBackward>)
[4093] Train Loss:    tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[4094] Train Loss:    tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4095] Train Loss:   tensor(0.8631, device='cuda:0', grad_fn=<NllLossBackward>)
[4096] Train Loss:   tensor(0.8588, device='cuda:0', grad_fn=<NllLossBackward>)
[4097] Train Loss:   tensor(0.8765, device='cuda:0', grad_fn=<NllLossBackward>)
[4098] Train Loss:   tensor(0.8436, device='cuda:0', grad_fn=<NllLossBackward>)
[4099] Train Loss:   tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[4100] Train Loss:   tensor(0.8361, device='cuda:0', grad_fn=<NllLossBackward>)
[4101] Train Loss:   tensor(0.8806, device='cuda:0', grad_fn=<NllLossBackward>)
[4102] Train Loss:   tensor(0.8983, device='cuda:0', grad_fn=<NllLossBackward>)
[4103] Train Loss:   tensor(0.8404, device='cuda:0', grad_fn=<NllLossBackward>)
[4104] Train Loss:   tensor(0.8810, device='cuda:0', grad_fn=<NllLossBackward>)
[4105] Train Loss:   tensor(0.8562, device='cuda:0', grad_fn=<NllLossBackward>)
[4106] Train Loss:   tensor(0.8737, device='cuda:0', grad_fn=<NllLossBackward>)
[4107] Train Loss:   tensor(0.8446, device='cuda:0', grad_fn=<NllLossBackward>)
[4108] Train Loss:   tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[4109] Train Loss:   tensor(0.8450, device='cuda:0', grad_fn=<NllLossBackward>)
[4110] Train Loss:   tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[4111] Train Loss:   tensor(0.8735, device='cuda:0', grad_fn=<NllLossBackward>)
[4112] Train Loss:   tensor(0.8491, device='cuda:0', grad_fn=<NllLossBackward>)
[4113] Train Loss:   tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[4114] Train Loss:   tensor(0.8479, device='cuda:0', grad_fn=<NllLossBackward>)
[4115] Train Loss:   tensor(0.8367, device='cuda:0', grad_fn=<NllLossBackward>)
[4116] Train Loss:   tensor(0.8348, device='cuda:0', grad_fn=<NllLossBackward>)
[4117] Train Loss:   tensor(0.8700, device='cuda:0', grad_fn=<NllLossBackward>)
[4118] Train Loss:   tensor(0.8773, device='cuda:0', grad_fn=<NllLossBackward>)
[4119] Train Loss:   tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[4120] Train Loss:   tensor(0.8343, device='cuda:0', grad_fn=<NllLossBackward>)
[4121] Train Loss:   tensor(0.8411, device='cuda:0', grad_fn=<NllLossBackward>)
[4122] Train Loss:   tensor(0.8693, device='cuda:0', grad_fn=<NllLossBackward>)
[4123] Train Loss:   tensor(0.8717, device='cuda:0', grad_fn=<NllLossBackward>)
[4124] Train Loss:   tensor(0.8791, device='cuda:0', grad_fn=<NllLossBackward>)
[4125] Train Loss:   tensor(0.8445, device='cuda:0', grad_fn=<NllLossBackward>)
[4126] Train Loss:   tensor(0.8827, device='cuda:0', grad_fn=<NllLossBackward>)
[4127] Train Loss:   tensor(0.8446, device='cuda:0', grad_fn=<NllLossBackward>)
[4128] Train Loss:   tensor(0.8696, device='cuda:0', grad_fn=<NllLossBackward>)
[4129] Train Loss:   tensor(0.8652, device='cuda:0', grad_fn=<NllLossBackward>)
[4130] Train Loss:   tensor(0.8842, device='cuda:0', grad_fn=<NllLossBackward>)
[4131] Train Loss:   tensor(0.8965, device='cuda:0', grad_fn=<NllLossBackward>)
[4132] Train Loss:   tensor(0.9042, device='cuda:0', grad_fn=<NllLossBackward>)
[4133] Train Loss:   tensor(0.8610, device='cuda:0', grad_fn=<NllLossBackward>)
[4134] Train Loss:   tensor(0.8616, device='cuda:0', grad_fn=<NllLossBackward>)
[4135] Train Loss:   tensor(0.8915, device='cuda:0', grad_fn=<NllLossBackward>)
[4136] Train Loss:   tensor(0.8677, device='cuda:0', grad_fn=<NllLossBackward>)
[4137] Train Loss:   tensor(0.8504, device='cuda:0', grad_fn=<NllLossBackward>)
[4138] Train Loss:   tensor(0.8819, device='cuda:0', grad_fn=<NllLossBackward>)
[4139] Train Loss:   tensor(0.8460, device='cuda:0', grad_fn=<NllLossBackward>)
[4140] Train Loss:   tensor(0.8244, device='cuda:0', grad_fn=<NllLossBackward>)
[4141] Train Loss:   tensor(0.8603, device='cuda:0', grad_fn=<NllLossBackward>)
[4142] Train Loss:   tensor(0.8942, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4143] Train Loss:    tensor(0.8641, device='cuda:0', grad_fn=<NllLossBackward>)
[4144] Train Loss:    tensor(0.8589, device='cuda:0', grad_fn=<NllLossBackward>)
[4145] Train Loss:    tensor(0.8378, device='cuda:0', grad_fn=<NllLossBackward>)
[4146] Train Loss:    tensor(0.9053, device='cuda:0', grad_fn=<NllLossBackward>)
[4147] Train Loss:    tensor(0.8565, device='cuda:0', grad_fn=<NllLossBackward>)
[4148] Train Loss:    tensor(0.8539, device='cuda:0', grad_fn=<NllLossBackward>)
[4149] Train Loss:    tensor(0.8996, device='cuda:0', grad_fn=<NllLossBackward>)
[4150] Train Loss:    tensor(0.8784, device='cuda:0', grad_fn=<NllLossBackward>)
[4151] Train Loss:    tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[4152] Train Loss:    tensor(0.8228, device='cuda:0', grad_fn=<NllLossBackward>)
[4153] Train Loss:    tensor(0.8310, device='cuda:0', grad_fn=<NllLossBackward>)
[4154] Train Loss:    tensor(0.8511, device='cuda:0', grad_fn=<NllLossBackward>)
[4155] Train Loss:    tensor(0.8797, device='cuda:0', grad_fn=<NllLossBackward>)
[4156] Train Loss:    tensor(0.8601, device='cuda:0', grad_fn=<NllLossBackward>)
[4157] Train Loss:    tensor(0.8563, device='cuda:0', grad_fn=<NllLossBackward>)
[4158] Train Loss:    tensor(0.8687, device='cuda:0', grad_fn=<NllLossBackward>)
[4159] Train Loss:    tensor(0.8669, device='cuda:0', grad_fn=<NllLossBackward>)
[4160] Train Loss:    tensor(0.8748, device='cuda:0', grad_fn=<NllLossBackward>)
[4161] Train Loss:    tensor(0.8685, device='cuda:0', grad_fn=<NllLossBackward>)
[4162] Train Loss:    tensor(0.8796, device='cuda:0', grad_fn=<NllLossBackward>)
[4163] Train Loss:    tensor(0.8761, device='cuda:0', grad_fn=<NllLossBackward>)
[4164] Train Loss:    tensor(0.8666, device='cuda:0', grad_fn=<NllLossBackward>)
[4165] Train Loss:    tensor(0.8540, device='cuda:0', grad_fn=<NllLossBackward>)
[4166] Train Loss:    tensor(0.8882, device='cuda:0', grad_fn=<NllLossBackward>)
[4167] Train Loss:    tensor(0.8623, device='cuda:0', grad_fn=<NllLossBackward>)
[4168] Train Loss:    tensor(0.8568, device='cuda:0', grad_fn=<NllLossBackward>)
[4169] Train Loss:    tensor(0.8496, device='cuda:0', grad_fn=<NllLossBackward>)
[4170] Train Loss:    tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[4171] Train Loss:    tensor(0.8383, device='cuda:0', grad_fn=<NllLossBackward>)
[4172] Train Loss:    tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[4173] Train Loss:    tensor(0.8748, device='cuda:0', grad_fn=<NllLossBackward>)
[4174] Train Loss:    tensor(0.8919, device='cuda:0', grad_fn=<NllLossBackward>)
[4175] Train Loss:    tensor(0.8412, device='cuda:0', grad_fn=<NllLossBackward>)
[4176] Train Loss:    tensor(0.8533, device='cuda:0', grad_fn=<NllLossBackward>)
[4177] Train Loss:    tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[4178] Train Loss:    tensor(0.8437, device='cuda:0', grad_fn=<NllLossBackward>)
[4179] Train Loss:    tensor(0.8997, device='cuda:0', grad_fn=<NllLossBackward>)
[4180] Train Loss:    tensor(0.8870, device='cuda:0', grad_fn=<NllLossBackward>)
[4181] Train Loss:    tensor(0.8414, device='cuda:0', grad_fn=<NllLossBackward>)
[4182] Train Loss:    tensor(0.9067, device='cuda:0', grad_fn=<NllLossBackward>)
[4183] Train Loss:    tensor(0.8617, device='cuda:0', grad_fn=<NllLossBackward>)
[4184] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[4185] Train Loss:    tensor(0.8664, device='cuda:0', grad_fn=<NllLossBackward>)
[4186] Train Loss:    tensor(0.8695, device='cuda:0', grad_fn=<NllLossBackward>)
[4187] Train Loss:    tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[4188] Train Loss:    tensor(0.8748, device='cuda:0', grad_fn=<NllLossBackward>)
[4189] Train Loss:    tensor(0.8944, device='cuda:0', grad_fn=<NllLossBackward>)
[4190] Train Loss:    tensor(0.9025, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4191] Train Loss:   tensor(0.8671, device='cuda:0', grad_fn=<NllLossBackward>)
[4192] Train Loss:   tensor(0.8532, device='cuda:0', grad_fn=<NllLossBackward>)
[4193] Train Loss:   tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[4194] Train Loss:   tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[4195] Train Loss:   tensor(0.8624, device='cuda:0', grad_fn=<NllLossBackward>)
[4196] Train Loss:   tensor(0.8356, device='cuda:0', grad_fn=<NllLossBackward>)
[4197] Train Loss:   tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[4198] Train Loss:   tensor(0.8730, device='cuda:0', grad_fn=<NllLossBackward>)
[4199] Train Loss:   tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[4200] Train Loss:   tensor(0.8893, device='cuda:0', grad_fn=<NllLossBackward>)
[4201] Train Loss:   tensor(0.8327, device='cuda:0', grad_fn=<NllLossBackward>)
[4202] Train Loss:   tensor(0.8586, device='cuda:0', grad_fn=<NllLossBackward>)
[4203] Train Loss:   tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[4204] Train Loss:   tensor(0.8621, device='cuda:0', grad_fn=<NllLossBackward>)
[4205] Train Loss:   tensor(0.8757, device='cuda:0', grad_fn=<NllLossBackward>)
[4206] Train Loss:   tensor(0.8669, device='cuda:0', grad_fn=<NllLossBackward>)
[4207] Train Loss:   tensor(0.8307, device='cuda:0', grad_fn=<NllLossBackward>)
[4208] Train Loss:   tensor(0.8975, device='cuda:0', grad_fn=<NllLossBackward>)
[4209] Train Loss:   tensor(0.9140, device='cuda:0', grad_fn=<NllLossBackward>)
[4210] Train Loss:   tensor(0.8689, device='cuda:0', grad_fn=<NllLossBackward>)
[4211] Train Loss:   tensor(0.8445, device='cuda:0', grad_fn=<NllLossBackward>)
[4212] Train Loss:   tensor(0.8589, device='cuda:0', grad_fn=<NllLossBackward>)
[4213] Train Loss:   tensor(0.9035, device='cuda:0', grad_fn=<NllLossBackward>)
[4214] Train Loss:   tensor(0.8651, device='cuda:0', grad_fn=<NllLossBackward>)
[4215] Train Loss:   tensor(0.8756, device='cuda:0', grad_fn=<NllLossBackward>)
[4216] Train Loss:   tensor(0.8749, device='cuda:0', grad_fn=<NllLossBackward>)
[4217] Train Loss:   tensor(0.8742, device='cuda:0', grad_fn=<NllLossBackward>)
[4218] Train Loss:   tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[4219] Train Loss:   tensor(0.8556, device='cuda:0', grad_fn=<NllLossBackward>)
[4220] Train Loss:   tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
[4221] Train Loss:   tensor(0.8510, device='cuda:0', grad_fn=<NllLossBackward>)
[4222] Train Loss:   tensor(0.8715, device='cuda:0', grad_fn=<NllLossBackward>)
[4223] Train Loss:   tensor(0.8793, device='cuda:0', grad_fn=<NllLossBackward>)
[4224] Train Loss:   tensor(0.8450, device='cuda:0', grad_fn=<NllLossBackward>)
[4225] Train Loss:   tensor(0.8809, device='cuda:0', grad_fn=<NllLossBackward>)
[4226] Train Loss:   tensor(0.8588, device='cuda:0', grad_fn=<NllLossBackward>)
[4227] Train Loss:   tensor(0.8663, device='cuda:0', grad_fn=<NllLossBackward>)
[4228] Train Loss:   tensor(0.8413, device='cuda:0', grad_fn=<NllLossBackward>)
[4229] Train Loss:   tensor(0.9066, device='cuda:0', grad_fn=<NllLossBackward>)
[4230] Train Loss:   tensor(0.8512, device='cuda:0', grad_fn=<NllLossBackward>)
[4231] Train Loss:   tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[4232] Train Loss:   tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[4233] Train Loss:   tensor(0.8803, device='cuda:0', grad_fn=<NllLossBackward>)
[4234] Train Loss:   tensor(0.8616, device='cuda:0', grad_fn=<NllLossBackward>)
[4235] Train Loss:   tensor(0.8870, device='cuda:0', grad_fn=<NllLossBackward>)
[4236] Train Loss:   tensor(0.8827, device='cuda:0', grad_fn=<NllLossBackward>)
[4237] Train Loss:   tensor(0.8505, device='cuda:0', grad_fn=<NllLossBackward>)
[4238] Train Loss:   tensor(0.8337, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4239] Train Loss:    tensor(0.8762, device='cuda:0', grad_fn=<NllLossBackward>)
[4240] Train Loss:    tensor(0.8553, device='cuda:0', grad_fn=<NllLossBackward>)
[4241] Train Loss:    tensor(0.8688, device='cuda:0', grad_fn=<NllLossBackward>)
[4242] Train Loss:    tensor(0.8597, device='cuda:0', grad_fn=<NllLossBackward>)
[4243] Train Loss:    tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[4244] Train Loss:    tensor(0.8659, device='cuda:0', grad_fn=<NllLossBackward>)
[4245] Train Loss:    tensor(0.8596, device='cuda:0', grad_fn=<NllLossBackward>)
[4246] Train Loss:    tensor(0.8964, device='cuda:0', grad_fn=<NllLossBackward>)
[4247] Train Loss:    tensor(0.8657, device='cuda:0', grad_fn=<NllLossBackward>)
[4248] Train Loss:    tensor(0.9120, device='cuda:0', grad_fn=<NllLossBackward>)
[4249] Train Loss:    tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[4250] Train Loss:    tensor(0.8600, device='cuda:0', grad_fn=<NllLossBackward>)
[4251] Train Loss:    tensor(0.8631, device='cuda:0', grad_fn=<NllLossBackward>)
[4252] Train Loss:    tensor(0.8645, device='cuda:0', grad_fn=<NllLossBackward>)
[4253] Train Loss:    tensor(0.8726, device='cuda:0', grad_fn=<NllLossBackward>)
[4254] Train Loss:    tensor(0.8711, device='cuda:0', grad_fn=<NllLossBackward>)
[4255] Train Loss:    tensor(0.8928, device='cuda:0', grad_fn=<NllLossBackward>)
[4256] Train Loss:    tensor(0.8825, device='cuda:0', grad_fn=<NllLossBackward>)
[4257] Train Loss:    tensor(0.8799, device='cuda:0', grad_fn=<NllLossBackward>)
[4258] Train Loss:    tensor(0.8794, device='cuda:0', grad_fn=<NllLossBackward>)
[4259] Train Loss:    tensor(0.8788, device='cuda:0', grad_fn=<NllLossBackward>)
[4260] Train Loss:    tensor(0.8629, device='cuda:0', grad_fn=<NllLossBackward>)
[4261] Train Loss:    tensor(0.8892, device='cuda:0', grad_fn=<NllLossBackward>)
[4262] Train Loss:    tensor(0.8307, device='cuda:0', grad_fn=<NllLossBackward>)
[4263] Train Loss:    tensor(0.8712, device='cuda:0', grad_fn=<NllLossBackward>)
[4264] Train Loss:    tensor(0.8452, device='cuda:0', grad_fn=<NllLossBackward>)
[4265] Train Loss:    tensor(0.8615, device='cuda:0', grad_fn=<NllLossBackward>)
[4266] Train Loss:    tensor(0.8488, device='cuda:0', grad_fn=<NllLossBackward>)
[4267] Train Loss:    tensor(0.8698, device='cuda:0', grad_fn=<NllLossBackward>)
[4268] Train Loss:    tensor(0.8767, device='cuda:0', grad_fn=<NllLossBackward>)
[4269] Train Loss:    tensor(0.8747, device='cuda:0', grad_fn=<NllLossBackward>)
[4270] Train Loss:    tensor(0.8680, device='cuda:0', grad_fn=<NllLossBackward>)
[4271] Train Loss:    tensor(0.8799, device='cuda:0', grad_fn=<NllLossBackward>)
[4272] Train Loss:    tensor(0.8663, device='cuda:0', grad_fn=<NllLossBackward>)
[4273] Train Loss:    tensor(0.8648, device='cuda:0', grad_fn=<NllLossBackward>)
[4274] Train Loss:    tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[4275] Train Loss:    tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[4276] Train Loss:    tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[4277] Train Loss:    tensor(0.8630, device='cuda:0', grad_fn=<NllLossBackward>)
[4278] Train Loss:    tensor(0.9135, device='cuda:0', grad_fn=<NllLossBackward>)
[4279] Train Loss:    tensor(0.8744, device='cuda:0', grad_fn=<NllLossBackward>)
[4280] Train Loss:    tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[4281] Train Loss:    tensor(0.8727, device='cuda:0', grad_fn=<NllLossBackward>)
[4282] Train Loss:    tensor(0.8690, device='cuda:0', grad_fn=<NllLossBackward>)
[4283] Train Loss:    tensor(0.8898, device='cuda:0', grad_fn=<NllLossBackward>)
[4284] Train Loss:    tensor(0.8852, device='cuda:0', grad_fn=<NllLossBackward>)
[4285] Train Loss:    tensor(0.8860, device='cuda:0', grad_fn=<NllLossBackward>)
[4286] Train Loss:    tensor(0.8803, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4287] Train Loss:   tensor(0.8684, device='cuda:0', grad_fn=<NllLossBackward>)
[4288] Train Loss:   tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
[4289] Train Loss:   tensor(0.8844, device='cuda:0', grad_fn=<NllLossBackward>)
[4290] Train Loss:   tensor(0.8379, device='cuda:0', grad_fn=<NllLossBackward>)
[4291] Train Loss:   tensor(0.8882, device='cuda:0', grad_fn=<NllLossBackward>)
[4292] Train Loss:   tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[4293] Train Loss:   tensor(0.8581, device='cuda:0', grad_fn=<NllLossBackward>)
[4294] Train Loss:   tensor(0.8808, device='cuda:0', grad_fn=<NllLossBackward>)
[4295] Train Loss:   tensor(0.8534, device='cuda:0', grad_fn=<NllLossBackward>)
[4296] Train Loss:   tensor(0.8509, device='cuda:0', grad_fn=<NllLossBackward>)
[4297] Train Loss:   tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[4298] Train Loss:   tensor(0.8703, device='cuda:0', grad_fn=<NllLossBackward>)
[4299] Train Loss:   tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[4300] Train Loss:   tensor(0.8745, device='cuda:0', grad_fn=<NllLossBackward>)
[4301] Train Loss:   tensor(0.8597, device='cuda:0', grad_fn=<NllLossBackward>)
[4302] Train Loss:   tensor(0.8740, device='cuda:0', grad_fn=<NllLossBackward>)
[4303] Train Loss:   tensor(0.8749, device='cuda:0', grad_fn=<NllLossBackward>)
[4304] Train Loss:   tensor(0.8540, device='cuda:0', grad_fn=<NllLossBackward>)
[4305] Train Loss:   tensor(0.8748, device='cuda:0', grad_fn=<NllLossBackward>)
[4306] Train Loss:   tensor(0.8473, device='cuda:0', grad_fn=<NllLossBackward>)
[4307] Train Loss:   tensor(0.8604, device='cuda:0', grad_fn=<NllLossBackward>)
[4308] Train Loss:   tensor(0.8700, device='cuda:0', grad_fn=<NllLossBackward>)
[4309] Train Loss:   tensor(0.8882, device='cuda:0', grad_fn=<NllLossBackward>)
[4310] Train Loss:   tensor(0.8569, device='cuda:0', grad_fn=<NllLossBackward>)
[4311] Train Loss:   tensor(0.9143, device='cuda:0', grad_fn=<NllLossBackward>)
[4312] Train Loss:   tensor(0.8736, device='cuda:0', grad_fn=<NllLossBackward>)
[4313] Train Loss:   tensor(0.8775, device='cuda:0', grad_fn=<NllLossBackward>)
[4314] Train Loss:   tensor(0.8566, device='cuda:0', grad_fn=<NllLossBackward>)
[4315] Train Loss:   tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[4316] Train Loss:   tensor(0.8414, device='cuda:0', grad_fn=<NllLossBackward>)
[4317] Train Loss:   tensor(0.8633, device='cuda:0', grad_fn=<NllLossBackward>)
[4318] Train Loss:   tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[4319] Train Loss:   tensor(0.8716, device='cuda:0', grad_fn=<NllLossBackward>)
[4320] Train Loss:   tensor(0.8663, device='cuda:0', grad_fn=<NllLossBackward>)
[4321] Train Loss:   tensor(0.8564, device='cuda:0', grad_fn=<NllLossBackward>)
[4322] Train Loss:   tensor(0.9010, device='cuda:0', grad_fn=<NllLossBackward>)
[4323] Train Loss:   tensor(0.8413, device='cuda:0', grad_fn=<NllLossBackward>)
[4324] Train Loss:   tensor(0.8769, device='cuda:0', grad_fn=<NllLossBackward>)
[4325] Train Loss:   tensor(0.8750, device='cuda:0', grad_fn=<NllLossBackward>)
[4326] Train Loss:   tensor(0.8538, device='cuda:0', grad_fn=<NllLossBackward>)
[4327] Train Loss:   tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[4328] Train Loss:   tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[4329] Train Loss:   tensor(0.8817, device='cuda:0', grad_fn=<NllLossBackward>)
[4330] Train Loss:   tensor(0.8909, device='cuda:0', grad_fn=<NllLossBackward>)
[4331] Train Loss:   tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[4332] Train Loss:   tensor(0.8741, device='cuda:0', grad_fn=<NllLossBackward>)
[4333] Train Loss:   tensor(0.8118, device='cuda:0', grad_fn=<NllLossBackward>)
[4334] Train Loss:   tensor(0.9030, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4335] Train Loss:    tensor(0.8685, device='cuda:0', grad_fn=<NllLossBackward>)
[4336] Train Loss:    tensor(0.8473, device='cuda:0', grad_fn=<NllLossBackward>)
[4337] Train Loss:    tensor(0.9031, device='cuda:0', grad_fn=<NllLossBackward>)
[4338] Train Loss:    tensor(0.8337, device='cuda:0', grad_fn=<NllLossBackward>)
[4339] Train Loss:    tensor(0.8954, device='cuda:0', grad_fn=<NllLossBackward>)
[4340] Train Loss:    tensor(0.8736, device='cuda:0', grad_fn=<NllLossBackward>)
[4341] Train Loss:    tensor(0.8646, device='cuda:0', grad_fn=<NllLossBackward>)
[4342] Train Loss:    tensor(0.8407, device='cuda:0', grad_fn=<NllLossBackward>)
[4343] Train Loss:    tensor(0.8577, device='cuda:0', grad_fn=<NllLossBackward>)
[4344] Train Loss:    tensor(0.8506, device='cuda:0', grad_fn=<NllLossBackward>)
[4345] Train Loss:    tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[4346] Train Loss:    tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[4347] Train Loss:    tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[4348] Train Loss:    tensor(0.9030, device='cuda:0', grad_fn=<NllLossBackward>)
[4349] Train Loss:    tensor(0.8352, device='cuda:0', grad_fn=<NllLossBackward>)
[4350] Train Loss:    tensor(0.8790, device='cuda:0', grad_fn=<NllLossBackward>)
[4351] Train Loss:    tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[4352] Train Loss:    tensor(0.8702, device='cuda:0', grad_fn=<NllLossBackward>)
[4353] Train Loss:    tensor(0.8649, device='cuda:0', grad_fn=<NllLossBackward>)
[4354] Train Loss:    tensor(0.8795, device='cuda:0', grad_fn=<NllLossBackward>)
[4355] Train Loss:    tensor(0.8455, device='cuda:0', grad_fn=<NllLossBackward>)
[4356] Train Loss:    tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[4357] Train Loss:    tensor(0.8942, device='cuda:0', grad_fn=<NllLossBackward>)
[4358] Train Loss:    tensor(0.8469, device='cuda:0', grad_fn=<NllLossBackward>)
[4359] Train Loss:    tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[4360] Train Loss:    tensor(0.8686, device='cuda:0', grad_fn=<NllLossBackward>)
[4361] Train Loss:    tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[4362] Train Loss:    tensor(0.8890, device='cuda:0', grad_fn=<NllLossBackward>)
[4363] Train Loss:    tensor(0.8764, device='cuda:0', grad_fn=<NllLossBackward>)
[4364] Train Loss:    tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[4365] Train Loss:    tensor(0.8411, device='cuda:0', grad_fn=<NllLossBackward>)
[4366] Train Loss:    tensor(0.8848, device='cuda:0', grad_fn=<NllLossBackward>)
[4367] Train Loss:    tensor(0.8657, device='cuda:0', grad_fn=<NllLossBackward>)
[4368] Train Loss:    tensor(0.8614, device='cuda:0', grad_fn=<NllLossBackward>)
[4369] Train Loss:    tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[4370] Train Loss:    tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[4371] Train Loss:    tensor(0.8771, device='cuda:0', grad_fn=<NllLossBackward>)
[4372] Train Loss:    tensor(0.8502, device='cuda:0', grad_fn=<NllLossBackward>)
[4373] Train Loss:    tensor(0.8656, device='cuda:0', grad_fn=<NllLossBackward>)
[4374] Train Loss:    tensor(0.8715, device='cuda:0', grad_fn=<NllLossBackward>)
[4375] Train Loss:    tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[4376] Train Loss:    tensor(0.8530, device='cuda:0', grad_fn=<NllLossBackward>)
[4377] Train Loss:    tensor(0.8737, device='cuda:0', grad_fn=<NllLossBackward>)
[4378] Train Loss:    tensor(0.8827, device='cuda:0', grad_fn=<NllLossBackward>)
[4379] Train Loss:    tensor(0.8560, device='cuda:0', grad_fn=<NllLossBackward>)
[4380] Train Loss:    tensor(0.8676, device='cuda:0', grad_fn=<NllLossBackward>)
[4381] Train Loss:    tensor(0.8671, device='cuda:0', grad_fn=<NllLossBackward>)
[4382] Train Loss:    tensor(0.8543, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4383] Train Loss:    tensor(0.8576, device='cuda:0', grad_fn=<NllLossBackward>)
[4384] Train Loss:    tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[4385] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[4386] Train Loss:    tensor(0.8501, device='cuda:0', grad_fn=<NllLossBackward>)
[4387] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[4388] Train Loss:    tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[4389] Train Loss:    tensor(0.8375, device='cuda:0', grad_fn=<NllLossBackward>)
[4390] Train Loss:    tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[4391] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[4392] Train Loss:    tensor(0.8885, device='cuda:0', grad_fn=<NllLossBackward>)
[4393] Train Loss:    tensor(0.8794, device='cuda:0', grad_fn=<NllLossBackward>)
[4394] Train Loss:    tensor(0.8485, device='cuda:0', grad_fn=<NllLossBackward>)
[4395] Train Loss:    tensor(0.8624, device='cuda:0', grad_fn=<NllLossBackward>)
[4396] Train Loss:    tensor(0.8674, device='cuda:0', grad_fn=<NllLossBackward>)
[4397] Train Loss:    tensor(0.8850, device='cuda:0', grad_fn=<NllLossBackward>)
[4398] Train Loss:    tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[4399] Train Loss:    tensor(0.8726, device='cuda:0', grad_fn=<NllLossBackward>)
[4400] Train Loss:    tensor(0.8650, device='cuda:0', grad_fn=<NllLossBackward>)
[4401] Train Loss:    tensor(0.8572, device='cuda:0', grad_fn=<NllLossBackward>)
[4402] Train Loss:    tensor(0.8558, device='cuda:0', grad_fn=<NllLossBackward>)
[4403] Train Loss:    tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[4404] Train Loss:    tensor(0.8570, device='cuda:0', grad_fn=<NllLossBackward>)
[4405] Train Loss:    tensor(0.8946, device='cuda:0', grad_fn=<NllLossBackward>)
[4406] Train Loss:    tensor(0.8428, device='cuda:0', grad_fn=<NllLossBackward>)
[4407] Train Loss:    tensor(0.8841, device='cuda:0', grad_fn=<NllLossBackward>)
[4408] Train Loss:    tensor(0.8462, device='cuda:0', grad_fn=<NllLossBackward>)
[4409] Train Loss:    tensor(0.8821, device='cuda:0', grad_fn=<NllLossBackward>)
[4410] Train Loss:    tensor(0.8710, device='cuda:0', grad_fn=<NllLossBackward>)
[4411] Train Loss:    tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[4412] Train Loss:    tensor(0.8709, device='cuda:0', grad_fn=<NllLossBackward>)
[4413] Train Loss:    tensor(0.9130, device='cuda:0', grad_fn=<NllLossBackward>)
[4414] Train Loss:    tensor(0.8863, device='cuda:0', grad_fn=<NllLossBackward>)
[4415] Train Loss:    tensor(0.8537, device='cuda:0', grad_fn=<NllLossBackward>)
[4416] Train Loss:    tensor(0.8538, device='cuda:0', grad_fn=<NllLossBackward>)
[4417] Train Loss:    tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[4418] Train Loss:    tensor(0.8438, device='cuda:0', grad_fn=<NllLossBackward>)
[4419] Train Loss:    tensor(0.8550, device='cuda:0', grad_fn=<NllLossBackward>)
[4420] Train Loss:    tensor(0.8370, device='cuda:0', grad_fn=<NllLossBackward>)
[4421] Train Loss:    tensor(0.8793, device='cuda:0', grad_fn=<NllLossBackward>)
[4422] Train Loss:    tensor(0.8609, device='cuda:0', grad_fn=<NllLossBackward>)
[4423] Train Loss:    tensor(0.8591, device='cuda:0', grad_fn=<NllLossBackward>)
[4424] Train Loss:    tensor(0.8497, device='cuda:0', grad_fn=<NllLossBackward>)
[4425] Train Loss:    tensor(0.8658, device='cuda:0', grad_fn=<NllLossBackward>)
[4426] Train Loss:    tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[4427] Train Loss:    tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[4428] Train Loss:    tensor(0.8771, device='cuda:0', grad_fn=<NllLossBackward>)
[4429] Train Loss:    tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[4430] Train Loss:    tensor(0.8532, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4431] Train Loss:    tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
[4432] Train Loss:    tensor(0.8637, device='cuda:0', grad_fn=<NllLossBackward>)
[4433] Train Loss:    tensor(0.8370, device='cuda:0', grad_fn=<NllLossBackward>)
[4434] Train Loss:    tensor(0.8787, device='cuda:0', grad_fn=<NllLossBackward>)
[4435] Train Loss:    tensor(0.8999, device='cuda:0', grad_fn=<NllLossBackward>)
[4436] Train Loss:    tensor(0.8513, device='cuda:0', grad_fn=<NllLossBackward>)
[4437] Train Loss:    tensor(0.8888, device='cuda:0', grad_fn=<NllLossBackward>)
[4438] Train Loss:    tensor(0.8514, device='cuda:0', grad_fn=<NllLossBackward>)
[4439] Train Loss:    tensor(0.8480, device='cuda:0', grad_fn=<NllLossBackward>)
[4440] Train Loss:    tensor(0.8915, device='cuda:0', grad_fn=<NllLossBackward>)
[4441] Train Loss:    tensor(0.8408, device='cuda:0', grad_fn=<NllLossBackward>)
[4442] Train Loss:    tensor(0.9091, device='cuda:0', grad_fn=<NllLossBackward>)
[4443] Train Loss:    tensor(0.8756, device='cuda:0', grad_fn=<NllLossBackward>)
[4444] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[4445] Train Loss:    tensor(0.8532, device='cuda:0', grad_fn=<NllLossBackward>)
[4446] Train Loss:    tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[4447] Train Loss:    tensor(0.8849, device='cuda:0', grad_fn=<NllLossBackward>)
[4448] Train Loss:    tensor(0.8883, device='cuda:0', grad_fn=<NllLossBackward>)
[4449] Train Loss:    tensor(0.8397, device='cuda:0', grad_fn=<NllLossBackward>)
[4450] Train Loss:    tensor(0.8432, device='cuda:0', grad_fn=<NllLossBackward>)
[4451] Train Loss:    tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[4452] Train Loss:    tensor(0.8077, device='cuda:0', grad_fn=<NllLossBackward>)
[4453] Train Loss:    tensor(0.8323, device='cuda:0', grad_fn=<NllLossBackward>)
[4454] Train Loss:    tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[4455] Train Loss:    tensor(0.8309, device='cuda:0', grad_fn=<NllLossBackward>)
[4456] Train Loss:    tensor(0.8912, device='cuda:0', grad_fn=<NllLossBackward>)
[4457] Train Loss:    tensor(0.8733, device='cuda:0', grad_fn=<NllLossBackward>)
[4458] Train Loss:    tensor(0.8641, device='cuda:0', grad_fn=<NllLossBackward>)
[4459] Train Loss:    tensor(0.8755, device='cuda:0', grad_fn=<NllLossBackward>)
[4460] Train Loss:    tensor(0.8491, device='cuda:0', grad_fn=<NllLossBackward>)
[4461] Train Loss:    tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[4462] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[4463] Train Loss:    tensor(0.8468, device='cuda:0', grad_fn=<NllLossBackward>)
[4464] Train Loss:    tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[4465] Train Loss:    tensor(0.8556, device='cuda:0', grad_fn=<NllLossBackward>)
[4466] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[4467] Train Loss:    tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[4468] Train Loss:    tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[4469] Train Loss:    tensor(0.8619, device='cuda:0', grad_fn=<NllLossBackward>)
[4470] Train Loss:    tensor(0.8238, device='cuda:0', grad_fn=<NllLossBackward>)
[4471] Train Loss:    tensor(0.8911, device='cuda:0', grad_fn=<NllLossBackward>)
[4472] Train Loss:    tensor(0.8804, device='cuda:0', grad_fn=<NllLossBackward>)
[4473] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[4474] Train Loss:    tensor(0.8382, device='cuda:0', grad_fn=<NllLossBackward>)
[4475] Train Loss:    tensor(0.8469, device='cuda:0', grad_fn=<NllLossBackward>)
[4476] Train Loss:    tensor(0.8586, device='cuda:0', grad_fn=<NllLossBackward>)
[4477] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[4478] Train Loss:    tensor(0.8248, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4479] Train Loss:  tensor(0.8525, device='cuda:0', grad_fn=<NllLossBackward>)
[4480] Train Loss:  tensor(0.8544, device='cuda:0', grad_fn=<NllLossBackward>)
[4481] Train Loss:  tensor(0.8587, device='cuda:0', grad_fn=<NllLossBackward>)
[4482] Train Loss:  tensor(0.8608, device='cuda:0', grad_fn=<NllLossBackward>)
[4483] Train Loss:  tensor(0.8370, device='cuda:0', grad_fn=<NllLossBackward>)
[4484] Train Loss:  tensor(0.8677, device='cuda:0', grad_fn=<NllLossBackward>)
[4485] Train Loss:  tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[4486] Train Loss:  tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[4487] Train Loss:  tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[4488] Train Loss:  tensor(0.8575, device='cuda:0', grad_fn=<NllLossBackward>)
[4489] Train Loss:  tensor(0.8639, device='cuda:0', grad_fn=<NllLossBackward>)
[4490] Train Loss:  tensor(0.8535, device='cuda:0', grad_fn=<NllLossBackward>)
[4491] Train Loss:  tensor(0.8603, device='cuda:0', grad_fn=<NllLossBackward>)
[4492] Train Loss:  tensor(0.8176, device='cuda:0', grad_fn=<NllLossBackward>)
[4493] Train Loss:  tensor(0.8951, device='cuda:0', grad_fn=<NllLossBackward>)
[4494] Train Loss:  tensor(0.8694, device='cuda:0', grad_fn=<NllLossBackward>)
[4495] Train Loss:  tensor(0.8794, device='cuda:0', grad_fn=<NllLossBackward>)
[4496] Train Loss:  tensor(0.8331, device='cuda:0', grad_fn=<NllLossBackward>)
[4497] Train Loss:  tensor(0.8532, device='cuda:0', grad_fn=<NllLossBackward>)
[4498] Train Loss:  tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[4499] Train Loss:  tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[4500] Train Loss:  tensor(0.9047, device='cuda:0', grad_fn=<NllLossBackward>)
[4501] Train Loss:  tensor(0.8477, device='cuda:0', grad_fn=<NllLossBackward>)
[4502] Train Loss:  tensor(0.8519, device='cuda:0', grad_fn=<NllLossBackward>)
[4503] Train Loss:  tensor(0.8290, device='cuda:0', grad_fn=<NllLossBackward>)
[4504] Train Loss:  tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[4505] Train Loss:  tensor(0.8388, device='cuda:0', grad_fn=<NllLossBackward>)
[4506] Train Loss:  tensor(0.8841, device='cuda:0', grad_fn=<NllLossBackward>)
[4507] Train Loss:  tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[4508] Train Loss:  tensor(0.8718, device='cuda:0', grad_fn=<NllLossBackward>)
[4509] Train Loss:  tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[4510] Train Loss:  tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[4511] Train Loss:  tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
Training Loss:  0.894009234581856 Time:  2528.6109669208527 s
Training Accuaracy:  72.90780358280696 %
[0] Valid Loss:  tensor(0.8032, device='cuda:0')
[1] Valid Loss:  tensor(0.9320, device='cuda:0')
[2] Valid Loss:  tensor(0.7954, device='cuda:0')
[3] Valid Loss:  tensor(0.6905, device='cuda:0')
[4] Valid Loss:  tensor(0.7945, device='cuda:0')
[5] Valid Loss:  tensor(0.7812, device='cuda:0')
[6] Valid Loss:  tensor(0.8128, device='cuda:0')
[7] Valid Loss:  tensor(0.9359, device='cuda:0')
[8] Valid Loss:  tensor(0.8561, device='cuda:0')
[9] Valid Loss:  tensor(0.8892, device='cuda:0')
[10] Valid Loss:  tensor(0.7667, device='cuda:0')
[11] Valid Loss:  tensor(0.7816, device='cuda:0')
[12] Valid Loss:  tensor(1.1899, device='cuda:0')
```

```
[13] Valid Loss:  tensor(1.0513, device='cuda:0')
[14] Valid Loss:  tensor(0.6985, device='cuda:0')
[15] Valid Loss:  tensor(0.8565, device='cuda:0')
[16] Valid Loss:  tensor(0.7231, device='cuda:0')
[17] Valid Loss:  tensor(0.8442, device='cuda:0')
[18] Valid Loss:  tensor(0.7929, device='cuda:0')
[19] Valid Loss:  tensor(0.6998, device='cuda:0')
[20] Valid Loss:  tensor(0.7612, device='cuda:0')
[21] Valid Loss:  tensor(0.6816, device='cuda:0')
[22] Valid Loss:  tensor(0.6803, device='cuda:0')
[23] Valid Loss:  tensor(0.9102, device='cuda:0')
[24] Valid Loss:  tensor(0.8274, device='cuda:0')
[25] Valid Loss:  tensor(1.0275, device='cuda:0')
[26] Valid Loss:  tensor(0.6166, device='cuda:0')
[27] Valid Loss:  tensor(0.9520, device='cuda:0')
[28] Valid Loss:  tensor(0.7718, device='cuda:0')
[29] Valid Loss:  tensor(0.8073, device='cuda:0')
[30] Valid Loss:  tensor(0.7004, device='cuda:0')
[31] Valid Loss:  tensor(0.8392, device='cuda:0')
[32] Valid Loss:  tensor(0.9295, device='cuda:0')
[33] Valid Loss:  tensor(0.9043, device='cuda:0')
[34] Valid Loss:  tensor(0.8571, device='cuda:0')
[35] Valid Loss:  tensor(0.9545, device='cuda:0')
[36] Valid Loss:  tensor(0.7981, device='cuda:0')
[37] Valid Loss:  tensor(1.0336, device='cuda:0')
[38] Valid Loss:  tensor(0.9294, device='cuda:0')
[39] Valid Loss:  tensor(0.7784, device='cuda:0')
[40] Valid Loss:  tensor(0.8673, device='cuda:0')
[41] Valid Loss:  tensor(0.9482, device='cuda:0')
[42] Valid Loss:  tensor(0.8278, device='cuda:0')
[43] Valid Loss:  tensor(0.6652, device='cuda:0')
[44] Valid Loss:  tensor(0.7604, device='cuda:0')
[45] Valid Loss:  tensor(0.6573, device='cuda:0')
[46] Valid Loss:  tensor(0.9539, device='cuda:0')
[47] Valid Loss:  tensor(0.8810, device='cuda:0')
[48] Valid Loss:  tensor(0.7687, device='cuda:0')
[49] Valid Loss:  tensor(0.8363, device='cuda:0')
[50] Valid Loss:  tensor(0.8270, device='cuda:0')
[51] Valid Loss:  tensor(0.7289, device='cuda:0')
[52] Valid Loss:  tensor(0.7032, device='cuda:0')
[53] Valid Loss:  tensor(0.9531, device='cuda:0')
[54] Valid Loss:  tensor(0.6957, device='cuda:0')
[55] Valid Loss:  tensor(0.6399, device='cuda:0')
[56] Valid Loss:  tensor(0.9662, device='cuda:0')
[57] Valid Loss:  tensor(0.8975, device='cuda:0')
[58] Valid Loss:  tensor(0.8570, device='cuda:0')
[59] Valid Loss:  tensor(0.7739, device='cuda:0')
[60] Valid Loss:  tensor(1.1175, device='cuda:0')
```

```
[61] Valid Loss:  tensor(0.7639, device='cuda:0')
[62] Valid Loss:  tensor(0.9129, device='cuda:0')
[63] Valid Loss:  tensor(0.7857, device='cuda:0')
[64] Valid Loss:  tensor(0.8409, device='cuda:0')
[65] Valid Loss:  tensor(0.9407, device='cuda:0')
[66] Valid Loss:  tensor(0.7923, device='cuda:0')
[67] Valid Loss:  tensor(0.8207, device='cuda:0')
[68] Valid Loss:  tensor(0.7913, device='cuda:0')
[69] Valid Loss:  tensor(0.6691, device='cuda:0')
[70] Valid Loss:  tensor(0.6874, device='cuda:0')
[71] Valid Loss:  tensor(0.8954, device='cuda:0')
[72] Valid Loss:  tensor(0.8226, device='cuda:0')
[73] Valid Loss:  tensor(0.7817, device='cuda:0')
[74] Valid Loss:  tensor(0.8395, device='cuda:0')
[75] Valid Loss:  tensor(0.9341, device='cuda:0')
[76] Valid Loss:  tensor(0.8732, device='cuda:0')
[77] Valid Loss:  tensor(0.8081, device='cuda:0')
[78] Valid Loss:  tensor(0.8692, device='cuda:0')
[79] Valid Loss:  tensor(0.7463, device='cuda:0')
[80] Valid Loss:  tensor(0.8346, device='cuda:0')
[81] Valid Loss:  tensor(0.6986, device='cuda:0')
[82] Valid Loss:  tensor(0.9802, device='cuda:0')
[83] Valid Loss:  tensor(0.8820, device='cuda:0')
[84] Valid Loss:  tensor(0.7238, device='cuda:0')
[85] Valid Loss:  tensor(0.8262, device='cuda:0')
[86] Valid Loss:  tensor(0.8316, device='cuda:0')
[87] Valid Loss:  tensor(0.8856, device='cuda:0')
[88] Valid Loss:  tensor(0.6626, device='cuda:0')
[89] Valid Loss:  tensor(0.8691, device='cuda:0')
[90] Valid Loss:  tensor(1.0051, device='cuda:0')
[91] Valid Loss:  tensor(0.8369, device='cuda:0')
[92] Valid Loss:  tensor(0.8803, device='cuda:0')
[93] Valid Loss:  tensor(0.8978, device='cuda:0')
[94] Valid Loss:  tensor(0.9181, device='cuda:0')
[95] Valid Loss:  tensor(0.7511, device='cuda:0')
[96] Valid Loss:  tensor(0.9354, device='cuda:0')
[97] Valid Loss:  tensor(0.8769, device='cuda:0')
[98] Valid Loss:  tensor(0.8963, device='cuda:0')
[99] Valid Loss:  tensor(0.7201, device='cuda:0')
[100] Valid Loss:  tensor(0.8198, device='cuda:0')
[101] Valid Loss:  tensor(0.8188, device='cuda:0')
[102] Valid Loss:  tensor(0.9995, device='cuda:0')
[103] Valid Loss:  tensor(0.9425, device='cuda:0')
[104] Valid Loss:  tensor(0.7637, device='cuda:0')
[105] Valid Loss:  tensor(0.7506, device='cuda:0')
[106] Valid Loss:  tensor(0.6899, device='cuda:0')
[107] Valid Loss:  tensor(0.8919, device='cuda:0')
[108] Valid Loss:  tensor(0.9148, device='cuda:0')
```

```
[109] Valid Loss:   tensor(0.8084, device='cuda:0')
[110] Valid Loss:   tensor(0.9079, device='cuda:0')
[111] Valid Loss:   tensor(0.8638, device='cuda:0')
[112] Valid Loss:   tensor(0.9062, device='cuda:0')
[113] Valid Loss:   tensor(0.8424, device='cuda:0')
[114] Valid Loss:   tensor(0.8108, device='cuda:0')
[115] Valid Loss:   tensor(0.8922, device='cuda:0')
[116] Valid Loss:   tensor(0.9040, device='cuda:0')
[117] Valid Loss:   tensor(0.9133, device='cuda:0')
[118] Valid Loss:   tensor(0.7653, device='cuda:0')
[119] Valid Loss:   tensor(0.7047, device='cuda:0')
[120] Valid Loss:   tensor(1.0486, device='cuda:0')
[121] Valid Loss:   tensor(0.7729, device='cuda:0')
[122] Valid Loss:   tensor(0.8454, device='cuda:0')
[123] Valid Loss:   tensor(0.7738, device='cuda:0')
[124] Valid Loss:   tensor(0.9833, device='cuda:0')
[125] Valid Loss:   tensor(1.0516, device='cuda:0')
[126] Valid Loss:   tensor(0.6456, device='cuda:0')
[127] Valid Loss:   tensor(0.7133, device='cuda:0')
[128] Valid Loss:   tensor(0.9357, device='cuda:0')
[129] Valid Loss:   tensor(0.8088, device='cuda:0')
[130] Valid Loss:   tensor(0.9267, device='cuda:0')
[131] Valid Loss:   tensor(0.7733, device='cuda:0')
[132] Valid Loss:   tensor(0.7941, device='cuda:0')
[133] Valid Loss:   tensor(0.8307, device='cuda:0')
[134] Valid Loss:   tensor(0.7583, device='cuda:0')
[135] Valid Loss:   tensor(0.6931, device='cuda:0')
[136] Valid Loss:   tensor(0.7238, device='cuda:0')
[137] Valid Loss:   tensor(1.0358, device='cuda:0')
[138] Valid Loss:   tensor(0.9061, device='cuda:0')
[139] Valid Loss:   tensor(1.1298, device='cuda:0')
[140] Valid Loss:   tensor(0.8860, device='cuda:0')
[141] Valid Loss:   tensor(0.7218, device='cuda:0')
[142] Valid Loss:   tensor(1.0463, device='cuda:0')
[143] Valid Loss:   tensor(0.7210, device='cuda:0')
[144] Valid Loss:   tensor(0.8501, device='cuda:0')
[145] Valid Loss:   tensor(1.0165, device='cuda:0')
[146] Valid Loss:   tensor(0.7495, device='cuda:0')
[147] Valid Loss:   tensor(0.8868, device='cuda:0')
[148] Valid Loss:   tensor(0.7864, device='cuda:0')
[149] Valid Loss:   tensor(0.8628, device='cuda:0')
[150] Valid Loss:   tensor(0.8702, device='cuda:0')
[151] Valid Loss:   tensor(0.8390, device='cuda:0')
[152] Valid Loss:   tensor(0.7859, device='cuda:0')
[153] Valid Loss:   tensor(0.7534, device='cuda:0')
[154] Valid Loss:   tensor(0.7874, device='cuda:0')
[155] Valid Loss:   tensor(0.8501, device='cuda:0')
[156] Valid Loss:   tensor(0.8865, device='cuda:0')
```

```
[157] Valid Loss:   tensor(0.7471, device='cuda:0')
[158] Valid Loss:   tensor(0.7834, device='cuda:0')
[159] Valid Loss:   tensor(0.8025, device='cuda:0')
[160] Valid Loss:   tensor(0.6914, device='cuda:0')
[161] Valid Loss:   tensor(0.9317, device='cuda:0')
[162] Valid Loss:   tensor(0.8608, device='cuda:0')
[163] Valid Loss:   tensor(0.7551, device='cuda:0')
[164] Valid Loss:   tensor(1.0412, device='cuda:0')
[165] Valid Loss:   tensor(0.8036, device='cuda:0')
[166] Valid Loss:   tensor(0.8112, device='cuda:0')
[167] Valid Loss:   tensor(0.7013, device='cuda:0')
[168] Valid Loss:   tensor(0.7780, device='cuda:0')
[169] Valid Loss:   tensor(0.7808, device='cuda:0')
[170] Valid Loss:   tensor(0.7482, device='cuda:0')
[171] Valid Loss:   tensor(0.6688, device='cuda:0')
[172] Valid Loss:   tensor(0.8616, device='cuda:0')
[173] Valid Loss:   tensor(0.7384, device='cuda:0')
[174] Valid Loss:   tensor(0.8251, device='cuda:0')
[175] Valid Loss:   tensor(0.7411, device='cuda:0')
[176] Valid Loss:   tensor(0.6935, device='cuda:0')
[177] Valid Loss:   tensor(0.8426, device='cuda:0')
[178] Valid Loss:   tensor(0.8867, device='cuda:0')
[179] Valid Loss:   tensor(0.8330, device='cuda:0')
[180] Valid Loss:   tensor(0.8802, device='cuda:0')
[181] Valid Loss:   tensor(0.8260, device='cuda:0')
[182] Valid Loss:   tensor(0.8607, device='cuda:0')
[183] Valid Loss:   tensor(0.8257, device='cuda:0')
[184] Valid Loss:   tensor(0.9730, device='cuda:0')
[185] Valid Loss:   tensor(0.9015, device='cuda:0')
[186] Valid Loss:   tensor(0.7983, device='cuda:0')
[187] Valid Loss:   tensor(0.9185, device='cuda:0')
[188] Valid Loss:   tensor(0.9904, device='cuda:0')
[189] Valid Loss:   tensor(0.8298, device='cuda:0')
[190] Valid Loss:   tensor(0.8203, device='cuda:0')
[191] Valid Loss:   tensor(0.7076, device='cuda:0')
[192] Valid Loss:   tensor(1.0782, device='cuda:0')
[193] Valid Loss:   tensor(0.8411, device='cuda:0')
[194] Valid Loss:   tensor(0.9109, device='cuda:0')
[195] Valid Loss:   tensor(0.9409, device='cuda:0')
[196] Valid Loss:   tensor(0.7048, device='cuda:0')
[197] Valid Loss:   tensor(0.8564, device='cuda:0')
[198] Valid Loss:   tensor(0.8251, device='cuda:0')
[199] Valid Loss:   tensor(0.8695, device='cuda:0')
[200] Valid Loss:   tensor(0.8240, device='cuda:0')
[201] Valid Loss:   tensor(0.7804, device='cuda:0')
[202] Valid Loss:   tensor(0.7557, device='cuda:0')
[203] Valid Loss:   tensor(0.9816, device='cuda:0')
[204] Valid Loss:   tensor(0.9260, device='cuda:0')
```

```
[205] Valid Loss:   tensor(0.9515, device='cuda:0')
[206] Valid Loss:   tensor(1.0099, device='cuda:0')
[207] Valid Loss:   tensor(0.8850, device='cuda:0')
[208] Valid Loss:   tensor(0.7076, device='cuda:0')
[209] Valid Loss:   tensor(0.8095, device='cuda:0')
[210] Valid Loss:   tensor(1.2064, device='cuda:0')
[211] Valid Loss:   tensor(0.8278, device='cuda:0')
[212] Valid Loss:   tensor(0.7447, device='cuda:0')
[213] Valid Loss:   tensor(0.8032, device='cuda:0')
[214] Valid Loss:   tensor(0.7514, device='cuda:0')
[215] Valid Loss:   tensor(0.8908, device='cuda:0')
[216] Valid Loss:   tensor(0.8110, device='cuda:0')
[217] Valid Loss:   tensor(0.8359, device='cuda:0')
[218] Valid Loss:   tensor(0.6752, device='cuda:0')
[219] Valid Loss:   tensor(0.7870, device='cuda:0')
[220] Valid Loss:   tensor(0.7746, device='cuda:0')
[221] Valid Loss:   tensor(0.8758, device='cuda:0')
[222] Valid Loss:   tensor(0.6526, device='cuda:0')
[223] Valid Loss:   tensor(0.7858, device='cuda:0')
[224] Valid Loss:   tensor(0.9130, device='cuda:0')
[225] Valid Loss:   tensor(0.7024, device='cuda:0')
[226] Valid Loss:   tensor(0.8314, device='cuda:0')
[227] Valid Loss:   tensor(0.7231, device='cuda:0')
[228] Valid Loss:   tensor(0.8946, device='cuda:0')
[229] Valid Loss:   tensor(0.7655, device='cuda:0')
[230] Valid Loss:   tensor(0.8675, device='cuda:0')
[231] Valid Loss:   tensor(0.7783, device='cuda:0')
[232] Valid Loss:   tensor(0.6606, device='cuda:0')
[233] Valid Loss:   tensor(0.7984, device='cuda:0')
[234] Valid Loss:   tensor(0.7027, device='cuda:0')
[235] Valid Loss:   tensor(0.7059, device='cuda:0')
[236] Valid Loss:   tensor(0.8197, device='cuda:0')
[237] Valid Loss:   tensor(0.8805, device='cuda:0')
[238] Valid Loss:   tensor(0.6284, device='cuda:0')
[239] Valid Loss:   tensor(0.8743, device='cuda:0')
[240] Valid Loss:   tensor(0.7896, device='cuda:0')
[241] Valid Loss:   tensor(0.7141, device='cuda:0')
[242] Valid Loss:   tensor(0.6206, device='cuda:0')
[243] Valid Loss:   tensor(1.2345, device='cuda:0')
[244] Valid Loss:   tensor(0.9066, device='cuda:0')
[245] Valid Loss:   tensor(0.7937, device='cuda:0')
[246] Valid Loss:   tensor(0.7702, device='cuda:0')
[247] Valid Loss:   tensor(0.7944, device='cuda:0')
[248] Valid Loss:   tensor(1.0195, device='cuda:0')
[249] Valid Loss:   tensor(0.7902, device='cuda:0')
[250] Valid Loss:   tensor(0.8698, device='cuda:0')
[251] Valid Loss:   tensor(0.9174, device='cuda:0')
[252] Valid Loss:   tensor(0.7239, device='cuda:0')
```

```
[253] Valid Loss:  tensor(0.7246, device='cuda:0')
[254] Valid Loss:  tensor(0.7665, device='cuda:0')
[255] Valid Loss:  tensor(0.8827, device='cuda:0')
[256] Valid Loss:  tensor(0.8995, device='cuda:0')
[257] Valid Loss:  tensor(0.8083, device='cuda:0')
[258] Valid Loss:  tensor(0.9194, device='cuda:0')
[259] Valid Loss:  tensor(0.8031, device='cuda:0')
[260] Valid Loss:  tensor(0.9755, device='cuda:0')
[261] Valid Loss:  tensor(0.8311, device='cuda:0')
[262] Valid Loss:  tensor(0.9822, device='cuda:0')
[263] Valid Loss:  tensor(0.9704, device='cuda:0')
[264] Valid Loss:  tensor(0.9387, device='cuda:0')
[265] Valid Loss:  tensor(0.9101, device='cuda:0')
[266] Valid Loss:  tensor(0.7710, device='cuda:0')
[267] Valid Loss:  tensor(0.7955, device='cuda:0')
[268] Valid Loss:  tensor(0.8298, device='cuda:0')
[269] Valid Loss:  tensor(0.7870, device='cuda:0')
[270] Valid Loss:  tensor(0.7239, device='cuda:0')
[271] Valid Loss:  tensor(1.0371, device='cuda:0')
[272] Valid Loss:  tensor(0.8236, device='cuda:0')
[273] Valid Loss:  tensor(0.9591, device='cuda:0')
[274] Valid Loss:  tensor(0.8472, device='cuda:0')
[275] Valid Loss:  tensor(0.8277, device='cuda:0')
[276] Valid Loss:  tensor(0.7541, device='cuda:0')
[277] Valid Loss:  tensor(0.7313, device='cuda:0')
[278] Valid Loss:  tensor(0.9842, device='cuda:0')
[279] Valid Loss:  tensor(0.9091, device='cuda:0')
[280] Valid Loss:  tensor(0.9491, device='cuda:0')
[281] Valid Loss:  tensor(0.7520, device='cuda:0')
[282] Valid Loss:  tensor(0.8886, device='cuda:0')
[283] Valid Loss:  tensor(0.9800, device='cuda:0')
[284] Valid Loss:  tensor(0.7611, device='cuda:0')
[285] Valid Loss:  tensor(0.6611, device='cuda:0')
[286] Valid Loss:  tensor(0.8893, device='cuda:0')
[287] Valid Loss:  tensor(1.0221, device='cuda:0')
[288] Valid Loss:  tensor(0.6664, device='cuda:0')
[289] Valid Loss:  tensor(0.8967, device='cuda:0')
[290] Valid Loss:  tensor(0.6873, device='cuda:0')
[291] Valid Loss:  tensor(0.7941, device='cuda:0')
[292] Valid Loss:  tensor(0.9580, device='cuda:0')
[293] Valid Loss:  tensor(0.7404, device='cuda:0')
[294] Valid Loss:  tensor(0.9024, device='cuda:0')
[295] Valid Loss:  tensor(0.7232, device='cuda:0')
[296] Valid Loss:  tensor(0.8327, device='cuda:0')
[297] Valid Loss:  tensor(0.8719, device='cuda:0')
[298] Valid Loss:  tensor(0.7594, device='cuda:0')
[299] Valid Loss:  tensor(0.8799, device='cuda:0')
[300] Valid Loss:  tensor(1.0203, device='cuda:0')
```

```
[301] Valid Loss:    tensor(0.6798, device='cuda:0')
[302] Valid Loss:    tensor(0.6427, device='cuda:0')
[303] Valid Loss:    tensor(0.7776, device='cuda:0')
[304] Valid Loss:    tensor(0.7374, device='cuda:0')
[305] Valid Loss:    tensor(0.8275, device='cuda:0')
[306] Valid Loss:    tensor(0.8534, device='cuda:0')
[307] Valid Loss:    tensor(0.6863, device='cuda:0')
[308] Valid Loss:    tensor(0.7112, device='cuda:0')
[309] Valid Loss:    tensor(0.7779, device='cuda:0')
[310] Valid Loss:    tensor(0.8075, device='cuda:0')
[311] Valid Loss:    tensor(0.7229, device='cuda:0')
[312] Valid Loss:    tensor(0.7103, device='cuda:0')
[313] Valid Loss:    tensor(0.8060, device='cuda:0')
[314] Valid Loss:    tensor(0.7707, device='cuda:0')
[315] Valid Loss:    tensor(1.0983, device='cuda:0')
[316] Valid Loss:    tensor(0.9208, device='cuda:0')
[317] Valid Loss:    tensor(0.7932, device='cuda:0')
[318] Valid Loss:    tensor(0.7847, device='cuda:0')
[319] Valid Loss:    tensor(0.7353, device='cuda:0')
[320] Valid Loss:    tensor(0.7796, device='cuda:0')
[321] Valid Loss:    tensor(0.7816, device='cuda:0')
[322] Valid Loss:    tensor(0.7656, device='cuda:0')
[323] Valid Loss:    tensor(0.9304, device='cuda:0')
[324] Valid Loss:    tensor(0.8335, device='cuda:0')
[325] Valid Loss:    tensor(0.7183, device='cuda:0')
[326] Valid Loss:    tensor(0.7984, device='cuda:0')
[327] Valid Loss:    tensor(0.5541, device='cuda:0')
[328] Valid Loss:    tensor(0.7857, device='cuda:0')
[329] Valid Loss:    tensor(0.8824, device='cuda:0')
[330] Valid Loss:    tensor(0.6877, device='cuda:0')
[331] Valid Loss:    tensor(1.0139, device='cuda:0')
[332] Valid Loss:    tensor(0.7057, device='cuda:0')
[333] Valid Loss:    tensor(0.6673, device='cuda:0')
[334] Valid Loss:    tensor(0.7013, device='cuda:0')
[335] Valid Loss:    tensor(0.7789, device='cuda:0')
[336] Valid Loss:    tensor(0.8686, device='cuda:0')
[337] Valid Loss:    tensor(0.9662, device='cuda:0')
[338] Valid Loss:    tensor(0.8589, device='cuda:0')
[339] Valid Loss:    tensor(0.7852, device='cuda:0')
[340] Valid Loss:    tensor(0.8533, device='cuda:0')
[341] Valid Loss:    tensor(0.9249, device='cuda:0')
[342] Valid Loss:    tensor(0.9991, device='cuda:0')
[343] Valid Loss:    tensor(0.8799, device='cuda:0')
[344] Valid Loss:    tensor(0.7314, device='cuda:0')
[345] Valid Loss:    tensor(0.6546, device='cuda:0')
[346] Valid Loss:    tensor(0.7859, device='cuda:0')
[347] Valid Loss:    tensor(0.9012, device='cuda:0')
[348] Valid Loss:    tensor(0.8049, device='cuda:0')
```

```
[349] Valid Loss:  tensor(0.7609, device='cuda:0')
[350] Valid Loss:  tensor(0.8046, device='cuda:0')
[351] Valid Loss:  tensor(0.9355, device='cuda:0')
[352] Valid Loss:  tensor(0.7306, device='cuda:0')
[353] Valid Loss:  tensor(0.8339, device='cuda:0')
[354] Valid Loss:  tensor(0.9887, device='cuda:0')
[355] Valid Loss:  tensor(0.6499, device='cuda:0')
[356] Valid Loss:  tensor(0.7495, device='cuda:0')
[357] Valid Loss:  tensor(0.8485, device='cuda:0')
[358] Valid Loss:  tensor(0.7729, device='cuda:0')
[359] Valid Loss:  tensor(0.8908, device='cuda:0')
[360] Valid Loss:  tensor(0.8037, device='cuda:0')
[361] Valid Loss:  tensor(0.8280, device='cuda:0')
[362] Valid Loss:  tensor(0.7861, device='cuda:0')
[363] Valid Loss:  tensor(0.8694, device='cuda:0')
[364] Valid Loss:  tensor(0.9881, device='cuda:0')
[365] Valid Loss:  tensor(0.7463, device='cuda:0')
[366] Valid Loss:  tensor(0.7771, device='cuda:0')
[367] Valid Loss:  tensor(0.9083, device='cuda:0')
[368] Valid Loss:  tensor(0.9076, device='cuda:0')
[369] Valid Loss:  tensor(0.7961, device='cuda:0')
[370] Valid Loss:  tensor(0.7033, device='cuda:0')
[371] Valid Loss:  tensor(0.7011, device='cuda:0')
[372] Valid Loss:  tensor(0.8980, device='cuda:0')
[373] Valid Loss:  tensor(0.6548, device='cuda:0')
[374] Valid Loss:  tensor(0.7754, device='cuda:0')
[375] Valid Loss:  tensor(0.8923, device='cuda:0')
[376] Valid Loss:  tensor(0.7229, device='cuda:0')
[377] Valid Loss:  tensor(1.0169, device='cuda:0')
[378] Valid Loss:  tensor(0.7400, device='cuda:0')
[379] Valid Loss:  tensor(0.8914, device='cuda:0')
[380] Valid Loss:  tensor(1.1060, device='cuda:0')
Testing Loss:  0.830471642217611
Testing Accuaracy:  74.34673652085844 %
Epoch Number:  2
[0] Train Loss:  tensor(0.8741, device='cuda:0', grad_fn=<NllLossBackward>)
[1] Train Loss:  tensor(0.8302, device='cuda:0', grad_fn=<NllLossBackward>)
[2] Train Loss:  tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[3] Train Loss:  tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
[4] Train Loss:  tensor(0.8589, device='cuda:0', grad_fn=<NllLossBackward>)
[5] Train Loss:  tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[6] Train Loss:  tensor(0.8405, device='cuda:0', grad_fn=<NllLossBackward>)
[7] Train Loss:  tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[8] Train Loss:  tensor(0.8552, device='cuda:0', grad_fn=<NllLossBackward>)
[9] Train Loss:  tensor(0.8315, device='cuda:0', grad_fn=<NllLossBackward>)
[10] Train Loss:  tensor(0.8767, device='cuda:0', grad_fn=<NllLossBackward>)
[11] Train Loss:  tensor(0.8882, device='cuda:0', grad_fn=<NllLossBackward>)
[12] Train Loss:  tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[13] Train Loss:   tensor(0.8760, device='cuda:0', grad_fn=<NllLossBackward>)
[14] Train Loss:   tensor(0.8533, device='cuda:0', grad_fn=<NllLossBackward>)
[15] Train Loss:   tensor(0.8359, device='cuda:0', grad_fn=<NllLossBackward>)
[16] Train Loss:   tensor(0.8463, device='cuda:0', grad_fn=<NllLossBackward>)
[17] Train Loss:   tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[18] Train Loss:   tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[19] Train Loss:   tensor(0.8344, device='cuda:0', grad_fn=<NllLossBackward>)
[20] Train Loss:   tensor(0.8195, device='cuda:0', grad_fn=<NllLossBackward>)
[21] Train Loss:   tensor(0.8790, device='cuda:0', grad_fn=<NllLossBackward>)
[22] Train Loss:   tensor(0.8656, device='cuda:0', grad_fn=<NllLossBackward>)
[23] Train Loss:   tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[24] Train Loss:   tensor(0.8459, device='cuda:0', grad_fn=<NllLossBackward>)
[25] Train Loss:   tensor(0.8428, device='cuda:0', grad_fn=<NllLossBackward>)
[26] Train Loss:   tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[27] Train Loss:   tensor(0.8409, device='cuda:0', grad_fn=<NllLossBackward>)
[28] Train Loss:   tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
[29] Train Loss:   tensor(0.8576, device='cuda:0', grad_fn=<NllLossBackward>)
[30] Train Loss:   tensor(0.8561, device='cuda:0', grad_fn=<NllLossBackward>)
[31] Train Loss:   tensor(0.8370, device='cuda:0', grad_fn=<NllLossBackward>)
[32] Train Loss:   tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[33] Train Loss:   tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[34] Train Loss:   tensor(0.8658, device='cuda:0', grad_fn=<NllLossBackward>)
[35] Train Loss:   tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[36] Train Loss:   tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[37] Train Loss:   tensor(0.8531, device='cuda:0', grad_fn=<NllLossBackward>)
[38] Train Loss:   tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[39] Train Loss:   tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[40] Train Loss:   tensor(0.8454, device='cuda:0', grad_fn=<NllLossBackward>)
[41] Train Loss:   tensor(0.8563, device='cuda:0', grad_fn=<NllLossBackward>)
[42] Train Loss:   tensor(0.8621, device='cuda:0', grad_fn=<NllLossBackward>)
[43] Train Loss:   tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[44] Train Loss:   tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[45] Train Loss:   tensor(0.8670, device='cuda:0', grad_fn=<NllLossBackward>)
[46] Train Loss:   tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[47] Train Loss:   tensor(0.8515, device='cuda:0', grad_fn=<NllLossBackward>)
[48] Train Loss:   tensor(0.8139, device='cuda:0', grad_fn=<NllLossBackward>)
[49] Train Loss:   tensor(0.8429, device='cuda:0', grad_fn=<NllLossBackward>)
[50] Train Loss:   tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[51] Train Loss:   tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[52] Train Loss:   tensor(0.8624, device='cuda:0', grad_fn=<NllLossBackward>)
[53] Train Loss:   tensor(0.8671, device='cuda:0', grad_fn=<NllLossBackward>)
[54] Train Loss:   tensor(0.8455, device='cuda:0', grad_fn=<NllLossBackward>)
[55] Train Loss:   tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
[56] Train Loss:   tensor(0.8639, device='cuda:0', grad_fn=<NllLossBackward>)
[57] Train Loss:   tensor(0.8317, device='cuda:0', grad_fn=<NllLossBackward>)
[58] Train Loss:   tensor(0.8631, device='cuda:0', grad_fn=<NllLossBackward>)
[59] Train Loss:   tensor(0.8268, device='cuda:0', grad_fn=<NllLossBackward>)
[60] Train Loss:   tensor(0.8411, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[61] Train Loss:    tensor(0.8562, device='cuda:0', grad_fn=<NllLossBackward>)
[62] Train Loss:    tensor(0.8201, device='cuda:0', grad_fn=<NllLossBackward>)
[63] Train Loss:    tensor(0.8518, device='cuda:0', grad_fn=<NllLossBackward>)
[64] Train Loss:    tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[65] Train Loss:    tensor(0.8913, device='cuda:0', grad_fn=<NllLossBackward>)
[66] Train Loss:    tensor(0.8092, device='cuda:0', grad_fn=<NllLossBackward>)
[67] Train Loss:    tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[68] Train Loss:    tensor(0.8467, device='cuda:0', grad_fn=<NllLossBackward>)
[69] Train Loss:    tensor(0.9079, device='cuda:0', grad_fn=<NllLossBackward>)
[70] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[71] Train Loss:    tensor(0.8423, device='cuda:0', grad_fn=<NllLossBackward>)
[72] Train Loss:    tensor(0.8450, device='cuda:0', grad_fn=<NllLossBackward>)
[73] Train Loss:    tensor(0.8336, device='cuda:0', grad_fn=<NllLossBackward>)
[74] Train Loss:    tensor(0.8556, device='cuda:0', grad_fn=<NllLossBackward>)
[75] Train Loss:    tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[76] Train Loss:    tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[77] Train Loss:    tensor(0.8294, device='cuda:0', grad_fn=<NllLossBackward>)
[78] Train Loss:    tensor(0.8167, device='cuda:0', grad_fn=<NllLossBackward>)
[79] Train Loss:    tensor(0.8518, device='cuda:0', grad_fn=<NllLossBackward>)
[80] Train Loss:    tensor(0.8169, device='cuda:0', grad_fn=<NllLossBackward>)
[81] Train Loss:    tensor(0.8608, device='cuda:0', grad_fn=<NllLossBackward>)
[82] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[83] Train Loss:    tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[84] Train Loss:    tensor(0.8387, device='cuda:0', grad_fn=<NllLossBackward>)
[85] Train Loss:    tensor(0.8320, device='cuda:0', grad_fn=<NllLossBackward>)
[86] Train Loss:    tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[87] Train Loss:    tensor(0.8263, device='cuda:0', grad_fn=<NllLossBackward>)
[88] Train Loss:    tensor(0.8506, device='cuda:0', grad_fn=<NllLossBackward>)
[89] Train Loss:    tensor(0.8581, device='cuda:0', grad_fn=<NllLossBackward>)
[90] Train Loss:    tensor(0.8438, device='cuda:0', grad_fn=<NllLossBackward>)
[91] Train Loss:    tensor(0.8522, device='cuda:0', grad_fn=<NllLossBackward>)
[92] Train Loss:    tensor(0.8525, device='cuda:0', grad_fn=<NllLossBackward>)
[93] Train Loss:    tensor(0.8336, device='cuda:0', grad_fn=<NllLossBackward>)
[94] Train Loss:    tensor(0.8605, device='cuda:0', grad_fn=<NllLossBackward>)
[95] Train Loss:    tensor(0.8623, device='cuda:0', grad_fn=<NllLossBackward>)
[96] Train Loss:    tensor(0.8349, device='cuda:0', grad_fn=<NllLossBackward>)
[97] Train Loss:    tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[98] Train Loss:    tensor(0.8416, device='cuda:0', grad_fn=<NllLossBackward>)
[99] Train Loss:    tensor(0.8659, device='cuda:0', grad_fn=<NllLossBackward>)
[100] Train Loss:    tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[101] Train Loss:    tensor(0.8503, device='cuda:0', grad_fn=<NllLossBackward>)
[102] Train Loss:    tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[103] Train Loss:    tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[104] Train Loss:    tensor(0.8367, device='cuda:0', grad_fn=<NllLossBackward>)
[105] Train Loss:    tensor(0.8137, device='cuda:0', grad_fn=<NllLossBackward>)
[106] Train Loss:    tensor(0.8515, device='cuda:0', grad_fn=<NllLossBackward>)
[107] Train Loss:    tensor(0.8717, device='cuda:0', grad_fn=<NllLossBackward>)
[108] Train Loss:    tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[109] Train Loss:   tensor(0.8522, device='cuda:0', grad_fn=<NllLossBackward>)
[110] Train Loss:   tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[111] Train Loss:   tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[112] Train Loss:   tensor(0.8515, device='cuda:0', grad_fn=<NllLossBackward>)
[113] Train Loss:   tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[114] Train Loss:   tensor(0.8533, device='cuda:0', grad_fn=<NllLossBackward>)
[115] Train Loss:   tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[116] Train Loss:   tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[117] Train Loss:   tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[118] Train Loss:   tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[119] Train Loss:   tensor(0.8488, device='cuda:0', grad_fn=<NllLossBackward>)
[120] Train Loss:   tensor(0.8378, device='cuda:0', grad_fn=<NllLossBackward>)
[121] Train Loss:   tensor(0.8594, device='cuda:0', grad_fn=<NllLossBackward>)
[122] Train Loss:   tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[123] Train Loss:   tensor(0.8480, device='cuda:0', grad_fn=<NllLossBackward>)
[124] Train Loss:   tensor(0.8648, device='cuda:0', grad_fn=<NllLossBackward>)
[125] Train Loss:   tensor(0.8704, device='cuda:0', grad_fn=<NllLossBackward>)
[126] Train Loss:   tensor(0.8405, device='cuda:0', grad_fn=<NllLossBackward>)
[127] Train Loss:   tensor(0.8454, device='cuda:0', grad_fn=<NllLossBackward>)
[128] Train Loss:   tensor(0.8639, device='cuda:0', grad_fn=<NllLossBackward>)
[129] Train Loss:   tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[130] Train Loss:   tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[131] Train Loss:   tensor(0.8723, device='cuda:0', grad_fn=<NllLossBackward>)
[132] Train Loss:   tensor(0.8521, device='cuda:0', grad_fn=<NllLossBackward>)
[133] Train Loss:   tensor(0.8742, device='cuda:0', grad_fn=<NllLossBackward>)
[134] Train Loss:   tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[135] Train Loss:   tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[136] Train Loss:   tensor(0.8587, device='cuda:0', grad_fn=<NllLossBackward>)
[137] Train Loss:   tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
[138] Train Loss:   tensor(0.8348, device='cuda:0', grad_fn=<NllLossBackward>)
[139] Train Loss:   tensor(0.8386, device='cuda:0', grad_fn=<NllLossBackward>)
[140] Train Loss:   tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[141] Train Loss:   tensor(0.8433, device='cuda:0', grad_fn=<NllLossBackward>)
[142] Train Loss:   tensor(0.8268, device='cuda:0', grad_fn=<NllLossBackward>)
[143] Train Loss:   tensor(0.8536, device='cuda:0', grad_fn=<NllLossBackward>)
[144] Train Loss:   tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[145] Train Loss:   tensor(0.8693, device='cuda:0', grad_fn=<NllLossBackward>)
[146] Train Loss:   tensor(0.8535, device='cuda:0', grad_fn=<NllLossBackward>)
[147] Train Loss:   tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[148] Train Loss:   tensor(0.8421, device='cuda:0', grad_fn=<NllLossBackward>)
[149] Train Loss:   tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[150] Train Loss:   tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[151] Train Loss:   tensor(0.8312, device='cuda:0', grad_fn=<NllLossBackward>)
[152] Train Loss:   tensor(0.8187, device='cuda:0', grad_fn=<NllLossBackward>)
[153] Train Loss:   tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[154] Train Loss:   tensor(0.8768, device='cuda:0', grad_fn=<NllLossBackward>)
[155] Train Loss:   tensor(0.8452, device='cuda:0', grad_fn=<NllLossBackward>)
[156] Train Loss:   tensor(0.8301, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[157] Train Loss:   tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[158] Train Loss:   tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[159] Train Loss:   tensor(0.8473, device='cuda:0', grad_fn=<NllLossBackward>)
[160] Train Loss:   tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[161] Train Loss:   tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[162] Train Loss:   tensor(0.8326, device='cuda:0', grad_fn=<NllLossBackward>)
[163] Train Loss:   tensor(0.8306, device='cuda:0', grad_fn=<NllLossBackward>)
[164] Train Loss:   tensor(0.8563, device='cuda:0', grad_fn=<NllLossBackward>)
[165] Train Loss:   tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[166] Train Loss:   tensor(0.8464, device='cuda:0', grad_fn=<NllLossBackward>)
[167] Train Loss:   tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[168] Train Loss:   tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[169] Train Loss:   tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[170] Train Loss:   tensor(0.8517, device='cuda:0', grad_fn=<NllLossBackward>)
[171] Train Loss:   tensor(0.8238, device='cuda:0', grad_fn=<NllLossBackward>)
[172] Train Loss:   tensor(0.8328, device='cuda:0', grad_fn=<NllLossBackward>)
[173] Train Loss:   tensor(0.8615, device='cuda:0', grad_fn=<NllLossBackward>)
[174] Train Loss:   tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[175] Train Loss:   tensor(0.8832, device='cuda:0', grad_fn=<NllLossBackward>)
[176] Train Loss:   tensor(0.8686, device='cuda:0', grad_fn=<NllLossBackward>)
[177] Train Loss:   tensor(0.8607, device='cuda:0', grad_fn=<NllLossBackward>)
[178] Train Loss:   tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[179] Train Loss:   tensor(0.8256, device='cuda:0', grad_fn=<NllLossBackward>)
[180] Train Loss:   tensor(0.8546, device='cuda:0', grad_fn=<NllLossBackward>)
[181] Train Loss:   tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[182] Train Loss:   tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[183] Train Loss:   tensor(0.8718, device='cuda:0', grad_fn=<NllLossBackward>)
[184] Train Loss:   tensor(0.8534, device='cuda:0', grad_fn=<NllLossBackward>)
[185] Train Loss:   tensor(0.8460, device='cuda:0', grad_fn=<NllLossBackward>)
[186] Train Loss:   tensor(0.8709, device='cuda:0', grad_fn=<NllLossBackward>)
[187] Train Loss:   tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[188] Train Loss:   tensor(0.8601, device='cuda:0', grad_fn=<NllLossBackward>)
[189] Train Loss:   tensor(0.8393, device='cuda:0', grad_fn=<NllLossBackward>)
[190] Train Loss:   tensor(0.8812, device='cuda:0', grad_fn=<NllLossBackward>)
[191] Train Loss:   tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[192] Train Loss:   tensor(0.8456, device='cuda:0', grad_fn=<NllLossBackward>)
[193] Train Loss:   tensor(0.8575, device='cuda:0', grad_fn=<NllLossBackward>)
[194] Train Loss:   tensor(0.8375, device='cuda:0', grad_fn=<NllLossBackward>)
[195] Train Loss:   tensor(0.8382, device='cuda:0', grad_fn=<NllLossBackward>)
[196] Train Loss:   tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[197] Train Loss:   tensor(0.8460, device='cuda:0', grad_fn=<NllLossBackward>)
[198] Train Loss:   tensor(0.8225, device='cuda:0', grad_fn=<NllLossBackward>)
[199] Train Loss:   tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[200] Train Loss:   tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[201] Train Loss:   tensor(0.8236, device='cuda:0', grad_fn=<NllLossBackward>)
[202] Train Loss:   tensor(0.8236, device='cuda:0', grad_fn=<NllLossBackward>)
[203] Train Loss:   tensor(0.8564, device='cuda:0', grad_fn=<NllLossBackward>)
[204] Train Loss:   tensor(0.8710, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[205] Train Loss:  tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[206] Train Loss:  tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[207] Train Loss:  tensor(0.8416, device='cuda:0', grad_fn=<NllLossBackward>)
[208] Train Loss:  tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[209] Train Loss:  tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[210] Train Loss:  tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[211] Train Loss:  tensor(0.8155, device='cuda:0', grad_fn=<NllLossBackward>)
[212] Train Loss:  tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[213] Train Loss:  tensor(0.8501, device='cuda:0', grad_fn=<NllLossBackward>)
[214] Train Loss:  tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[215] Train Loss:  tensor(0.8569, device='cuda:0', grad_fn=<NllLossBackward>)
[216] Train Loss:  tensor(0.8547, device='cuda:0', grad_fn=<NllLossBackward>)
[217] Train Loss:  tensor(0.8693, device='cuda:0', grad_fn=<NllLossBackward>)
[218] Train Loss:  tensor(0.8515, device='cuda:0', grad_fn=<NllLossBackward>)
[219] Train Loss:  tensor(0.8505, device='cuda:0', grad_fn=<NllLossBackward>)
[220] Train Loss:  tensor(0.8565, device='cuda:0', grad_fn=<NllLossBackward>)
[221] Train Loss:  tensor(0.8672, device='cuda:0', grad_fn=<NllLossBackward>)
[222] Train Loss:  tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[223] Train Loss:  tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[224] Train Loss:  tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
[225] Train Loss:  tensor(0.8871, device='cuda:0', grad_fn=<NllLossBackward>)
[226] Train Loss:  tensor(0.8337, device='cuda:0', grad_fn=<NllLossBackward>)
[227] Train Loss:  tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
[228] Train Loss:  tensor(0.8624, device='cuda:0', grad_fn=<NllLossBackward>)
[229] Train Loss:  tensor(0.8477, device='cuda:0', grad_fn=<NllLossBackward>)
[230] Train Loss:  tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[231] Train Loss:  tensor(0.8352, device='cuda:0', grad_fn=<NllLossBackward>)
[232] Train Loss:  tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[233] Train Loss:  tensor(0.8315, device='cuda:0', grad_fn=<NllLossBackward>)
[234] Train Loss:  tensor(0.8309, device='cuda:0', grad_fn=<NllLossBackward>)
[235] Train Loss:  tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[236] Train Loss:  tensor(0.8606, device='cuda:0', grad_fn=<NllLossBackward>)
[237] Train Loss:  tensor(0.8658, device='cuda:0', grad_fn=<NllLossBackward>)
[238] Train Loss:  tensor(0.8344, device='cuda:0', grad_fn=<NllLossBackward>)
[239] Train Loss:  tensor(0.8306, device='cuda:0', grad_fn=<NllLossBackward>)
[240] Train Loss:  tensor(0.8591, device='cuda:0', grad_fn=<NllLossBackward>)
[241] Train Loss:  tensor(0.8373, device='cuda:0', grad_fn=<NllLossBackward>)
[242] Train Loss:  tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[243] Train Loss:  tensor(0.8639, device='cuda:0', grad_fn=<NllLossBackward>)
[244] Train Loss:  tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[245] Train Loss:  tensor(0.8625, device='cuda:0', grad_fn=<NllLossBackward>)
[246] Train Loss:  tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[247] Train Loss:  tensor(0.8728, device='cuda:0', grad_fn=<NllLossBackward>)
[248] Train Loss:  tensor(0.8824, device='cuda:0', grad_fn=<NllLossBackward>)
[249] Train Loss:  tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[250] Train Loss:  tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[251] Train Loss:  tensor(0.8579, device='cuda:0', grad_fn=<NllLossBackward>)
[252] Train Loss:  tensor(0.8670, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[253] Train Loss:   tensor(0.8251, device='cuda:0', grad_fn=<NllLossBackward>)
[254] Train Loss:   tensor(0.8609, device='cuda:0', grad_fn=<NllLossBackward>)
[255] Train Loss:   tensor(0.8462, device='cuda:0', grad_fn=<NllLossBackward>)
[256] Train Loss:   tensor(0.8481, device='cuda:0', grad_fn=<NllLossBackward>)
[257] Train Loss:   tensor(0.8518, device='cuda:0', grad_fn=<NllLossBackward>)
[258] Train Loss:   tensor(0.8170, device='cuda:0', grad_fn=<NllLossBackward>)
[259] Train Loss:   tensor(0.8780, device='cuda:0', grad_fn=<NllLossBackward>)
[260] Train Loss:   tensor(0.8530, device='cuda:0', grad_fn=<NllLossBackward>)
[261] Train Loss:   tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[262] Train Loss:   tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[263] Train Loss:   tensor(0.8612, device='cuda:0', grad_fn=<NllLossBackward>)
[264] Train Loss:   tensor(0.8304, device='cuda:0', grad_fn=<NllLossBackward>)
[265] Train Loss:   tensor(0.8597, device='cuda:0', grad_fn=<NllLossBackward>)
[266] Train Loss:   tensor(0.8341, device='cuda:0', grad_fn=<NllLossBackward>)
[267] Train Loss:   tensor(0.8843, device='cuda:0', grad_fn=<NllLossBackward>)
[268] Train Loss:   tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[269] Train Loss:   tensor(0.8462, device='cuda:0', grad_fn=<NllLossBackward>)
[270] Train Loss:   tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[271] Train Loss:   tensor(0.8331, device='cuda:0', grad_fn=<NllLossBackward>)
[272] Train Loss:   tensor(0.8218, device='cuda:0', grad_fn=<NllLossBackward>)
[273] Train Loss:   tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[274] Train Loss:   tensor(0.8029, device='cuda:0', grad_fn=<NllLossBackward>)
[275] Train Loss:   tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[276] Train Loss:   tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[277] Train Loss:   tensor(0.8373, device='cuda:0', grad_fn=<NllLossBackward>)
[278] Train Loss:   tensor(0.8531, device='cuda:0', grad_fn=<NllLossBackward>)
[279] Train Loss:   tensor(0.8744, device='cuda:0', grad_fn=<NllLossBackward>)
[280] Train Loss:   tensor(0.8368, device='cuda:0', grad_fn=<NllLossBackward>)
[281] Train Loss:   tensor(0.8825, device='cuda:0', grad_fn=<NllLossBackward>)
[282] Train Loss:   tensor(0.8434, device='cuda:0', grad_fn=<NllLossBackward>)
[283] Train Loss:   tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[284] Train Loss:   tensor(0.8525, device='cuda:0', grad_fn=<NllLossBackward>)
[285] Train Loss:   tensor(0.8361, device='cuda:0', grad_fn=<NllLossBackward>)
[286] Train Loss:   tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[287] Train Loss:   tensor(0.8669, device='cuda:0', grad_fn=<NllLossBackward>)
[288] Train Loss:   tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[289] Train Loss:   tensor(0.8286, device='cuda:0', grad_fn=<NllLossBackward>)
[290] Train Loss:   tensor(0.8191, device='cuda:0', grad_fn=<NllLossBackward>)
[291] Train Loss:   tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[292] Train Loss:   tensor(0.8810, device='cuda:0', grad_fn=<NllLossBackward>)
[293] Train Loss:   tensor(0.8580, device='cuda:0', grad_fn=<NllLossBackward>)
[294] Train Loss:   tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[295] Train Loss:   tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[296] Train Loss:   tensor(0.8555, device='cuda:0', grad_fn=<NllLossBackward>)
[297] Train Loss:   tensor(0.8885, device='cuda:0', grad_fn=<NllLossBackward>)
[298] Train Loss:   tensor(0.8166, device='cuda:0', grad_fn=<NllLossBackward>)
[299] Train Loss:   tensor(0.8551, device='cuda:0', grad_fn=<NllLossBackward>)
[300] Train Loss:   tensor(0.8670, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[301] Train Loss:    tensor(0.8685, device='cuda:0', grad_fn=<NllLossBackward>)
[302] Train Loss:    tensor(0.8212, device='cuda:0', grad_fn=<NllLossBackward>)
[303] Train Loss:    tensor(0.8544, device='cuda:0', grad_fn=<NllLossBackward>)
[304] Train Loss:    tensor(0.8772, device='cuda:0', grad_fn=<NllLossBackward>)
[305] Train Loss:    tensor(0.8163, device='cuda:0', grad_fn=<NllLossBackward>)
[306] Train Loss:    tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[307] Train Loss:    tensor(0.8365, device='cuda:0', grad_fn=<NllLossBackward>)
[308] Train Loss:    tensor(0.8702, device='cuda:0', grad_fn=<NllLossBackward>)
[309] Train Loss:    tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[310] Train Loss:    tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[311] Train Loss:    tensor(0.8225, device='cuda:0', grad_fn=<NllLossBackward>)
[312] Train Loss:    tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[313] Train Loss:    tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[314] Train Loss:    tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[315] Train Loss:    tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[316] Train Loss:    tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[317] Train Loss:    tensor(0.8298, device='cuda:0', grad_fn=<NllLossBackward>)
[318] Train Loss:    tensor(0.8728, device='cuda:0', grad_fn=<NllLossBackward>)
[319] Train Loss:    tensor(0.8486, device='cuda:0', grad_fn=<NllLossBackward>)
[320] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[321] Train Loss:    tensor(0.8581, device='cuda:0', grad_fn=<NllLossBackward>)
[322] Train Loss:    tensor(0.8179, device='cuda:0', grad_fn=<NllLossBackward>)
[323] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[324] Train Loss:    tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[325] Train Loss:    tensor(0.8468, device='cuda:0', grad_fn=<NllLossBackward>)
[326] Train Loss:    tensor(0.8630, device='cuda:0', grad_fn=<NllLossBackward>)
[327] Train Loss:    tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[328] Train Loss:    tensor(0.8657, device='cuda:0', grad_fn=<NllLossBackward>)
[329] Train Loss:    tensor(0.8371, device='cuda:0', grad_fn=<NllLossBackward>)
[330] Train Loss:    tensor(0.8538, device='cuda:0', grad_fn=<NllLossBackward>)
[331] Train Loss:    tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[332] Train Loss:    tensor(0.8263, device='cuda:0', grad_fn=<NllLossBackward>)
[333] Train Loss:    tensor(0.8334, device='cuda:0', grad_fn=<NllLossBackward>)
[334] Train Loss:    tensor(0.8392, device='cuda:0', grad_fn=<NllLossBackward>)
[335] Train Loss:    tensor(0.8635, device='cuda:0', grad_fn=<NllLossBackward>)
[336] Train Loss:    tensor(0.8568, device='cuda:0', grad_fn=<NllLossBackward>)
[337] Train Loss:    tensor(0.8374, device='cuda:0', grad_fn=<NllLossBackward>)
[338] Train Loss:    tensor(0.8427, device='cuda:0', grad_fn=<NllLossBackward>)
[339] Train Loss:    tensor(0.8459, device='cuda:0', grad_fn=<NllLossBackward>)
[340] Train Loss:    tensor(0.8823, device='cuda:0', grad_fn=<NllLossBackward>)
[341] Train Loss:    tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[342] Train Loss:    tensor(0.8225, device='cuda:0', grad_fn=<NllLossBackward>)
[343] Train Loss:    tensor(0.8490, device='cuda:0', grad_fn=<NllLossBackward>)
[344] Train Loss:    tensor(0.8617, device='cuda:0', grad_fn=<NllLossBackward>)
[345] Train Loss:    tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[346] Train Loss:    tensor(0.8044, device='cuda:0', grad_fn=<NllLossBackward>)
[347] Train Loss:    tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
[348] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[349] Train Loss:    tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[350] Train Loss:    tensor(0.8404, device='cuda:0', grad_fn=<NllLossBackward>)
[351] Train Loss:    tensor(0.8597, device='cuda:0', grad_fn=<NllLossBackward>)
[352] Train Loss:    tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[353] Train Loss:    tensor(0.8621, device='cuda:0', grad_fn=<NllLossBackward>)
[354] Train Loss:    tensor(0.8493, device='cuda:0', grad_fn=<NllLossBackward>)
[355] Train Loss:    tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[356] Train Loss:    tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[357] Train Loss:    tensor(0.8409, device='cuda:0', grad_fn=<NllLossBackward>)
[358] Train Loss:    tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[359] Train Loss:    tensor(0.8244, device='cuda:0', grad_fn=<NllLossBackward>)
[360] Train Loss:    tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[361] Train Loss:    tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[362] Train Loss:    tensor(0.8391, device='cuda:0', grad_fn=<NllLossBackward>)
[363] Train Loss:    tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[364] Train Loss:    tensor(0.8447, device='cuda:0', grad_fn=<NllLossBackward>)
[365] Train Loss:    tensor(0.8446, device='cuda:0', grad_fn=<NllLossBackward>)
[366] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[367] Train Loss:    tensor(0.8519, device='cuda:0', grad_fn=<NllLossBackward>)
[368] Train Loss:    tensor(0.8436, device='cuda:0', grad_fn=<NllLossBackward>)
[369] Train Loss:    tensor(0.8529, device='cuda:0', grad_fn=<NllLossBackward>)
[370] Train Loss:    tensor(0.8401, device='cuda:0', grad_fn=<NllLossBackward>)
[371] Train Loss:    tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[372] Train Loss:    tensor(0.8419, device='cuda:0', grad_fn=<NllLossBackward>)
[373] Train Loss:    tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[374] Train Loss:    tensor(0.8651, device='cuda:0', grad_fn=<NllLossBackward>)
[375] Train Loss:    tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
[376] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[377] Train Loss:    tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[378] Train Loss:    tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[379] Train Loss:    tensor(0.8427, device='cuda:0', grad_fn=<NllLossBackward>)
[380] Train Loss:    tensor(0.8474, device='cuda:0', grad_fn=<NllLossBackward>)
[381] Train Loss:    tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[382] Train Loss:    tensor(0.8538, device='cuda:0', grad_fn=<NllLossBackward>)
[383] Train Loss:    tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[384] Train Loss:    tensor(0.8201, device='cuda:0', grad_fn=<NllLossBackward>)
[385] Train Loss:    tensor(0.8344, device='cuda:0', grad_fn=<NllLossBackward>)
[386] Train Loss:    tensor(0.8406, device='cuda:0', grad_fn=<NllLossBackward>)
[387] Train Loss:    tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[388] Train Loss:    tensor(0.8523, device='cuda:0', grad_fn=<NllLossBackward>)
[389] Train Loss:    tensor(0.8358, device='cuda:0', grad_fn=<NllLossBackward>)
[390] Train Loss:    tensor(0.8367, device='cuda:0', grad_fn=<NllLossBackward>)
[391] Train Loss:    tensor(0.8813, device='cuda:0', grad_fn=<NllLossBackward>)
[392] Train Loss:    tensor(0.8421, device='cuda:0', grad_fn=<NllLossBackward>)
[393] Train Loss:    tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
[394] Train Loss:    tensor(0.8454, device='cuda:0', grad_fn=<NllLossBackward>)
[395] Train Loss:    tensor(0.8822, device='cuda:0', grad_fn=<NllLossBackward>)
[396] Train Loss:    tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[397] Train Loss:    tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[398] Train Loss:    tensor(0.8654, device='cuda:0', grad_fn=<NllLossBackward>)
[399] Train Loss:    tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[400] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[401] Train Loss:    tensor(0.8533, device='cuda:0', grad_fn=<NllLossBackward>)
[402] Train Loss:    tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[403] Train Loss:    tensor(0.8627, device='cuda:0', grad_fn=<NllLossBackward>)
[404] Train Loss:    tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[405] Train Loss:    tensor(0.8794, device='cuda:0', grad_fn=<NllLossBackward>)
[406] Train Loss:    tensor(0.8766, device='cuda:0', grad_fn=<NllLossBackward>)
[407] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[408] Train Loss:    tensor(0.8080, device='cuda:0', grad_fn=<NllLossBackward>)
[409] Train Loss:    tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[410] Train Loss:    tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[411] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[412] Train Loss:    tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[413] Train Loss:    tensor(0.8602, device='cuda:0', grad_fn=<NllLossBackward>)
[414] Train Loss:    tensor(0.8088, device='cuda:0', grad_fn=<NllLossBackward>)
[415] Train Loss:    tensor(0.8097, device='cuda:0', grad_fn=<NllLossBackward>)
[416] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[417] Train Loss:    tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[418] Train Loss:    tensor(0.8370, device='cuda:0', grad_fn=<NllLossBackward>)
[419] Train Loss:    tensor(0.8727, device='cuda:0', grad_fn=<NllLossBackward>)
[420] Train Loss:    tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[421] Train Loss:    tensor(0.8465, device='cuda:0', grad_fn=<NllLossBackward>)
[422] Train Loss:    tensor(0.8683, device='cuda:0', grad_fn=<NllLossBackward>)
[423] Train Loss:    tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[424] Train Loss:    tensor(0.8684, device='cuda:0', grad_fn=<NllLossBackward>)
[425] Train Loss:    tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[426] Train Loss:    tensor(0.8320, device='cuda:0', grad_fn=<NllLossBackward>)
[427] Train Loss:    tensor(0.8276, device='cuda:0', grad_fn=<NllLossBackward>)
[428] Train Loss:    tensor(0.8570, device='cuda:0', grad_fn=<NllLossBackward>)
[429] Train Loss:    tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[430] Train Loss:    tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[431] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[432] Train Loss:    tensor(0.8564, device='cuda:0', grad_fn=<NllLossBackward>)
[433] Train Loss:    tensor(0.8310, device='cuda:0', grad_fn=<NllLossBackward>)
[434] Train Loss:    tensor(0.8524, device='cuda:0', grad_fn=<NllLossBackward>)
[435] Train Loss:    tensor(0.8467, device='cuda:0', grad_fn=<NllLossBackward>)
[436] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[437] Train Loss:    tensor(0.7986, device='cuda:0', grad_fn=<NllLossBackward>)
[438] Train Loss:    tensor(0.8453, device='cuda:0', grad_fn=<NllLossBackward>)
[439] Train Loss:    tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[440] Train Loss:    tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[441] Train Loss:    tensor(0.8271, device='cuda:0', grad_fn=<NllLossBackward>)
[442] Train Loss:    tensor(0.8573, device='cuda:0', grad_fn=<NllLossBackward>)
[443] Train Loss:    tensor(0.8655, device='cuda:0', grad_fn=<NllLossBackward>)
[444] Train Loss:    tensor(0.8428, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[445] Train Loss:    tensor(0.8660, device='cuda:0', grad_fn=<NllLossBackward>)
[446] Train Loss:    tensor(0.8199, device='cuda:0', grad_fn=<NllLossBackward>)
[447] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[448] Train Loss:    tensor(0.8602, device='cuda:0', grad_fn=<NllLossBackward>)
[449] Train Loss:    tensor(0.8659, device='cuda:0', grad_fn=<NllLossBackward>)
[450] Train Loss:    tensor(0.8374, device='cuda:0', grad_fn=<NllLossBackward>)
[451] Train Loss:    tensor(0.8362, device='cuda:0', grad_fn=<NllLossBackward>)
[452] Train Loss:    tensor(0.8336, device='cuda:0', grad_fn=<NllLossBackward>)
[453] Train Loss:    tensor(0.8923, device='cuda:0', grad_fn=<NllLossBackward>)
[454] Train Loss:    tensor(0.8595, device='cuda:0', grad_fn=<NllLossBackward>)
[455] Train Loss:    tensor(0.8301, device='cuda:0', grad_fn=<NllLossBackward>)
[456] Train Loss:    tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[457] Train Loss:    tensor(0.8306, device='cuda:0', grad_fn=<NllLossBackward>)
[458] Train Loss:    tensor(0.8832, device='cuda:0', grad_fn=<NllLossBackward>)
[459] Train Loss:    tensor(0.8363, device='cuda:0', grad_fn=<NllLossBackward>)
[460] Train Loss:    tensor(0.8504, device='cuda:0', grad_fn=<NllLossBackward>)
[461] Train Loss:    tensor(0.8673, device='cuda:0', grad_fn=<NllLossBackward>)
[462] Train Loss:    tensor(0.8183, device='cuda:0', grad_fn=<NllLossBackward>)
[463] Train Loss:    tensor(0.8315, device='cuda:0', grad_fn=<NllLossBackward>)
[464] Train Loss:    tensor(0.8411, device='cuda:0', grad_fn=<NllLossBackward>)
[465] Train Loss:    tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[466] Train Loss:    tensor(0.8481, device='cuda:0', grad_fn=<NllLossBackward>)
[467] Train Loss:    tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[468] Train Loss:    tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[469] Train Loss:    tensor(0.8154, device='cuda:0', grad_fn=<NllLossBackward>)
[470] Train Loss:    tensor(0.8397, device='cuda:0', grad_fn=<NllLossBackward>)
[471] Train Loss:    tensor(0.8271, device='cuda:0', grad_fn=<NllLossBackward>)
[472] Train Loss:    tensor(0.8570, device='cuda:0', grad_fn=<NllLossBackward>)
[473] Train Loss:    tensor(0.8327, device='cuda:0', grad_fn=<NllLossBackward>)
[474] Train Loss:    tensor(0.8854, device='cuda:0', grad_fn=<NllLossBackward>)
[475] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[476] Train Loss:    tensor(0.8511, device='cuda:0', grad_fn=<NllLossBackward>)
[477] Train Loss:    tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[478] Train Loss:    tensor(0.8493, device='cuda:0', grad_fn=<NllLossBackward>)
[479] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[480] Train Loss:    tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[481] Train Loss:    tensor(0.8356, device='cuda:0', grad_fn=<NllLossBackward>)
[482] Train Loss:    tensor(0.8520, device='cuda:0', grad_fn=<NllLossBackward>)
[483] Train Loss:    tensor(0.8547, device='cuda:0', grad_fn=<NllLossBackward>)
[484] Train Loss:    tensor(0.8336, device='cuda:0', grad_fn=<NllLossBackward>)
[485] Train Loss:    tensor(0.8706, device='cuda:0', grad_fn=<NllLossBackward>)
[486] Train Loss:    tensor(0.8349, device='cuda:0', grad_fn=<NllLossBackward>)
[487] Train Loss:    tensor(0.8603, device='cuda:0', grad_fn=<NllLossBackward>)
[488] Train Loss:    tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[489] Train Loss:    tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[490] Train Loss:    tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[491] Train Loss:    tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[492] Train Loss:    tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[493] Train Loss:   tensor(0.8479, device='cuda:0', grad_fn=<NllLossBackward>)
[494] Train Loss:   tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[495] Train Loss:   tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[496] Train Loss:   tensor(0.8347, device='cuda:0', grad_fn=<NllLossBackward>)
[497] Train Loss:   tensor(0.8260, device='cuda:0', grad_fn=<NllLossBackward>)
[498] Train Loss:   tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
[499] Train Loss:   tensor(0.8518, device='cuda:0', grad_fn=<NllLossBackward>)
[500] Train Loss:   tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[501] Train Loss:   tensor(0.8563, device='cuda:0', grad_fn=<NllLossBackward>)
[502] Train Loss:   tensor(0.8364, device='cuda:0', grad_fn=<NllLossBackward>)
[503] Train Loss:   tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[504] Train Loss:   tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[505] Train Loss:   tensor(0.8206, device='cuda:0', grad_fn=<NllLossBackward>)
[506] Train Loss:   tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[507] Train Loss:   tensor(0.8036, device='cuda:0', grad_fn=<NllLossBackward>)
[508] Train Loss:   tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[509] Train Loss:   tensor(0.8545, device='cuda:0', grad_fn=<NllLossBackward>)
[510] Train Loss:   tensor(0.8251, device='cuda:0', grad_fn=<NllLossBackward>)
[511] Train Loss:   tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[512] Train Loss:   tensor(0.8577, device='cuda:0', grad_fn=<NllLossBackward>)
[513] Train Loss:   tensor(0.8587, device='cuda:0', grad_fn=<NllLossBackward>)
[514] Train Loss:   tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
[515] Train Loss:   tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[516] Train Loss:   tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
[517] Train Loss:   tensor(0.8280, device='cuda:0', grad_fn=<NllLossBackward>)
[518] Train Loss:   tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[519] Train Loss:   tensor(0.8358, device='cuda:0', grad_fn=<NllLossBackward>)
[520] Train Loss:   tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[521] Train Loss:   tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[522] Train Loss:   tensor(0.8558, device='cuda:0', grad_fn=<NllLossBackward>)
[523] Train Loss:   tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[524] Train Loss:   tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[525] Train Loss:   tensor(0.8866, device='cuda:0', grad_fn=<NllLossBackward>)
[526] Train Loss:   tensor(0.8187, device='cuda:0', grad_fn=<NllLossBackward>)
[527] Train Loss:   tensor(0.8235, device='cuda:0', grad_fn=<NllLossBackward>)
[528] Train Loss:   tensor(0.8429, device='cuda:0', grad_fn=<NllLossBackward>)
[529] Train Loss:   tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[530] Train Loss:   tensor(0.8541, device='cuda:0', grad_fn=<NllLossBackward>)
[531] Train Loss:   tensor(0.8442, device='cuda:0', grad_fn=<NllLossBackward>)
[532] Train Loss:   tensor(0.8572, device='cuda:0', grad_fn=<NllLossBackward>)
[533] Train Loss:   tensor(0.8170, device='cuda:0', grad_fn=<NllLossBackward>)
[534] Train Loss:   tensor(0.8494, device='cuda:0', grad_fn=<NllLossBackward>)
[535] Train Loss:   tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[536] Train Loss:   tensor(0.8342, device='cuda:0', grad_fn=<NllLossBackward>)
[537] Train Loss:   tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
[538] Train Loss:   tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[539] Train Loss:   tensor(0.8724, device='cuda:0', grad_fn=<NllLossBackward>)
[540] Train Loss:   tensor(0.8506, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[541] Train Loss:    tensor(0.8469, device='cuda:0', grad_fn=<NllLossBackward>)
[542] Train Loss:    tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[543] Train Loss:    tensor(0.8685, device='cuda:0', grad_fn=<NllLossBackward>)
[544] Train Loss:    tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[545] Train Loss:    tensor(0.8703, device='cuda:0', grad_fn=<NllLossBackward>)
[546] Train Loss:    tensor(0.8378, device='cuda:0', grad_fn=<NllLossBackward>)
[547] Train Loss:    tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[548] Train Loss:    tensor(0.8476, device='cuda:0', grad_fn=<NllLossBackward>)
[549] Train Loss:    tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[550] Train Loss:    tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[551] Train Loss:    tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[552] Train Loss:    tensor(0.8629, device='cuda:0', grad_fn=<NllLossBackward>)
[553] Train Loss:    tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[554] Train Loss:    tensor(0.8170, device='cuda:0', grad_fn=<NllLossBackward>)
[555] Train Loss:    tensor(0.8407, device='cuda:0', grad_fn=<NllLossBackward>)
[556] Train Loss:    tensor(0.8726, device='cuda:0', grad_fn=<NllLossBackward>)
[557] Train Loss:    tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[558] Train Loss:    tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[559] Train Loss:    tensor(0.8624, device='cuda:0', grad_fn=<NllLossBackward>)
[560] Train Loss:    tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[561] Train Loss:    tensor(0.8488, device='cuda:0', grad_fn=<NllLossBackward>)
[562] Train Loss:    tensor(0.8373, device='cuda:0', grad_fn=<NllLossBackward>)
[563] Train Loss:    tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[564] Train Loss:    tensor(0.8206, device='cuda:0', grad_fn=<NllLossBackward>)
[565] Train Loss:    tensor(0.8553, device='cuda:0', grad_fn=<NllLossBackward>)
[566] Train Loss:    tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[567] Train Loss:    tensor(0.8370, device='cuda:0', grad_fn=<NllLossBackward>)
[568] Train Loss:    tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[569] Train Loss:    tensor(0.8659, device='cuda:0', grad_fn=<NllLossBackward>)
[570] Train Loss:    tensor(0.8596, device='cuda:0', grad_fn=<NllLossBackward>)
[571] Train Loss:    tensor(0.8396, device='cuda:0', grad_fn=<NllLossBackward>)
[572] Train Loss:    tensor(0.8714, device='cuda:0', grad_fn=<NllLossBackward>)
[573] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[574] Train Loss:    tensor(0.8545, device='cuda:0', grad_fn=<NllLossBackward>)
[575] Train Loss:    tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[576] Train Loss:    tensor(0.8558, device='cuda:0', grad_fn=<NllLossBackward>)
[577] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[578] Train Loss:    tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[579] Train Loss:    tensor(0.8341, device='cuda:0', grad_fn=<NllLossBackward>)
[580] Train Loss:    tensor(0.8516, device='cuda:0', grad_fn=<NllLossBackward>)
[581] Train Loss:    tensor(0.8118, device='cuda:0', grad_fn=<NllLossBackward>)
[582] Train Loss:    tensor(0.8651, device='cuda:0', grad_fn=<NllLossBackward>)
[583] Train Loss:    tensor(0.8719, device='cuda:0', grad_fn=<NllLossBackward>)
[584] Train Loss:    tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[585] Train Loss:    tensor(0.8622, device='cuda:0', grad_fn=<NllLossBackward>)
[586] Train Loss:    tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[587] Train Loss:    tensor(0.8235, device='cuda:0', grad_fn=<NllLossBackward>)
[588] Train Loss:    tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[589] Train Loss:   tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[590] Train Loss:   tensor(0.8140, device='cuda:0', grad_fn=<NllLossBackward>)
[591] Train Loss:   tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[592] Train Loss:   tensor(0.8243, device='cuda:0', grad_fn=<NllLossBackward>)
[593] Train Loss:   tensor(0.8314, device='cuda:0', grad_fn=<NllLossBackward>)
[594] Train Loss:   tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[595] Train Loss:   tensor(0.8337, device='cuda:0', grad_fn=<NllLossBackward>)
[596] Train Loss:   tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
[597] Train Loss:   tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[598] Train Loss:   tensor(0.8737, device='cuda:0', grad_fn=<NllLossBackward>)
[599] Train Loss:   tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[600] Train Loss:   tensor(0.8143, device='cuda:0', grad_fn=<NllLossBackward>)
[601] Train Loss:   tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[602] Train Loss:   tensor(0.8611, device='cuda:0', grad_fn=<NllLossBackward>)
[603] Train Loss:   tensor(0.8488, device='cuda:0', grad_fn=<NllLossBackward>)
[604] Train Loss:   tensor(0.8878, device='cuda:0', grad_fn=<NllLossBackward>)
[605] Train Loss:   tensor(0.8825, device='cuda:0', grad_fn=<NllLossBackward>)
[606] Train Loss:   tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[607] Train Loss:   tensor(0.8811, device='cuda:0', grad_fn=<NllLossBackward>)
[608] Train Loss:   tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[609] Train Loss:   tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[610] Train Loss:   tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[611] Train Loss:   tensor(0.8303, device='cuda:0', grad_fn=<NllLossBackward>)
[612] Train Loss:   tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[613] Train Loss:   tensor(0.8566, device='cuda:0', grad_fn=<NllLossBackward>)
[614] Train Loss:   tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[615] Train Loss:   tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[616] Train Loss:   tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[617] Train Loss:   tensor(0.8452, device='cuda:0', grad_fn=<NllLossBackward>)
[618] Train Loss:   tensor(0.8238, device='cuda:0', grad_fn=<NllLossBackward>)
[619] Train Loss:   tensor(0.8206, device='cuda:0', grad_fn=<NllLossBackward>)
[620] Train Loss:   tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[621] Train Loss:   tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[622] Train Loss:   tensor(0.8390, device='cuda:0', grad_fn=<NllLossBackward>)
[623] Train Loss:   tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[624] Train Loss:   tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[625] Train Loss:   tensor(0.8521, device='cuda:0', grad_fn=<NllLossBackward>)
[626] Train Loss:   tensor(0.8516, device='cuda:0', grad_fn=<NllLossBackward>)
[627] Train Loss:   tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[628] Train Loss:   tensor(0.8778, device='cuda:0', grad_fn=<NllLossBackward>)
[629] Train Loss:   tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[630] Train Loss:   tensor(0.8538, device='cuda:0', grad_fn=<NllLossBackward>)
[631] Train Loss:   tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[632] Train Loss:   tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[633] Train Loss:   tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[634] Train Loss:   tensor(0.8828, device='cuda:0', grad_fn=<NllLossBackward>)
[635] Train Loss:   tensor(0.8520, device='cuda:0', grad_fn=<NllLossBackward>)
[636] Train Loss:   tensor(0.8409, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[637] Train Loss:   tensor(0.7934, device='cuda:0', grad_fn=<NllLossBackward>)
[638] Train Loss:   tensor(0.8432, device='cuda:0', grad_fn=<NllLossBackward>)
[639] Train Loss:   tensor(0.8587, device='cuda:0', grad_fn=<NllLossBackward>)
[640] Train Loss:   tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[641] Train Loss:   tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[642] Train Loss:   tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[643] Train Loss:   tensor(0.8726, device='cuda:0', grad_fn=<NllLossBackward>)
[644] Train Loss:   tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[645] Train Loss:   tensor(0.8336, device='cuda:0', grad_fn=<NllLossBackward>)
[646] Train Loss:   tensor(0.8694, device='cuda:0', grad_fn=<NllLossBackward>)
[647] Train Loss:   tensor(0.8307, device='cuda:0', grad_fn=<NllLossBackward>)
[648] Train Loss:   tensor(0.8390, device='cuda:0', grad_fn=<NllLossBackward>)
[649] Train Loss:   tensor(0.8100, device='cuda:0', grad_fn=<NllLossBackward>)
[650] Train Loss:   tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[651] Train Loss:   tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
[652] Train Loss:   tensor(0.8228, device='cuda:0', grad_fn=<NllLossBackward>)
[653] Train Loss:   tensor(0.8382, device='cuda:0', grad_fn=<NllLossBackward>)
[654] Train Loss:   tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[655] Train Loss:   tensor(0.8593, device='cuda:0', grad_fn=<NllLossBackward>)
[656] Train Loss:   tensor(0.8467, device='cuda:0', grad_fn=<NllLossBackward>)
[657] Train Loss:   tensor(0.8647, device='cuda:0', grad_fn=<NllLossBackward>)
[658] Train Loss:   tensor(0.8152, device='cuda:0', grad_fn=<NllLossBackward>)
[659] Train Loss:   tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[660] Train Loss:   tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[661] Train Loss:   tensor(0.8357, device='cuda:0', grad_fn=<NllLossBackward>)
[662] Train Loss:   tensor(0.8795, device='cuda:0', grad_fn=<NllLossBackward>)
[663] Train Loss:   tensor(0.8788, device='cuda:0', grad_fn=<NllLossBackward>)
[664] Train Loss:   tensor(0.8412, device='cuda:0', grad_fn=<NllLossBackward>)
[665] Train Loss:   tensor(0.8447, device='cuda:0', grad_fn=<NllLossBackward>)
[666] Train Loss:   tensor(0.8737, device='cuda:0', grad_fn=<NllLossBackward>)
[667] Train Loss:   tensor(0.8833, device='cuda:0', grad_fn=<NllLossBackward>)
[668] Train Loss:   tensor(0.8503, device='cuda:0', grad_fn=<NllLossBackward>)
[669] Train Loss:   tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[670] Train Loss:   tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[671] Train Loss:   tensor(0.8808, device='cuda:0', grad_fn=<NllLossBackward>)
[672] Train Loss:   tensor(0.8415, device='cuda:0', grad_fn=<NllLossBackward>)
[673] Train Loss:   tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[674] Train Loss:   tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[675] Train Loss:   tensor(0.8317, device='cuda:0', grad_fn=<NllLossBackward>)
[676] Train Loss:   tensor(0.8365, device='cuda:0', grad_fn=<NllLossBackward>)
[677] Train Loss:   tensor(0.8256, device='cuda:0', grad_fn=<NllLossBackward>)
[678] Train Loss:   tensor(0.8531, device='cuda:0', grad_fn=<NllLossBackward>)
[679] Train Loss:   tensor(0.8769, device='cuda:0', grad_fn=<NllLossBackward>)
[680] Train Loss:   tensor(0.8550, device='cuda:0', grad_fn=<NllLossBackward>)
[681] Train Loss:   tensor(0.8310, device='cuda:0', grad_fn=<NllLossBackward>)
[682] Train Loss:   tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[683] Train Loss:   tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[684] Train Loss:   tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[685] Train Loss:   tensor(0.8657, device='cuda:0', grad_fn=<NllLossBackward>)
[686] Train Loss:   tensor(0.8271, device='cuda:0', grad_fn=<NllLossBackward>)
[687] Train Loss:   tensor(0.8386, device='cuda:0', grad_fn=<NllLossBackward>)
[688] Train Loss:   tensor(0.8604, device='cuda:0', grad_fn=<NllLossBackward>)
[689] Train Loss:   tensor(0.8365, device='cuda:0', grad_fn=<NllLossBackward>)
[690] Train Loss:   tensor(0.8593, device='cuda:0', grad_fn=<NllLossBackward>)
[691] Train Loss:   tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[692] Train Loss:   tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[693] Train Loss:   tensor(0.8825, device='cuda:0', grad_fn=<NllLossBackward>)
[694] Train Loss:   tensor(0.8449, device='cuda:0', grad_fn=<NllLossBackward>)
[695] Train Loss:   tensor(0.8560, device='cuda:0', grad_fn=<NllLossBackward>)
[696] Train Loss:   tensor(0.8580, device='cuda:0', grad_fn=<NllLossBackward>)
[697] Train Loss:   tensor(0.8508, device='cuda:0', grad_fn=<NllLossBackward>)
[698] Train Loss:   tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[699] Train Loss:   tensor(0.8564, device='cuda:0', grad_fn=<NllLossBackward>)
[700] Train Loss:   tensor(0.8311, device='cuda:0', grad_fn=<NllLossBackward>)
[701] Train Loss:   tensor(0.8608, device='cuda:0', grad_fn=<NllLossBackward>)
[702] Train Loss:   tensor(0.8347, device='cuda:0', grad_fn=<NllLossBackward>)
[703] Train Loss:   tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[704] Train Loss:   tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
[705] Train Loss:   tensor(0.8409, device='cuda:0', grad_fn=<NllLossBackward>)
[706] Train Loss:   tensor(0.8393, device='cuda:0', grad_fn=<NllLossBackward>)
[707] Train Loss:   tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[708] Train Loss:   tensor(0.8201, device='cuda:0', grad_fn=<NllLossBackward>)
[709] Train Loss:   tensor(0.8389, device='cuda:0', grad_fn=<NllLossBackward>)
[710] Train Loss:   tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[711] Train Loss:   tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[712] Train Loss:   tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[713] Train Loss:   tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[714] Train Loss:   tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[715] Train Loss:   tensor(0.8407, device='cuda:0', grad_fn=<NllLossBackward>)
[716] Train Loss:   tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[717] Train Loss:   tensor(0.8546, device='cuda:0', grad_fn=<NllLossBackward>)
[718] Train Loss:   tensor(0.8352, device='cuda:0', grad_fn=<NllLossBackward>)
[719] Train Loss:   tensor(0.8745, device='cuda:0', grad_fn=<NllLossBackward>)
[720] Train Loss:   tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[721] Train Loss:   tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[722] Train Loss:   tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[723] Train Loss:   tensor(0.8797, device='cuda:0', grad_fn=<NllLossBackward>)
[724] Train Loss:   tensor(0.8408, device='cuda:0', grad_fn=<NllLossBackward>)
[725] Train Loss:   tensor(0.8474, device='cuda:0', grad_fn=<NllLossBackward>)
[726] Train Loss:   tensor(0.8557, device='cuda:0', grad_fn=<NllLossBackward>)
[727] Train Loss:   tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[728] Train Loss:   tensor(0.8097, device='cuda:0', grad_fn=<NllLossBackward>)
[729] Train Loss:   tensor(0.8356, device='cuda:0', grad_fn=<NllLossBackward>)
[730] Train Loss:   tensor(0.8344, device='cuda:0', grad_fn=<NllLossBackward>)
[731] Train Loss:   tensor(0.8544, device='cuda:0', grad_fn=<NllLossBackward>)
[732] Train Loss:   tensor(0.8162, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[733] Train Loss:   tensor(0.8763, device='cuda:0', grad_fn=<NllLossBackward>)
[734] Train Loss:   tensor(0.8747, device='cuda:0', grad_fn=<NllLossBackward>)
[735] Train Loss:   tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[736] Train Loss:   tensor(0.8824, device='cuda:0', grad_fn=<NllLossBackward>)
[737] Train Loss:   tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
[738] Train Loss:   tensor(0.8373, device='cuda:0', grad_fn=<NllLossBackward>)
[739] Train Loss:   tensor(0.8552, device='cuda:0', grad_fn=<NllLossBackward>)
[740] Train Loss:   tensor(0.8579, device='cuda:0', grad_fn=<NllLossBackward>)
[741] Train Loss:   tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[742] Train Loss:   tensor(0.8612, device='cuda:0', grad_fn=<NllLossBackward>)
[743] Train Loss:   tensor(0.8500, device='cuda:0', grad_fn=<NllLossBackward>)
[744] Train Loss:   tensor(0.8181, device='cuda:0', grad_fn=<NllLossBackward>)
[745] Train Loss:   tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[746] Train Loss:   tensor(0.8703, device='cuda:0', grad_fn=<NllLossBackward>)
[747] Train Loss:   tensor(0.8494, device='cuda:0', grad_fn=<NllLossBackward>)
[748] Train Loss:   tensor(0.8341, device='cuda:0', grad_fn=<NllLossBackward>)
[749] Train Loss:   tensor(0.8500, device='cuda:0', grad_fn=<NllLossBackward>)
[750] Train Loss:   tensor(0.8360, device='cuda:0', grad_fn=<NllLossBackward>)
[751] Train Loss:   tensor(0.8937, device='cuda:0', grad_fn=<NllLossBackward>)
[752] Train Loss:   tensor(0.8690, device='cuda:0', grad_fn=<NllLossBackward>)
[753] Train Loss:   tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[754] Train Loss:   tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[755] Train Loss:   tensor(0.8390, device='cuda:0', grad_fn=<NllLossBackward>)
[756] Train Loss:   tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[757] Train Loss:   tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[758] Train Loss:   tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[759] Train Loss:   tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[760] Train Loss:   tensor(0.8287, device='cuda:0', grad_fn=<NllLossBackward>)
[761] Train Loss:   tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[762] Train Loss:   tensor(0.8267, device='cuda:0', grad_fn=<NllLossBackward>)
[763] Train Loss:   tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[764] Train Loss:   tensor(0.8451, device='cuda:0', grad_fn=<NllLossBackward>)
[765] Train Loss:   tensor(0.8309, device='cuda:0', grad_fn=<NllLossBackward>)
[766] Train Loss:   tensor(0.8588, device='cuda:0', grad_fn=<NllLossBackward>)
[767] Train Loss:   tensor(0.8571, device='cuda:0', grad_fn=<NllLossBackward>)
[768] Train Loss:   tensor(0.8429, device='cuda:0', grad_fn=<NllLossBackward>)
[769] Train Loss:   tensor(0.8859, device='cuda:0', grad_fn=<NllLossBackward>)
[770] Train Loss:   tensor(0.8590, device='cuda:0', grad_fn=<NllLossBackward>)
[771] Train Loss:   tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[772] Train Loss:   tensor(0.8285, device='cuda:0', grad_fn=<NllLossBackward>)
[773] Train Loss:   tensor(0.8365, device='cuda:0', grad_fn=<NllLossBackward>)
[774] Train Loss:   tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[775] Train Loss:   tensor(0.8165, device='cuda:0', grad_fn=<NllLossBackward>)
[776] Train Loss:   tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[777] Train Loss:   tensor(0.8467, device='cuda:0', grad_fn=<NllLossBackward>)
[778] Train Loss:   tensor(0.8387, device='cuda:0', grad_fn=<NllLossBackward>)
[779] Train Loss:   tensor(0.8541, device='cuda:0', grad_fn=<NllLossBackward>)
[780] Train Loss:   tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[781] Train Loss:   tensor(0.8431, device='cuda:0', grad_fn=<NllLossBackward>)
[782] Train Loss:   tensor(0.8328, device='cuda:0', grad_fn=<NllLossBackward>)
[783] Train Loss:   tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[784] Train Loss:   tensor(0.8465, device='cuda:0', grad_fn=<NllLossBackward>)
[785] Train Loss:   tensor(0.8449, device='cuda:0', grad_fn=<NllLossBackward>)
[786] Train Loss:   tensor(0.8554, device='cuda:0', grad_fn=<NllLossBackward>)
[787] Train Loss:   tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[788] Train Loss:   tensor(0.8266, device='cuda:0', grad_fn=<NllLossBackward>)
[789] Train Loss:   tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[790] Train Loss:   tensor(0.8357, device='cuda:0', grad_fn=<NllLossBackward>)
[791] Train Loss:   tensor(0.8280, device='cuda:0', grad_fn=<NllLossBackward>)
[792] Train Loss:   tensor(0.8402, device='cuda:0', grad_fn=<NllLossBackward>)
[793] Train Loss:   tensor(0.8784, device='cuda:0', grad_fn=<NllLossBackward>)
[794] Train Loss:   tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[795] Train Loss:   tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[796] Train Loss:   tensor(0.8450, device='cuda:0', grad_fn=<NllLossBackward>)
[797] Train Loss:   tensor(0.8575, device='cuda:0', grad_fn=<NllLossBackward>)
[798] Train Loss:   tensor(0.8168, device='cuda:0', grad_fn=<NllLossBackward>)
[799] Train Loss:   tensor(0.8317, device='cuda:0', grad_fn=<NllLossBackward>)
[800] Train Loss:   tensor(0.8729, device='cuda:0', grad_fn=<NllLossBackward>)
[801] Train Loss:   tensor(0.8506, device='cuda:0', grad_fn=<NllLossBackward>)
[802] Train Loss:   tensor(0.8356, device='cuda:0', grad_fn=<NllLossBackward>)
[803] Train Loss:   tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[804] Train Loss:   tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[805] Train Loss:   tensor(0.8531, device='cuda:0', grad_fn=<NllLossBackward>)
[806] Train Loss:   tensor(0.8505, device='cuda:0', grad_fn=<NllLossBackward>)
[807] Train Loss:   tensor(0.8467, device='cuda:0', grad_fn=<NllLossBackward>)
[808] Train Loss:   tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[809] Train Loss:   tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[810] Train Loss:   tensor(0.8861, device='cuda:0', grad_fn=<NllLossBackward>)
[811] Train Loss:   tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[812] Train Loss:   tensor(0.8745, device='cuda:0', grad_fn=<NllLossBackward>)
[813] Train Loss:   tensor(0.8610, device='cuda:0', grad_fn=<NllLossBackward>)
[814] Train Loss:   tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[815] Train Loss:   tensor(0.8206, device='cuda:0', grad_fn=<NllLossBackward>)
[816] Train Loss:   tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[817] Train Loss:   tensor(0.8344, device='cuda:0', grad_fn=<NllLossBackward>)
[818] Train Loss:   tensor(0.8279, device='cuda:0', grad_fn=<NllLossBackward>)
[819] Train Loss:   tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[820] Train Loss:   tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[821] Train Loss:   tensor(0.8460, device='cuda:0', grad_fn=<NllLossBackward>)
[822] Train Loss:   tensor(0.8307, device='cuda:0', grad_fn=<NllLossBackward>)
[823] Train Loss:   tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[824] Train Loss:   tensor(0.8286, device='cuda:0', grad_fn=<NllLossBackward>)
[825] Train Loss:   tensor(0.8541, device='cuda:0', grad_fn=<NllLossBackward>)
[826] Train Loss:   tensor(0.8734, device='cuda:0', grad_fn=<NllLossBackward>)
[827] Train Loss:   tensor(0.8530, device='cuda:0', grad_fn=<NllLossBackward>)
[828] Train Loss:   tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[829] Train Loss:    tensor(0.8636, device='cuda:0', grad_fn=<NllLossBackward>)
[830] Train Loss:    tensor(0.8500, device='cuda:0', grad_fn=<NllLossBackward>)
[831] Train Loss:    tensor(0.8490, device='cuda:0', grad_fn=<NllLossBackward>)
[832] Train Loss:    tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[833] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[834] Train Loss:    tensor(0.8628, device='cuda:0', grad_fn=<NllLossBackward>)
[835] Train Loss:    tensor(0.8479, device='cuda:0', grad_fn=<NllLossBackward>)
[836] Train Loss:    tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[837] Train Loss:    tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[838] Train Loss:    tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[839] Train Loss:    tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[840] Train Loss:    tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[841] Train Loss:    tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[842] Train Loss:    tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[843] Train Loss:    tensor(0.8671, device='cuda:0', grad_fn=<NllLossBackward>)
[844] Train Loss:    tensor(0.8474, device='cuda:0', grad_fn=<NllLossBackward>)
[845] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[846] Train Loss:    tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
[847] Train Loss:    tensor(0.8541, device='cuda:0', grad_fn=<NllLossBackward>)
[848] Train Loss:    tensor(0.8869, device='cuda:0', grad_fn=<NllLossBackward>)
[849] Train Loss:    tensor(0.8186, device='cuda:0', grad_fn=<NllLossBackward>)
[850] Train Loss:    tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[851] Train Loss:    tensor(0.8510, device='cuda:0', grad_fn=<NllLossBackward>)
[852] Train Loss:    tensor(0.8449, device='cuda:0', grad_fn=<NllLossBackward>)
[853] Train Loss:    tensor(0.8349, device='cuda:0', grad_fn=<NllLossBackward>)
[854] Train Loss:    tensor(0.8648, device='cuda:0', grad_fn=<NllLossBackward>)
[855] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[856] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[857] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[858] Train Loss:    tensor(0.8434, device='cuda:0', grad_fn=<NllLossBackward>)
[859] Train Loss:    tensor(0.8418, device='cuda:0', grad_fn=<NllLossBackward>)
[860] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[861] Train Loss:    tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[862] Train Loss:    tensor(0.8531, device='cuda:0', grad_fn=<NllLossBackward>)
[863] Train Loss:    tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[864] Train Loss:    tensor(0.8614, device='cuda:0', grad_fn=<NllLossBackward>)
[865] Train Loss:    tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[866] Train Loss:    tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[867] Train Loss:    tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[868] Train Loss:    tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[869] Train Loss:    tensor(0.8167, device='cuda:0', grad_fn=<NllLossBackward>)
[870] Train Loss:    tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[871] Train Loss:    tensor(0.8300, device='cuda:0', grad_fn=<NllLossBackward>)
[872] Train Loss:    tensor(0.8408, device='cuda:0', grad_fn=<NllLossBackward>)
[873] Train Loss:    tensor(0.8474, device='cuda:0', grad_fn=<NllLossBackward>)
[874] Train Loss:    tensor(0.8730, device='cuda:0', grad_fn=<NllLossBackward>)
[875] Train Loss:    tensor(0.8543, device='cuda:0', grad_fn=<NllLossBackward>)
[876] Train Loss:    tensor(0.8562, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[877] Train Loss:   tensor(0.8611, device='cuda:0', grad_fn=<NllLossBackward>)
[878] Train Loss:   tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[879] Train Loss:   tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[880] Train Loss:   tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[881] Train Loss:   tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[882] Train Loss:   tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
[883] Train Loss:   tensor(0.8706, device='cuda:0', grad_fn=<NllLossBackward>)
[884] Train Loss:   tensor(0.8391, device='cuda:0', grad_fn=<NllLossBackward>)
[885] Train Loss:   tensor(0.8591, device='cuda:0', grad_fn=<NllLossBackward>)
[886] Train Loss:   tensor(0.8392, device='cuda:0', grad_fn=<NllLossBackward>)
[887] Train Loss:   tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[888] Train Loss:   tensor(0.8623, device='cuda:0', grad_fn=<NllLossBackward>)
[889] Train Loss:   tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[890] Train Loss:   tensor(0.8478, device='cuda:0', grad_fn=<NllLossBackward>)
[891] Train Loss:   tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[892] Train Loss:   tensor(0.8627, device='cuda:0', grad_fn=<NllLossBackward>)
[893] Train Loss:   tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[894] Train Loss:   tensor(0.8332, device='cuda:0', grad_fn=<NllLossBackward>)
[895] Train Loss:   tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[896] Train Loss:   tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[897] Train Loss:   tensor(0.8330, device='cuda:0', grad_fn=<NllLossBackward>)
[898] Train Loss:   tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[899] Train Loss:   tensor(0.8118, device='cuda:0', grad_fn=<NllLossBackward>)
[900] Train Loss:   tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[901] Train Loss:   tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[902] Train Loss:   tensor(0.8719, device='cuda:0', grad_fn=<NllLossBackward>)
[903] Train Loss:   tensor(0.8531, device='cuda:0', grad_fn=<NllLossBackward>)
[904] Train Loss:   tensor(0.8453, device='cuda:0', grad_fn=<NllLossBackward>)
[905] Train Loss:   tensor(0.8416, device='cuda:0', grad_fn=<NllLossBackward>)
[906] Train Loss:   tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[907] Train Loss:   tensor(0.8579, device='cuda:0', grad_fn=<NllLossBackward>)
[908] Train Loss:   tensor(0.8209, device='cuda:0', grad_fn=<NllLossBackward>)
[909] Train Loss:   tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[910] Train Loss:   tensor(0.8352, device='cuda:0', grad_fn=<NllLossBackward>)
[911] Train Loss:   tensor(0.8323, device='cuda:0', grad_fn=<NllLossBackward>)
[912] Train Loss:   tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[913] Train Loss:   tensor(0.8407, device='cuda:0', grad_fn=<NllLossBackward>)
[914] Train Loss:   tensor(0.8917, device='cuda:0', grad_fn=<NllLossBackward>)
[915] Train Loss:   tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[916] Train Loss:   tensor(0.8683, device='cuda:0', grad_fn=<NllLossBackward>)
[917] Train Loss:   tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[918] Train Loss:   tensor(0.8554, device='cuda:0', grad_fn=<NllLossBackward>)
[919] Train Loss:   tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[920] Train Loss:   tensor(0.8688, device='cuda:0', grad_fn=<NllLossBackward>)
[921] Train Loss:   tensor(0.8695, device='cuda:0', grad_fn=<NllLossBackward>)
[922] Train Loss:   tensor(0.8323, device='cuda:0', grad_fn=<NllLossBackward>)
[923] Train Loss:   tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[924] Train Loss:   tensor(0.8469, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[925] Train Loss:   tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[926] Train Loss:   tensor(0.8632, device='cuda:0', grad_fn=<NllLossBackward>)
[927] Train Loss:   tensor(0.8405, device='cuda:0', grad_fn=<NllLossBackward>)
[928] Train Loss:   tensor(0.8363, device='cuda:0', grad_fn=<NllLossBackward>)
[929] Train Loss:   tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[930] Train Loss:   tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[931] Train Loss:   tensor(0.8543, device='cuda:0', grad_fn=<NllLossBackward>)
[932] Train Loss:   tensor(0.8331, device='cuda:0', grad_fn=<NllLossBackward>)
[933] Train Loss:   tensor(0.8361, device='cuda:0', grad_fn=<NllLossBackward>)
[934] Train Loss:   tensor(0.8186, device='cuda:0', grad_fn=<NllLossBackward>)
[935] Train Loss:   tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[936] Train Loss:   tensor(0.8178, device='cuda:0', grad_fn=<NllLossBackward>)
[937] Train Loss:   tensor(0.8589, device='cuda:0', grad_fn=<NllLossBackward>)
[938] Train Loss:   tensor(0.8405, device='cuda:0', grad_fn=<NllLossBackward>)
[939] Train Loss:   tensor(0.8169, device='cuda:0', grad_fn=<NllLossBackward>)
[940] Train Loss:   tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[941] Train Loss:   tensor(0.8589, device='cuda:0', grad_fn=<NllLossBackward>)
[942] Train Loss:   tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[943] Train Loss:   tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[944] Train Loss:   tensor(0.8266, device='cuda:0', grad_fn=<NllLossBackward>)
[945] Train Loss:   tensor(0.8378, device='cuda:0', grad_fn=<NllLossBackward>)
[946] Train Loss:   tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[947] Train Loss:   tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[948] Train Loss:   tensor(0.8577, device='cuda:0', grad_fn=<NllLossBackward>)
[949] Train Loss:   tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[950] Train Loss:   tensor(0.8534, device='cuda:0', grad_fn=<NllLossBackward>)
[951] Train Loss:   tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
[952] Train Loss:   tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[953] Train Loss:   tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[954] Train Loss:   tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[955] Train Loss:   tensor(0.8466, device='cuda:0', grad_fn=<NllLossBackward>)
[956] Train Loss:   tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[957] Train Loss:   tensor(0.8328, device='cuda:0', grad_fn=<NllLossBackward>)
[958] Train Loss:   tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[959] Train Loss:   tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[960] Train Loss:   tensor(0.7856, device='cuda:0', grad_fn=<NllLossBackward>)
[961] Train Loss:   tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[962] Train Loss:   tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[963] Train Loss:   tensor(0.8361, device='cuda:0', grad_fn=<NllLossBackward>)
[964] Train Loss:   tensor(0.8616, device='cuda:0', grad_fn=<NllLossBackward>)
[965] Train Loss:   tensor(0.8373, device='cuda:0', grad_fn=<NllLossBackward>)
[966] Train Loss:   tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[967] Train Loss:   tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[968] Train Loss:   tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[969] Train Loss:   tensor(0.8533, device='cuda:0', grad_fn=<NllLossBackward>)
[970] Train Loss:   tensor(0.8160, device='cuda:0', grad_fn=<NllLossBackward>)
[971] Train Loss:   tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[972] Train Loss:   tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[973] Train Loss:  tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[974] Train Loss:  tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[975] Train Loss:  tensor(0.8353, device='cuda:0', grad_fn=<NllLossBackward>)
[976] Train Loss:  tensor(0.8653, device='cuda:0', grad_fn=<NllLossBackward>)
[977] Train Loss:  tensor(0.8090, device='cuda:0', grad_fn=<NllLossBackward>)
[978] Train Loss:  tensor(0.8890, device='cuda:0', grad_fn=<NllLossBackward>)
[979] Train Loss:  tensor(0.8629, device='cuda:0', grad_fn=<NllLossBackward>)
[980] Train Loss:  tensor(0.8491, device='cuda:0', grad_fn=<NllLossBackward>)
[981] Train Loss:  tensor(0.8487, device='cuda:0', grad_fn=<NllLossBackward>)
[982] Train Loss:  tensor(0.8497, device='cuda:0', grad_fn=<NllLossBackward>)
[983] Train Loss:  tensor(0.8794, device='cuda:0', grad_fn=<NllLossBackward>)
[984] Train Loss:  tensor(0.8412, device='cuda:0', grad_fn=<NllLossBackward>)
[985] Train Loss:  tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[986] Train Loss:  tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[987] Train Loss:  tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[988] Train Loss:  tensor(0.8688, device='cuda:0', grad_fn=<NllLossBackward>)
[989] Train Loss:  tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[990] Train Loss:  tensor(0.8445, device='cuda:0', grad_fn=<NllLossBackward>)
[991] Train Loss:  tensor(0.8525, device='cuda:0', grad_fn=<NllLossBackward>)
[992] Train Loss:  tensor(0.8496, device='cuda:0', grad_fn=<NllLossBackward>)
[993] Train Loss:  tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[994] Train Loss:  tensor(0.8539, device='cuda:0', grad_fn=<NllLossBackward>)
[995] Train Loss:  tensor(0.8477, device='cuda:0', grad_fn=<NllLossBackward>)
[996] Train Loss:  tensor(0.8320, device='cuda:0', grad_fn=<NllLossBackward>)
[997] Train Loss:  tensor(0.8317, device='cuda:0', grad_fn=<NllLossBackward>)
[998] Train Loss:  tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[999] Train Loss:  tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[1000] Train Loss:  tensor(0.8780, device='cuda:0', grad_fn=<NllLossBackward>)
[1001] Train Loss:  tensor(0.8803, device='cuda:0', grad_fn=<NllLossBackward>)
[1002] Train Loss:  tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[1003] Train Loss:  tensor(0.8386, device='cuda:0', grad_fn=<NllLossBackward>)
[1004] Train Loss:  tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[1005] Train Loss:  tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[1006] Train Loss:  tensor(0.8447, device='cuda:0', grad_fn=<NllLossBackward>)
[1007] Train Loss:  tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[1008] Train Loss:  tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[1009] Train Loss:  tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[1010] Train Loss:  tensor(0.8150, device='cuda:0', grad_fn=<NllLossBackward>)
[1011] Train Loss:  tensor(0.8497, device='cuda:0', grad_fn=<NllLossBackward>)
[1012] Train Loss:  tensor(0.8464, device='cuda:0', grad_fn=<NllLossBackward>)
[1013] Train Loss:  tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[1014] Train Loss:  tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[1015] Train Loss:  tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[1016] Train Loss:  tensor(0.8576, device='cuda:0', grad_fn=<NllLossBackward>)
[1017] Train Loss:  tensor(0.8396, device='cuda:0', grad_fn=<NllLossBackward>)
[1018] Train Loss:  tensor(0.8250, device='cuda:0', grad_fn=<NllLossBackward>)
[1019] Train Loss:  tensor(0.8425, device='cuda:0', grad_fn=<NllLossBackward>)
[1020] Train Loss:  tensor(0.8431, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1021] Train Loss:    tensor(0.8792, device='cuda:0', grad_fn=<NllLossBackward>)
[1022] Train Loss:    tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
[1023] Train Loss:    tensor(0.8193, device='cuda:0', grad_fn=<NllLossBackward>)
[1024] Train Loss:    tensor(0.8365, device='cuda:0', grad_fn=<NllLossBackward>)
[1025] Train Loss:    tensor(0.8232, device='cuda:0', grad_fn=<NllLossBackward>)
[1026] Train Loss:    tensor(0.8415, device='cuda:0', grad_fn=<NllLossBackward>)
[1027] Train Loss:    tensor(0.8311, device='cuda:0', grad_fn=<NllLossBackward>)
[1028] Train Loss:    tensor(0.8437, device='cuda:0', grad_fn=<NllLossBackward>)
[1029] Train Loss:    tensor(0.8268, device='cuda:0', grad_fn=<NllLossBackward>)
[1030] Train Loss:    tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[1031] Train Loss:    tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[1032] Train Loss:    tensor(0.8176, device='cuda:0', grad_fn=<NllLossBackward>)
[1033] Train Loss:    tensor(0.8625, device='cuda:0', grad_fn=<NllLossBackward>)
[1034] Train Loss:    tensor(0.8235, device='cuda:0', grad_fn=<NllLossBackward>)
[1035] Train Loss:    tensor(0.8291, device='cuda:0', grad_fn=<NllLossBackward>)
[1036] Train Loss:    tensor(0.8333, device='cuda:0', grad_fn=<NllLossBackward>)
[1037] Train Loss:    tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[1038] Train Loss:    tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[1039] Train Loss:    tensor(0.8764, device='cuda:0', grad_fn=<NllLossBackward>)
[1040] Train Loss:    tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
[1041] Train Loss:    tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[1042] Train Loss:    tensor(0.8588, device='cuda:0', grad_fn=<NllLossBackward>)
[1043] Train Loss:    tensor(0.8346, device='cuda:0', grad_fn=<NllLossBackward>)
[1044] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[1045] Train Loss:    tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
[1046] Train Loss:    tensor(0.8570, device='cuda:0', grad_fn=<NllLossBackward>)
[1047] Train Loss:    tensor(0.8168, device='cuda:0', grad_fn=<NllLossBackward>)
[1048] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[1049] Train Loss:    tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[1050] Train Loss:    tensor(0.8568, device='cuda:0', grad_fn=<NllLossBackward>)
[1051] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[1052] Train Loss:    tensor(0.8958, device='cuda:0', grad_fn=<NllLossBackward>)
[1053] Train Loss:    tensor(0.8206, device='cuda:0', grad_fn=<NllLossBackward>)
[1054] Train Loss:    tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[1055] Train Loss:    tensor(0.8889, device='cuda:0', grad_fn=<NllLossBackward>)
[1056] Train Loss:    tensor(0.8294, device='cuda:0', grad_fn=<NllLossBackward>)
[1057] Train Loss:    tensor(0.8143, device='cuda:0', grad_fn=<NllLossBackward>)
[1058] Train Loss:    tensor(0.8689, device='cuda:0', grad_fn=<NllLossBackward>)
[1059] Train Loss:    tensor(0.8294, device='cuda:0', grad_fn=<NllLossBackward>)
[1060] Train Loss:    tensor(0.8298, device='cuda:0', grad_fn=<NllLossBackward>)
[1061] Train Loss:    tensor(0.8659, device='cuda:0', grad_fn=<NllLossBackward>)
[1062] Train Loss:    tensor(0.8315, device='cuda:0', grad_fn=<NllLossBackward>)
[1063] Train Loss:    tensor(0.8796, device='cuda:0', grad_fn=<NllLossBackward>)
[1064] Train Loss:    tensor(0.8236, device='cuda:0', grad_fn=<NllLossBackward>)
[1065] Train Loss:    tensor(0.8095, device='cuda:0', grad_fn=<NllLossBackward>)
[1066] Train Loss:    tensor(0.7945, device='cuda:0', grad_fn=<NllLossBackward>)
[1067] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[1068] Train Loss:    tensor(0.8684, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1069] Train Loss:   tensor(0.8695, device='cuda:0', grad_fn=<NllLossBackward>)
[1070] Train Loss:   tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[1071] Train Loss:   tensor(0.8459, device='cuda:0', grad_fn=<NllLossBackward>)
[1072] Train Loss:   tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[1073] Train Loss:   tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[1074] Train Loss:   tensor(0.8170, device='cuda:0', grad_fn=<NllLossBackward>)
[1075] Train Loss:   tensor(0.8426, device='cuda:0', grad_fn=<NllLossBackward>)
[1076] Train Loss:   tensor(0.8458, device='cuda:0', grad_fn=<NllLossBackward>)
[1077] Train Loss:   tensor(0.8358, device='cuda:0', grad_fn=<NllLossBackward>)
[1078] Train Loss:   tensor(0.8605, device='cuda:0', grad_fn=<NllLossBackward>)
[1079] Train Loss:   tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[1080] Train Loss:   tensor(0.8401, device='cuda:0', grad_fn=<NllLossBackward>)
[1081] Train Loss:   tensor(0.8339, device='cuda:0', grad_fn=<NllLossBackward>)
[1082] Train Loss:   tensor(0.8477, device='cuda:0', grad_fn=<NllLossBackward>)
[1083] Train Loss:   tensor(0.8543, device='cuda:0', grad_fn=<NllLossBackward>)
[1084] Train Loss:   tensor(0.8413, device='cuda:0', grad_fn=<NllLossBackward>)
[1085] Train Loss:   tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[1086] Train Loss:   tensor(0.8250, device='cuda:0', grad_fn=<NllLossBackward>)
[1087] Train Loss:   tensor(0.8543, device='cuda:0', grad_fn=<NllLossBackward>)
[1088] Train Loss:   tensor(0.8420, device='cuda:0', grad_fn=<NllLossBackward>)
[1089] Train Loss:   tensor(0.8438, device='cuda:0', grad_fn=<NllLossBackward>)
[1090] Train Loss:   tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[1091] Train Loss:   tensor(0.8226, device='cuda:0', grad_fn=<NllLossBackward>)
[1092] Train Loss:   tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[1093] Train Loss:   tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[1094] Train Loss:   tensor(0.8469, device='cuda:0', grad_fn=<NllLossBackward>)
[1095] Train Loss:   tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[1096] Train Loss:   tensor(0.8201, device='cuda:0', grad_fn=<NllLossBackward>)
[1097] Train Loss:   tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[1098] Train Loss:   tensor(0.8486, device='cuda:0', grad_fn=<NllLossBackward>)
[1099] Train Loss:   tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[1100] Train Loss:   tensor(0.8500, device='cuda:0', grad_fn=<NllLossBackward>)
[1101] Train Loss:   tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[1102] Train Loss:   tensor(0.8428, device='cuda:0', grad_fn=<NllLossBackward>)
[1103] Train Loss:   tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[1104] Train Loss:   tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[1105] Train Loss:   tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[1106] Train Loss:   tensor(0.8491, device='cuda:0', grad_fn=<NllLossBackward>)
[1107] Train Loss:   tensor(0.8758, device='cuda:0', grad_fn=<NllLossBackward>)
[1108] Train Loss:   tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[1109] Train Loss:   tensor(0.8414, device='cuda:0', grad_fn=<NllLossBackward>)
[1110] Train Loss:   tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[1111] Train Loss:   tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[1112] Train Loss:   tensor(0.8557, device='cuda:0', grad_fn=<NllLossBackward>)
[1113] Train Loss:   tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[1114] Train Loss:   tensor(0.8347, device='cuda:0', grad_fn=<NllLossBackward>)
[1115] Train Loss:   tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[1116] Train Loss:   tensor(0.8241, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1117] Train Loss:   tensor(0.8414, device='cuda:0', grad_fn=<NllLossBackward>)
[1118] Train Loss:   tensor(0.8328, device='cuda:0', grad_fn=<NllLossBackward>)
[1119] Train Loss:   tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[1120] Train Loss:   tensor(0.8551, device='cuda:0', grad_fn=<NllLossBackward>)
[1121] Train Loss:   tensor(0.8802, device='cuda:0', grad_fn=<NllLossBackward>)
[1122] Train Loss:   tensor(0.8362, device='cuda:0', grad_fn=<NllLossBackward>)
[1123] Train Loss:   tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[1124] Train Loss:   tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[1125] Train Loss:   tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[1126] Train Loss:   tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[1127] Train Loss:   tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[1128] Train Loss:   tensor(0.8637, device='cuda:0', grad_fn=<NllLossBackward>)
[1129] Train Loss:   tensor(0.8348, device='cuda:0', grad_fn=<NllLossBackward>)
[1130] Train Loss:   tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[1131] Train Loss:   tensor(0.8554, device='cuda:0', grad_fn=<NllLossBackward>)
[1132] Train Loss:   tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[1133] Train Loss:   tensor(0.8388, device='cuda:0', grad_fn=<NllLossBackward>)
[1134] Train Loss:   tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[1135] Train Loss:   tensor(0.8699, device='cuda:0', grad_fn=<NllLossBackward>)
[1136] Train Loss:   tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[1137] Train Loss:   tensor(0.8127, device='cuda:0', grad_fn=<NllLossBackward>)
[1138] Train Loss:   tensor(0.8336, device='cuda:0', grad_fn=<NllLossBackward>)
[1139] Train Loss:   tensor(0.8353, device='cuda:0', grad_fn=<NllLossBackward>)
[1140] Train Loss:   tensor(0.8467, device='cuda:0', grad_fn=<NllLossBackward>)
[1141] Train Loss:   tensor(0.8787, device='cuda:0', grad_fn=<NllLossBackward>)
[1142] Train Loss:   tensor(0.8460, device='cuda:0', grad_fn=<NllLossBackward>)
[1143] Train Loss:   tensor(0.8551, device='cuda:0', grad_fn=<NllLossBackward>)
[1144] Train Loss:   tensor(0.8636, device='cuda:0', grad_fn=<NllLossBackward>)
[1145] Train Loss:   tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[1146] Train Loss:   tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[1147] Train Loss:   tensor(0.8452, device='cuda:0', grad_fn=<NllLossBackward>)
[1148] Train Loss:   tensor(0.8538, device='cuda:0', grad_fn=<NllLossBackward>)
[1149] Train Loss:   tensor(0.8368, device='cuda:0', grad_fn=<NllLossBackward>)
[1150] Train Loss:   tensor(0.8526, device='cuda:0', grad_fn=<NllLossBackward>)
[1151] Train Loss:   tensor(0.8334, device='cuda:0', grad_fn=<NllLossBackward>)
[1152] Train Loss:   tensor(0.8464, device='cuda:0', grad_fn=<NllLossBackward>)
[1153] Train Loss:   tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[1154] Train Loss:   tensor(0.8408, device='cuda:0', grad_fn=<NllLossBackward>)
[1155] Train Loss:   tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[1156] Train Loss:   tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[1157] Train Loss:   tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[1158] Train Loss:   tensor(0.8271, device='cuda:0', grad_fn=<NllLossBackward>)
[1159] Train Loss:   tensor(0.8514, device='cuda:0', grad_fn=<NllLossBackward>)
[1160] Train Loss:   tensor(0.8118, device='cuda:0', grad_fn=<NllLossBackward>)
[1161] Train Loss:   tensor(0.7859, device='cuda:0', grad_fn=<NllLossBackward>)
[1162] Train Loss:   tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[1163] Train Loss:   tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[1164] Train Loss:   tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1165] Train Loss:    tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[1166] Train Loss:    tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[1167] Train Loss:    tensor(0.8256, device='cuda:0', grad_fn=<NllLossBackward>)
[1168] Train Loss:    tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[1169] Train Loss:    tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[1170] Train Loss:    tensor(0.8392, device='cuda:0', grad_fn=<NllLossBackward>)
[1171] Train Loss:    tensor(0.8327, device='cuda:0', grad_fn=<NllLossBackward>)
[1172] Train Loss:    tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[1173] Train Loss:    tensor(0.8732, device='cuda:0', grad_fn=<NllLossBackward>)
[1174] Train Loss:    tensor(0.8422, device='cuda:0', grad_fn=<NllLossBackward>)
[1175] Train Loss:    tensor(0.8629, device='cuda:0', grad_fn=<NllLossBackward>)
[1176] Train Loss:    tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[1177] Train Loss:    tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[1178] Train Loss:    tensor(0.8363, device='cuda:0', grad_fn=<NllLossBackward>)
[1179] Train Loss:    tensor(0.8657, device='cuda:0', grad_fn=<NllLossBackward>)
[1180] Train Loss:    tensor(0.8370, device='cuda:0', grad_fn=<NllLossBackward>)
[1181] Train Loss:    tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[1182] Train Loss:    tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[1183] Train Loss:    tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[1184] Train Loss:    tensor(0.8571, device='cuda:0', grad_fn=<NllLossBackward>)
[1185] Train Loss:    tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[1186] Train Loss:    tensor(0.8193, device='cuda:0', grad_fn=<NllLossBackward>)
[1187] Train Loss:    tensor(0.8346, device='cuda:0', grad_fn=<NllLossBackward>)
[1188] Train Loss:    tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[1189] Train Loss:    tensor(0.8235, device='cuda:0', grad_fn=<NllLossBackward>)
[1190] Train Loss:    tensor(0.8621, device='cuda:0', grad_fn=<NllLossBackward>)
[1191] Train Loss:    tensor(0.8307, device='cuda:0', grad_fn=<NllLossBackward>)
[1192] Train Loss:    tensor(0.8345, device='cuda:0', grad_fn=<NllLossBackward>)
[1193] Train Loss:    tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
[1194] Train Loss:    tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[1195] Train Loss:    tensor(0.8606, device='cuda:0', grad_fn=<NllLossBackward>)
[1196] Train Loss:    tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[1197] Train Loss:    tensor(0.8438, device='cuda:0', grad_fn=<NllLossBackward>)
[1198] Train Loss:    tensor(0.8642, device='cuda:0', grad_fn=<NllLossBackward>)
[1199] Train Loss:    tensor(0.8447, device='cuda:0', grad_fn=<NllLossBackward>)
[1200] Train Loss:    tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[1201] Train Loss:    tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[1202] Train Loss:    tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[1203] Train Loss:    tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[1204] Train Loss:    tensor(0.8228, device='cuda:0', grad_fn=<NllLossBackward>)
[1205] Train Loss:    tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[1206] Train Loss:    tensor(0.8156, device='cuda:0', grad_fn=<NllLossBackward>)
[1207] Train Loss:    tensor(0.8721, device='cuda:0', grad_fn=<NllLossBackward>)
[1208] Train Loss:    tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[1209] Train Loss:    tensor(0.8350, device='cuda:0', grad_fn=<NllLossBackward>)
[1210] Train Loss:    tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[1211] Train Loss:    tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[1212] Train Loss:    tensor(0.8399, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1213] Train Loss:   tensor(0.8781, device='cuda:0', grad_fn=<NllLossBackward>)
[1214] Train Loss:   tensor(0.8390, device='cuda:0', grad_fn=<NllLossBackward>)
[1215] Train Loss:   tensor(0.8357, device='cuda:0', grad_fn=<NllLossBackward>)
[1216] Train Loss:   tensor(0.8550, device='cuda:0', grad_fn=<NllLossBackward>)
[1217] Train Loss:   tensor(0.8495, device='cuda:0', grad_fn=<NllLossBackward>)
[1218] Train Loss:   tensor(0.8678, device='cuda:0', grad_fn=<NllLossBackward>)
[1219] Train Loss:   tensor(0.8719, device='cuda:0', grad_fn=<NllLossBackward>)
[1220] Train Loss:   tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[1221] Train Loss:   tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[1222] Train Loss:   tensor(0.8212, device='cuda:0', grad_fn=<NllLossBackward>)
[1223] Train Loss:   tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[1224] Train Loss:   tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[1225] Train Loss:   tensor(0.8353, device='cuda:0', grad_fn=<NllLossBackward>)
[1226] Train Loss:   tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[1227] Train Loss:   tensor(0.8543, device='cuda:0', grad_fn=<NllLossBackward>)
[1228] Train Loss:   tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[1229] Train Loss:   tensor(0.8358, device='cuda:0', grad_fn=<NllLossBackward>)
[1230] Train Loss:   tensor(0.8491, device='cuda:0', grad_fn=<NllLossBackward>)
[1231] Train Loss:   tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[1232] Train Loss:   tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[1233] Train Loss:   tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[1234] Train Loss:   tensor(0.8446, device='cuda:0', grad_fn=<NllLossBackward>)
[1235] Train Loss:   tensor(0.8307, device='cuda:0', grad_fn=<NllLossBackward>)
[1236] Train Loss:   tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[1237] Train Loss:   tensor(0.8572, device='cuda:0', grad_fn=<NllLossBackward>)
[1238] Train Loss:   tensor(0.8221, device='cuda:0', grad_fn=<NllLossBackward>)
[1239] Train Loss:   tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[1240] Train Loss:   tensor(0.8610, device='cuda:0', grad_fn=<NllLossBackward>)
[1241] Train Loss:   tensor(0.8100, device='cuda:0', grad_fn=<NllLossBackward>)
[1242] Train Loss:   tensor(0.8161, device='cuda:0', grad_fn=<NllLossBackward>)
[1243] Train Loss:   tensor(0.8724, device='cuda:0', grad_fn=<NllLossBackward>)
[1244] Train Loss:   tensor(0.8678, device='cuda:0', grad_fn=<NllLossBackward>)
[1245] Train Loss:   tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[1246] Train Loss:   tensor(0.8192, device='cuda:0', grad_fn=<NllLossBackward>)
[1247] Train Loss:   tensor(0.9006, device='cuda:0', grad_fn=<NllLossBackward>)
[1248] Train Loss:   tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[1249] Train Loss:   tensor(0.8661, device='cuda:0', grad_fn=<NllLossBackward>)
[1250] Train Loss:   tensor(0.8152, device='cuda:0', grad_fn=<NllLossBackward>)
[1251] Train Loss:   tensor(0.8313, device='cuda:0', grad_fn=<NllLossBackward>)
[1252] Train Loss:   tensor(0.8742, device='cuda:0', grad_fn=<NllLossBackward>)
[1253] Train Loss:   tensor(0.8040, device='cuda:0', grad_fn=<NllLossBackward>)
[1254] Train Loss:   tensor(0.8566, device='cuda:0', grad_fn=<NllLossBackward>)
[1255] Train Loss:   tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[1256] Train Loss:   tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[1257] Train Loss:   tensor(0.8488, device='cuda:0', grad_fn=<NllLossBackward>)
[1258] Train Loss:   tensor(0.8462, device='cuda:0', grad_fn=<NllLossBackward>)
[1259] Train Loss:   tensor(0.8679, device='cuda:0', grad_fn=<NllLossBackward>)
[1260] Train Loss:   tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1261] Train Loss:    tensor(0.8267, device='cuda:0', grad_fn=<NllLossBackward>)
[1262] Train Loss:    tensor(0.8326, device='cuda:0', grad_fn=<NllLossBackward>)
[1263] Train Loss:    tensor(0.8445, device='cuda:0', grad_fn=<NllLossBackward>)
[1264] Train Loss:    tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[1265] Train Loss:    tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[1266] Train Loss:    tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[1267] Train Loss:    tensor(0.8333, device='cuda:0', grad_fn=<NllLossBackward>)
[1268] Train Loss:    tensor(0.8293, device='cuda:0', grad_fn=<NllLossBackward>)
[1269] Train Loss:    tensor(0.8285, device='cuda:0', grad_fn=<NllLossBackward>)
[1270] Train Loss:    tensor(0.8564, device='cuda:0', grad_fn=<NllLossBackward>)
[1271] Train Loss:    tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[1272] Train Loss:    tensor(0.8579, device='cuda:0', grad_fn=<NllLossBackward>)
[1273] Train Loss:    tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[1274] Train Loss:    tensor(0.8737, device='cuda:0', grad_fn=<NllLossBackward>)
[1275] Train Loss:    tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[1276] Train Loss:    tensor(0.8229, device='cuda:0', grad_fn=<NllLossBackward>)
[1277] Train Loss:    tensor(0.8421, device='cuda:0', grad_fn=<NllLossBackward>)
[1278] Train Loss:    tensor(0.8343, device='cuda:0', grad_fn=<NllLossBackward>)
[1279] Train Loss:    tensor(0.8193, device='cuda:0', grad_fn=<NllLossBackward>)
[1280] Train Loss:    tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[1281] Train Loss:    tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[1282] Train Loss:    tensor(0.8381, device='cuda:0', grad_fn=<NllLossBackward>)
[1283] Train Loss:    tensor(0.8187, device='cuda:0', grad_fn=<NllLossBackward>)
[1284] Train Loss:    tensor(0.8268, device='cuda:0', grad_fn=<NllLossBackward>)
[1285] Train Loss:    tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
[1286] Train Loss:    tensor(0.8513, device='cuda:0', grad_fn=<NllLossBackward>)
[1287] Train Loss:    tensor(0.8488, device='cuda:0', grad_fn=<NllLossBackward>)
[1288] Train Loss:    tensor(0.8487, device='cuda:0', grad_fn=<NllLossBackward>)
[1289] Train Loss:    tensor(0.8307, device='cuda:0', grad_fn=<NllLossBackward>)
[1290] Train Loss:    tensor(0.8432, device='cuda:0', grad_fn=<NllLossBackward>)
[1291] Train Loss:    tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[1292] Train Loss:    tensor(0.8328, device='cuda:0', grad_fn=<NllLossBackward>)
[1293] Train Loss:    tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[1294] Train Loss:    tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[1295] Train Loss:    tensor(0.8643, device='cuda:0', grad_fn=<NllLossBackward>)
[1296] Train Loss:    tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[1297] Train Loss:    tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[1298] Train Loss:    tensor(0.8462, device='cuda:0', grad_fn=<NllLossBackward>)
[1299] Train Loss:    tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[1300] Train Loss:    tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[1301] Train Loss:    tensor(0.8510, device='cuda:0', grad_fn=<NllLossBackward>)
[1302] Train Loss:    tensor(0.8652, device='cuda:0', grad_fn=<NllLossBackward>)
[1303] Train Loss:    tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[1304] Train Loss:    tensor(0.8392, device='cuda:0', grad_fn=<NllLossBackward>)
[1305] Train Loss:    tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[1306] Train Loss:    tensor(0.8546, device='cuda:0', grad_fn=<NllLossBackward>)
[1307] Train Loss:    tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[1308] Train Loss:    tensor(0.8400, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1309] Train Loss:    tensor(0.8599, device='cuda:0', grad_fn=<NllLossBackward>)
[1310] Train Loss:    tensor(0.8315, device='cuda:0', grad_fn=<NllLossBackward>)
[1311] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[1312] Train Loss:    tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[1313] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[1314] Train Loss:    tensor(0.8326, device='cuda:0', grad_fn=<NllLossBackward>)
[1315] Train Loss:    tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[1316] Train Loss:    tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[1317] Train Loss:    tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
[1318] Train Loss:    tensor(0.8212, device='cuda:0', grad_fn=<NllLossBackward>)
[1319] Train Loss:    tensor(0.8152, device='cuda:0', grad_fn=<NllLossBackward>)
[1320] Train Loss:    tensor(0.8541, device='cuda:0', grad_fn=<NllLossBackward>)
[1321] Train Loss:    tensor(0.8556, device='cuda:0', grad_fn=<NllLossBackward>)
[1322] Train Loss:    tensor(0.8152, device='cuda:0', grad_fn=<NllLossBackward>)
[1323] Train Loss:    tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[1324] Train Loss:    tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[1325] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[1326] Train Loss:    tensor(0.8731, device='cuda:0', grad_fn=<NllLossBackward>)
[1327] Train Loss:    tensor(0.8775, device='cuda:0', grad_fn=<NllLossBackward>)
[1328] Train Loss:    tensor(0.8422, device='cuda:0', grad_fn=<NllLossBackward>)
[1329] Train Loss:    tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
[1330] Train Loss:    tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
[1331] Train Loss:    tensor(0.8086, device='cuda:0', grad_fn=<NllLossBackward>)
[1332] Train Loss:    tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[1333] Train Loss:    tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[1334] Train Loss:    tensor(0.8415, device='cuda:0', grad_fn=<NllLossBackward>)
[1335] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[1336] Train Loss:    tensor(0.8481, device='cuda:0', grad_fn=<NllLossBackward>)
[1337] Train Loss:    tensor(0.8768, device='cuda:0', grad_fn=<NllLossBackward>)
[1338] Train Loss:    tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[1339] Train Loss:    tensor(0.8218, device='cuda:0', grad_fn=<NllLossBackward>)
[1340] Train Loss:    tensor(0.8627, device='cuda:0', grad_fn=<NllLossBackward>)
[1341] Train Loss:    tensor(0.8565, device='cuda:0', grad_fn=<NllLossBackward>)
[1342] Train Loss:    tensor(0.8446, device='cuda:0', grad_fn=<NllLossBackward>)
[1343] Train Loss:    tensor(0.8717, device='cuda:0', grad_fn=<NllLossBackward>)
[1344] Train Loss:    tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[1345] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[1346] Train Loss:    tensor(0.8400, device='cuda:0', grad_fn=<NllLossBackward>)
[1347] Train Loss:    tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[1348] Train Loss:    tensor(0.8119, device='cuda:0', grad_fn=<NllLossBackward>)
[1349] Train Loss:    tensor(0.8342, device='cuda:0', grad_fn=<NllLossBackward>)
[1350] Train Loss:    tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[1351] Train Loss:    tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[1352] Train Loss:    tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[1353] Train Loss:    tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[1354] Train Loss:    tensor(0.8505, device='cuda:0', grad_fn=<NllLossBackward>)
[1355] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[1356] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1357] Train Loss:    tensor(0.8422, device='cuda:0', grad_fn=<NllLossBackward>)
[1358] Train Loss:    tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[1359] Train Loss:    tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
[1360] Train Loss:    tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[1361] Train Loss:    tensor(0.8512, device='cuda:0', grad_fn=<NllLossBackward>)
[1362] Train Loss:    tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[1363] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[1364] Train Loss:    tensor(0.8343, device='cuda:0', grad_fn=<NllLossBackward>)
[1365] Train Loss:    tensor(0.8202, device='cuda:0', grad_fn=<NllLossBackward>)
[1366] Train Loss:    tensor(0.8626, device='cuda:0', grad_fn=<NllLossBackward>)
[1367] Train Loss:    tensor(0.8626, device='cuda:0', grad_fn=<NllLossBackward>)
[1368] Train Loss:    tensor(0.8675, device='cuda:0', grad_fn=<NllLossBackward>)
[1369] Train Loss:    tensor(0.8486, device='cuda:0', grad_fn=<NllLossBackward>)
[1370] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[1371] Train Loss:    tensor(0.8526, device='cuda:0', grad_fn=<NllLossBackward>)
[1372] Train Loss:    tensor(0.8774, device='cuda:0', grad_fn=<NllLossBackward>)
[1373] Train Loss:    tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[1374] Train Loss:    tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[1375] Train Loss:    tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
[1376] Train Loss:    tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[1377] Train Loss:    tensor(0.8716, device='cuda:0', grad_fn=<NllLossBackward>)
[1378] Train Loss:    tensor(0.8283, device='cuda:0', grad_fn=<NllLossBackward>)
[1379] Train Loss:    tensor(0.8371, device='cuda:0', grad_fn=<NllLossBackward>)
[1380] Train Loss:    tensor(0.8393, device='cuda:0', grad_fn=<NllLossBackward>)
[1381] Train Loss:    tensor(0.8056, device='cuda:0', grad_fn=<NllLossBackward>)
[1382] Train Loss:    tensor(0.8715, device='cuda:0', grad_fn=<NllLossBackward>)
[1383] Train Loss:    tensor(0.8581, device='cuda:0', grad_fn=<NllLossBackward>)
[1384] Train Loss:    tensor(0.8549, device='cuda:0', grad_fn=<NllLossBackward>)
[1385] Train Loss:    tensor(0.8252, device='cuda:0', grad_fn=<NllLossBackward>)
[1386] Train Loss:    tensor(0.8153, device='cuda:0', grad_fn=<NllLossBackward>)
[1387] Train Loss:    tensor(0.8341, device='cuda:0', grad_fn=<NllLossBackward>)
[1388] Train Loss:    tensor(0.8506, device='cuda:0', grad_fn=<NllLossBackward>)
[1389] Train Loss:    tensor(0.8260, device='cuda:0', grad_fn=<NllLossBackward>)
[1390] Train Loss:    tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[1391] Train Loss:    tensor(0.8410, device='cuda:0', grad_fn=<NllLossBackward>)
[1392] Train Loss:    tensor(0.8391, device='cuda:0', grad_fn=<NllLossBackward>)
[1393] Train Loss:    tensor(0.8332, device='cuda:0', grad_fn=<NllLossBackward>)
[1394] Train Loss:    tensor(0.8405, device='cuda:0', grad_fn=<NllLossBackward>)
[1395] Train Loss:    tensor(0.8416, device='cuda:0', grad_fn=<NllLossBackward>)
[1396] Train Loss:    tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[1397] Train Loss:    tensor(0.8663, device='cuda:0', grad_fn=<NllLossBackward>)
[1398] Train Loss:    tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[1399] Train Loss:    tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[1400] Train Loss:    tensor(0.8611, device='cuda:0', grad_fn=<NllLossBackward>)
[1401] Train Loss:    tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[1402] Train Loss:    tensor(0.8178, device='cuda:0', grad_fn=<NllLossBackward>)
[1403] Train Loss:    tensor(0.8622, device='cuda:0', grad_fn=<NllLossBackward>)
[1404] Train Loss:    tensor(0.8139, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1405] Train Loss:    tensor(0.8252, device='cuda:0', grad_fn=<NllLossBackward>)
[1406] Train Loss:    tensor(0.8203, device='cuda:0', grad_fn=<NllLossBackward>)
[1407] Train Loss:    tensor(0.8127, device='cuda:0', grad_fn=<NllLossBackward>)
[1408] Train Loss:    tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[1409] Train Loss:    tensor(0.8717, device='cuda:0', grad_fn=<NllLossBackward>)
[1410] Train Loss:    tensor(0.8225, device='cuda:0', grad_fn=<NllLossBackward>)
[1411] Train Loss:    tensor(0.8225, device='cuda:0', grad_fn=<NllLossBackward>)
[1412] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[1413] Train Loss:    tensor(0.8824, device='cuda:0', grad_fn=<NllLossBackward>)
[1414] Train Loss:    tensor(0.8412, device='cuda:0', grad_fn=<NllLossBackward>)
[1415] Train Loss:    tensor(0.8020, device='cuda:0', grad_fn=<NllLossBackward>)
[1416] Train Loss:    tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[1417] Train Loss:    tensor(0.8643, device='cuda:0', grad_fn=<NllLossBackward>)
[1418] Train Loss:    tensor(0.8490, device='cuda:0', grad_fn=<NllLossBackward>)
[1419] Train Loss:    tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[1420] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[1421] Train Loss:    tensor(0.8328, device='cuda:0', grad_fn=<NllLossBackward>)
[1422] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[1423] Train Loss:    tensor(0.8643, device='cuda:0', grad_fn=<NllLossBackward>)
[1424] Train Loss:    tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[1425] Train Loss:    tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[1426] Train Loss:    tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[1427] Train Loss:    tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[1428] Train Loss:    tensor(0.8759, device='cuda:0', grad_fn=<NllLossBackward>)
[1429] Train Loss:    tensor(0.8495, device='cuda:0', grad_fn=<NllLossBackward>)
[1430] Train Loss:    tensor(0.8356, device='cuda:0', grad_fn=<NllLossBackward>)
[1431] Train Loss:    tensor(0.8907, device='cuda:0', grad_fn=<NllLossBackward>)
[1432] Train Loss:    tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[1433] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[1434] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[1435] Train Loss:    tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[1436] Train Loss:    tensor(0.8715, device='cuda:0', grad_fn=<NllLossBackward>)
[1437] Train Loss:    tensor(0.8342, device='cuda:0', grad_fn=<NllLossBackward>)
[1438] Train Loss:    tensor(0.8351, device='cuda:0', grad_fn=<NllLossBackward>)
[1439] Train Loss:    tensor(0.8263, device='cuda:0', grad_fn=<NllLossBackward>)
[1440] Train Loss:    tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[1441] Train Loss:    tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[1442] Train Loss:    tensor(0.8248, device='cuda:0', grad_fn=<NllLossBackward>)
[1443] Train Loss:    tensor(0.8176, device='cuda:0', grad_fn=<NllLossBackward>)
[1444] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[1445] Train Loss:    tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[1446] Train Loss:    tensor(0.8252, device='cuda:0', grad_fn=<NllLossBackward>)
[1447] Train Loss:    tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[1448] Train Loss:    tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[1449] Train Loss:    tensor(0.7996, device='cuda:0', grad_fn=<NllLossBackward>)
[1450] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[1451] Train Loss:    tensor(0.8907, device='cuda:0', grad_fn=<NllLossBackward>)
[1452] Train Loss:    tensor(0.8459, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1453] Train Loss:    tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[1454] Train Loss:    tensor(0.8534, device='cuda:0', grad_fn=<NllLossBackward>)
[1455] Train Loss:    tensor(0.8291, device='cuda:0', grad_fn=<NllLossBackward>)
[1456] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[1457] Train Loss:    tensor(0.8382, device='cuda:0', grad_fn=<NllLossBackward>)
[1458] Train Loss:    tensor(0.8300, device='cuda:0', grad_fn=<NllLossBackward>)
[1459] Train Loss:    tensor(0.8092, device='cuda:0', grad_fn=<NllLossBackward>)
[1460] Train Loss:    tensor(0.8412, device='cuda:0', grad_fn=<NllLossBackward>)
[1461] Train Loss:    tensor(0.8153, device='cuda:0', grad_fn=<NllLossBackward>)
[1462] Train Loss:    tensor(0.8560, device='cuda:0', grad_fn=<NllLossBackward>)
[1463] Train Loss:    tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[1464] Train Loss:    tensor(0.8723, device='cuda:0', grad_fn=<NllLossBackward>)
[1465] Train Loss:    tensor(0.8029, device='cuda:0', grad_fn=<NllLossBackward>)
[1466] Train Loss:    tensor(0.8701, device='cuda:0', grad_fn=<NllLossBackward>)
[1467] Train Loss:    tensor(0.8263, device='cuda:0', grad_fn=<NllLossBackward>)
[1468] Train Loss:    tensor(0.8191, device='cuda:0', grad_fn=<NllLossBackward>)
[1469] Train Loss:    tensor(0.8317, device='cuda:0', grad_fn=<NllLossBackward>)
[1470] Train Loss:    tensor(0.8540, device='cuda:0', grad_fn=<NllLossBackward>)
[1471] Train Loss:    tensor(0.8593, device='cuda:0', grad_fn=<NllLossBackward>)
[1472] Train Loss:    tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[1473] Train Loss:    tensor(0.8579, device='cuda:0', grad_fn=<NllLossBackward>)
[1474] Train Loss:    tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[1475] Train Loss:    tensor(0.8280, device='cuda:0', grad_fn=<NllLossBackward>)
[1476] Train Loss:    tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[1477] Train Loss:    tensor(0.8510, device='cuda:0', grad_fn=<NllLossBackward>)
[1478] Train Loss:    tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[1479] Train Loss:    tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[1480] Train Loss:    tensor(0.8628, device='cuda:0', grad_fn=<NllLossBackward>)
[1481] Train Loss:    tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[1482] Train Loss:    tensor(0.8012, device='cuda:0', grad_fn=<NllLossBackward>)
[1483] Train Loss:    tensor(0.8510, device='cuda:0', grad_fn=<NllLossBackward>)
[1484] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[1485] Train Loss:    tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[1486] Train Loss:    tensor(0.8556, device='cuda:0', grad_fn=<NllLossBackward>)
[1487] Train Loss:    tensor(0.8128, device='cuda:0', grad_fn=<NllLossBackward>)
[1488] Train Loss:    tensor(0.8203, device='cuda:0', grad_fn=<NllLossBackward>)
[1489] Train Loss:    tensor(0.8495, device='cuda:0', grad_fn=<NllLossBackward>)
[1490] Train Loss:    tensor(0.8375, device='cuda:0', grad_fn=<NllLossBackward>)
[1491] Train Loss:    tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[1492] Train Loss:    tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[1493] Train Loss:    tensor(0.8710, device='cuda:0', grad_fn=<NllLossBackward>)
[1494] Train Loss:    tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[1495] Train Loss:    tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[1496] Train Loss:    tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[1497] Train Loss:    tensor(0.8485, device='cuda:0', grad_fn=<NllLossBackward>)
[1498] Train Loss:    tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[1499] Train Loss:    tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[1500] Train Loss:    tensor(0.7985, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1501] Train Loss:    tensor(0.8076, device='cuda:0', grad_fn=<NllLossBackward>)
[1502] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[1503] Train Loss:    tensor(0.8696, device='cuda:0', grad_fn=<NllLossBackward>)
[1504] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[1505] Train Loss:    tensor(0.8348, device='cuda:0', grad_fn=<NllLossBackward>)
[1506] Train Loss:    tensor(0.8501, device='cuda:0', grad_fn=<NllLossBackward>)
[1507] Train Loss:    tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
[1508] Train Loss:    tensor(0.8525, device='cuda:0', grad_fn=<NllLossBackward>)
[1509] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[1510] Train Loss:    tensor(0.8420, device='cuda:0', grad_fn=<NllLossBackward>)
[1511] Train Loss:    tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[1512] Train Loss:    tensor(0.8788, device='cuda:0', grad_fn=<NllLossBackward>)
[1513] Train Loss:    tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
[1514] Train Loss:    tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[1515] Train Loss:    tensor(0.8555, device='cuda:0', grad_fn=<NllLossBackward>)
[1516] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[1517] Train Loss:    tensor(0.7898, device='cuda:0', grad_fn=<NllLossBackward>)
[1518] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[1519] Train Loss:    tensor(0.8458, device='cuda:0', grad_fn=<NllLossBackward>)
[1520] Train Loss:    tensor(0.8406, device='cuda:0', grad_fn=<NllLossBackward>)
[1521] Train Loss:    tensor(0.8762, device='cuda:0', grad_fn=<NllLossBackward>)
[1522] Train Loss:    tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[1523] Train Loss:    tensor(0.8367, device='cuda:0', grad_fn=<NllLossBackward>)
[1524] Train Loss:    tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[1525] Train Loss:    tensor(0.8749, device='cuda:0', grad_fn=<NllLossBackward>)
[1526] Train Loss:    tensor(0.8643, device='cuda:0', grad_fn=<NllLossBackward>)
[1527] Train Loss:    tensor(0.8134, device='cuda:0', grad_fn=<NllLossBackward>)
[1528] Train Loss:    tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[1529] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[1530] Train Loss:    tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[1531] Train Loss:    tensor(0.8486, device='cuda:0', grad_fn=<NllLossBackward>)
[1532] Train Loss:    tensor(0.8310, device='cuda:0', grad_fn=<NllLossBackward>)
[1533] Train Loss:    tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[1534] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[1535] Train Loss:    tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[1536] Train Loss:    tensor(0.8465, device='cuda:0', grad_fn=<NllLossBackward>)
[1537] Train Loss:    tensor(0.8627, device='cuda:0', grad_fn=<NllLossBackward>)
[1538] Train Loss:    tensor(0.8310, device='cuda:0', grad_fn=<NllLossBackward>)
[1539] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[1540] Train Loss:    tensor(0.8525, device='cuda:0', grad_fn=<NllLossBackward>)
[1541] Train Loss:    tensor(0.8331, device='cuda:0', grad_fn=<NllLossBackward>)
[1542] Train Loss:    tensor(0.8595, device='cuda:0', grad_fn=<NllLossBackward>)
[1543] Train Loss:    tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[1544] Train Loss:    tensor(0.8408, device='cuda:0', grad_fn=<NllLossBackward>)
[1545] Train Loss:    tensor(0.8545, device='cuda:0', grad_fn=<NllLossBackward>)
[1546] Train Loss:    tensor(0.8488, device='cuda:0', grad_fn=<NllLossBackward>)
[1547] Train Loss:    tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[1548] Train Loss:    tensor(0.8476, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1549] Train Loss:   tensor(0.8615, device='cuda:0', grad_fn=<NllLossBackward>)
[1550] Train Loss:   tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[1551] Train Loss:   tensor(0.8165, device='cuda:0', grad_fn=<NllLossBackward>)
[1552] Train Loss:   tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[1553] Train Loss:   tensor(0.8505, device='cuda:0', grad_fn=<NllLossBackward>)
[1554] Train Loss:   tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[1555] Train Loss:   tensor(0.8323, device='cuda:0', grad_fn=<NllLossBackward>)
[1556] Train Loss:   tensor(0.8313, device='cuda:0', grad_fn=<NllLossBackward>)
[1557] Train Loss:   tensor(0.8129, device='cuda:0', grad_fn=<NllLossBackward>)
[1558] Train Loss:   tensor(0.8320, device='cuda:0', grad_fn=<NllLossBackward>)
[1559] Train Loss:   tensor(0.8202, device='cuda:0', grad_fn=<NllLossBackward>)
[1560] Train Loss:   tensor(0.8361, device='cuda:0', grad_fn=<NllLossBackward>)
[1561] Train Loss:   tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[1562] Train Loss:   tensor(0.8564, device='cuda:0', grad_fn=<NllLossBackward>)
[1563] Train Loss:   tensor(0.8433, device='cuda:0', grad_fn=<NllLossBackward>)
[1564] Train Loss:   tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[1565] Train Loss:   tensor(0.8333, device='cuda:0', grad_fn=<NllLossBackward>)
[1566] Train Loss:   tensor(0.8636, device='cuda:0', grad_fn=<NllLossBackward>)
[1567] Train Loss:   tensor(0.8025, device='cuda:0', grad_fn=<NllLossBackward>)
[1568] Train Loss:   tensor(0.8206, device='cuda:0', grad_fn=<NllLossBackward>)
[1569] Train Loss:   tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[1570] Train Loss:   tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[1571] Train Loss:   tensor(0.8283, device='cuda:0', grad_fn=<NllLossBackward>)
[1572] Train Loss:   tensor(0.8404, device='cuda:0', grad_fn=<NllLossBackward>)
[1573] Train Loss:   tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[1574] Train Loss:   tensor(0.8179, device='cuda:0', grad_fn=<NllLossBackward>)
[1575] Train Loss:   tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[1576] Train Loss:   tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[1577] Train Loss:   tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[1578] Train Loss:   tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[1579] Train Loss:   tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[1580] Train Loss:   tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[1581] Train Loss:   tensor(0.8516, device='cuda:0', grad_fn=<NllLossBackward>)
[1582] Train Loss:   tensor(0.8595, device='cuda:0', grad_fn=<NllLossBackward>)
[1583] Train Loss:   tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[1584] Train Loss:   tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[1585] Train Loss:   tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[1586] Train Loss:   tensor(0.8442, device='cuda:0', grad_fn=<NllLossBackward>)
[1587] Train Loss:   tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[1588] Train Loss:   tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[1589] Train Loss:   tensor(0.8567, device='cuda:0', grad_fn=<NllLossBackward>)
[1590] Train Loss:   tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[1591] Train Loss:   tensor(0.8412, device='cuda:0', grad_fn=<NllLossBackward>)
[1592] Train Loss:   tensor(0.8181, device='cuda:0', grad_fn=<NllLossBackward>)
[1593] Train Loss:   tensor(0.8599, device='cuda:0', grad_fn=<NllLossBackward>)
[1594] Train Loss:   tensor(0.8442, device='cuda:0', grad_fn=<NllLossBackward>)
[1595] Train Loss:   tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[1596] Train Loss:   tensor(0.8327, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1597] Train Loss:   tensor(0.8440, device='cuda:0', grad_fn=<NllLossBackward>)
[1598] Train Loss:   tensor(0.8711, device='cuda:0', grad_fn=<NllLossBackward>)
[1599] Train Loss:   tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[1600] Train Loss:   tensor(0.8407, device='cuda:0', grad_fn=<NllLossBackward>)
[1601] Train Loss:   tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[1602] Train Loss:   tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[1603] Train Loss:   tensor(0.8755, device='cuda:0', grad_fn=<NllLossBackward>)
[1604] Train Loss:   tensor(0.8358, device='cuda:0', grad_fn=<NllLossBackward>)
[1605] Train Loss:   tensor(0.8293, device='cuda:0', grad_fn=<NllLossBackward>)
[1606] Train Loss:   tensor(0.8587, device='cuda:0', grad_fn=<NllLossBackward>)
[1607] Train Loss:   tensor(0.8025, device='cuda:0', grad_fn=<NllLossBackward>)
[1608] Train Loss:   tensor(0.8480, device='cuda:0', grad_fn=<NllLossBackward>)
[1609] Train Loss:   tensor(0.8433, device='cuda:0', grad_fn=<NllLossBackward>)
[1610] Train Loss:   tensor(0.8418, device='cuda:0', grad_fn=<NllLossBackward>)
[1611] Train Loss:   tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[1612] Train Loss:   tensor(0.8573, device='cuda:0', grad_fn=<NllLossBackward>)
[1613] Train Loss:   tensor(0.8487, device='cuda:0', grad_fn=<NllLossBackward>)
[1614] Train Loss:   tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[1615] Train Loss:   tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[1616] Train Loss:   tensor(0.8268, device='cuda:0', grad_fn=<NllLossBackward>)
[1617] Train Loss:   tensor(0.8382, device='cuda:0', grad_fn=<NllLossBackward>)
[1618] Train Loss:   tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[1619] Train Loss:   tensor(0.8264, device='cuda:0', grad_fn=<NllLossBackward>)
[1620] Train Loss:   tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
[1621] Train Loss:   tensor(0.8293, device='cuda:0', grad_fn=<NllLossBackward>)
[1622] Train Loss:   tensor(0.8445, device='cuda:0', grad_fn=<NllLossBackward>)
[1623] Train Loss:   tensor(0.8490, device='cuda:0', grad_fn=<NllLossBackward>)
[1624] Train Loss:   tensor(0.8192, device='cuda:0', grad_fn=<NllLossBackward>)
[1625] Train Loss:   tensor(0.8421, device='cuda:0', grad_fn=<NllLossBackward>)
[1626] Train Loss:   tensor(0.8352, device='cuda:0', grad_fn=<NllLossBackward>)
[1627] Train Loss:   tensor(0.8201, device='cuda:0', grad_fn=<NllLossBackward>)
[1628] Train Loss:   tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[1629] Train Loss:   tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[1630] Train Loss:   tensor(0.8427, device='cuda:0', grad_fn=<NllLossBackward>)
[1631] Train Loss:   tensor(0.8304, device='cuda:0', grad_fn=<NllLossBackward>)
[1632] Train Loss:   tensor(0.8524, device='cuda:0', grad_fn=<NllLossBackward>)
[1633] Train Loss:   tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[1634] Train Loss:   tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[1635] Train Loss:   tensor(0.8236, device='cuda:0', grad_fn=<NllLossBackward>)
[1636] Train Loss:   tensor(0.8707, device='cuda:0', grad_fn=<NllLossBackward>)
[1637] Train Loss:   tensor(0.8442, device='cuda:0', grad_fn=<NllLossBackward>)
[1638] Train Loss:   tensor(0.8442, device='cuda:0', grad_fn=<NllLossBackward>)
[1639] Train Loss:   tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[1640] Train Loss:   tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[1641] Train Loss:   tensor(0.8400, device='cuda:0', grad_fn=<NllLossBackward>)
[1642] Train Loss:   tensor(0.8228, device='cuda:0', grad_fn=<NllLossBackward>)
[1643] Train Loss:   tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[1644] Train Loss:   tensor(0.8160, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1645] Train Loss:    tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[1646] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[1647] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[1648] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[1649] Train Loss:    tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[1650] Train Loss:    tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[1651] Train Loss:    tensor(0.8500, device='cuda:0', grad_fn=<NllLossBackward>)
[1652] Train Loss:    tensor(0.8658, device='cuda:0', grad_fn=<NllLossBackward>)
[1653] Train Loss:    tensor(0.8360, device='cuda:0', grad_fn=<NllLossBackward>)
[1654] Train Loss:    tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[1655] Train Loss:    tensor(0.8481, device='cuda:0', grad_fn=<NllLossBackward>)
[1656] Train Loss:    tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[1657] Train Loss:    tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[1658] Train Loss:    tensor(0.8362, device='cuda:0', grad_fn=<NllLossBackward>)
[1659] Train Loss:    tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[1660] Train Loss:    tensor(0.8412, device='cuda:0', grad_fn=<NllLossBackward>)
[1661] Train Loss:    tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[1662] Train Loss:    tensor(0.8348, device='cuda:0', grad_fn=<NllLossBackward>)
[1663] Train Loss:    tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[1664] Train Loss:    tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[1665] Train Loss:    tensor(0.8193, device='cuda:0', grad_fn=<NllLossBackward>)
[1666] Train Loss:    tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[1667] Train Loss:    tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[1668] Train Loss:    tensor(0.8397, device='cuda:0', grad_fn=<NllLossBackward>)
[1669] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[1670] Train Loss:    tensor(0.8346, device='cuda:0', grad_fn=<NllLossBackward>)
[1671] Train Loss:    tensor(0.8437, device='cuda:0', grad_fn=<NllLossBackward>)
[1672] Train Loss:    tensor(0.8186, device='cuda:0', grad_fn=<NllLossBackward>)
[1673] Train Loss:    tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[1674] Train Loss:    tensor(0.8558, device='cuda:0', grad_fn=<NllLossBackward>)
[1675] Train Loss:    tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[1676] Train Loss:    tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[1677] Train Loss:    tensor(0.8733, device='cuda:0', grad_fn=<NllLossBackward>)
[1678] Train Loss:    tensor(0.7975, device='cuda:0', grad_fn=<NllLossBackward>)
[1679] Train Loss:    tensor(0.8411, device='cuda:0', grad_fn=<NllLossBackward>)
[1680] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[1681] Train Loss:    tensor(0.8311, device='cuda:0', grad_fn=<NllLossBackward>)
[1682] Train Loss:    tensor(0.8477, device='cuda:0', grad_fn=<NllLossBackward>)
[1683] Train Loss:    tensor(0.8253, device='cuda:0', grad_fn=<NllLossBackward>)
[1684] Train Loss:    tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[1685] Train Loss:    tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[1686] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[1687] Train Loss:    tensor(0.8584, device='cuda:0', grad_fn=<NllLossBackward>)
[1688] Train Loss:    tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[1689] Train Loss:    tensor(0.8393, device='cuda:0', grad_fn=<NllLossBackward>)
[1690] Train Loss:    tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[1691] Train Loss:    tensor(0.8454, device='cuda:0', grad_fn=<NllLossBackward>)
[1692] Train Loss:    tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1693] Train Loss:    tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[1694] Train Loss:    tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[1695] Train Loss:    tensor(0.7948, device='cuda:0', grad_fn=<NllLossBackward>)
[1696] Train Loss:    tensor(0.8803, device='cuda:0', grad_fn=<NllLossBackward>)
[1697] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[1698] Train Loss:    tensor(0.8464, device='cuda:0', grad_fn=<NllLossBackward>)
[1699] Train Loss:    tensor(0.8581, device='cuda:0', grad_fn=<NllLossBackward>)
[1700] Train Loss:    tensor(0.8127, device='cuda:0', grad_fn=<NllLossBackward>)
[1701] Train Loss:    tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[1702] Train Loss:    tensor(0.8421, device='cuda:0', grad_fn=<NllLossBackward>)
[1703] Train Loss:    tensor(0.8343, device='cuda:0', grad_fn=<NllLossBackward>)
[1704] Train Loss:    tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[1705] Train Loss:    tensor(0.8358, device='cuda:0', grad_fn=<NllLossBackward>)
[1706] Train Loss:    tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[1707] Train Loss:    tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[1708] Train Loss:    tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[1709] Train Loss:    tensor(0.8256, device='cuda:0', grad_fn=<NllLossBackward>)
[1710] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[1711] Train Loss:    tensor(0.8651, device='cuda:0', grad_fn=<NllLossBackward>)
[1712] Train Loss:    tensor(0.8461, device='cuda:0', grad_fn=<NllLossBackward>)
[1713] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[1714] Train Loss:    tensor(0.8287, device='cuda:0', grad_fn=<NllLossBackward>)
[1715] Train Loss:    tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[1716] Train Loss:    tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[1717] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[1718] Train Loss:    tensor(0.8347, device='cuda:0', grad_fn=<NllLossBackward>)
[1719] Train Loss:    tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[1720] Train Loss:    tensor(0.8312, device='cuda:0', grad_fn=<NllLossBackward>)
[1721] Train Loss:    tensor(0.8303, device='cuda:0', grad_fn=<NllLossBackward>)
[1722] Train Loss:    tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[1723] Train Loss:    tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[1724] Train Loss:    tensor(0.8496, device='cuda:0', grad_fn=<NllLossBackward>)
[1725] Train Loss:    tensor(0.8127, device='cuda:0', grad_fn=<NllLossBackward>)
[1726] Train Loss:    tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[1727] Train Loss:    tensor(0.8646, device='cuda:0', grad_fn=<NllLossBackward>)
[1728] Train Loss:    tensor(0.8612, device='cuda:0', grad_fn=<NllLossBackward>)
[1729] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[1730] Train Loss:    tensor(0.8306, device='cuda:0', grad_fn=<NllLossBackward>)
[1731] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[1732] Train Loss:    tensor(0.8362, device='cuda:0', grad_fn=<NllLossBackward>)
[1733] Train Loss:    tensor(0.9118, device='cuda:0', grad_fn=<NllLossBackward>)
[1734] Train Loss:    tensor(0.8286, device='cuda:0', grad_fn=<NllLossBackward>)
[1735] Train Loss:    tensor(0.8382, device='cuda:0', grad_fn=<NllLossBackward>)
[1736] Train Loss:    tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[1737] Train Loss:    tensor(0.8499, device='cuda:0', grad_fn=<NllLossBackward>)
[1738] Train Loss:    tensor(0.8547, device='cuda:0', grad_fn=<NllLossBackward>)
[1739] Train Loss:    tensor(0.8226, device='cuda:0', grad_fn=<NllLossBackward>)
[1740] Train Loss:    tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1741] Train Loss:    tensor(0.8458, device='cuda:0', grad_fn=<NllLossBackward>)
[1742] Train Loss:    tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[1743] Train Loss:    tensor(0.8691, device='cuda:0', grad_fn=<NllLossBackward>)
[1744] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[1745] Train Loss:    tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[1746] Train Loss:    tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[1747] Train Loss:    tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[1748] Train Loss:    tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[1749] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[1750] Train Loss:    tensor(0.8633, device='cuda:0', grad_fn=<NllLossBackward>)
[1751] Train Loss:    tensor(0.8201, device='cuda:0', grad_fn=<NllLossBackward>)
[1752] Train Loss:    tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[1753] Train Loss:    tensor(0.8503, device='cuda:0', grad_fn=<NllLossBackward>)
[1754] Train Loss:    tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[1755] Train Loss:    tensor(0.8549, device='cuda:0', grad_fn=<NllLossBackward>)
[1756] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[1757] Train Loss:    tensor(0.8402, device='cuda:0', grad_fn=<NllLossBackward>)
[1758] Train Loss:    tensor(0.8328, device='cuda:0', grad_fn=<NllLossBackward>)
[1759] Train Loss:    tensor(0.8119, device='cuda:0', grad_fn=<NllLossBackward>)
[1760] Train Loss:    tensor(0.8363, device='cuda:0', grad_fn=<NllLossBackward>)
[1761] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[1762] Train Loss:    tensor(0.8418, device='cuda:0', grad_fn=<NllLossBackward>)
[1763] Train Loss:    tensor(0.8331, device='cuda:0', grad_fn=<NllLossBackward>)
[1764] Train Loss:    tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[1765] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[1766] Train Loss:    tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[1767] Train Loss:    tensor(0.8367, device='cuda:0', grad_fn=<NllLossBackward>)
[1768] Train Loss:    tensor(0.8425, device='cuda:0', grad_fn=<NllLossBackward>)
[1769] Train Loss:    tensor(0.8438, device='cuda:0', grad_fn=<NllLossBackward>)
[1770] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[1771] Train Loss:    tensor(0.8402, device='cuda:0', grad_fn=<NllLossBackward>)
[1772] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[1773] Train Loss:    tensor(0.8097, device='cuda:0', grad_fn=<NllLossBackward>)
[1774] Train Loss:    tensor(0.8143, device='cuda:0', grad_fn=<NllLossBackward>)
[1775] Train Loss:    tensor(0.8087, device='cuda:0', grad_fn=<NllLossBackward>)
[1776] Train Loss:    tensor(0.8552, device='cuda:0', grad_fn=<NllLossBackward>)
[1777] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[1778] Train Loss:    tensor(0.8427, device='cuda:0', grad_fn=<NllLossBackward>)
[1779] Train Loss:    tensor(0.8167, device='cuda:0', grad_fn=<NllLossBackward>)
[1780] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[1781] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[1782] Train Loss:    tensor(0.8513, device='cuda:0', grad_fn=<NllLossBackward>)
[1783] Train Loss:    tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[1784] Train Loss:    tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[1785] Train Loss:    tensor(0.8360, device='cuda:0', grad_fn=<NllLossBackward>)
[1786] Train Loss:    tensor(0.8163, device='cuda:0', grad_fn=<NllLossBackward>)
[1787] Train Loss:    tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[1788] Train Loss:    tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1789] Train Loss:    tensor(0.8199, device='cuda:0', grad_fn=<NllLossBackward>)
[1790] Train Loss:    tensor(0.8407, device='cuda:0', grad_fn=<NllLossBackward>)
[1791] Train Loss:    tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[1792] Train Loss:    tensor(0.8480, device='cuda:0', grad_fn=<NllLossBackward>)
[1793] Train Loss:    tensor(0.8374, device='cuda:0', grad_fn=<NllLossBackward>)
[1794] Train Loss:    tensor(0.8567, device='cuda:0', grad_fn=<NllLossBackward>)
[1795] Train Loss:    tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[1796] Train Loss:    tensor(0.8478, device='cuda:0', grad_fn=<NllLossBackward>)
[1797] Train Loss:    tensor(0.8256, device='cuda:0', grad_fn=<NllLossBackward>)
[1798] Train Loss:    tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[1799] Train Loss:    tensor(0.8396, device='cuda:0', grad_fn=<NllLossBackward>)
[1800] Train Loss:    tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[1801] Train Loss:    tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[1802] Train Loss:    tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[1803] Train Loss:    tensor(0.8626, device='cuda:0', grad_fn=<NllLossBackward>)
[1804] Train Loss:    tensor(0.8414, device='cuda:0', grad_fn=<NllLossBackward>)
[1805] Train Loss:    tensor(0.8402, device='cuda:0', grad_fn=<NllLossBackward>)
[1806] Train Loss:    tensor(0.8203, device='cuda:0', grad_fn=<NllLossBackward>)
[1807] Train Loss:    tensor(0.8134, device='cuda:0', grad_fn=<NllLossBackward>)
[1808] Train Loss:    tensor(0.8169, device='cuda:0', grad_fn=<NllLossBackward>)
[1809] Train Loss:    tensor(0.8600, device='cuda:0', grad_fn=<NllLossBackward>)
[1810] Train Loss:    tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[1811] Train Loss:    tensor(0.8059, device='cuda:0', grad_fn=<NllLossBackward>)
[1812] Train Loss:    tensor(0.8268, device='cuda:0', grad_fn=<NllLossBackward>)
[1813] Train Loss:    tensor(0.8464, device='cuda:0', grad_fn=<NllLossBackward>)
[1814] Train Loss:    tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[1815] Train Loss:    tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[1816] Train Loss:    tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[1817] Train Loss:    tensor(0.7898, device='cuda:0', grad_fn=<NllLossBackward>)
[1818] Train Loss:    tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[1819] Train Loss:    tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[1820] Train Loss:    tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
[1821] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[1822] Train Loss:    tensor(0.8361, device='cuda:0', grad_fn=<NllLossBackward>)
[1823] Train Loss:    tensor(0.8557, device='cuda:0', grad_fn=<NllLossBackward>)
[1824] Train Loss:    tensor(0.8271, device='cuda:0', grad_fn=<NllLossBackward>)
[1825] Train Loss:    tensor(0.8348, device='cuda:0', grad_fn=<NllLossBackward>)
[1826] Train Loss:    tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[1827] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[1828] Train Loss:    tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[1829] Train Loss:    tensor(0.8267, device='cuda:0', grad_fn=<NllLossBackward>)
[1830] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[1831] Train Loss:    tensor(0.8648, device='cuda:0', grad_fn=<NllLossBackward>)
[1832] Train Loss:    tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
[1833] Train Loss:    tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[1834] Train Loss:    tensor(0.8415, device='cuda:0', grad_fn=<NllLossBackward>)
[1835] Train Loss:    tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[1836] Train Loss:    tensor(0.8791, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1837] Train Loss:    tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[1838] Train Loss:    tensor(0.8209, device='cuda:0', grad_fn=<NllLossBackward>)
[1839] Train Loss:    tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[1840] Train Loss:    tensor(0.8631, device='cuda:0', grad_fn=<NllLossBackward>)
[1841] Train Loss:    tensor(0.8382, device='cuda:0', grad_fn=<NllLossBackward>)
[1842] Train Loss:    tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[1843] Train Loss:    tensor(0.8287, device='cuda:0', grad_fn=<NllLossBackward>)
[1844] Train Loss:    tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[1845] Train Loss:    tensor(0.8343, device='cuda:0', grad_fn=<NllLossBackward>)
[1846] Train Loss:    tensor(0.8514, device='cuda:0', grad_fn=<NllLossBackward>)
[1847] Train Loss:    tensor(0.8044, device='cuda:0', grad_fn=<NllLossBackward>)
[1848] Train Loss:    tensor(0.8256, device='cuda:0', grad_fn=<NllLossBackward>)
[1849] Train Loss:    tensor(0.8186, device='cuda:0', grad_fn=<NllLossBackward>)
[1850] Train Loss:    tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[1851] Train Loss:    tensor(0.8306, device='cuda:0', grad_fn=<NllLossBackward>)
[1852] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[1853] Train Loss:    tensor(0.8488, device='cuda:0', grad_fn=<NllLossBackward>)
[1854] Train Loss:    tensor(0.8304, device='cuda:0', grad_fn=<NllLossBackward>)
[1855] Train Loss:    tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[1856] Train Loss:    tensor(0.8314, device='cuda:0', grad_fn=<NllLossBackward>)
[1857] Train Loss:    tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[1858] Train Loss:    tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[1859] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[1860] Train Loss:    tensor(0.8452, device='cuda:0', grad_fn=<NllLossBackward>)
[1861] Train Loss:    tensor(0.8666, device='cuda:0', grad_fn=<NllLossBackward>)
[1862] Train Loss:    tensor(0.8353, device='cuda:0', grad_fn=<NllLossBackward>)
[1863] Train Loss:    tensor(0.8690, device='cuda:0', grad_fn=<NllLossBackward>)
[1864] Train Loss:    tensor(0.8438, device='cuda:0', grad_fn=<NllLossBackward>)
[1865] Train Loss:    tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[1866] Train Loss:    tensor(0.8025, device='cuda:0', grad_fn=<NllLossBackward>)
[1867] Train Loss:    tensor(0.8309, device='cuda:0', grad_fn=<NllLossBackward>)
[1868] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[1869] Train Loss:    tensor(0.8618, device='cuda:0', grad_fn=<NllLossBackward>)
[1870] Train Loss:    tensor(0.7970, device='cuda:0', grad_fn=<NllLossBackward>)
[1871] Train Loss:    tensor(0.8591, device='cuda:0', grad_fn=<NllLossBackward>)
[1872] Train Loss:    tensor(0.8309, device='cuda:0', grad_fn=<NllLossBackward>)
[1873] Train Loss:    tensor(0.7970, device='cuda:0', grad_fn=<NllLossBackward>)
[1874] Train Loss:    tensor(0.8626, device='cuda:0', grad_fn=<NllLossBackward>)
[1875] Train Loss:    tensor(0.8462, device='cuda:0', grad_fn=<NllLossBackward>)
[1876] Train Loss:    tensor(0.8303, device='cuda:0', grad_fn=<NllLossBackward>)
[1877] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[1878] Train Loss:    tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[1879] Train Loss:    tensor(0.8263, device='cuda:0', grad_fn=<NllLossBackward>)
[1880] Train Loss:    tensor(0.8347, device='cuda:0', grad_fn=<NllLossBackward>)
[1881] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[1882] Train Loss:    tensor(0.8271, device='cuda:0', grad_fn=<NllLossBackward>)
[1883] Train Loss:    tensor(0.8361, device='cuda:0', grad_fn=<NllLossBackward>)
[1884] Train Loss:    tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1885] Train Loss:    tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[1886] Train Loss:    tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[1887] Train Loss:    tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[1888] Train Loss:    tensor(0.8665, device='cuda:0', grad_fn=<NllLossBackward>)
[1889] Train Loss:    tensor(0.8085, device='cuda:0', grad_fn=<NllLossBackward>)
[1890] Train Loss:    tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[1891] Train Loss:    tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[1892] Train Loss:    tensor(0.8530, device='cuda:0', grad_fn=<NllLossBackward>)
[1893] Train Loss:    tensor(0.8418, device='cuda:0', grad_fn=<NllLossBackward>)
[1894] Train Loss:    tensor(0.8537, device='cuda:0', grad_fn=<NllLossBackward>)
[1895] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[1896] Train Loss:    tensor(0.8253, device='cuda:0', grad_fn=<NllLossBackward>)
[1897] Train Loss:    tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[1898] Train Loss:    tensor(0.8315, device='cuda:0', grad_fn=<NllLossBackward>)
[1899] Train Loss:    tensor(0.8465, device='cuda:0', grad_fn=<NllLossBackward>)
[1900] Train Loss:    tensor(0.8294, device='cuda:0', grad_fn=<NllLossBackward>)
[1901] Train Loss:    tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[1902] Train Loss:    tensor(0.8471, device='cuda:0', grad_fn=<NllLossBackward>)
[1903] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[1904] Train Loss:    tensor(0.8440, device='cuda:0', grad_fn=<NllLossBackward>)
[1905] Train Loss:    tensor(0.8770, device='cuda:0', grad_fn=<NllLossBackward>)
[1906] Train Loss:    tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[1907] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[1908] Train Loss:    tensor(0.8411, device='cuda:0', grad_fn=<NllLossBackward>)
[1909] Train Loss:    tensor(0.8302, device='cuda:0', grad_fn=<NllLossBackward>)
[1910] Train Loss:    tensor(0.8561, device='cuda:0', grad_fn=<NllLossBackward>)
[1911] Train Loss:    tensor(0.8521, device='cuda:0', grad_fn=<NllLossBackward>)
[1912] Train Loss:    tensor(0.8253, device='cuda:0', grad_fn=<NllLossBackward>)
[1913] Train Loss:    tensor(0.8589, device='cuda:0', grad_fn=<NllLossBackward>)
[1914] Train Loss:    tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[1915] Train Loss:    tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
[1916] Train Loss:    tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[1917] Train Loss:    tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[1918] Train Loss:    tensor(0.8312, device='cuda:0', grad_fn=<NllLossBackward>)
[1919] Train Loss:    tensor(0.8405, device='cuda:0', grad_fn=<NllLossBackward>)
[1920] Train Loss:    tensor(0.8781, device='cuda:0', grad_fn=<NllLossBackward>)
[1921] Train Loss:    tensor(0.8485, device='cuda:0', grad_fn=<NllLossBackward>)
[1922] Train Loss:    tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
[1923] Train Loss:    tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[1924] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[1925] Train Loss:    tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[1926] Train Loss:    tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[1927] Train Loss:    tensor(0.8578, device='cuda:0', grad_fn=<NllLossBackward>)
[1928] Train Loss:    tensor(0.8436, device='cuda:0', grad_fn=<NllLossBackward>)
[1929] Train Loss:    tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[1930] Train Loss:    tensor(0.8179, device='cuda:0', grad_fn=<NllLossBackward>)
[1931] Train Loss:    tensor(0.8367, device='cuda:0', grad_fn=<NllLossBackward>)
[1932] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1933] Train Loss:    tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[1934] Train Loss:    tensor(0.8473, device='cuda:0', grad_fn=<NllLossBackward>)
[1935] Train Loss:    tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[1936] Train Loss:    tensor(0.8306, device='cuda:0', grad_fn=<NllLossBackward>)
[1937] Train Loss:    tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[1938] Train Loss:    tensor(0.8739, device='cuda:0', grad_fn=<NllLossBackward>)
[1939] Train Loss:    tensor(0.8440, device='cuda:0', grad_fn=<NllLossBackward>)
[1940] Train Loss:    tensor(0.8371, device='cuda:0', grad_fn=<NllLossBackward>)
[1941] Train Loss:    tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[1942] Train Loss:    tensor(0.8530, device='cuda:0', grad_fn=<NllLossBackward>)
[1943] Train Loss:    tensor(0.8127, device='cuda:0', grad_fn=<NllLossBackward>)
[1944] Train Loss:    tensor(0.8154, device='cuda:0', grad_fn=<NllLossBackward>)
[1945] Train Loss:    tensor(0.8361, device='cuda:0', grad_fn=<NllLossBackward>)
[1946] Train Loss:    tensor(0.8187, device='cuda:0', grad_fn=<NllLossBackward>)
[1947] Train Loss:    tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[1948] Train Loss:    tensor(0.8357, device='cuda:0', grad_fn=<NllLossBackward>)
[1949] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[1950] Train Loss:    tensor(0.8445, device='cuda:0', grad_fn=<NllLossBackward>)
[1951] Train Loss:    tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[1952] Train Loss:    tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
[1953] Train Loss:    tensor(0.8345, device='cuda:0', grad_fn=<NllLossBackward>)
[1954] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[1955] Train Loss:    tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[1956] Train Loss:    tensor(0.8368, device='cuda:0', grad_fn=<NllLossBackward>)
[1957] Train Loss:    tensor(0.8397, device='cuda:0', grad_fn=<NllLossBackward>)
[1958] Train Loss:    tensor(0.8368, device='cuda:0', grad_fn=<NllLossBackward>)
[1959] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[1960] Train Loss:    tensor(0.8033, device='cuda:0', grad_fn=<NllLossBackward>)
[1961] Train Loss:    tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[1962] Train Loss:    tensor(0.8342, device='cuda:0', grad_fn=<NllLossBackward>)
[1963] Train Loss:    tensor(0.8761, device='cuda:0', grad_fn=<NllLossBackward>)
[1964] Train Loss:    tensor(0.8538, device='cuda:0', grad_fn=<NllLossBackward>)
[1965] Train Loss:    tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[1966] Train Loss:    tensor(0.8127, device='cuda:0', grad_fn=<NllLossBackward>)
[1967] Train Loss:    tensor(0.7858, device='cuda:0', grad_fn=<NllLossBackward>)
[1968] Train Loss:    tensor(0.8166, device='cuda:0', grad_fn=<NllLossBackward>)
[1969] Train Loss:    tensor(0.8671, device='cuda:0', grad_fn=<NllLossBackward>)
[1970] Train Loss:    tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
[1971] Train Loss:    tensor(0.8139, device='cuda:0', grad_fn=<NllLossBackward>)
[1972] Train Loss:    tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[1973] Train Loss:    tensor(0.8368, device='cuda:0', grad_fn=<NllLossBackward>)
[1974] Train Loss:    tensor(0.8293, device='cuda:0', grad_fn=<NllLossBackward>)
[1975] Train Loss:    tensor(0.8567, device='cuda:0', grad_fn=<NllLossBackward>)
[1976] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[1977] Train Loss:    tensor(0.8664, device='cuda:0', grad_fn=<NllLossBackward>)
[1978] Train Loss:    tensor(0.8518, device='cuda:0', grad_fn=<NllLossBackward>)
[1979] Train Loss:    tensor(0.8303, device='cuda:0', grad_fn=<NllLossBackward>)
[1980] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1981] Train Loss:    tensor(0.8409, device='cuda:0', grad_fn=<NllLossBackward>)
[1982] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[1983] Train Loss:    tensor(0.8525, device='cuda:0', grad_fn=<NllLossBackward>)
[1984] Train Loss:    tensor(0.8550, device='cuda:0', grad_fn=<NllLossBackward>)
[1985] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[1986] Train Loss:    tensor(0.8129, device='cuda:0', grad_fn=<NllLossBackward>)
[1987] Train Loss:    tensor(0.8440, device='cuda:0', grad_fn=<NllLossBackward>)
[1988] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[1989] Train Loss:    tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[1990] Train Loss:    tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[1991] Train Loss:    tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[1992] Train Loss:    tensor(0.8782, device='cuda:0', grad_fn=<NllLossBackward>)
[1993] Train Loss:    tensor(0.8344, device='cuda:0', grad_fn=<NllLossBackward>)
[1994] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[1995] Train Loss:    tensor(0.8341, device='cuda:0', grad_fn=<NllLossBackward>)
[1996] Train Loss:    tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[1997] Train Loss:    tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[1998] Train Loss:    tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[1999] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[2000] Train Loss:    tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[2001] Train Loss:    tensor(0.8790, device='cuda:0', grad_fn=<NllLossBackward>)
[2002] Train Loss:    tensor(0.8715, device='cuda:0', grad_fn=<NllLossBackward>)
[2003] Train Loss:    tensor(0.8095, device='cuda:0', grad_fn=<NllLossBackward>)
[2004] Train Loss:    tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[2005] Train Loss:    tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[2006] Train Loss:    tensor(0.8326, device='cuda:0', grad_fn=<NllLossBackward>)
[2007] Train Loss:    tensor(0.8454, device='cuda:0', grad_fn=<NllLossBackward>)
[2008] Train Loss:    tensor(0.8562, device='cuda:0', grad_fn=<NllLossBackward>)
[2009] Train Loss:    tensor(0.8055, device='cuda:0', grad_fn=<NllLossBackward>)
[2010] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[2011] Train Loss:    tensor(0.8315, device='cuda:0', grad_fn=<NllLossBackward>)
[2012] Train Loss:    tensor(0.8360, device='cuda:0', grad_fn=<NllLossBackward>)
[2013] Train Loss:    tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[2014] Train Loss:    tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[2015] Train Loss:    tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[2016] Train Loss:    tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[2017] Train Loss:    tensor(0.8408, device='cuda:0', grad_fn=<NllLossBackward>)
[2018] Train Loss:    tensor(0.8570, device='cuda:0', grad_fn=<NllLossBackward>)
[2019] Train Loss:    tensor(0.8121, device='cuda:0', grad_fn=<NllLossBackward>)
[2020] Train Loss:    tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[2021] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[2022] Train Loss:    tensor(0.8502, device='cuda:0', grad_fn=<NllLossBackward>)
[2023] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[2024] Train Loss:    tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[2025] Train Loss:    tensor(0.8668, device='cuda:0', grad_fn=<NllLossBackward>)
[2026] Train Loss:    tensor(0.8437, device='cuda:0', grad_fn=<NllLossBackward>)
[2027] Train Loss:    tensor(0.8307, device='cuda:0', grad_fn=<NllLossBackward>)
[2028] Train Loss:    tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2029] Train Loss:    tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[2030] Train Loss:    tensor(0.8490, device='cuda:0', grad_fn=<NllLossBackward>)
[2031] Train Loss:    tensor(0.8328, device='cuda:0', grad_fn=<NllLossBackward>)
[2032] Train Loss:    tensor(0.8360, device='cuda:0', grad_fn=<NllLossBackward>)
[2033] Train Loss:    tensor(0.9018, device='cuda:0', grad_fn=<NllLossBackward>)
[2034] Train Loss:    tensor(0.8762, device='cuda:0', grad_fn=<NllLossBackward>)
[2035] Train Loss:    tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[2036] Train Loss:    tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[2037] Train Loss:    tensor(0.8456, device='cuda:0', grad_fn=<NllLossBackward>)
[2038] Train Loss:    tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[2039] Train Loss:    tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[2040] Train Loss:    tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[2041] Train Loss:    tensor(0.8189, device='cuda:0', grad_fn=<NllLossBackward>)
[2042] Train Loss:    tensor(0.8829, device='cuda:0', grad_fn=<NllLossBackward>)
[2043] Train Loss:    tensor(0.8313, device='cuda:0', grad_fn=<NllLossBackward>)
[2044] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[2045] Train Loss:    tensor(0.8294, device='cuda:0', grad_fn=<NllLossBackward>)
[2046] Train Loss:    tensor(0.8187, device='cuda:0', grad_fn=<NllLossBackward>)
[2047] Train Loss:    tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[2048] Train Loss:    tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[2049] Train Loss:    tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[2050] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[2051] Train Loss:    tensor(0.8033, device='cuda:0', grad_fn=<NllLossBackward>)
[2052] Train Loss:    tensor(0.8414, device='cuda:0', grad_fn=<NllLossBackward>)
[2053] Train Loss:    tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[2054] Train Loss:    tensor(0.8212, device='cuda:0', grad_fn=<NllLossBackward>)
[2055] Train Loss:    tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[2056] Train Loss:    tensor(0.8480, device='cuda:0', grad_fn=<NllLossBackward>)
[2057] Train Loss:    tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[2058] Train Loss:    tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[2059] Train Loss:    tensor(0.8423, device='cuda:0', grad_fn=<NllLossBackward>)
[2060] Train Loss:    tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[2061] Train Loss:    tensor(0.8622, device='cuda:0', grad_fn=<NllLossBackward>)
[2062] Train Loss:    tensor(0.8337, device='cuda:0', grad_fn=<NllLossBackward>)
[2063] Train Loss:    tensor(0.8150, device='cuda:0', grad_fn=<NllLossBackward>)
[2064] Train Loss:    tensor(0.8256, device='cuda:0', grad_fn=<NllLossBackward>)
[2065] Train Loss:    tensor(0.8396, device='cuda:0', grad_fn=<NllLossBackward>)
[2066] Train Loss:    tensor(0.8285, device='cuda:0', grad_fn=<NllLossBackward>)
[2067] Train Loss:    tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[2068] Train Loss:    tensor(0.8504, device='cuda:0', grad_fn=<NllLossBackward>)
[2069] Train Loss:    tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[2070] Train Loss:    tensor(0.8345, device='cuda:0', grad_fn=<NllLossBackward>)
[2071] Train Loss:    tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[2072] Train Loss:    tensor(0.8584, device='cuda:0', grad_fn=<NllLossBackward>)
[2073] Train Loss:    tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[2074] Train Loss:    tensor(0.8485, device='cuda:0', grad_fn=<NllLossBackward>)
[2075] Train Loss:    tensor(0.8712, device='cuda:0', grad_fn=<NllLossBackward>)
[2076] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2077] Train Loss:    tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
[2078] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[2079] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[2080] Train Loss:    tensor(0.8747, device='cuda:0', grad_fn=<NllLossBackward>)
[2081] Train Loss:    tensor(0.8033, device='cuda:0', grad_fn=<NllLossBackward>)
[2082] Train Loss:    tensor(0.8479, device='cuda:0', grad_fn=<NllLossBackward>)
[2083] Train Loss:    tensor(0.8191, device='cuda:0', grad_fn=<NllLossBackward>)
[2084] Train Loss:    tensor(0.8459, device='cuda:0', grad_fn=<NllLossBackward>)
[2085] Train Loss:    tensor(0.8283, device='cuda:0', grad_fn=<NllLossBackward>)
[2086] Train Loss:    tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[2087] Train Loss:    tensor(0.8159, device='cuda:0', grad_fn=<NllLossBackward>)
[2088] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[2089] Train Loss:    tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[2090] Train Loss:    tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[2091] Train Loss:    tensor(0.8638, device='cuda:0', grad_fn=<NllLossBackward>)
[2092] Train Loss:    tensor(0.8688, device='cuda:0', grad_fn=<NllLossBackward>)
[2093] Train Loss:    tensor(0.8557, device='cuda:0', grad_fn=<NllLossBackward>)
[2094] Train Loss:    tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[2095] Train Loss:    tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[2096] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[2097] Train Loss:    tensor(0.8415, device='cuda:0', grad_fn=<NllLossBackward>)
[2098] Train Loss:    tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[2099] Train Loss:    tensor(0.8187, device='cuda:0', grad_fn=<NllLossBackward>)
[2100] Train Loss:    tensor(0.8285, device='cuda:0', grad_fn=<NllLossBackward>)
[2101] Train Loss:    tensor(0.8314, device='cuda:0', grad_fn=<NllLossBackward>)
[2102] Train Loss:    tensor(0.8499, device='cuda:0', grad_fn=<NllLossBackward>)
[2103] Train Loss:    tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[2104] Train Loss:    tensor(0.8484, device='cuda:0', grad_fn=<NllLossBackward>)
[2105] Train Loss:    tensor(0.8410, device='cuda:0', grad_fn=<NllLossBackward>)
[2106] Train Loss:    tensor(0.8167, device='cuda:0', grad_fn=<NllLossBackward>)
[2107] Train Loss:    tensor(0.8547, device='cuda:0', grad_fn=<NllLossBackward>)
[2108] Train Loss:    tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[2109] Train Loss:    tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[2110] Train Loss:    tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[2111] Train Loss:    tensor(0.8406, device='cuda:0', grad_fn=<NllLossBackward>)
[2112] Train Loss:    tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[2113] Train Loss:    tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
[2114] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[2115] Train Loss:    tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[2116] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[2117] Train Loss:    tensor(0.8525, device='cuda:0', grad_fn=<NllLossBackward>)
[2118] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[2119] Train Loss:    tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
[2120] Train Loss:    tensor(0.8726, device='cuda:0', grad_fn=<NllLossBackward>)
[2121] Train Loss:    tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[2122] Train Loss:    tensor(0.8518, device='cuda:0', grad_fn=<NllLossBackward>)
[2123] Train Loss:    tensor(0.8332, device='cuda:0', grad_fn=<NllLossBackward>)
[2124] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2125] Train Loss:    tensor(0.8232, device='cuda:0', grad_fn=<NllLossBackward>)
[2126] Train Loss:    tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[2127] Train Loss:    tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[2128] Train Loss:    tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[2129] Train Loss:    tensor(0.8422, device='cuda:0', grad_fn=<NllLossBackward>)
[2130] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[2131] Train Loss:    tensor(0.8538, device='cuda:0', grad_fn=<NllLossBackward>)
[2132] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[2133] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[2134] Train Loss:    tensor(0.8405, device='cuda:0', grad_fn=<NllLossBackward>)
[2135] Train Loss:    tensor(0.8499, device='cuda:0', grad_fn=<NllLossBackward>)
[2136] Train Loss:    tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[2137] Train Loss:    tensor(0.8431, device='cuda:0', grad_fn=<NllLossBackward>)
[2138] Train Loss:    tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[2139] Train Loss:    tensor(0.8243, device='cuda:0', grad_fn=<NllLossBackward>)
[2140] Train Loss:    tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[2141] Train Loss:    tensor(0.8170, device='cuda:0', grad_fn=<NllLossBackward>)
[2142] Train Loss:    tensor(0.8229, device='cuda:0', grad_fn=<NllLossBackward>)
[2143] Train Loss:    tensor(0.7975, device='cuda:0', grad_fn=<NllLossBackward>)
[2144] Train Loss:    tensor(0.8464, device='cuda:0', grad_fn=<NllLossBackward>)
[2145] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[2146] Train Loss:    tensor(0.8149, device='cuda:0', grad_fn=<NllLossBackward>)
[2147] Train Loss:    tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[2148] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[2149] Train Loss:    tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[2150] Train Loss:    tensor(0.8268, device='cuda:0', grad_fn=<NllLossBackward>)
[2151] Train Loss:    tensor(0.8689, device='cuda:0', grad_fn=<NllLossBackward>)
[2152] Train Loss:    tensor(0.8313, device='cuda:0', grad_fn=<NllLossBackward>)
[2153] Train Loss:    tensor(0.8487, device='cuda:0', grad_fn=<NllLossBackward>)
[2154] Train Loss:    tensor(0.8515, device='cuda:0', grad_fn=<NllLossBackward>)
[2155] Train Loss:    tensor(0.8129, device='cuda:0', grad_fn=<NllLossBackward>)
[2156] Train Loss:    tensor(0.8612, device='cuda:0', grad_fn=<NllLossBackward>)
[2157] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[2158] Train Loss:    tensor(0.8558, device='cuda:0', grad_fn=<NllLossBackward>)
[2159] Train Loss:    tensor(0.8508, device='cuda:0', grad_fn=<NllLossBackward>)
[2160] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[2161] Train Loss:    tensor(0.8777, device='cuda:0', grad_fn=<NllLossBackward>)
[2162] Train Loss:    tensor(0.8425, device='cuda:0', grad_fn=<NllLossBackward>)
[2163] Train Loss:    tensor(0.8086, device='cuda:0', grad_fn=<NllLossBackward>)
[2164] Train Loss:    tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[2165] Train Loss:    tensor(0.8537, device='cuda:0', grad_fn=<NllLossBackward>)
[2166] Train Loss:    tensor(0.8320, device='cuda:0', grad_fn=<NllLossBackward>)
[2167] Train Loss:    tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[2168] Train Loss:    tensor(0.8301, device='cuda:0', grad_fn=<NllLossBackward>)
[2169] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[2170] Train Loss:    tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[2171] Train Loss:    tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[2172] Train Loss:    tensor(0.8109, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2173] Train Loss:   tensor(0.8080, device='cuda:0', grad_fn=<NllLossBackward>)
[2174] Train Loss:   tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[2175] Train Loss:   tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[2176] Train Loss:   tensor(0.7982, device='cuda:0', grad_fn=<NllLossBackward>)
[2177] Train Loss:   tensor(0.8722, device='cuda:0', grad_fn=<NllLossBackward>)
[2178] Train Loss:   tensor(0.8285, device='cuda:0', grad_fn=<NllLossBackward>)
[2179] Train Loss:   tensor(0.8252, device='cuda:0', grad_fn=<NllLossBackward>)
[2180] Train Loss:   tensor(0.8330, device='cuda:0', grad_fn=<NllLossBackward>)
[2181] Train Loss:   tensor(0.8454, device='cuda:0', grad_fn=<NllLossBackward>)
[2182] Train Loss:   tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[2183] Train Loss:   tensor(0.8279, device='cuda:0', grad_fn=<NllLossBackward>)
[2184] Train Loss:   tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[2185] Train Loss:   tensor(0.8546, device='cuda:0', grad_fn=<NllLossBackward>)
[2186] Train Loss:   tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[2187] Train Loss:   tensor(0.8586, device='cuda:0', grad_fn=<NllLossBackward>)
[2188] Train Loss:   tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[2189] Train Loss:   tensor(0.8636, device='cuda:0', grad_fn=<NllLossBackward>)
[2190] Train Loss:   tensor(0.8359, device='cuda:0', grad_fn=<NllLossBackward>)
[2191] Train Loss:   tensor(0.8613, device='cuda:0', grad_fn=<NllLossBackward>)
[2192] Train Loss:   tensor(0.8294, device='cuda:0', grad_fn=<NllLossBackward>)
[2193] Train Loss:   tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[2194] Train Loss:   tensor(0.8085, device='cuda:0', grad_fn=<NllLossBackward>)
[2195] Train Loss:   tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[2196] Train Loss:   tensor(0.8701, device='cuda:0', grad_fn=<NllLossBackward>)
[2197] Train Loss:   tensor(0.8568, device='cuda:0', grad_fn=<NllLossBackward>)
[2198] Train Loss:   tensor(0.8357, device='cuda:0', grad_fn=<NllLossBackward>)
[2199] Train Loss:   tensor(0.7903, device='cuda:0', grad_fn=<NllLossBackward>)
[2200] Train Loss:   tensor(0.8309, device='cuda:0', grad_fn=<NllLossBackward>)
[2201] Train Loss:   tensor(0.8293, device='cuda:0', grad_fn=<NllLossBackward>)
[2202] Train Loss:   tensor(0.8436, device='cuda:0', grad_fn=<NllLossBackward>)
[2203] Train Loss:   tensor(0.8606, device='cuda:0', grad_fn=<NllLossBackward>)
[2204] Train Loss:   tensor(0.8478, device='cuda:0', grad_fn=<NllLossBackward>)
[2205] Train Loss:   tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[2206] Train Loss:   tensor(0.8252, device='cuda:0', grad_fn=<NllLossBackward>)
[2207] Train Loss:   tensor(0.8404, device='cuda:0', grad_fn=<NllLossBackward>)
[2208] Train Loss:   tensor(0.8195, device='cuda:0', grad_fn=<NllLossBackward>)
[2209] Train Loss:   tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[2210] Train Loss:   tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[2211] Train Loss:   tensor(0.8287, device='cuda:0', grad_fn=<NllLossBackward>)
[2212] Train Loss:   tensor(0.8294, device='cuda:0', grad_fn=<NllLossBackward>)
[2213] Train Loss:   tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[2214] Train Loss:   tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[2215] Train Loss:   tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[2216] Train Loss:   tensor(0.7948, device='cuda:0', grad_fn=<NllLossBackward>)
[2217] Train Loss:   tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[2218] Train Loss:   tensor(0.8157, device='cuda:0', grad_fn=<NllLossBackward>)
[2219] Train Loss:   tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[2220] Train Loss:   tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2221] Train Loss:   tensor(0.8453, device='cuda:0', grad_fn=<NllLossBackward>)
[2222] Train Loss:   tensor(0.8352, device='cuda:0', grad_fn=<NllLossBackward>)
[2223] Train Loss:   tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[2224] Train Loss:   tensor(0.8183, device='cuda:0', grad_fn=<NllLossBackward>)
[2225] Train Loss:   tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[2226] Train Loss:   tensor(0.8415, device='cuda:0', grad_fn=<NllLossBackward>)
[2227] Train Loss:   tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[2228] Train Loss:   tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[2229] Train Loss:   tensor(0.8393, device='cuda:0', grad_fn=<NllLossBackward>)
[2230] Train Loss:   tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[2231] Train Loss:   tensor(0.8327, device='cuda:0', grad_fn=<NllLossBackward>)
[2232] Train Loss:   tensor(0.7969, device='cuda:0', grad_fn=<NllLossBackward>)
[2233] Train Loss:   tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
[2234] Train Loss:   tensor(0.8365, device='cuda:0', grad_fn=<NllLossBackward>)
[2235] Train Loss:   tensor(0.8411, device='cuda:0', grad_fn=<NllLossBackward>)
[2236] Train Loss:   tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[2237] Train Loss:   tensor(0.8506, device='cuda:0', grad_fn=<NllLossBackward>)
[2238] Train Loss:   tensor(0.8567, device='cuda:0', grad_fn=<NllLossBackward>)
[2239] Train Loss:   tensor(0.8663, device='cuda:0', grad_fn=<NllLossBackward>)
[2240] Train Loss:   tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[2241] Train Loss:   tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[2242] Train Loss:   tensor(0.8391, device='cuda:0', grad_fn=<NllLossBackward>)
[2243] Train Loss:   tensor(0.8326, device='cuda:0', grad_fn=<NllLossBackward>)
[2244] Train Loss:   tensor(0.8434, device='cuda:0', grad_fn=<NllLossBackward>)
[2245] Train Loss:   tensor(0.8580, device='cuda:0', grad_fn=<NllLossBackward>)
[2246] Train Loss:   tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[2247] Train Loss:   tensor(0.8331, device='cuda:0', grad_fn=<NllLossBackward>)
[2248] Train Loss:   tensor(0.8506, device='cuda:0', grad_fn=<NllLossBackward>)
[2249] Train Loss:   tensor(0.8501, device='cuda:0', grad_fn=<NllLossBackward>)
[2250] Train Loss:   tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[2251] Train Loss:   tensor(0.8451, device='cuda:0', grad_fn=<NllLossBackward>)
[2252] Train Loss:   tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[2253] Train Loss:   tensor(0.8280, device='cuda:0', grad_fn=<NllLossBackward>)
[2254] Train Loss:   tensor(0.8298, device='cuda:0', grad_fn=<NllLossBackward>)
[2255] Train Loss:   tensor(0.8808, device='cuda:0', grad_fn=<NllLossBackward>)
[2256] Train Loss:   tensor(0.8471, device='cuda:0', grad_fn=<NllLossBackward>)
[2257] Train Loss:   tensor(0.8320, device='cuda:0', grad_fn=<NllLossBackward>)
[2258] Train Loss:   tensor(0.8189, device='cuda:0', grad_fn=<NllLossBackward>)
[2259] Train Loss:   tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[2260] Train Loss:   tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[2261] Train Loss:   tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[2262] Train Loss:   tensor(0.8465, device='cuda:0', grad_fn=<NllLossBackward>)
[2263] Train Loss:   tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[2264] Train Loss:   tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[2265] Train Loss:   tensor(0.8383, device='cuda:0', grad_fn=<NllLossBackward>)
[2266] Train Loss:   tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[2267] Train Loss:   tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[2268] Train Loss:   tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2269] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[2270] Train Loss:    tensor(0.8502, device='cuda:0', grad_fn=<NllLossBackward>)
[2271] Train Loss:    tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[2272] Train Loss:    tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[2273] Train Loss:    tensor(0.8087, device='cuda:0', grad_fn=<NllLossBackward>)
[2274] Train Loss:    tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[2275] Train Loss:    tensor(0.8624, device='cuda:0', grad_fn=<NllLossBackward>)
[2276] Train Loss:    tensor(0.8719, device='cuda:0', grad_fn=<NllLossBackward>)
[2277] Train Loss:    tensor(0.8391, device='cuda:0', grad_fn=<NllLossBackward>)
[2278] Train Loss:    tensor(0.8520, device='cuda:0', grad_fn=<NllLossBackward>)
[2279] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[2280] Train Loss:    tensor(0.8560, device='cuda:0', grad_fn=<NllLossBackward>)
[2281] Train Loss:    tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[2282] Train Loss:    tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[2283] Train Loss:    tensor(0.8560, device='cuda:0', grad_fn=<NllLossBackward>)
[2284] Train Loss:    tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
[2285] Train Loss:    tensor(0.8345, device='cuda:0', grad_fn=<NllLossBackward>)
[2286] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[2287] Train Loss:    tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[2288] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[2289] Train Loss:    tensor(0.8209, device='cuda:0', grad_fn=<NllLossBackward>)
[2290] Train Loss:    tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
[2291] Train Loss:    tensor(0.8416, device='cuda:0', grad_fn=<NllLossBackward>)
[2292] Train Loss:    tensor(0.8181, device='cuda:0', grad_fn=<NllLossBackward>)
[2293] Train Loss:    tensor(0.8407, device='cuda:0', grad_fn=<NllLossBackward>)
[2294] Train Loss:    tensor(0.8389, device='cuda:0', grad_fn=<NllLossBackward>)
[2295] Train Loss:    tensor(0.8100, device='cuda:0', grad_fn=<NllLossBackward>)
[2296] Train Loss:    tensor(0.8506, device='cuda:0', grad_fn=<NllLossBackward>)
[2297] Train Loss:    tensor(0.8502, device='cuda:0', grad_fn=<NllLossBackward>)
[2298] Train Loss:    tensor(0.8392, device='cuda:0', grad_fn=<NllLossBackward>)
[2299] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[2300] Train Loss:    tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[2301] Train Loss:    tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[2302] Train Loss:    tensor(0.8271, device='cuda:0', grad_fn=<NllLossBackward>)
[2303] Train Loss:    tensor(0.8359, device='cuda:0', grad_fn=<NllLossBackward>)
[2304] Train Loss:    tensor(0.8371, device='cuda:0', grad_fn=<NllLossBackward>)
[2305] Train Loss:    tensor(0.8640, device='cuda:0', grad_fn=<NllLossBackward>)
[2306] Train Loss:    tensor(0.8462, device='cuda:0', grad_fn=<NllLossBackward>)
[2307] Train Loss:    tensor(0.8721, device='cuda:0', grad_fn=<NllLossBackward>)
[2308] Train Loss:    tensor(0.8568, device='cuda:0', grad_fn=<NllLossBackward>)
[2309] Train Loss:    tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[2310] Train Loss:    tensor(0.8311, device='cuda:0', grad_fn=<NllLossBackward>)
[2311] Train Loss:    tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[2312] Train Loss:    tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
[2313] Train Loss:    tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
[2314] Train Loss:    tensor(0.8521, device='cuda:0', grad_fn=<NllLossBackward>)
[2315] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[2316] Train Loss:    tensor(0.7980, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2317] Train Loss:    tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[2318] Train Loss:    tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[2319] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[2320] Train Loss:    tensor(0.8186, device='cuda:0', grad_fn=<NllLossBackward>)
[2321] Train Loss:    tensor(0.8283, device='cuda:0', grad_fn=<NllLossBackward>)
[2322] Train Loss:    tensor(0.8485, device='cuda:0', grad_fn=<NllLossBackward>)
[2323] Train Loss:    tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[2324] Train Loss:    tensor(0.8432, device='cuda:0', grad_fn=<NllLossBackward>)
[2325] Train Loss:    tensor(0.8570, device='cuda:0', grad_fn=<NllLossBackward>)
[2326] Train Loss:    tensor(0.8402, device='cuda:0', grad_fn=<NllLossBackward>)
[2327] Train Loss:    tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[2328] Train Loss:    tensor(0.8290, device='cuda:0', grad_fn=<NllLossBackward>)
[2329] Train Loss:    tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[2330] Train Loss:    tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[2331] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[2332] Train Loss:    tensor(0.8561, device='cuda:0', grad_fn=<NllLossBackward>)
[2333] Train Loss:    tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[2334] Train Loss:    tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
[2335] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[2336] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[2337] Train Loss:    tensor(0.8249, device='cuda:0', grad_fn=<NllLossBackward>)
[2338] Train Loss:    tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[2339] Train Loss:    tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[2340] Train Loss:    tensor(0.8268, device='cuda:0', grad_fn=<NllLossBackward>)
[2341] Train Loss:    tensor(0.8487, device='cuda:0', grad_fn=<NllLossBackward>)
[2342] Train Loss:    tensor(0.8502, device='cuda:0', grad_fn=<NllLossBackward>)
[2343] Train Loss:    tensor(0.8858, device='cuda:0', grad_fn=<NllLossBackward>)
[2344] Train Loss:    tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[2345] Train Loss:    tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[2346] Train Loss:    tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
[2347] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[2348] Train Loss:    tensor(0.8174, device='cuda:0', grad_fn=<NllLossBackward>)
[2349] Train Loss:    tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[2350] Train Loss:    tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[2351] Train Loss:    tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[2352] Train Loss:    tensor(0.8149, device='cuda:0', grad_fn=<NllLossBackward>)
[2353] Train Loss:    tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[2354] Train Loss:    tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[2355] Train Loss:    tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[2356] Train Loss:    tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[2357] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[2358] Train Loss:    tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[2359] Train Loss:    tensor(0.8379, device='cuda:0', grad_fn=<NllLossBackward>)
[2360] Train Loss:    tensor(0.8253, device='cuda:0', grad_fn=<NllLossBackward>)
[2361] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[2362] Train Loss:    tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[2363] Train Loss:    tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[2364] Train Loss:    tensor(0.8059, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2365] Train Loss:   tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[2366] Train Loss:   tensor(0.8343, device='cuda:0', grad_fn=<NllLossBackward>)
[2367] Train Loss:   tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
[2368] Train Loss:   tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[2369] Train Loss:   tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[2370] Train Loss:   tensor(0.8201, device='cuda:0', grad_fn=<NllLossBackward>)
[2371] Train Loss:   tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[2372] Train Loss:   tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[2373] Train Loss:   tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[2374] Train Loss:   tensor(0.8300, device='cuda:0', grad_fn=<NllLossBackward>)
[2375] Train Loss:   tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[2376] Train Loss:   tensor(0.8381, device='cuda:0', grad_fn=<NllLossBackward>)
[2377] Train Loss:   tensor(0.8687, device='cuda:0', grad_fn=<NllLossBackward>)
[2378] Train Loss:   tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[2379] Train Loss:   tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[2380] Train Loss:   tensor(0.8383, device='cuda:0', grad_fn=<NllLossBackward>)
[2381] Train Loss:   tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[2382] Train Loss:   tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[2383] Train Loss:   tensor(0.8283, device='cuda:0', grad_fn=<NllLossBackward>)
[2384] Train Loss:   tensor(0.8253, device='cuda:0', grad_fn=<NllLossBackward>)
[2385] Train Loss:   tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[2386] Train Loss:   tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[2387] Train Loss:   tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[2388] Train Loss:   tensor(0.8030, device='cuda:0', grad_fn=<NllLossBackward>)
[2389] Train Loss:   tensor(0.8097, device='cuda:0', grad_fn=<NllLossBackward>)
[2390] Train Loss:   tensor(0.8226, device='cuda:0', grad_fn=<NllLossBackward>)
[2391] Train Loss:   tensor(0.8478, device='cuda:0', grad_fn=<NllLossBackward>)
[2392] Train Loss:   tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[2393] Train Loss:   tensor(0.8333, device='cuda:0', grad_fn=<NllLossBackward>)
[2394] Train Loss:   tensor(0.8050, device='cuda:0', grad_fn=<NllLossBackward>)
[2395] Train Loss:   tensor(0.8392, device='cuda:0', grad_fn=<NllLossBackward>)
[2396] Train Loss:   tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[2397] Train Loss:   tensor(0.8741, device='cuda:0', grad_fn=<NllLossBackward>)
[2398] Train Loss:   tensor(0.8266, device='cuda:0', grad_fn=<NllLossBackward>)
[2399] Train Loss:   tensor(0.8290, device='cuda:0', grad_fn=<NllLossBackward>)
[2400] Train Loss:   tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[2401] Train Loss:   tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[2402] Train Loss:   tensor(0.8381, device='cuda:0', grad_fn=<NllLossBackward>)
[2403] Train Loss:   tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[2404] Train Loss:   tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[2405] Train Loss:   tensor(0.8147, device='cuda:0', grad_fn=<NllLossBackward>)
[2406] Train Loss:   tensor(0.8334, device='cuda:0', grad_fn=<NllLossBackward>)
[2407] Train Loss:   tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[2408] Train Loss:   tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[2409] Train Loss:   tensor(0.8469, device='cuda:0', grad_fn=<NllLossBackward>)
[2410] Train Loss:   tensor(0.8456, device='cuda:0', grad_fn=<NllLossBackward>)
[2411] Train Loss:   tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[2412] Train Loss:   tensor(0.8449, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2413] Train Loss:   tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[2414] Train Loss:   tensor(0.8189, device='cuda:0', grad_fn=<NllLossBackward>)
[2415] Train Loss:   tensor(0.8481, device='cuda:0', grad_fn=<NllLossBackward>)
[2416] Train Loss:   tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[2417] Train Loss:   tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[2418] Train Loss:   tensor(0.8225, device='cuda:0', grad_fn=<NllLossBackward>)
[2419] Train Loss:   tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[2420] Train Loss:   tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[2421] Train Loss:   tensor(0.8462, device='cuda:0', grad_fn=<NllLossBackward>)
[2422] Train Loss:   tensor(0.8191, device='cuda:0', grad_fn=<NllLossBackward>)
[2423] Train Loss:   tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[2424] Train Loss:   tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[2425] Train Loss:   tensor(0.8704, device='cuda:0', grad_fn=<NllLossBackward>)
[2426] Train Loss:   tensor(0.8010, device='cuda:0', grad_fn=<NllLossBackward>)
[2427] Train Loss:   tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[2428] Train Loss:   tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[2429] Train Loss:   tensor(0.8024, device='cuda:0', grad_fn=<NllLossBackward>)
[2430] Train Loss:   tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[2431] Train Loss:   tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[2432] Train Loss:   tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[2433] Train Loss:   tensor(0.8331, device='cuda:0', grad_fn=<NllLossBackward>)
[2434] Train Loss:   tensor(0.8524, device='cuda:0', grad_fn=<NllLossBackward>)
[2435] Train Loss:   tensor(0.8520, device='cuda:0', grad_fn=<NllLossBackward>)
[2436] Train Loss:   tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[2437] Train Loss:   tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[2438] Train Loss:   tensor(0.8285, device='cuda:0', grad_fn=<NllLossBackward>)
[2439] Train Loss:   tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[2440] Train Loss:   tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[2441] Train Loss:   tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[2442] Train Loss:   tensor(0.8388, device='cuda:0', grad_fn=<NllLossBackward>)
[2443] Train Loss:   tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
[2444] Train Loss:   tensor(0.8565, device='cuda:0', grad_fn=<NllLossBackward>)
[2445] Train Loss:   tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[2446] Train Loss:   tensor(0.8381, device='cuda:0', grad_fn=<NllLossBackward>)
[2447] Train Loss:   tensor(0.8333, device='cuda:0', grad_fn=<NllLossBackward>)
[2448] Train Loss:   tensor(0.8439, device='cuda:0', grad_fn=<NllLossBackward>)
[2449] Train Loss:   tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[2450] Train Loss:   tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[2451] Train Loss:   tensor(0.8371, device='cuda:0', grad_fn=<NllLossBackward>)
[2452] Train Loss:   tensor(0.8691, device='cuda:0', grad_fn=<NllLossBackward>)
[2453] Train Loss:   tensor(0.8510, device='cuda:0', grad_fn=<NllLossBackward>)
[2454] Train Loss:   tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
[2455] Train Loss:   tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[2456] Train Loss:   tensor(0.8256, device='cuda:0', grad_fn=<NllLossBackward>)
[2457] Train Loss:   tensor(0.8426, device='cuda:0', grad_fn=<NllLossBackward>)
[2458] Train Loss:   tensor(0.8280, device='cuda:0', grad_fn=<NllLossBackward>)
[2459] Train Loss:   tensor(0.8687, device='cuda:0', grad_fn=<NllLossBackward>)
[2460] Train Loss:   tensor(0.8423, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2461] Train Loss:    tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[2462] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[2463] Train Loss:    tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[2464] Train Loss:    tensor(0.8591, device='cuda:0', grad_fn=<NllLossBackward>)
[2465] Train Loss:    tensor(0.8271, device='cuda:0', grad_fn=<NllLossBackward>)
[2466] Train Loss:    tensor(0.8389, device='cuda:0', grad_fn=<NllLossBackward>)
[2467] Train Loss:    tensor(0.8508, device='cuda:0', grad_fn=<NllLossBackward>)
[2468] Train Loss:    tensor(0.8323, device='cuda:0', grad_fn=<NllLossBackward>)
[2469] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[2470] Train Loss:    tensor(0.8534, device='cuda:0', grad_fn=<NllLossBackward>)
[2471] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[2472] Train Loss:    tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[2473] Train Loss:    tensor(0.8157, device='cuda:0', grad_fn=<NllLossBackward>)
[2474] Train Loss:    tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
[2475] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[2476] Train Loss:    tensor(0.8798, device='cuda:0', grad_fn=<NllLossBackward>)
[2477] Train Loss:    tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[2478] Train Loss:    tensor(0.8238, device='cuda:0', grad_fn=<NllLossBackward>)
[2479] Train Loss:    tensor(0.8409, device='cuda:0', grad_fn=<NllLossBackward>)
[2480] Train Loss:    tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[2481] Train Loss:    tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[2482] Train Loss:    tensor(0.8529, device='cuda:0', grad_fn=<NllLossBackward>)
[2483] Train Loss:    tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
[2484] Train Loss:    tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[2485] Train Loss:    tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[2486] Train Loss:    tensor(0.8350, device='cuda:0', grad_fn=<NllLossBackward>)
[2487] Train Loss:    tensor(0.8505, device='cuda:0', grad_fn=<NllLossBackward>)
[2488] Train Loss:    tensor(0.8127, device='cuda:0', grad_fn=<NllLossBackward>)
[2489] Train Loss:    tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[2490] Train Loss:    tensor(0.8279, device='cuda:0', grad_fn=<NllLossBackward>)
[2491] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[2492] Train Loss:    tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[2493] Train Loss:    tensor(0.8477, device='cuda:0', grad_fn=<NllLossBackward>)
[2494] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[2495] Train Loss:    tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[2496] Train Loss:    tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[2497] Train Loss:    tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[2498] Train Loss:    tensor(0.8243, device='cuda:0', grad_fn=<NllLossBackward>)
[2499] Train Loss:    tensor(0.8423, device='cuda:0', grad_fn=<NllLossBackward>)
[2500] Train Loss:    tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[2501] Train Loss:    tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
[2502] Train Loss:    tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[2503] Train Loss:    tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[2504] Train Loss:    tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[2505] Train Loss:    tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[2506] Train Loss:    tensor(0.8496, device='cuda:0', grad_fn=<NllLossBackward>)
[2507] Train Loss:    tensor(0.8286, device='cuda:0', grad_fn=<NllLossBackward>)
[2508] Train Loss:    tensor(0.8503, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2509] Train Loss:    tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[2510] Train Loss:    tensor(0.8469, device='cuda:0', grad_fn=<NllLossBackward>)
[2511] Train Loss:    tensor(0.8536, device='cuda:0', grad_fn=<NllLossBackward>)
[2512] Train Loss:    tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[2513] Train Loss:    tensor(0.8531, device='cuda:0', grad_fn=<NllLossBackward>)
[2514] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[2515] Train Loss:    tensor(0.8347, device='cuda:0', grad_fn=<NllLossBackward>)
[2516] Train Loss:    tensor(0.8351, device='cuda:0', grad_fn=<NllLossBackward>)
[2517] Train Loss:    tensor(0.8371, device='cuda:0', grad_fn=<NllLossBackward>)
[2518] Train Loss:    tensor(0.8249, device='cuda:0', grad_fn=<NllLossBackward>)
[2519] Train Loss:    tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[2520] Train Loss:    tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[2521] Train Loss:    tensor(0.8503, device='cuda:0', grad_fn=<NllLossBackward>)
[2522] Train Loss:    tensor(0.8401, device='cuda:0', grad_fn=<NllLossBackward>)
[2523] Train Loss:    tensor(0.7909, device='cuda:0', grad_fn=<NllLossBackward>)
[2524] Train Loss:    tensor(0.8621, device='cuda:0', grad_fn=<NllLossBackward>)
[2525] Train Loss:    tensor(0.8153, device='cuda:0', grad_fn=<NllLossBackward>)
[2526] Train Loss:    tensor(0.8145, device='cuda:0', grad_fn=<NllLossBackward>)
[2527] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[2528] Train Loss:    tensor(0.8334, device='cuda:0', grad_fn=<NllLossBackward>)
[2529] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[2530] Train Loss:    tensor(0.8356, device='cuda:0', grad_fn=<NllLossBackward>)
[2531] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[2532] Train Loss:    tensor(0.8440, device='cuda:0', grad_fn=<NllLossBackward>)
[2533] Train Loss:    tensor(0.8583, device='cuda:0', grad_fn=<NllLossBackward>)
[2534] Train Loss:    tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[2535] Train Loss:    tensor(0.8263, device='cuda:0', grad_fn=<NllLossBackward>)
[2536] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[2537] Train Loss:    tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[2538] Train Loss:    tensor(0.8532, device='cuda:0', grad_fn=<NllLossBackward>)
[2539] Train Loss:    tensor(0.8631, device='cuda:0', grad_fn=<NllLossBackward>)
[2540] Train Loss:    tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[2541] Train Loss:    tensor(0.8263, device='cuda:0', grad_fn=<NllLossBackward>)
[2542] Train Loss:    tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[2543] Train Loss:    tensor(0.8229, device='cuda:0', grad_fn=<NllLossBackward>)
[2544] Train Loss:    tensor(0.8432, device='cuda:0', grad_fn=<NllLossBackward>)
[2545] Train Loss:    tensor(0.8524, device='cuda:0', grad_fn=<NllLossBackward>)
[2546] Train Loss:    tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[2547] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[2548] Train Loss:    tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[2549] Train Loss:    tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[2550] Train Loss:    tensor(0.8129, device='cuda:0', grad_fn=<NllLossBackward>)
[2551] Train Loss:    tensor(0.8050, device='cuda:0', grad_fn=<NllLossBackward>)
[2552] Train Loss:    tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
[2553] Train Loss:    tensor(0.8371, device='cuda:0', grad_fn=<NllLossBackward>)
[2554] Train Loss:    tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[2555] Train Loss:    tensor(0.8201, device='cuda:0', grad_fn=<NllLossBackward>)
[2556] Train Loss:    tensor(0.8085, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2557] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[2558] Train Loss:    tensor(0.8555, device='cuda:0', grad_fn=<NllLossBackward>)
[2559] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[2560] Train Loss:    tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[2561] Train Loss:    tensor(0.8294, device='cuda:0', grad_fn=<NllLossBackward>)
[2562] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[2563] Train Loss:    tensor(0.8166, device='cuda:0', grad_fn=<NllLossBackward>)
[2564] Train Loss:    tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[2565] Train Loss:    tensor(0.8378, device='cuda:0', grad_fn=<NllLossBackward>)
[2566] Train Loss:    tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[2567] Train Loss:    tensor(0.8359, device='cuda:0', grad_fn=<NllLossBackward>)
[2568] Train Loss:    tensor(0.8442, device='cuda:0', grad_fn=<NllLossBackward>)
[2569] Train Loss:    tensor(0.8436, device='cuda:0', grad_fn=<NllLossBackward>)
[2570] Train Loss:    tensor(0.8566, device='cuda:0', grad_fn=<NllLossBackward>)
[2571] Train Loss:    tensor(0.8438, device='cuda:0', grad_fn=<NllLossBackward>)
[2572] Train Loss:    tensor(0.8249, device='cuda:0', grad_fn=<NllLossBackward>)
[2573] Train Loss:    tensor(0.8603, device='cuda:0', grad_fn=<NllLossBackward>)
[2574] Train Loss:    tensor(0.8412, device='cuda:0', grad_fn=<NllLossBackward>)
[2575] Train Loss:    tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[2576] Train Loss:    tensor(0.8507, device='cuda:0', grad_fn=<NllLossBackward>)
[2577] Train Loss:    tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[2578] Train Loss:    tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[2579] Train Loss:    tensor(0.8314, device='cuda:0', grad_fn=<NllLossBackward>)
[2580] Train Loss:    tensor(0.8451, device='cuda:0', grad_fn=<NllLossBackward>)
[2581] Train Loss:    tensor(0.7930, device='cuda:0', grad_fn=<NllLossBackward>)
[2582] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[2583] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[2584] Train Loss:    tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[2585] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[2586] Train Loss:    tensor(0.8471, device='cuda:0', grad_fn=<NllLossBackward>)
[2587] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[2588] Train Loss:    tensor(0.8421, device='cuda:0', grad_fn=<NllLossBackward>)
[2589] Train Loss:    tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[2590] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[2591] Train Loss:    tensor(0.8819, device='cuda:0', grad_fn=<NllLossBackward>)
[2592] Train Loss:    tensor(0.8339, device='cuda:0', grad_fn=<NllLossBackward>)
[2593] Train Loss:    tensor(0.8065, device='cuda:0', grad_fn=<NllLossBackward>)
[2594] Train Loss:    tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
[2595] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[2596] Train Loss:    tensor(0.8481, device='cuda:0', grad_fn=<NllLossBackward>)
[2597] Train Loss:    tensor(0.8508, device='cuda:0', grad_fn=<NllLossBackward>)
[2598] Train Loss:    tensor(0.8358, device='cuda:0', grad_fn=<NllLossBackward>)
[2599] Train Loss:    tensor(0.8429, device='cuda:0', grad_fn=<NllLossBackward>)
[2600] Train Loss:    tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[2601] Train Loss:    tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[2602] Train Loss:    tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[2603] Train Loss:    tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[2604] Train Loss:    tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2605] Train Loss:   tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[2606] Train Loss:   tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
[2607] Train Loss:   tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[2608] Train Loss:   tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[2609] Train Loss:   tensor(0.8004, device='cuda:0', grad_fn=<NllLossBackward>)
[2610] Train Loss:   tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[2611] Train Loss:   tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[2612] Train Loss:   tensor(0.8432, device='cuda:0', grad_fn=<NllLossBackward>)
[2613] Train Loss:   tensor(0.8401, device='cuda:0', grad_fn=<NllLossBackward>)
[2614] Train Loss:   tensor(0.8218, device='cuda:0', grad_fn=<NllLossBackward>)
[2615] Train Loss:   tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[2616] Train Loss:   tensor(0.8401, device='cuda:0', grad_fn=<NllLossBackward>)
[2617] Train Loss:   tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[2618] Train Loss:   tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[2619] Train Loss:   tensor(0.8081, device='cuda:0', grad_fn=<NllLossBackward>)
[2620] Train Loss:   tensor(0.8615, device='cuda:0', grad_fn=<NllLossBackward>)
[2621] Train Loss:   tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[2622] Train Loss:   tensor(0.8287, device='cuda:0', grad_fn=<NllLossBackward>)
[2623] Train Loss:   tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[2624] Train Loss:   tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[2625] Train Loss:   tensor(0.8649, device='cuda:0', grad_fn=<NllLossBackward>)
[2626] Train Loss:   tensor(0.7865, device='cuda:0', grad_fn=<NllLossBackward>)
[2627] Train Loss:   tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[2628] Train Loss:   tensor(0.8356, device='cuda:0', grad_fn=<NllLossBackward>)
[2629] Train Loss:   tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[2630] Train Loss:   tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[2631] Train Loss:   tensor(0.8243, device='cuda:0', grad_fn=<NllLossBackward>)
[2632] Train Loss:   tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[2633] Train Loss:   tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[2634] Train Loss:   tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[2635] Train Loss:   tensor(0.8667, device='cuda:0', grad_fn=<NllLossBackward>)
[2636] Train Loss:   tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[2637] Train Loss:   tensor(0.8418, device='cuda:0', grad_fn=<NllLossBackward>)
[2638] Train Loss:   tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[2639] Train Loss:   tensor(0.8714, device='cuda:0', grad_fn=<NllLossBackward>)
[2640] Train Loss:   tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[2641] Train Loss:   tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[2642] Train Loss:   tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[2643] Train Loss:   tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[2644] Train Loss:   tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[2645] Train Loss:   tensor(0.8189, device='cuda:0', grad_fn=<NllLossBackward>)
[2646] Train Loss:   tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[2647] Train Loss:   tensor(0.8418, device='cuda:0', grad_fn=<NllLossBackward>)
[2648] Train Loss:   tensor(0.8256, device='cuda:0', grad_fn=<NllLossBackward>)
[2649] Train Loss:   tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[2650] Train Loss:   tensor(0.8766, device='cuda:0', grad_fn=<NllLossBackward>)
[2651] Train Loss:   tensor(0.8617, device='cuda:0', grad_fn=<NllLossBackward>)
[2652] Train Loss:   tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2653] Train Loss:    tensor(0.8341, device='cuda:0', grad_fn=<NllLossBackward>)
[2654] Train Loss:    tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[2655] Train Loss:    tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[2656] Train Loss:    tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[2657] Train Loss:    tensor(0.8633, device='cuda:0', grad_fn=<NllLossBackward>)
[2658] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[2659] Train Loss:    tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[2660] Train Loss:    tensor(0.8460, device='cuda:0', grad_fn=<NllLossBackward>)
[2661] Train Loss:    tensor(0.8244, device='cuda:0', grad_fn=<NllLossBackward>)
[2662] Train Loss:    tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[2663] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[2664] Train Loss:    tensor(0.8350, device='cuda:0', grad_fn=<NllLossBackward>)
[2665] Train Loss:    tensor(0.8101, device='cuda:0', grad_fn=<NllLossBackward>)
[2666] Train Loss:    tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[2667] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[2668] Train Loss:    tensor(0.8496, device='cuda:0', grad_fn=<NllLossBackward>)
[2669] Train Loss:    tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
[2670] Train Loss:    tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[2671] Train Loss:    tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[2672] Train Loss:    tensor(0.8353, device='cuda:0', grad_fn=<NllLossBackward>)
[2673] Train Loss:    tensor(0.8365, device='cuda:0', grad_fn=<NllLossBackward>)
[2674] Train Loss:    tensor(0.8158, device='cuda:0', grad_fn=<NllLossBackward>)
[2675] Train Loss:    tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[2676] Train Loss:    tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[2677] Train Loss:    tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[2678] Train Loss:    tensor(0.8558, device='cuda:0', grad_fn=<NllLossBackward>)
[2679] Train Loss:    tensor(0.8386, device='cuda:0', grad_fn=<NllLossBackward>)
[2680] Train Loss:    tensor(0.8481, device='cuda:0', grad_fn=<NllLossBackward>)
[2681] Train Loss:    tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[2682] Train Loss:    tensor(0.8476, device='cuda:0', grad_fn=<NllLossBackward>)
[2683] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[2684] Train Loss:    tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[2685] Train Loss:    tensor(0.8251, device='cuda:0', grad_fn=<NllLossBackward>)
[2686] Train Loss:    tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[2687] Train Loss:    tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
[2688] Train Loss:    tensor(0.8238, device='cuda:0', grad_fn=<NllLossBackward>)
[2689] Train Loss:    tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[2690] Train Loss:    tensor(0.8362, device='cuda:0', grad_fn=<NllLossBackward>)
[2691] Train Loss:    tensor(0.8642, device='cuda:0', grad_fn=<NllLossBackward>)
[2692] Train Loss:    tensor(0.8409, device='cuda:0', grad_fn=<NllLossBackward>)
[2693] Train Loss:    tensor(0.8519, device='cuda:0', grad_fn=<NllLossBackward>)
[2694] Train Loss:    tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[2695] Train Loss:    tensor(0.8485, device='cuda:0', grad_fn=<NllLossBackward>)
[2696] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[2697] Train Loss:    tensor(0.8167, device='cuda:0', grad_fn=<NllLossBackward>)
[2698] Train Loss:    tensor(0.8137, device='cuda:0', grad_fn=<NllLossBackward>)
[2699] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[2700] Train Loss:    tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2701] Train Loss:   tensor(0.8320, device='cuda:0', grad_fn=<NllLossBackward>)
[2702] Train Loss:   tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[2703] Train Loss:   tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[2704] Train Loss:   tensor(0.8499, device='cuda:0', grad_fn=<NllLossBackward>)
[2705] Train Loss:   tensor(0.8706, device='cuda:0', grad_fn=<NllLossBackward>)
[2706] Train Loss:   tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[2707] Train Loss:   tensor(0.8203, device='cuda:0', grad_fn=<NllLossBackward>)
[2708] Train Loss:   tensor(0.8161, device='cuda:0', grad_fn=<NllLossBackward>)
[2709] Train Loss:   tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[2710] Train Loss:   tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[2711] Train Loss:   tensor(0.8085, device='cuda:0', grad_fn=<NllLossBackward>)
[2712] Train Loss:   tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[2713] Train Loss:   tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[2714] Train Loss:   tensor(0.8206, device='cuda:0', grad_fn=<NllLossBackward>)
[2715] Train Loss:   tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[2716] Train Loss:   tensor(0.8332, device='cuda:0', grad_fn=<NllLossBackward>)
[2717] Train Loss:   tensor(0.7901, device='cuda:0', grad_fn=<NllLossBackward>)
[2718] Train Loss:   tensor(0.8749, device='cuda:0', grad_fn=<NllLossBackward>)
[2719] Train Loss:   tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[2720] Train Loss:   tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[2721] Train Loss:   tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[2722] Train Loss:   tensor(0.8206, device='cuda:0', grad_fn=<NllLossBackward>)
[2723] Train Loss:   tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[2724] Train Loss:   tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[2725] Train Loss:   tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[2726] Train Loss:   tensor(0.8362, device='cuda:0', grad_fn=<NllLossBackward>)
[2727] Train Loss:   tensor(0.8447, device='cuda:0', grad_fn=<NllLossBackward>)
[2728] Train Loss:   tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[2729] Train Loss:   tensor(0.8559, device='cuda:0', grad_fn=<NllLossBackward>)
[2730] Train Loss:   tensor(0.8137, device='cuda:0', grad_fn=<NllLossBackward>)
[2731] Train Loss:   tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[2732] Train Loss:   tensor(0.8167, device='cuda:0', grad_fn=<NllLossBackward>)
[2733] Train Loss:   tensor(0.8353, device='cuda:0', grad_fn=<NllLossBackward>)
[2734] Train Loss:   tensor(0.8028, device='cuda:0', grad_fn=<NllLossBackward>)
[2735] Train Loss:   tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[2736] Train Loss:   tensor(0.8407, device='cuda:0', grad_fn=<NllLossBackward>)
[2737] Train Loss:   tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[2738] Train Loss:   tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[2739] Train Loss:   tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[2740] Train Loss:   tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[2741] Train Loss:   tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[2742] Train Loss:   tensor(0.8363, device='cuda:0', grad_fn=<NllLossBackward>)
[2743] Train Loss:   tensor(0.8665, device='cuda:0', grad_fn=<NllLossBackward>)
[2744] Train Loss:   tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[2745] Train Loss:   tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[2746] Train Loss:   tensor(0.8365, device='cuda:0', grad_fn=<NllLossBackward>)
[2747] Train Loss:   tensor(0.8383, device='cuda:0', grad_fn=<NllLossBackward>)
[2748] Train Loss:   tensor(0.8519, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2749] Train Loss:    tensor(0.8578, device='cuda:0', grad_fn=<NllLossBackward>)
[2750] Train Loss:    tensor(0.8471, device='cuda:0', grad_fn=<NllLossBackward>)
[2751] Train Loss:    tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[2752] Train Loss:    tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[2753] Train Loss:    tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[2754] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[2755] Train Loss:    tensor(0.8161, device='cuda:0', grad_fn=<NllLossBackward>)
[2756] Train Loss:    tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[2757] Train Loss:    tensor(0.8263, device='cuda:0', grad_fn=<NllLossBackward>)
[2758] Train Loss:    tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[2759] Train Loss:    tensor(0.8294, device='cuda:0', grad_fn=<NllLossBackward>)
[2760] Train Loss:    tensor(0.8171, device='cuda:0', grad_fn=<NllLossBackward>)
[2761] Train Loss:    tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[2762] Train Loss:    tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[2763] Train Loss:    tensor(0.8649, device='cuda:0', grad_fn=<NllLossBackward>)
[2764] Train Loss:    tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[2765] Train Loss:    tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[2766] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[2767] Train Loss:    tensor(0.8400, device='cuda:0', grad_fn=<NllLossBackward>)
[2768] Train Loss:    tensor(0.8379, device='cuda:0', grad_fn=<NllLossBackward>)
[2769] Train Loss:    tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[2770] Train Loss:    tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[2771] Train Loss:    tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[2772] Train Loss:    tensor(0.8539, device='cuda:0', grad_fn=<NllLossBackward>)
[2773] Train Loss:    tensor(0.8530, device='cuda:0', grad_fn=<NllLossBackward>)
[2774] Train Loss:    tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[2775] Train Loss:    tensor(0.7991, device='cuda:0', grad_fn=<NllLossBackward>)
[2776] Train Loss:    tensor(0.8243, device='cuda:0', grad_fn=<NllLossBackward>)
[2777] Train Loss:    tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[2778] Train Loss:    tensor(0.8248, device='cuda:0', grad_fn=<NllLossBackward>)
[2779] Train Loss:    tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
[2780] Train Loss:    tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[2781] Train Loss:    tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[2782] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[2783] Train Loss:    tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[2784] Train Loss:    tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
[2785] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[2786] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[2787] Train Loss:    tensor(0.8212, device='cuda:0', grad_fn=<NllLossBackward>)
[2788] Train Loss:    tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[2789] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[2790] Train Loss:    tensor(0.8501, device='cuda:0', grad_fn=<NllLossBackward>)
[2791] Train Loss:    tensor(0.8574, device='cuda:0', grad_fn=<NllLossBackward>)
[2792] Train Loss:    tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[2793] Train Loss:    tensor(0.8033, device='cuda:0', grad_fn=<NllLossBackward>)
[2794] Train Loss:    tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[2795] Train Loss:    tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[2796] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2797] Train Loss:    tensor(0.8477, device='cuda:0', grad_fn=<NllLossBackward>)
[2798] Train Loss:    tensor(0.8137, device='cuda:0', grad_fn=<NllLossBackward>)
[2799] Train Loss:    tensor(0.8228, device='cuda:0', grad_fn=<NllLossBackward>)
[2800] Train Loss:    tensor(0.8076, device='cuda:0', grad_fn=<NllLossBackward>)
[2801] Train Loss:    tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[2802] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[2803] Train Loss:    tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[2804] Train Loss:    tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
[2805] Train Loss:    tensor(0.7996, device='cuda:0', grad_fn=<NllLossBackward>)
[2806] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[2807] Train Loss:    tensor(0.8260, device='cuda:0', grad_fn=<NllLossBackward>)
[2808] Train Loss:    tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[2809] Train Loss:    tensor(0.8225, device='cuda:0', grad_fn=<NllLossBackward>)
[2810] Train Loss:    tensor(0.8549, device='cuda:0', grad_fn=<NllLossBackward>)
[2811] Train Loss:    tensor(0.8374, device='cuda:0', grad_fn=<NllLossBackward>)
[2812] Train Loss:    tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[2813] Train Loss:    tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[2814] Train Loss:    tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[2815] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[2816] Train Loss:    tensor(0.8780, device='cuda:0', grad_fn=<NllLossBackward>)
[2817] Train Loss:    tensor(0.8326, device='cuda:0', grad_fn=<NllLossBackward>)
[2818] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[2819] Train Loss:    tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[2820] Train Loss:    tensor(0.8347, device='cuda:0', grad_fn=<NllLossBackward>)
[2821] Train Loss:    tensor(0.8202, device='cuda:0', grad_fn=<NllLossBackward>)
[2822] Train Loss:    tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[2823] Train Loss:    tensor(0.8549, device='cuda:0', grad_fn=<NllLossBackward>)
[2824] Train Loss:    tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[2825] Train Loss:    tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
[2826] Train Loss:    tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
[2827] Train Loss:    tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[2828] Train Loss:    tensor(0.8423, device='cuda:0', grad_fn=<NllLossBackward>)
[2829] Train Loss:    tensor(0.8474, device='cuda:0', grad_fn=<NllLossBackward>)
[2830] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[2831] Train Loss:    tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[2832] Train Loss:    tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[2833] Train Loss:    tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[2834] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[2835] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[2836] Train Loss:    tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[2837] Train Loss:    tensor(0.8330, device='cuda:0', grad_fn=<NllLossBackward>)
[2838] Train Loss:    tensor(0.8267, device='cuda:0', grad_fn=<NllLossBackward>)
[2839] Train Loss:    tensor(0.7945, device='cuda:0', grad_fn=<NllLossBackward>)
[2840] Train Loss:    tensor(0.8245, device='cuda:0', grad_fn=<NllLossBackward>)
[2841] Train Loss:    tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[2842] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[2843] Train Loss:    tensor(0.8367, device='cuda:0', grad_fn=<NllLossBackward>)
[2844] Train Loss:    tensor(0.8326, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2845] Train Loss:   tensor(0.8249, device='cuda:0', grad_fn=<NllLossBackward>)
[2846] Train Loss:   tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[2847] Train Loss:   tensor(0.8339, device='cuda:0', grad_fn=<NllLossBackward>)
[2848] Train Loss:   tensor(0.8556, device='cuda:0', grad_fn=<NllLossBackward>)
[2849] Train Loss:   tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[2850] Train Loss:   tensor(0.8128, device='cuda:0', grad_fn=<NllLossBackward>)
[2851] Train Loss:   tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[2852] Train Loss:   tensor(0.8389, device='cuda:0', grad_fn=<NllLossBackward>)
[2853] Train Loss:   tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[2854] Train Loss:   tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[2855] Train Loss:   tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[2856] Train Loss:   tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[2857] Train Loss:   tensor(0.8147, device='cuda:0', grad_fn=<NllLossBackward>)
[2858] Train Loss:   tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[2859] Train Loss:   tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[2860] Train Loss:   tensor(0.7980, device='cuda:0', grad_fn=<NllLossBackward>)
[2861] Train Loss:   tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[2862] Train Loss:   tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[2863] Train Loss:   tensor(0.8201, device='cuda:0', grad_fn=<NllLossBackward>)
[2864] Train Loss:   tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[2865] Train Loss:   tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[2866] Train Loss:   tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[2867] Train Loss:   tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[2868] Train Loss:   tensor(0.8401, device='cuda:0', grad_fn=<NllLossBackward>)
[2869] Train Loss:   tensor(0.8232, device='cuda:0', grad_fn=<NllLossBackward>)
[2870] Train Loss:   tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[2871] Train Loss:   tensor(0.8456, device='cuda:0', grad_fn=<NllLossBackward>)
[2872] Train Loss:   tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[2873] Train Loss:   tensor(0.8336, device='cuda:0', grad_fn=<NllLossBackward>)
[2874] Train Loss:   tensor(0.8285, device='cuda:0', grad_fn=<NllLossBackward>)
[2875] Train Loss:   tensor(0.8400, device='cuda:0', grad_fn=<NllLossBackward>)
[2876] Train Loss:   tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[2877] Train Loss:   tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[2878] Train Loss:   tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[2879] Train Loss:   tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[2880] Train Loss:   tensor(0.8179, device='cuda:0', grad_fn=<NllLossBackward>)
[2881] Train Loss:   tensor(0.8557, device='cuda:0', grad_fn=<NllLossBackward>)
[2882] Train Loss:   tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[2883] Train Loss:   tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[2884] Train Loss:   tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[2885] Train Loss:   tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[2886] Train Loss:   tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[2887] Train Loss:   tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[2888] Train Loss:   tensor(0.8012, device='cuda:0', grad_fn=<NllLossBackward>)
[2889] Train Loss:   tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[2890] Train Loss:   tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[2891] Train Loss:   tensor(0.8454, device='cuda:0', grad_fn=<NllLossBackward>)
[2892] Train Loss:   tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2893] Train Loss:    tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[2894] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[2895] Train Loss:    tensor(0.8473, device='cuda:0', grad_fn=<NllLossBackward>)
[2896] Train Loss:    tensor(0.8286, device='cuda:0', grad_fn=<NllLossBackward>)
[2897] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[2898] Train Loss:    tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[2899] Train Loss:    tensor(0.8465, device='cuda:0', grad_fn=<NllLossBackward>)
[2900] Train Loss:    tensor(0.8193, device='cuda:0', grad_fn=<NllLossBackward>)
[2901] Train Loss:    tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[2902] Train Loss:    tensor(0.8410, device='cuda:0', grad_fn=<NllLossBackward>)
[2903] Train Loss:    tensor(0.8086, device='cuda:0', grad_fn=<NllLossBackward>)
[2904] Train Loss:    tensor(0.7986, device='cuda:0', grad_fn=<NllLossBackward>)
[2905] Train Loss:    tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[2906] Train Loss:    tensor(0.8353, device='cuda:0', grad_fn=<NllLossBackward>)
[2907] Train Loss:    tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[2908] Train Loss:    tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[2909] Train Loss:    tensor(0.8021, device='cuda:0', grad_fn=<NllLossBackward>)
[2910] Train Loss:    tensor(0.8221, device='cuda:0', grad_fn=<NllLossBackward>)
[2911] Train Loss:    tensor(0.8159, device='cuda:0', grad_fn=<NllLossBackward>)
[2912] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[2913] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[2914] Train Loss:    tensor(0.8042, device='cuda:0', grad_fn=<NllLossBackward>)
[2915] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[2916] Train Loss:    tensor(0.8392, device='cuda:0', grad_fn=<NllLossBackward>)
[2917] Train Loss:    tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[2918] Train Loss:    tensor(0.8388, device='cuda:0', grad_fn=<NllLossBackward>)
[2919] Train Loss:    tensor(0.8446, device='cuda:0', grad_fn=<NllLossBackward>)
[2920] Train Loss:    tensor(0.8467, device='cuda:0', grad_fn=<NllLossBackward>)
[2921] Train Loss:    tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[2922] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[2923] Train Loss:    tensor(0.8606, device='cuda:0', grad_fn=<NllLossBackward>)
[2924] Train Loss:    tensor(0.8526, device='cuda:0', grad_fn=<NllLossBackward>)
[2925] Train Loss:    tensor(0.8286, device='cuda:0', grad_fn=<NllLossBackward>)
[2926] Train Loss:    tensor(0.8601, device='cuda:0', grad_fn=<NllLossBackward>)
[2927] Train Loss:    tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[2928] Train Loss:    tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[2929] Train Loss:    tensor(0.8353, device='cuda:0', grad_fn=<NllLossBackward>)
[2930] Train Loss:    tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[2931] Train Loss:    tensor(0.8165, device='cuda:0', grad_fn=<NllLossBackward>)
[2932] Train Loss:    tensor(0.8267, device='cuda:0', grad_fn=<NllLossBackward>)
[2933] Train Loss:    tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[2934] Train Loss:    tensor(0.8279, device='cuda:0', grad_fn=<NllLossBackward>)
[2935] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[2936] Train Loss:    tensor(0.8426, device='cuda:0', grad_fn=<NllLossBackward>)
[2937] Train Loss:    tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[2938] Train Loss:    tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[2939] Train Loss:    tensor(0.8077, device='cuda:0', grad_fn=<NllLossBackward>)
[2940] Train Loss:    tensor(0.8253, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2941] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[2942] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[2943] Train Loss:    tensor(0.8097, device='cuda:0', grad_fn=<NllLossBackward>)
[2944] Train Loss:    tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[2945] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[2946] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[2947] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[2948] Train Loss:    tensor(0.8547, device='cuda:0', grad_fn=<NllLossBackward>)
[2949] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[2950] Train Loss:    tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[2951] Train Loss:    tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[2952] Train Loss:    tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[2953] Train Loss:    tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[2954] Train Loss:    tensor(0.8202, device='cuda:0', grad_fn=<NllLossBackward>)
[2955] Train Loss:    tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[2956] Train Loss:    tensor(0.8235, device='cuda:0', grad_fn=<NllLossBackward>)
[2957] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[2958] Train Loss:    tensor(0.8303, device='cuda:0', grad_fn=<NllLossBackward>)
[2959] Train Loss:    tensor(0.8081, device='cuda:0', grad_fn=<NllLossBackward>)
[2960] Train Loss:    tensor(0.8055, device='cuda:0', grad_fn=<NllLossBackward>)
[2961] Train Loss:    tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[2962] Train Loss:    tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[2963] Train Loss:    tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[2964] Train Loss:    tensor(0.8333, device='cuda:0', grad_fn=<NllLossBackward>)
[2965] Train Loss:    tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[2966] Train Loss:    tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[2967] Train Loss:    tensor(0.8585, device='cuda:0', grad_fn=<NllLossBackward>)
[2968] Train Loss:    tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[2969] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[2970] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[2971] Train Loss:    tensor(0.8251, device='cuda:0', grad_fn=<NllLossBackward>)
[2972] Train Loss:    tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[2973] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[2974] Train Loss:    tensor(0.8202, device='cuda:0', grad_fn=<NllLossBackward>)
[2975] Train Loss:    tensor(0.8314, device='cuda:0', grad_fn=<NllLossBackward>)
[2976] Train Loss:    tensor(0.8228, device='cuda:0', grad_fn=<NllLossBackward>)
[2977] Train Loss:    tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[2978] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[2979] Train Loss:    tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[2980] Train Loss:    tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[2981] Train Loss:    tensor(0.8425, device='cuda:0', grad_fn=<NllLossBackward>)
[2982] Train Loss:    tensor(0.8293, device='cuda:0', grad_fn=<NllLossBackward>)
[2983] Train Loss:    tensor(0.8287, device='cuda:0', grad_fn=<NllLossBackward>)
[2984] Train Loss:    tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[2985] Train Loss:    tensor(0.8101, device='cuda:0', grad_fn=<NllLossBackward>)
[2986] Train Loss:    tensor(0.8134, device='cuda:0', grad_fn=<NllLossBackward>)
[2987] Train Loss:    tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[2988] Train Loss:    tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2989] Train Loss:    tensor(0.8469, device='cuda:0', grad_fn=<NllLossBackward>)
[2990] Train Loss:    tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
[2991] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[2992] Train Loss:    tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[2993] Train Loss:    tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[2994] Train Loss:    tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[2995] Train Loss:    tensor(0.8127, device='cuda:0', grad_fn=<NllLossBackward>)
[2996] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[2997] Train Loss:    tensor(0.8404, device='cuda:0', grad_fn=<NllLossBackward>)
[2998] Train Loss:    tensor(0.8359, device='cuda:0', grad_fn=<NllLossBackward>)
[2999] Train Loss:    tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[3000] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[3001] Train Loss:    tensor(0.8104, device='cuda:0', grad_fn=<NllLossBackward>)
[3002] Train Loss:    tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[3003] Train Loss:    tensor(0.8347, device='cuda:0', grad_fn=<NllLossBackward>)
[3004] Train Loss:    tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[3005] Train Loss:    tensor(0.8414, device='cuda:0', grad_fn=<NllLossBackward>)
[3006] Train Loss:    tensor(0.8150, device='cuda:0', grad_fn=<NllLossBackward>)
[3007] Train Loss:    tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[3008] Train Loss:    tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[3009] Train Loss:    tensor(0.7903, device='cuda:0', grad_fn=<NllLossBackward>)
[3010] Train Loss:    tensor(0.8243, device='cuda:0', grad_fn=<NllLossBackward>)
[3011] Train Loss:    tensor(0.8505, device='cuda:0', grad_fn=<NllLossBackward>)
[3012] Train Loss:    tensor(0.8311, device='cuda:0', grad_fn=<NllLossBackward>)
[3013] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[3014] Train Loss:    tensor(0.8279, device='cuda:0', grad_fn=<NllLossBackward>)
[3015] Train Loss:    tensor(0.8167, device='cuda:0', grad_fn=<NllLossBackward>)
[3016] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[3017] Train Loss:    tensor(0.8391, device='cuda:0', grad_fn=<NllLossBackward>)
[3018] Train Loss:    tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[3019] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[3020] Train Loss:    tensor(0.8493, device='cuda:0', grad_fn=<NllLossBackward>)
[3021] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[3022] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[3023] Train Loss:    tensor(0.8283, device='cuda:0', grad_fn=<NllLossBackward>)
[3024] Train Loss:    tensor(0.8104, device='cuda:0', grad_fn=<NllLossBackward>)
[3025] Train Loss:    tensor(0.8416, device='cuda:0', grad_fn=<NllLossBackward>)
[3026] Train Loss:    tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[3027] Train Loss:    tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[3028] Train Loss:    tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[3029] Train Loss:    tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[3030] Train Loss:    tensor(0.8153, device='cuda:0', grad_fn=<NllLossBackward>)
[3031] Train Loss:    tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[3032] Train Loss:    tensor(0.8346, device='cuda:0', grad_fn=<NllLossBackward>)
[3033] Train Loss:    tensor(0.8367, device='cuda:0', grad_fn=<NllLossBackward>)
[3034] Train Loss:    tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[3035] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[3036] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3037] Train Loss:    tensor(0.8218, device='cuda:0', grad_fn=<NllLossBackward>)
[3038] Train Loss:    tensor(0.8634, device='cuda:0', grad_fn=<NllLossBackward>)
[3039] Train Loss:    tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[3040] Train Loss:    tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[3041] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[3042] Train Loss:    tensor(0.8193, device='cuda:0', grad_fn=<NllLossBackward>)
[3043] Train Loss:    tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[3044] Train Loss:    tensor(0.8615, device='cuda:0', grad_fn=<NllLossBackward>)
[3045] Train Loss:    tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[3046] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[3047] Train Loss:    tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[3048] Train Loss:    tensor(0.8253, device='cuda:0', grad_fn=<NllLossBackward>)
[3049] Train Loss:    tensor(0.8422, device='cuda:0', grad_fn=<NllLossBackward>)
[3050] Train Loss:    tensor(0.8387, device='cuda:0', grad_fn=<NllLossBackward>)
[3051] Train Loss:    tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[3052] Train Loss:    tensor(0.8567, device='cuda:0', grad_fn=<NllLossBackward>)
[3053] Train Loss:    tensor(0.8249, device='cuda:0', grad_fn=<NllLossBackward>)
[3054] Train Loss:    tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[3055] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[3056] Train Loss:    tensor(0.8465, device='cuda:0', grad_fn=<NllLossBackward>)
[3057] Train Loss:    tensor(0.7865, device='cuda:0', grad_fn=<NllLossBackward>)
[3058] Train Loss:    tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[3059] Train Loss:    tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[3060] Train Loss:    tensor(0.8121, device='cuda:0', grad_fn=<NllLossBackward>)
[3061] Train Loss:    tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[3062] Train Loss:    tensor(0.8453, device='cuda:0', grad_fn=<NllLossBackward>)
[3063] Train Loss:    tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[3064] Train Loss:    tensor(0.8267, device='cuda:0', grad_fn=<NllLossBackward>)
[3065] Train Loss:    tensor(0.8481, device='cuda:0', grad_fn=<NllLossBackward>)
[3066] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[3067] Train Loss:    tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[3068] Train Loss:    tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[3069] Train Loss:    tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[3070] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[3071] Train Loss:    tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[3072] Train Loss:    tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[3073] Train Loss:    tensor(0.8290, device='cuda:0', grad_fn=<NllLossBackward>)
[3074] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[3075] Train Loss:    tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[3076] Train Loss:    tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[3077] Train Loss:    tensor(0.8527, device='cuda:0', grad_fn=<NllLossBackward>)
[3078] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[3079] Train Loss:    tensor(0.8358, device='cuda:0', grad_fn=<NllLossBackward>)
[3080] Train Loss:    tensor(0.8286, device='cuda:0', grad_fn=<NllLossBackward>)
[3081] Train Loss:    tensor(0.8363, device='cuda:0', grad_fn=<NllLossBackward>)
[3082] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[3083] Train Loss:    tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[3084] Train Loss:    tensor(0.8595, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3085] Train Loss:    tensor(0.8931, device='cuda:0', grad_fn=<NllLossBackward>)
[3086] Train Loss:    tensor(0.8179, device='cuda:0', grad_fn=<NllLossBackward>)
[3087] Train Loss:    tensor(0.8333, device='cuda:0', grad_fn=<NllLossBackward>)
[3088] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[3089] Train Loss:    tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[3090] Train Loss:    tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[3091] Train Loss:    tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[3092] Train Loss:    tensor(0.8356, device='cuda:0', grad_fn=<NllLossBackward>)
[3093] Train Loss:    tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[3094] Train Loss:    tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[3095] Train Loss:    tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[3096] Train Loss:    tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[3097] Train Loss:    tensor(0.7856, device='cuda:0', grad_fn=<NllLossBackward>)
[3098] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[3099] Train Loss:    tensor(0.8028, device='cuda:0', grad_fn=<NllLossBackward>)
[3100] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[3101] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[3102] Train Loss:    tensor(0.8160, device='cuda:0', grad_fn=<NllLossBackward>)
[3103] Train Loss:    tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[3104] Train Loss:    tensor(0.8287, device='cuda:0', grad_fn=<NllLossBackward>)
[3105] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[3106] Train Loss:    tensor(0.8720, device='cuda:0', grad_fn=<NllLossBackward>)
[3107] Train Loss:    tensor(0.8312, device='cuda:0', grad_fn=<NllLossBackward>)
[3108] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[3109] Train Loss:    tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[3110] Train Loss:    tensor(0.8301, device='cuda:0', grad_fn=<NllLossBackward>)
[3111] Train Loss:    tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[3112] Train Loss:    tensor(0.8109, device='cuda:0', grad_fn=<NllLossBackward>)
[3113] Train Loss:    tensor(0.8312, device='cuda:0', grad_fn=<NllLossBackward>)
[3114] Train Loss:    tensor(0.8209, device='cuda:0', grad_fn=<NllLossBackward>)
[3115] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[3116] Train Loss:    tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[3117] Train Loss:    tensor(0.8162, device='cuda:0', grad_fn=<NllLossBackward>)
[3118] Train Loss:    tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[3119] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[3120] Train Loss:    tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[3121] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[3122] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[3123] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[3124] Train Loss:    tensor(0.8056, device='cuda:0', grad_fn=<NllLossBackward>)
[3125] Train Loss:    tensor(0.8314, device='cuda:0', grad_fn=<NllLossBackward>)
[3126] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[3127] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[3128] Train Loss:    tensor(0.8248, device='cuda:0', grad_fn=<NllLossBackward>)
[3129] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[3130] Train Loss:    tensor(0.8087, device='cuda:0', grad_fn=<NllLossBackward>)
[3131] Train Loss:    tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[3132] Train Loss:    tensor(0.7986, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3133] Train Loss:    tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[3134] Train Loss:    tensor(0.8140, device='cuda:0', grad_fn=<NllLossBackward>)
[3135] Train Loss:    tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[3136] Train Loss:    tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[3137] Train Loss:    tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[3138] Train Loss:    tensor(0.8534, device='cuda:0', grad_fn=<NllLossBackward>)
[3139] Train Loss:    tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[3140] Train Loss:    tensor(0.8249, device='cuda:0', grad_fn=<NllLossBackward>)
[3141] Train Loss:    tensor(0.8388, device='cuda:0', grad_fn=<NllLossBackward>)
[3142] Train Loss:    tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[3143] Train Loss:    tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[3144] Train Loss:    tensor(0.8174, device='cuda:0', grad_fn=<NllLossBackward>)
[3145] Train Loss:    tensor(0.8301, device='cuda:0', grad_fn=<NllLossBackward>)
[3146] Train Loss:    tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[3147] Train Loss:    tensor(0.8021, device='cuda:0', grad_fn=<NllLossBackward>)
[3148] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[3149] Train Loss:    tensor(0.8302, device='cuda:0', grad_fn=<NllLossBackward>)
[3150] Train Loss:    tensor(0.8036, device='cuda:0', grad_fn=<NllLossBackward>)
[3151] Train Loss:    tensor(0.8304, device='cuda:0', grad_fn=<NllLossBackward>)
[3152] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[3153] Train Loss:    tensor(0.8153, device='cuda:0', grad_fn=<NllLossBackward>)
[3154] Train Loss:    tensor(0.8459, device='cuda:0', grad_fn=<NllLossBackward>)
[3155] Train Loss:    tensor(0.8422, device='cuda:0', grad_fn=<NllLossBackward>)
[3156] Train Loss:    tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[3157] Train Loss:    tensor(0.8364, device='cuda:0', grad_fn=<NllLossBackward>)
[3158] Train Loss:    tensor(0.8562, device='cuda:0', grad_fn=<NllLossBackward>)
[3159] Train Loss:    tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[3160] Train Loss:    tensor(0.8152, device='cuda:0', grad_fn=<NllLossBackward>)
[3161] Train Loss:    tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[3162] Train Loss:    tensor(0.8012, device='cuda:0', grad_fn=<NllLossBackward>)
[3163] Train Loss:    tensor(0.8009, device='cuda:0', grad_fn=<NllLossBackward>)
[3164] Train Loss:    tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[3165] Train Loss:    tensor(0.8586, device='cuda:0', grad_fn=<NllLossBackward>)
[3166] Train Loss:    tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
[3167] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[3168] Train Loss:    tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[3169] Train Loss:    tensor(0.8303, device='cuda:0', grad_fn=<NllLossBackward>)
[3170] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[3171] Train Loss:    tensor(0.8059, device='cuda:0', grad_fn=<NllLossBackward>)
[3172] Train Loss:    tensor(0.8189, device='cuda:0', grad_fn=<NllLossBackward>)
[3173] Train Loss:    tensor(0.8483, device='cuda:0', grad_fn=<NllLossBackward>)
[3174] Train Loss:    tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[3175] Train Loss:    tensor(0.8291, device='cuda:0', grad_fn=<NllLossBackward>)
[3176] Train Loss:    tensor(0.8156, device='cuda:0', grad_fn=<NllLossBackward>)
[3177] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[3178] Train Loss:    tensor(0.8426, device='cuda:0', grad_fn=<NllLossBackward>)
[3179] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[3180] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3181] Train Loss:    tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[3182] Train Loss:    tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[3183] Train Loss:    tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[3184] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[3185] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[3186] Train Loss:    tensor(0.8209, device='cuda:0', grad_fn=<NllLossBackward>)
[3187] Train Loss:    tensor(0.8121, device='cuda:0', grad_fn=<NllLossBackward>)
[3188] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[3189] Train Loss:    tensor(0.8029, device='cuda:0', grad_fn=<NllLossBackward>)
[3190] Train Loss:    tensor(0.8143, device='cuda:0', grad_fn=<NllLossBackward>)
[3191] Train Loss:    tensor(0.8508, device='cuda:0', grad_fn=<NllLossBackward>)
[3192] Train Loss:    tensor(0.8238, device='cuda:0', grad_fn=<NllLossBackward>)
[3193] Train Loss:    tensor(0.8327, device='cuda:0', grad_fn=<NllLossBackward>)
[3194] Train Loss:    tensor(0.8332, device='cuda:0', grad_fn=<NllLossBackward>)
[3195] Train Loss:    tensor(0.8532, device='cuda:0', grad_fn=<NllLossBackward>)
[3196] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[3197] Train Loss:    tensor(0.8410, device='cuda:0', grad_fn=<NllLossBackward>)
[3198] Train Loss:    tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[3199] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[3200] Train Loss:    tensor(0.8140, device='cuda:0', grad_fn=<NllLossBackward>)
[3201] Train Loss:    tensor(0.8378, device='cuda:0', grad_fn=<NllLossBackward>)
[3202] Train Loss:    tensor(0.8288, device='cuda:0', grad_fn=<NllLossBackward>)
[3203] Train Loss:    tensor(0.8076, device='cuda:0', grad_fn=<NllLossBackward>)
[3204] Train Loss:    tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[3205] Train Loss:    tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[3206] Train Loss:    tensor(0.8346, device='cuda:0', grad_fn=<NllLossBackward>)
[3207] Train Loss:    tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[3208] Train Loss:    tensor(0.8252, device='cuda:0', grad_fn=<NllLossBackward>)
[3209] Train Loss:    tensor(0.8056, device='cuda:0', grad_fn=<NllLossBackward>)
[3210] Train Loss:    tensor(0.8392, device='cuda:0', grad_fn=<NllLossBackward>)
[3211] Train Loss:    tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[3212] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[3213] Train Loss:    tensor(0.8424, device='cuda:0', grad_fn=<NllLossBackward>)
[3214] Train Loss:    tensor(0.8266, device='cuda:0', grad_fn=<NllLossBackward>)
[3215] Train Loss:    tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[3216] Train Loss:    tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[3217] Train Loss:    tensor(0.8167, device='cuda:0', grad_fn=<NllLossBackward>)
[3218] Train Loss:    tensor(0.8525, device='cuda:0', grad_fn=<NllLossBackward>)
[3219] Train Loss:    tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[3220] Train Loss:    tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[3221] Train Loss:    tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[3222] Train Loss:    tensor(0.8631, device='cuda:0', grad_fn=<NllLossBackward>)
[3223] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[3224] Train Loss:    tensor(0.8496, device='cuda:0', grad_fn=<NllLossBackward>)
[3225] Train Loss:    tensor(0.8189, device='cuda:0', grad_fn=<NllLossBackward>)
[3226] Train Loss:    tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[3227] Train Loss:    tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[3228] Train Loss:    tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3229] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[3230] Train Loss:    tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[3231] Train Loss:    tensor(0.8390, device='cuda:0', grad_fn=<NllLossBackward>)
[3232] Train Loss:    tensor(0.8450, device='cuda:0', grad_fn=<NllLossBackward>)
[3233] Train Loss:    tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[3234] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[3235] Train Loss:    tensor(0.8212, device='cuda:0', grad_fn=<NllLossBackward>)
[3236] Train Loss:    tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[3237] Train Loss:    tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[3238] Train Loss:    tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[3239] Train Loss:    tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[3240] Train Loss:    tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[3241] Train Loss:    tensor(0.8028, device='cuda:0', grad_fn=<NllLossBackward>)
[3242] Train Loss:    tensor(0.8388, device='cuda:0', grad_fn=<NllLossBackward>)
[3243] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[3244] Train Loss:    tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[3245] Train Loss:    tensor(0.8492, device='cuda:0', grad_fn=<NllLossBackward>)
[3246] Train Loss:    tensor(0.8267, device='cuda:0', grad_fn=<NllLossBackward>)
[3247] Train Loss:    tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[3248] Train Loss:    tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[3249] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[3250] Train Loss:    tensor(0.8228, device='cuda:0', grad_fn=<NllLossBackward>)
[3251] Train Loss:    tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[3252] Train Loss:    tensor(0.8357, device='cuda:0', grad_fn=<NllLossBackward>)
[3253] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[3254] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[3255] Train Loss:    tensor(0.8307, device='cuda:0', grad_fn=<NllLossBackward>)
[3256] Train Loss:    tensor(0.8725, device='cuda:0', grad_fn=<NllLossBackward>)
[3257] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[3258] Train Loss:    tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
[3259] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[3260] Train Loss:    tensor(0.8389, device='cuda:0', grad_fn=<NllLossBackward>)
[3261] Train Loss:    tensor(0.8391, device='cuda:0', grad_fn=<NllLossBackward>)
[3262] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[3263] Train Loss:    tensor(0.8314, device='cuda:0', grad_fn=<NllLossBackward>)
[3264] Train Loss:    tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[3265] Train Loss:    tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[3266] Train Loss:    tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[3267] Train Loss:    tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[3268] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[3269] Train Loss:    tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[3270] Train Loss:    tensor(0.8140, device='cuda:0', grad_fn=<NllLossBackward>)
[3271] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[3272] Train Loss:    tensor(0.8154, device='cuda:0', grad_fn=<NllLossBackward>)
[3273] Train Loss:    tensor(0.8304, device='cuda:0', grad_fn=<NllLossBackward>)
[3274] Train Loss:    tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[3275] Train Loss:    tensor(0.8368, device='cuda:0', grad_fn=<NllLossBackward>)
[3276] Train Loss:    tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3277] Train Loss:    tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[3278] Train Loss:    tensor(0.8477, device='cuda:0', grad_fn=<NllLossBackward>)
[3279] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[3280] Train Loss:    tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[3281] Train Loss:    tensor(0.8347, device='cuda:0', grad_fn=<NllLossBackward>)
[3282] Train Loss:    tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[3283] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[3284] Train Loss:    tensor(0.8056, device='cuda:0', grad_fn=<NllLossBackward>)
[3285] Train Loss:    tensor(0.8365, device='cuda:0', grad_fn=<NllLossBackward>)
[3286] Train Loss:    tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[3287] Train Loss:    tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[3288] Train Loss:    tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[3289] Train Loss:    tensor(0.8455, device='cuda:0', grad_fn=<NllLossBackward>)
[3290] Train Loss:    tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[3291] Train Loss:    tensor(0.8461, device='cuda:0', grad_fn=<NllLossBackward>)
[3292] Train Loss:    tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[3293] Train Loss:    tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[3294] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[3295] Train Loss:    tensor(0.8568, device='cuda:0', grad_fn=<NllLossBackward>)
[3296] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[3297] Train Loss:    tensor(0.8333, device='cuda:0', grad_fn=<NllLossBackward>)
[3298] Train Loss:    tensor(0.8388, device='cuda:0', grad_fn=<NllLossBackward>)
[3299] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[3300] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[3301] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[3302] Train Loss:    tensor(0.8323, device='cuda:0', grad_fn=<NllLossBackward>)
[3303] Train Loss:    tensor(0.7942, device='cuda:0', grad_fn=<NllLossBackward>)
[3304] Train Loss:    tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[3305] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[3306] Train Loss:    tensor(0.8342, device='cuda:0', grad_fn=<NllLossBackward>)
[3307] Train Loss:    tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[3308] Train Loss:    tensor(0.8744, device='cuda:0', grad_fn=<NllLossBackward>)
[3309] Train Loss:    tensor(0.8311, device='cuda:0', grad_fn=<NllLossBackward>)
[3310] Train Loss:    tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[3311] Train Loss:    tensor(0.8157, device='cuda:0', grad_fn=<NllLossBackward>)
[3312] Train Loss:    tensor(0.8109, device='cuda:0', grad_fn=<NllLossBackward>)
[3313] Train Loss:    tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[3314] Train Loss:    tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[3315] Train Loss:    tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[3316] Train Loss:    tensor(0.8383, device='cuda:0', grad_fn=<NllLossBackward>)
[3317] Train Loss:    tensor(0.8501, device='cuda:0', grad_fn=<NllLossBackward>)
[3318] Train Loss:    tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[3319] Train Loss:    tensor(0.8309, device='cuda:0', grad_fn=<NllLossBackward>)
[3320] Train Loss:    tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[3321] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[3322] Train Loss:    tensor(0.8236, device='cuda:0', grad_fn=<NllLossBackward>)
[3323] Train Loss:    tensor(0.8191, device='cuda:0', grad_fn=<NllLossBackward>)
[3324] Train Loss:    tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3325] Train Loss:    tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[3326] Train Loss:    tensor(0.8479, device='cuda:0', grad_fn=<NllLossBackward>)
[3327] Train Loss:    tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[3328] Train Loss:    tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[3329] Train Loss:    tensor(0.8015, device='cuda:0', grad_fn=<NllLossBackward>)
[3330] Train Loss:    tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[3331] Train Loss:    tensor(0.7784, device='cuda:0', grad_fn=<NllLossBackward>)
[3332] Train Loss:    tensor(0.8578, device='cuda:0', grad_fn=<NllLossBackward>)
[3333] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[3334] Train Loss:    tensor(0.8199, device='cuda:0', grad_fn=<NllLossBackward>)
[3335] Train Loss:    tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[3336] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[3337] Train Loss:    tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[3338] Train Loss:    tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[3339] Train Loss:    tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[3340] Train Loss:    tensor(0.8543, device='cuda:0', grad_fn=<NllLossBackward>)
[3341] Train Loss:    tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[3342] Train Loss:    tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[3343] Train Loss:    tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[3344] Train Loss:    tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[3345] Train Loss:    tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[3346] Train Loss:    tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[3347] Train Loss:    tensor(0.8337, device='cuda:0', grad_fn=<NllLossBackward>)
[3348] Train Loss:    tensor(0.8580, device='cuda:0', grad_fn=<NllLossBackward>)
[3349] Train Loss:    tensor(0.8332, device='cuda:0', grad_fn=<NllLossBackward>)
[3350] Train Loss:    tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[3351] Train Loss:    tensor(0.8498, device='cuda:0', grad_fn=<NllLossBackward>)
[3352] Train Loss:    tensor(0.8386, device='cuda:0', grad_fn=<NllLossBackward>)
[3353] Train Loss:    tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[3354] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[3355] Train Loss:    tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[3356] Train Loss:    tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[3357] Train Loss:    tensor(0.8378, device='cuda:0', grad_fn=<NllLossBackward>)
[3358] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[3359] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[3360] Train Loss:    tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[3361] Train Loss:    tensor(0.8819, device='cuda:0', grad_fn=<NllLossBackward>)
[3362] Train Loss:    tensor(0.8028, device='cuda:0', grad_fn=<NllLossBackward>)
[3363] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[3364] Train Loss:    tensor(0.8343, device='cuda:0', grad_fn=<NllLossBackward>)
[3365] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[3366] Train Loss:    tensor(0.8090, device='cuda:0', grad_fn=<NllLossBackward>)
[3367] Train Loss:    tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[3368] Train Loss:    tensor(0.8248, device='cuda:0', grad_fn=<NllLossBackward>)
[3369] Train Loss:    tensor(0.8523, device='cuda:0', grad_fn=<NllLossBackward>)
[3370] Train Loss:    tensor(0.8291, device='cuda:0', grad_fn=<NllLossBackward>)
[3371] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[3372] Train Loss:    tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3373] Train Loss:    tensor(0.8081, device='cuda:0', grad_fn=<NllLossBackward>)
[3374] Train Loss:    tensor(0.8364, device='cuda:0', grad_fn=<NllLossBackward>)
[3375] Train Loss:    tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[3376] Train Loss:    tensor(0.8176, device='cuda:0', grad_fn=<NllLossBackward>)
[3377] Train Loss:    tensor(0.8346, device='cuda:0', grad_fn=<NllLossBackward>)
[3378] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[3379] Train Loss:    tensor(0.8232, device='cuda:0', grad_fn=<NllLossBackward>)
[3380] Train Loss:    tensor(0.8077, device='cuda:0', grad_fn=<NllLossBackward>)
[3381] Train Loss:    tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[3382] Train Loss:    tensor(0.8225, device='cuda:0', grad_fn=<NllLossBackward>)
[3383] Train Loss:    tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[3384] Train Loss:    tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[3385] Train Loss:    tensor(0.8280, device='cuda:0', grad_fn=<NllLossBackward>)
[3386] Train Loss:    tensor(0.8534, device='cuda:0', grad_fn=<NllLossBackward>)
[3387] Train Loss:    tensor(0.8186, device='cuda:0', grad_fn=<NllLossBackward>)
[3388] Train Loss:    tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[3389] Train Loss:    tensor(0.8298, device='cuda:0', grad_fn=<NllLossBackward>)
[3390] Train Loss:    tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
[3391] Train Loss:    tensor(0.8551, device='cuda:0', grad_fn=<NllLossBackward>)
[3392] Train Loss:    tensor(0.8362, device='cuda:0', grad_fn=<NllLossBackward>)
[3393] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[3394] Train Loss:    tensor(0.8312, device='cuda:0', grad_fn=<NllLossBackward>)
[3395] Train Loss:    tensor(0.8393, device='cuda:0', grad_fn=<NllLossBackward>)
[3396] Train Loss:    tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[3397] Train Loss:    tensor(0.8029, device='cuda:0', grad_fn=<NllLossBackward>)
[3398] Train Loss:    tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[3399] Train Loss:    tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[3400] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[3401] Train Loss:    tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
[3402] Train Loss:    tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[3403] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[3404] Train Loss:    tensor(0.8159, device='cuda:0', grad_fn=<NllLossBackward>)
[3405] Train Loss:    tensor(0.8377, device='cuda:0', grad_fn=<NllLossBackward>)
[3406] Train Loss:    tensor(0.8471, device='cuda:0', grad_fn=<NllLossBackward>)
[3407] Train Loss:    tensor(0.8021, device='cuda:0', grad_fn=<NllLossBackward>)
[3408] Train Loss:    tensor(0.8456, device='cuda:0', grad_fn=<NllLossBackward>)
[3409] Train Loss:    tensor(0.8451, device='cuda:0', grad_fn=<NllLossBackward>)
[3410] Train Loss:    tensor(0.8352, device='cuda:0', grad_fn=<NllLossBackward>)
[3411] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[3412] Train Loss:    tensor(0.8343, device='cuda:0', grad_fn=<NllLossBackward>)
[3413] Train Loss:    tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[3414] Train Loss:    tensor(0.8260, device='cuda:0', grad_fn=<NllLossBackward>)
[3415] Train Loss:    tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[3416] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[3417] Train Loss:    tensor(0.8339, device='cuda:0', grad_fn=<NllLossBackward>)
[3418] Train Loss:    tensor(0.8336, device='cuda:0', grad_fn=<NllLossBackward>)
[3419] Train Loss:    tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[3420] Train Loss:    tensor(0.8407, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3421] Train Loss:    tensor(0.8691, device='cuda:0', grad_fn=<NllLossBackward>)
[3422] Train Loss:    tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[3423] Train Loss:    tensor(0.8448, device='cuda:0', grad_fn=<NllLossBackward>)
[3424] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[3425] Train Loss:    tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[3426] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[3427] Train Loss:    tensor(0.8787, device='cuda:0', grad_fn=<NllLossBackward>)
[3428] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[3429] Train Loss:    tensor(0.8227, device='cuda:0', grad_fn=<NllLossBackward>)
[3430] Train Loss:    tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[3431] Train Loss:    tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[3432] Train Loss:    tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[3433] Train Loss:    tensor(0.8118, device='cuda:0', grad_fn=<NllLossBackward>)
[3434] Train Loss:    tensor(0.7945, device='cuda:0', grad_fn=<NllLossBackward>)
[3435] Train Loss:    tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[3436] Train Loss:    tensor(0.8458, device='cuda:0', grad_fn=<NllLossBackward>)
[3437] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[3438] Train Loss:    tensor(0.8404, device='cuda:0', grad_fn=<NllLossBackward>)
[3439] Train Loss:    tensor(0.8379, device='cuda:0', grad_fn=<NllLossBackward>)
[3440] Train Loss:    tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[3441] Train Loss:    tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[3442] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[3443] Train Loss:    tensor(0.8413, device='cuda:0', grad_fn=<NllLossBackward>)
[3444] Train Loss:    tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[3445] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[3446] Train Loss:    tensor(0.8209, device='cuda:0', grad_fn=<NllLossBackward>)
[3447] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[3448] Train Loss:    tensor(0.8244, device='cuda:0', grad_fn=<NllLossBackward>)
[3449] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[3450] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[3451] Train Loss:    tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[3452] Train Loss:    tensor(0.8364, device='cuda:0', grad_fn=<NllLossBackward>)
[3453] Train Loss:    tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[3454] Train Loss:    tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[3455] Train Loss:    tensor(0.8509, device='cuda:0', grad_fn=<NllLossBackward>)
[3456] Train Loss:    tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[3457] Train Loss:    tensor(0.8153, device='cuda:0', grad_fn=<NllLossBackward>)
[3458] Train Loss:    tensor(0.8021, device='cuda:0', grad_fn=<NllLossBackward>)
[3459] Train Loss:    tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[3460] Train Loss:    tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[3461] Train Loss:    tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[3462] Train Loss:    tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[3463] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[3464] Train Loss:    tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[3465] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[3466] Train Loss:    tensor(0.8574, device='cuda:0', grad_fn=<NllLossBackward>)
[3467] Train Loss:    tensor(0.8427, device='cuda:0', grad_fn=<NllLossBackward>)
[3468] Train Loss:    tensor(0.8307, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3469] Train Loss:   tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[3470] Train Loss:   tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[3471] Train Loss:   tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[3472] Train Loss:   tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[3473] Train Loss:   tensor(0.8449, device='cuda:0', grad_fn=<NllLossBackward>)
[3474] Train Loss:   tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[3475] Train Loss:   tensor(0.8330, device='cuda:0', grad_fn=<NllLossBackward>)
[3476] Train Loss:   tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[3477] Train Loss:   tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[3478] Train Loss:   tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[3479] Train Loss:   tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[3480] Train Loss:   tensor(0.8427, device='cuda:0', grad_fn=<NllLossBackward>)
[3481] Train Loss:   tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[3482] Train Loss:   tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[3483] Train Loss:   tensor(0.8445, device='cuda:0', grad_fn=<NllLossBackward>)
[3484] Train Loss:   tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[3485] Train Loss:   tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[3486] Train Loss:   tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[3487] Train Loss:   tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[3488] Train Loss:   tensor(0.8181, device='cuda:0', grad_fn=<NllLossBackward>)
[3489] Train Loss:   tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[3490] Train Loss:   tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[3491] Train Loss:   tensor(0.8129, device='cuda:0', grad_fn=<NllLossBackward>)
[3492] Train Loss:   tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[3493] Train Loss:   tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[3494] Train Loss:   tensor(0.8156, device='cuda:0', grad_fn=<NllLossBackward>)
[3495] Train Loss:   tensor(0.8292, device='cuda:0', grad_fn=<NllLossBackward>)
[3496] Train Loss:   tensor(0.8333, device='cuda:0', grad_fn=<NllLossBackward>)
[3497] Train Loss:   tensor(0.8235, device='cuda:0', grad_fn=<NllLossBackward>)
[3498] Train Loss:   tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[3499] Train Loss:   tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[3500] Train Loss:   tensor(0.8232, device='cuda:0', grad_fn=<NllLossBackward>)
[3501] Train Loss:   tensor(0.8131, device='cuda:0', grad_fn=<NllLossBackward>)
[3502] Train Loss:   tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[3503] Train Loss:   tensor(0.8119, device='cuda:0', grad_fn=<NllLossBackward>)
[3504] Train Loss:   tensor(0.8104, device='cuda:0', grad_fn=<NllLossBackward>)
[3505] Train Loss:   tensor(0.8131, device='cuda:0', grad_fn=<NllLossBackward>)
[3506] Train Loss:   tensor(0.8161, device='cuda:0', grad_fn=<NllLossBackward>)
[3507] Train Loss:   tensor(0.8493, device='cuda:0', grad_fn=<NllLossBackward>)
[3508] Train Loss:   tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[3509] Train Loss:   tensor(0.8429, device='cuda:0', grad_fn=<NllLossBackward>)
[3510] Train Loss:   tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[3511] Train Loss:   tensor(0.8119, device='cuda:0', grad_fn=<NllLossBackward>)
[3512] Train Loss:   tensor(0.8302, device='cuda:0', grad_fn=<NllLossBackward>)
[3513] Train Loss:   tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[3514] Train Loss:   tensor(0.8639, device='cuda:0', grad_fn=<NllLossBackward>)
[3515] Train Loss:   tensor(0.8339, device='cuda:0', grad_fn=<NllLossBackward>)
[3516] Train Loss:   tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3517] Train Loss:    tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[3518] Train Loss:    tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
[3519] Train Loss:    tensor(0.8491, device='cuda:0', grad_fn=<NllLossBackward>)
[3520] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[3521] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[3522] Train Loss:    tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[3523] Train Loss:    tensor(0.7991, device='cuda:0', grad_fn=<NllLossBackward>)
[3524] Train Loss:    tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[3525] Train Loss:    tensor(0.7875, device='cuda:0', grad_fn=<NllLossBackward>)
[3526] Train Loss:    tensor(0.8412, device='cuda:0', grad_fn=<NllLossBackward>)
[3527] Train Loss:    tensor(0.8243, device='cuda:0', grad_fn=<NllLossBackward>)
[3528] Train Loss:    tensor(0.8458, device='cuda:0', grad_fn=<NllLossBackward>)
[3529] Train Loss:    tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[3530] Train Loss:    tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[3531] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[3532] Train Loss:    tensor(0.8330, device='cuda:0', grad_fn=<NllLossBackward>)
[3533] Train Loss:    tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[3534] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[3535] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[3536] Train Loss:    tensor(0.8174, device='cuda:0', grad_fn=<NllLossBackward>)
[3537] Train Loss:    tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[3538] Train Loss:    tensor(0.8092, device='cuda:0', grad_fn=<NllLossBackward>)
[3539] Train Loss:    tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[3540] Train Loss:    tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[3541] Train Loss:    tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[3542] Train Loss:    tensor(0.8129, device='cuda:0', grad_fn=<NllLossBackward>)
[3543] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[3544] Train Loss:    tensor(0.8161, device='cuda:0', grad_fn=<NllLossBackward>)
[3545] Train Loss:    tensor(0.8253, device='cuda:0', grad_fn=<NllLossBackward>)
[3546] Train Loss:    tensor(0.8550, device='cuda:0', grad_fn=<NllLossBackward>)
[3547] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[3548] Train Loss:    tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[3549] Train Loss:    tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[3550] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[3551] Train Loss:    tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[3552] Train Loss:    tensor(0.8134, device='cuda:0', grad_fn=<NllLossBackward>)
[3553] Train Loss:    tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[3554] Train Loss:    tensor(0.8121, device='cuda:0', grad_fn=<NllLossBackward>)
[3555] Train Loss:    tensor(0.8419, device='cuda:0', grad_fn=<NllLossBackward>)
[3556] Train Loss:    tensor(0.8548, device='cuda:0', grad_fn=<NllLossBackward>)
[3557] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[3558] Train Loss:    tensor(0.8359, device='cuda:0', grad_fn=<NllLossBackward>)
[3559] Train Loss:    tensor(0.8475, device='cuda:0', grad_fn=<NllLossBackward>)
[3560] Train Loss:    tensor(0.9026, device='cuda:0', grad_fn=<NllLossBackward>)
[3561] Train Loss:    tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[3562] Train Loss:    tensor(0.7930, device='cuda:0', grad_fn=<NllLossBackward>)
[3563] Train Loss:    tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[3564] Train Loss:    tensor(0.8334, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3565] Train Loss:    tensor(0.8192, device='cuda:0', grad_fn=<NllLossBackward>)
[3566] Train Loss:    tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[3567] Train Loss:    tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
[3568] Train Loss:    tensor(0.8169, device='cuda:0', grad_fn=<NllLossBackward>)
[3569] Train Loss:    tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[3570] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[3571] Train Loss:    tensor(0.8487, device='cuda:0', grad_fn=<NllLossBackward>)
[3572] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[3573] Train Loss:    tensor(0.8050, device='cuda:0', grad_fn=<NllLossBackward>)
[3574] Train Loss:    tensor(0.8557, device='cuda:0', grad_fn=<NllLossBackward>)
[3575] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[3576] Train Loss:    tensor(0.8405, device='cuda:0', grad_fn=<NllLossBackward>)
[3577] Train Loss:    tensor(0.8401, device='cuda:0', grad_fn=<NllLossBackward>)
[3578] Train Loss:    tensor(0.8304, device='cuda:0', grad_fn=<NllLossBackward>)
[3579] Train Loss:    tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[3580] Train Loss:    tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[3581] Train Loss:    tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[3582] Train Loss:    tensor(0.8232, device='cuda:0', grad_fn=<NllLossBackward>)
[3583] Train Loss:    tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[3584] Train Loss:    tensor(0.8341, device='cuda:0', grad_fn=<NllLossBackward>)
[3585] Train Loss:    tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[3586] Train Loss:    tensor(0.8280, device='cuda:0', grad_fn=<NllLossBackward>)
[3587] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[3588] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[3589] Train Loss:    tensor(0.8155, device='cuda:0', grad_fn=<NllLossBackward>)
[3590] Train Loss:    tensor(0.8524, device='cuda:0', grad_fn=<NllLossBackward>)
[3591] Train Loss:    tensor(0.8490, device='cuda:0', grad_fn=<NllLossBackward>)
[3592] Train Loss:    tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[3593] Train Loss:    tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[3594] Train Loss:    tensor(0.7982, device='cuda:0', grad_fn=<NllLossBackward>)
[3595] Train Loss:    tensor(0.8249, device='cuda:0', grad_fn=<NllLossBackward>)
[3596] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[3597] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[3598] Train Loss:    tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[3599] Train Loss:    tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[3600] Train Loss:    tensor(0.8421, device='cuda:0', grad_fn=<NllLossBackward>)
[3601] Train Loss:    tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[3602] Train Loss:    tensor(0.8137, device='cuda:0', grad_fn=<NllLossBackward>)
[3603] Train Loss:    tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[3604] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[3605] Train Loss:    tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[3606] Train Loss:    tensor(0.8310, device='cuda:0', grad_fn=<NllLossBackward>)
[3607] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[3608] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[3609] Train Loss:    tensor(0.8081, device='cuda:0', grad_fn=<NllLossBackward>)
[3610] Train Loss:    tensor(0.7996, device='cuda:0', grad_fn=<NllLossBackward>)
[3611] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[3612] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3613] Train Loss:   tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[3614] Train Loss:   tensor(0.8297, device='cuda:0', grad_fn=<NllLossBackward>)
[3615] Train Loss:   tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[3616] Train Loss:   tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[3617] Train Loss:   tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[3618] Train Loss:   tensor(0.8293, device='cuda:0', grad_fn=<NllLossBackward>)
[3619] Train Loss:   tensor(0.8036, device='cuda:0', grad_fn=<NllLossBackward>)
[3620] Train Loss:   tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[3621] Train Loss:   tensor(0.8218, device='cuda:0', grad_fn=<NllLossBackward>)
[3622] Train Loss:   tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[3623] Train Loss:   tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[3624] Train Loss:   tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[3625] Train Loss:   tensor(0.8578, device='cuda:0', grad_fn=<NllLossBackward>)
[3626] Train Loss:   tensor(0.8487, device='cuda:0', grad_fn=<NllLossBackward>)
[3627] Train Loss:   tensor(0.8437, device='cuda:0', grad_fn=<NllLossBackward>)
[3628] Train Loss:   tensor(0.8158, device='cuda:0', grad_fn=<NllLossBackward>)
[3629] Train Loss:   tensor(0.8121, device='cuda:0', grad_fn=<NllLossBackward>)
[3630] Train Loss:   tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[3631] Train Loss:   tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[3632] Train Loss:   tensor(0.8542, device='cuda:0', grad_fn=<NllLossBackward>)
[3633] Train Loss:   tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[3634] Train Loss:   tensor(0.8075, device='cuda:0', grad_fn=<NllLossBackward>)
[3635] Train Loss:   tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[3636] Train Loss:   tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[3637] Train Loss:   tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[3638] Train Loss:   tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[3639] Train Loss:   tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[3640] Train Loss:   tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[3641] Train Loss:   tensor(0.8428, device='cuda:0', grad_fn=<NllLossBackward>)
[3642] Train Loss:   tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[3643] Train Loss:   tensor(0.8400, device='cuda:0', grad_fn=<NllLossBackward>)
[3644] Train Loss:   tensor(0.8361, device='cuda:0', grad_fn=<NllLossBackward>)
[3645] Train Loss:   tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[3646] Train Loss:   tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[3647] Train Loss:   tensor(0.8552, device='cuda:0', grad_fn=<NllLossBackward>)
[3648] Train Loss:   tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[3649] Train Loss:   tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[3650] Train Loss:   tensor(0.8387, device='cuda:0', grad_fn=<NllLossBackward>)
[3651] Train Loss:   tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[3652] Train Loss:   tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[3653] Train Loss:   tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[3654] Train Loss:   tensor(0.8140, device='cuda:0', grad_fn=<NllLossBackward>)
[3655] Train Loss:   tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[3656] Train Loss:   tensor(0.8332, device='cuda:0', grad_fn=<NllLossBackward>)
[3657] Train Loss:   tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[3658] Train Loss:   tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
[3659] Train Loss:   tensor(0.8303, device='cuda:0', grad_fn=<NllLossBackward>)
[3660] Train Loss:   tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3661] Train Loss:   tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[3662] Train Loss:   tensor(0.8458, device='cuda:0', grad_fn=<NllLossBackward>)
[3663] Train Loss:   tensor(0.8227, device='cuda:0', grad_fn=<NllLossBackward>)
[3664] Train Loss:   tensor(0.8764, device='cuda:0', grad_fn=<NllLossBackward>)
[3665] Train Loss:   tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[3666] Train Loss:   tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[3667] Train Loss:   tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[3668] Train Loss:   tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[3669] Train Loss:   tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[3670] Train Loss:   tensor(0.8312, device='cuda:0', grad_fn=<NllLossBackward>)
[3671] Train Loss:   tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
[3672] Train Loss:   tensor(0.8432, device='cuda:0', grad_fn=<NllLossBackward>)
[3673] Train Loss:   tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[3674] Train Loss:   tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[3675] Train Loss:   tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[3676] Train Loss:   tensor(0.8364, device='cuda:0', grad_fn=<NllLossBackward>)
[3677] Train Loss:   tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[3678] Train Loss:   tensor(0.8443, device='cuda:0', grad_fn=<NllLossBackward>)
[3679] Train Loss:   tensor(0.8071, device='cuda:0', grad_fn=<NllLossBackward>)
[3680] Train Loss:   tensor(0.8227, device='cuda:0', grad_fn=<NllLossBackward>)
[3681] Train Loss:   tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[3682] Train Loss:   tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[3683] Train Loss:   tensor(0.8181, device='cuda:0', grad_fn=<NllLossBackward>)
[3684] Train Loss:   tensor(0.8170, device='cuda:0', grad_fn=<NllLossBackward>)
[3685] Train Loss:   tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[3686] Train Loss:   tensor(0.8390, device='cuda:0', grad_fn=<NllLossBackward>)
[3687] Train Loss:   tensor(0.8333, device='cuda:0', grad_fn=<NllLossBackward>)
[3688] Train Loss:   tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[3689] Train Loss:   tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[3690] Train Loss:   tensor(0.8401, device='cuda:0', grad_fn=<NllLossBackward>)
[3691] Train Loss:   tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[3692] Train Loss:   tensor(0.8426, device='cuda:0', grad_fn=<NllLossBackward>)
[3693] Train Loss:   tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[3694] Train Loss:   tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[3695] Train Loss:   tensor(0.8118, device='cuda:0', grad_fn=<NllLossBackward>)
[3696] Train Loss:   tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[3697] Train Loss:   tensor(0.7816, device='cuda:0', grad_fn=<NllLossBackward>)
[3698] Train Loss:   tensor(0.8399, device='cuda:0', grad_fn=<NllLossBackward>)
[3699] Train Loss:   tensor(0.8455, device='cuda:0', grad_fn=<NllLossBackward>)
[3700] Train Loss:   tensor(0.8544, device='cuda:0', grad_fn=<NllLossBackward>)
[3701] Train Loss:   tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[3702] Train Loss:   tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[3703] Train Loss:   tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[3704] Train Loss:   tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[3705] Train Loss:   tensor(0.7888, device='cuda:0', grad_fn=<NllLossBackward>)
[3706] Train Loss:   tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[3707] Train Loss:   tensor(0.8033, device='cuda:0', grad_fn=<NllLossBackward>)
[3708] Train Loss:   tensor(0.8466, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3709] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[3710] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[3711] Train Loss:    tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[3712] Train Loss:    tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[3713] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[3714] Train Loss:    tensor(0.8243, device='cuda:0', grad_fn=<NllLossBackward>)
[3715] Train Loss:    tensor(0.8080, device='cuda:0', grad_fn=<NllLossBackward>)
[3716] Train Loss:    tensor(0.7930, device='cuda:0', grad_fn=<NllLossBackward>)
[3717] Train Loss:    tensor(0.8301, device='cuda:0', grad_fn=<NllLossBackward>)
[3718] Train Loss:    tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[3719] Train Loss:    tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[3720] Train Loss:    tensor(0.8541, device='cuda:0', grad_fn=<NllLossBackward>)
[3721] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[3722] Train Loss:    tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[3723] Train Loss:    tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[3724] Train Loss:    tensor(0.8421, device='cuda:0', grad_fn=<NllLossBackward>)
[3725] Train Loss:    tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[3726] Train Loss:    tensor(0.8315, device='cuda:0', grad_fn=<NllLossBackward>)
[3727] Train Loss:    tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[3728] Train Loss:    tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[3729] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[3730] Train Loss:    tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[3731] Train Loss:    tensor(0.8499, device='cuda:0', grad_fn=<NllLossBackward>)
[3732] Train Loss:    tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[3733] Train Loss:    tensor(0.7875, device='cuda:0', grad_fn=<NllLossBackward>)
[3734] Train Loss:    tensor(0.8330, device='cuda:0', grad_fn=<NllLossBackward>)
[3735] Train Loss:    tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[3736] Train Loss:    tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[3737] Train Loss:    tensor(0.8702, device='cuda:0', grad_fn=<NllLossBackward>)
[3738] Train Loss:    tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[3739] Train Loss:    tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[3740] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[3741] Train Loss:    tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[3742] Train Loss:    tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[3743] Train Loss:    tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[3744] Train Loss:    tensor(0.8086, device='cuda:0', grad_fn=<NllLossBackward>)
[3745] Train Loss:    tensor(0.8092, device='cuda:0', grad_fn=<NllLossBackward>)
[3746] Train Loss:    tensor(0.8512, device='cuda:0', grad_fn=<NllLossBackward>)
[3747] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[3748] Train Loss:    tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[3749] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[3750] Train Loss:    tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[3751] Train Loss:    tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[3752] Train Loss:    tensor(0.8095, device='cuda:0', grad_fn=<NllLossBackward>)
[3753] Train Loss:    tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[3754] Train Loss:    tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[3755] Train Loss:    tensor(0.8147, device='cuda:0', grad_fn=<NllLossBackward>)
[3756] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3757] Train Loss:    tensor(0.8374, device='cuda:0', grad_fn=<NllLossBackward>)
[3758] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[3759] Train Loss:    tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[3760] Train Loss:    tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[3761] Train Loss:    tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[3762] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[3763] Train Loss:    tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[3764] Train Loss:    tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[3765] Train Loss:    tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[3766] Train Loss:    tensor(0.8088, device='cuda:0', grad_fn=<NllLossBackward>)
[3767] Train Loss:    tensor(0.7969, device='cuda:0', grad_fn=<NllLossBackward>)
[3768] Train Loss:    tensor(0.8549, device='cuda:0', grad_fn=<NllLossBackward>)
[3769] Train Loss:    tensor(0.8412, device='cuda:0', grad_fn=<NllLossBackward>)
[3770] Train Loss:    tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[3771] Train Loss:    tensor(0.8350, device='cuda:0', grad_fn=<NllLossBackward>)
[3772] Train Loss:    tensor(0.8300, device='cuda:0', grad_fn=<NllLossBackward>)
[3773] Train Loss:    tensor(0.8409, device='cuda:0', grad_fn=<NllLossBackward>)
[3774] Train Loss:    tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[3775] Train Loss:    tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[3776] Train Loss:    tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[3777] Train Loss:    tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[3778] Train Loss:    tensor(0.8046, device='cuda:0', grad_fn=<NllLossBackward>)
[3779] Train Loss:    tensor(0.8306, device='cuda:0', grad_fn=<NllLossBackward>)
[3780] Train Loss:    tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[3781] Train Loss:    tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[3782] Train Loss:    tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[3783] Train Loss:    tensor(0.7895, device='cuda:0', grad_fn=<NllLossBackward>)
[3784] Train Loss:    tensor(0.8317, device='cuda:0', grad_fn=<NllLossBackward>)
[3785] Train Loss:    tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[3786] Train Loss:    tensor(0.8515, device='cuda:0', grad_fn=<NllLossBackward>)
[3787] Train Loss:    tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[3788] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[3789] Train Loss:    tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[3790] Train Loss:    tensor(0.8359, device='cuda:0', grad_fn=<NllLossBackward>)
[3791] Train Loss:    tensor(0.7838, device='cuda:0', grad_fn=<NllLossBackward>)
[3792] Train Loss:    tensor(0.8287, device='cuda:0', grad_fn=<NllLossBackward>)
[3793] Train Loss:    tensor(0.7934, device='cuda:0', grad_fn=<NllLossBackward>)
[3794] Train Loss:    tensor(0.8059, device='cuda:0', grad_fn=<NllLossBackward>)
[3795] Train Loss:    tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[3796] Train Loss:    tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[3797] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[3798] Train Loss:    tensor(0.8330, device='cuda:0', grad_fn=<NllLossBackward>)
[3799] Train Loss:    tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[3800] Train Loss:    tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[3801] Train Loss:    tensor(0.8391, device='cuda:0', grad_fn=<NllLossBackward>)
[3802] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[3803] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[3804] Train Loss:    tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3805] Train Loss:   tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[3806] Train Loss:   tensor(0.8380, device='cuda:0', grad_fn=<NllLossBackward>)
[3807] Train Loss:   tensor(0.8379, device='cuda:0', grad_fn=<NllLossBackward>)
[3808] Train Loss:   tensor(0.8362, device='cuda:0', grad_fn=<NllLossBackward>)
[3809] Train Loss:   tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[3810] Train Loss:   tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[3811] Train Loss:   tensor(0.8370, device='cuda:0', grad_fn=<NllLossBackward>)
[3812] Train Loss:   tensor(0.8337, device='cuda:0', grad_fn=<NllLossBackward>)
[3813] Train Loss:   tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[3814] Train Loss:   tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[3815] Train Loss:   tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[3816] Train Loss:   tensor(0.8221, device='cuda:0', grad_fn=<NllLossBackward>)
[3817] Train Loss:   tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[3818] Train Loss:   tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[3819] Train Loss:   tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[3820] Train Loss:   tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[3821] Train Loss:   tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[3822] Train Loss:   tensor(0.8399, device='cuda:0', grad_fn=<NllLossBackward>)
[3823] Train Loss:   tensor(0.8375, device='cuda:0', grad_fn=<NllLossBackward>)
[3824] Train Loss:   tensor(0.8243, device='cuda:0', grad_fn=<NllLossBackward>)
[3825] Train Loss:   tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[3826] Train Loss:   tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[3827] Train Loss:   tensor(0.8300, device='cuda:0', grad_fn=<NllLossBackward>)
[3828] Train Loss:   tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[3829] Train Loss:   tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[3830] Train Loss:   tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[3831] Train Loss:   tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[3832] Train Loss:   tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[3833] Train Loss:   tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[3834] Train Loss:   tensor(0.8432, device='cuda:0', grad_fn=<NllLossBackward>)
[3835] Train Loss:   tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[3836] Train Loss:   tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[3837] Train Loss:   tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[3838] Train Loss:   tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[3839] Train Loss:   tensor(0.8156, device='cuda:0', grad_fn=<NllLossBackward>)
[3840] Train Loss:   tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[3841] Train Loss:   tensor(0.8187, device='cuda:0', grad_fn=<NllLossBackward>)
[3842] Train Loss:   tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[3843] Train Loss:   tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[3844] Train Loss:   tensor(0.8486, device='cuda:0', grad_fn=<NllLossBackward>)
[3845] Train Loss:   tensor(0.8189, device='cuda:0', grad_fn=<NllLossBackward>)
[3846] Train Loss:   tensor(0.8160, device='cuda:0', grad_fn=<NllLossBackward>)
[3847] Train Loss:   tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[3848] Train Loss:   tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[3849] Train Loss:   tensor(0.8617, device='cuda:0', grad_fn=<NllLossBackward>)
[3850] Train Loss:   tensor(0.8312, device='cuda:0', grad_fn=<NllLossBackward>)
[3851] Train Loss:   tensor(0.8488, device='cuda:0', grad_fn=<NllLossBackward>)
[3852] Train Loss:   tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3853] Train Loss:    tensor(0.7784, device='cuda:0', grad_fn=<NllLossBackward>)
[3854] Train Loss:    tensor(0.8477, device='cuda:0', grad_fn=<NllLossBackward>)
[3855] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[3856] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[3857] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[3858] Train Loss:    tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[3859] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[3860] Train Loss:    tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[3861] Train Loss:    tensor(0.8399, device='cuda:0', grad_fn=<NllLossBackward>)
[3862] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[3863] Train Loss:    tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[3864] Train Loss:    tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
[3865] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[3866] Train Loss:    tensor(0.8081, device='cuda:0', grad_fn=<NllLossBackward>)
[3867] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[3868] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[3869] Train Loss:    tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[3870] Train Loss:    tensor(0.8150, device='cuda:0', grad_fn=<NllLossBackward>)
[3871] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[3872] Train Loss:    tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[3873] Train Loss:    tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[3874] Train Loss:    tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[3875] Train Loss:    tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[3876] Train Loss:    tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[3877] Train Loss:    tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[3878] Train Loss:    tensor(0.8015, device='cuda:0', grad_fn=<NllLossBackward>)
[3879] Train Loss:    tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[3880] Train Loss:    tensor(0.8304, device='cuda:0', grad_fn=<NllLossBackward>)
[3881] Train Loss:    tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[3882] Train Loss:    tensor(0.8418, device='cuda:0', grad_fn=<NllLossBackward>)
[3883] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[3884] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[3885] Train Loss:    tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[3886] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[3887] Train Loss:    tensor(0.8152, device='cuda:0', grad_fn=<NllLossBackward>)
[3888] Train Loss:    tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[3889] Train Loss:    tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[3890] Train Loss:    tensor(0.8410, device='cuda:0', grad_fn=<NllLossBackward>)
[3891] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[3892] Train Loss:    tensor(0.8179, device='cuda:0', grad_fn=<NllLossBackward>)
[3893] Train Loss:    tensor(0.8487, device='cuda:0', grad_fn=<NllLossBackward>)
[3894] Train Loss:    tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[3895] Train Loss:    tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[3896] Train Loss:    tensor(0.8403, device='cuda:0', grad_fn=<NllLossBackward>)
[3897] Train Loss:    tensor(0.8163, device='cuda:0', grad_fn=<NllLossBackward>)
[3898] Train Loss:    tensor(0.8513, device='cuda:0', grad_fn=<NllLossBackward>)
[3899] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[3900] Train Loss:    tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3901] Train Loss:    tensor(0.8545, device='cuda:0', grad_fn=<NllLossBackward>)
[3902] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[3903] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[3904] Train Loss:    tensor(0.8549, device='cuda:0', grad_fn=<NllLossBackward>)
[3905] Train Loss:    tensor(0.8169, device='cuda:0', grad_fn=<NllLossBackward>)
[3906] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[3907] Train Loss:    tensor(0.8235, device='cuda:0', grad_fn=<NllLossBackward>)
[3908] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[3909] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[3910] Train Loss:    tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[3911] Train Loss:    tensor(0.8381, device='cuda:0', grad_fn=<NllLossBackward>)
[3912] Train Loss:    tensor(0.8370, device='cuda:0', grad_fn=<NllLossBackward>)
[3913] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[3914] Train Loss:    tensor(0.8154, device='cuda:0', grad_fn=<NllLossBackward>)
[3915] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[3916] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[3917] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[3918] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[3919] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[3920] Train Loss:    tensor(0.8357, device='cuda:0', grad_fn=<NllLossBackward>)
[3921] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[3922] Train Loss:    tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[3923] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[3924] Train Loss:    tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[3925] Train Loss:    tensor(0.8128, device='cuda:0', grad_fn=<NllLossBackward>)
[3926] Train Loss:    tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[3927] Train Loss:    tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[3928] Train Loss:    tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[3929] Train Loss:    tensor(0.8285, device='cuda:0', grad_fn=<NllLossBackward>)
[3930] Train Loss:    tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[3931] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[3932] Train Loss:    tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[3933] Train Loss:    tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[3934] Train Loss:    tensor(0.8166, device='cuda:0', grad_fn=<NllLossBackward>)
[3935] Train Loss:    tensor(0.8156, device='cuda:0', grad_fn=<NllLossBackward>)
[3936] Train Loss:    tensor(0.8348, device='cuda:0', grad_fn=<NllLossBackward>)
[3937] Train Loss:    tensor(0.7856, device='cuda:0', grad_fn=<NllLossBackward>)
[3938] Train Loss:    tensor(0.8203, device='cuda:0', grad_fn=<NllLossBackward>)
[3939] Train Loss:    tensor(0.8360, device='cuda:0', grad_fn=<NllLossBackward>)
[3940] Train Loss:    tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[3941] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[3942] Train Loss:    tensor(0.7991, device='cuda:0', grad_fn=<NllLossBackward>)
[3943] Train Loss:    tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[3944] Train Loss:    tensor(0.8080, device='cuda:0', grad_fn=<NllLossBackward>)
[3945] Train Loss:    tensor(0.8364, device='cuda:0', grad_fn=<NllLossBackward>)
[3946] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[3947] Train Loss:    tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[3948] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3949] Train Loss:    tensor(0.8158, device='cuda:0', grad_fn=<NllLossBackward>)
[3950] Train Loss:    tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[3951] Train Loss:    tensor(0.8104, device='cuda:0', grad_fn=<NllLossBackward>)
[3952] Train Loss:    tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
[3953] Train Loss:    tensor(0.8358, device='cuda:0', grad_fn=<NllLossBackward>)
[3954] Train Loss:    tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[3955] Train Loss:    tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[3956] Train Loss:    tensor(0.7865, device='cuda:0', grad_fn=<NllLossBackward>)
[3957] Train Loss:    tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[3958] Train Loss:    tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[3959] Train Loss:    tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
[3960] Train Loss:    tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[3961] Train Loss:    tensor(0.7895, device='cuda:0', grad_fn=<NllLossBackward>)
[3962] Train Loss:    tensor(0.8010, device='cuda:0', grad_fn=<NllLossBackward>)
[3963] Train Loss:    tensor(0.8531, device='cuda:0', grad_fn=<NllLossBackward>)
[3964] Train Loss:    tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[3965] Train Loss:    tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[3966] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[3967] Train Loss:    tensor(0.8283, device='cuda:0', grad_fn=<NllLossBackward>)
[3968] Train Loss:    tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[3969] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[3970] Train Loss:    tensor(0.8195, device='cuda:0', grad_fn=<NllLossBackward>)
[3971] Train Loss:    tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[3972] Train Loss:    tensor(0.8538, device='cuda:0', grad_fn=<NllLossBackward>)
[3973] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[3974] Train Loss:    tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[3975] Train Loss:    tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[3976] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[3977] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[3978] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[3979] Train Loss:    tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
[3980] Train Loss:    tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[3981] Train Loss:    tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[3982] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[3983] Train Loss:    tensor(0.8539, device='cuda:0', grad_fn=<NllLossBackward>)
[3984] Train Loss:    tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[3985] Train Loss:    tensor(0.7975, device='cuda:0', grad_fn=<NllLossBackward>)
[3986] Train Loss:    tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[3987] Train Loss:    tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[3988] Train Loss:    tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[3989] Train Loss:    tensor(0.8528, device='cuda:0', grad_fn=<NllLossBackward>)
[3990] Train Loss:    tensor(0.8621, device='cuda:0', grad_fn=<NllLossBackward>)
[3991] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[3992] Train Loss:    tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[3993] Train Loss:    tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[3994] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[3995] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[3996] Train Loss:    tensor(0.8592, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3997] Train Loss:   tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[3998] Train Loss:   tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[3999] Train Loss:   tensor(0.8183, device='cuda:0', grad_fn=<NllLossBackward>)
[4000] Train Loss:   tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[4001] Train Loss:   tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[4002] Train Loss:   tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[4003] Train Loss:   tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[4004] Train Loss:   tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[4005] Train Loss:   tensor(0.8461, device='cuda:0', grad_fn=<NllLossBackward>)
[4006] Train Loss:   tensor(0.8459, device='cuda:0', grad_fn=<NllLossBackward>)
[4007] Train Loss:   tensor(0.8087, device='cuda:0', grad_fn=<NllLossBackward>)
[4008] Train Loss:   tensor(0.8482, device='cuda:0', grad_fn=<NllLossBackward>)
[4009] Train Loss:   tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[4010] Train Loss:   tensor(0.8225, device='cuda:0', grad_fn=<NllLossBackward>)
[4011] Train Loss:   tensor(0.8328, device='cuda:0', grad_fn=<NllLossBackward>)
[4012] Train Loss:   tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[4013] Train Loss:   tensor(0.8170, device='cuda:0', grad_fn=<NllLossBackward>)
[4014] Train Loss:   tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[4015] Train Loss:   tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[4016] Train Loss:   tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[4017] Train Loss:   tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[4018] Train Loss:   tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[4019] Train Loss:   tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[4020] Train Loss:   tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[4021] Train Loss:   tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[4022] Train Loss:   tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[4023] Train Loss:   tensor(0.8375, device='cuda:0', grad_fn=<NllLossBackward>)
[4024] Train Loss:   tensor(0.7948, device='cuda:0', grad_fn=<NllLossBackward>)
[4025] Train Loss:   tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[4026] Train Loss:   tensor(0.8734, device='cuda:0', grad_fn=<NllLossBackward>)
[4027] Train Loss:   tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[4028] Train Loss:   tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[4029] Train Loss:   tensor(0.8160, device='cuda:0', grad_fn=<NllLossBackward>)
[4030] Train Loss:   tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[4031] Train Loss:   tensor(0.8174, device='cuda:0', grad_fn=<NllLossBackward>)
[4032] Train Loss:   tensor(0.8168, device='cuda:0', grad_fn=<NllLossBackward>)
[4033] Train Loss:   tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[4034] Train Loss:   tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[4035] Train Loss:   tensor(0.7982, device='cuda:0', grad_fn=<NllLossBackward>)
[4036] Train Loss:   tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[4037] Train Loss:   tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[4038] Train Loss:   tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[4039] Train Loss:   tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[4040] Train Loss:   tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[4041] Train Loss:   tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[4042] Train Loss:   tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[4043] Train Loss:   tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[4044] Train Loss:   tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4045] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[4046] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[4047] Train Loss:    tensor(0.8029, device='cuda:0', grad_fn=<NllLossBackward>)
[4048] Train Loss:    tensor(0.8166, device='cuda:0', grad_fn=<NllLossBackward>)
[4049] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[4050] Train Loss:    tensor(0.8203, device='cuda:0', grad_fn=<NllLossBackward>)
[4051] Train Loss:    tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[4052] Train Loss:    tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[4053] Train Loss:    tensor(0.8322, device='cuda:0', grad_fn=<NllLossBackward>)
[4054] Train Loss:    tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[4055] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[4056] Train Loss:    tensor(0.8101, device='cuda:0', grad_fn=<NllLossBackward>)
[4057] Train Loss:    tensor(0.8134, device='cuda:0', grad_fn=<NllLossBackward>)
[4058] Train Loss:    tensor(0.8174, device='cuda:0', grad_fn=<NllLossBackward>)
[4059] Train Loss:    tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[4060] Train Loss:    tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[4061] Train Loss:    tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[4062] Train Loss:    tensor(0.8195, device='cuda:0', grad_fn=<NllLossBackward>)
[4063] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[4064] Train Loss:    tensor(0.8312, device='cuda:0', grad_fn=<NllLossBackward>)
[4065] Train Loss:    tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[4066] Train Loss:    tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[4067] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[4068] Train Loss:    tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[4069] Train Loss:    tensor(0.8417, device='cuda:0', grad_fn=<NllLossBackward>)
[4070] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[4071] Train Loss:    tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[4072] Train Loss:    tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
[4073] Train Loss:    tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[4074] Train Loss:    tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[4075] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[4076] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[4077] Train Loss:    tensor(0.8160, device='cuda:0', grad_fn=<NllLossBackward>)
[4078] Train Loss:    tensor(0.8264, device='cuda:0', grad_fn=<NllLossBackward>)
[4079] Train Loss:    tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[4080] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[4081] Train Loss:    tensor(0.8306, device='cuda:0', grad_fn=<NllLossBackward>)
[4082] Train Loss:    tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[4083] Train Loss:    tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[4084] Train Loss:    tensor(0.8486, device='cuda:0', grad_fn=<NllLossBackward>)
[4085] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[4086] Train Loss:    tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[4087] Train Loss:    tensor(0.8368, device='cuda:0', grad_fn=<NllLossBackward>)
[4088] Train Loss:    tensor(0.8159, device='cuda:0', grad_fn=<NllLossBackward>)
[4089] Train Loss:    tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[4090] Train Loss:    tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[4091] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[4092] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4093] Train Loss:    tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[4094] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[4095] Train Loss:    tensor(0.8512, device='cuda:0', grad_fn=<NllLossBackward>)
[4096] Train Loss:    tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[4097] Train Loss:    tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[4098] Train Loss:    tensor(0.8389, device='cuda:0', grad_fn=<NllLossBackward>)
[4099] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[4100] Train Loss:    tensor(0.8558, device='cuda:0', grad_fn=<NllLossBackward>)
[4101] Train Loss:    tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[4102] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[4103] Train Loss:    tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[4104] Train Loss:    tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[4105] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[4106] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[4107] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[4108] Train Loss:    tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[4109] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[4110] Train Loss:    tensor(0.8320, device='cuda:0', grad_fn=<NllLossBackward>)
[4111] Train Loss:    tensor(0.8388, device='cuda:0', grad_fn=<NllLossBackward>)
[4112] Train Loss:    tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[4113] Train Loss:    tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[4114] Train Loss:    tensor(0.8010, device='cuda:0', grad_fn=<NllLossBackward>)
[4115] Train Loss:    tensor(0.8364, device='cuda:0', grad_fn=<NllLossBackward>)
[4116] Train Loss:    tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[4117] Train Loss:    tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[4118] Train Loss:    tensor(0.8071, device='cuda:0', grad_fn=<NllLossBackward>)
[4119] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[4120] Train Loss:    tensor(0.8266, device='cuda:0', grad_fn=<NllLossBackward>)
[4121] Train Loss:    tensor(0.8349, device='cuda:0', grad_fn=<NllLossBackward>)
[4122] Train Loss:    tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[4123] Train Loss:    tensor(0.8276, device='cuda:0', grad_fn=<NllLossBackward>)
[4124] Train Loss:    tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[4125] Train Loss:    tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[4126] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[4127] Train Loss:    tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[4128] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[4129] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[4130] Train Loss:    tensor(0.8447, device='cuda:0', grad_fn=<NllLossBackward>)
[4131] Train Loss:    tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[4132] Train Loss:    tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[4133] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[4134] Train Loss:    tensor(0.8071, device='cuda:0', grad_fn=<NllLossBackward>)
[4135] Train Loss:    tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[4136] Train Loss:    tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[4137] Train Loss:    tensor(0.8156, device='cuda:0', grad_fn=<NllLossBackward>)
[4138] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[4139] Train Loss:    tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
[4140] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4141] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[4142] Train Loss:    tensor(0.8030, device='cuda:0', grad_fn=<NllLossBackward>)
[4143] Train Loss:    tensor(0.8266, device='cuda:0', grad_fn=<NllLossBackward>)
[4144] Train Loss:    tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[4145] Train Loss:    tensor(0.8241, device='cuda:0', grad_fn=<NllLossBackward>)
[4146] Train Loss:    tensor(0.8497, device='cuda:0', grad_fn=<NllLossBackward>)
[4147] Train Loss:    tensor(0.8251, device='cuda:0', grad_fn=<NllLossBackward>)
[4148] Train Loss:    tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[4149] Train Loss:    tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[4150] Train Loss:    tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[4151] Train Loss:    tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[4152] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[4153] Train Loss:    tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[4154] Train Loss:    tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[4155] Train Loss:    tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[4156] Train Loss:    tensor(0.8131, device='cuda:0', grad_fn=<NllLossBackward>)
[4157] Train Loss:    tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[4158] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[4159] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[4160] Train Loss:    tensor(0.7986, device='cuda:0', grad_fn=<NllLossBackward>)
[4161] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[4162] Train Loss:    tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[4163] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[4164] Train Loss:    tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[4165] Train Loss:    tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[4166] Train Loss:    tensor(0.8139, device='cuda:0', grad_fn=<NllLossBackward>)
[4167] Train Loss:    tensor(0.8165, device='cuda:0', grad_fn=<NllLossBackward>)
[4168] Train Loss:    tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[4169] Train Loss:    tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[4170] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[4171] Train Loss:    tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[4172] Train Loss:    tensor(0.8119, device='cuda:0', grad_fn=<NllLossBackward>)
[4173] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[4174] Train Loss:    tensor(0.8441, device='cuda:0', grad_fn=<NllLossBackward>)
[4175] Train Loss:    tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[4176] Train Loss:    tensor(0.8386, device='cuda:0', grad_fn=<NllLossBackward>)
[4177] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[4178] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[4179] Train Loss:    tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[4180] Train Loss:    tensor(0.8428, device='cuda:0', grad_fn=<NllLossBackward>)
[4181] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[4182] Train Loss:    tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[4183] Train Loss:    tensor(0.8042, device='cuda:0', grad_fn=<NllLossBackward>)
[4184] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[4185] Train Loss:    tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[4186] Train Loss:    tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[4187] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[4188] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4189] Train Loss:    tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[4190] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[4191] Train Loss:    tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[4192] Train Loss:    tensor(0.8209, device='cuda:0', grad_fn=<NllLossBackward>)
[4193] Train Loss:    tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[4194] Train Loss:    tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[4195] Train Loss:    tensor(0.7731, device='cuda:0', grad_fn=<NllLossBackward>)
[4196] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[4197] Train Loss:    tensor(0.8024, device='cuda:0', grad_fn=<NllLossBackward>)
[4198] Train Loss:    tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[4199] Train Loss:    tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[4200] Train Loss:    tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[4201] Train Loss:    tensor(0.8145, device='cuda:0', grad_fn=<NllLossBackward>)
[4202] Train Loss:    tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[4203] Train Loss:    tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[4204] Train Loss:    tensor(0.8474, device='cuda:0', grad_fn=<NllLossBackward>)
[4205] Train Loss:    tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[4206] Train Loss:    tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[4207] Train Loss:    tensor(0.8437, device='cuda:0', grad_fn=<NllLossBackward>)
[4208] Train Loss:    tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[4209] Train Loss:    tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[4210] Train Loss:    tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[4211] Train Loss:    tensor(0.8564, device='cuda:0', grad_fn=<NllLossBackward>)
[4212] Train Loss:    tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[4213] Train Loss:    tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[4214] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[4215] Train Loss:    tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[4216] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[4217] Train Loss:    tensor(0.8405, device='cuda:0', grad_fn=<NllLossBackward>)
[4218] Train Loss:    tensor(0.8189, device='cuda:0', grad_fn=<NllLossBackward>)
[4219] Train Loss:    tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
[4220] Train Loss:    tensor(0.8432, device='cuda:0', grad_fn=<NllLossBackward>)
[4221] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[4222] Train Loss:    tensor(0.8276, device='cuda:0', grad_fn=<NllLossBackward>)
[4223] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[4224] Train Loss:    tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
[4225] Train Loss:    tensor(0.8495, device='cuda:0', grad_fn=<NllLossBackward>)
[4226] Train Loss:    tensor(0.8228, device='cuda:0', grad_fn=<NllLossBackward>)
[4227] Train Loss:    tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[4228] Train Loss:    tensor(0.7986, device='cuda:0', grad_fn=<NllLossBackward>)
[4229] Train Loss:    tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[4230] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[4231] Train Loss:    tensor(0.7909, device='cuda:0', grad_fn=<NllLossBackward>)
[4232] Train Loss:    tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[4233] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[4234] Train Loss:    tensor(0.8406, device='cuda:0', grad_fn=<NllLossBackward>)
[4235] Train Loss:    tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[4236] Train Loss:    tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4237] Train Loss:    tensor(0.8280, device='cuda:0', grad_fn=<NllLossBackward>)
[4238] Train Loss:    tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[4239] Train Loss:    tensor(0.8155, device='cuda:0', grad_fn=<NllLossBackward>)
[4240] Train Loss:    tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[4241] Train Loss:    tensor(0.8086, device='cuda:0', grad_fn=<NllLossBackward>)
[4242] Train Loss:    tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[4243] Train Loss:    tensor(0.8415, device='cuda:0', grad_fn=<NllLossBackward>)
[4244] Train Loss:    tensor(0.8195, device='cuda:0', grad_fn=<NllLossBackward>)
[4245] Train Loss:    tensor(0.8332, device='cuda:0', grad_fn=<NllLossBackward>)
[4246] Train Loss:    tensor(0.8398, device='cuda:0', grad_fn=<NllLossBackward>)
[4247] Train Loss:    tensor(0.8451, device='cuda:0', grad_fn=<NllLossBackward>)
[4248] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[4249] Train Loss:    tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[4250] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[4251] Train Loss:    tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[4252] Train Loss:    tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[4253] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[4254] Train Loss:    tensor(0.8360, device='cuda:0', grad_fn=<NllLossBackward>)
[4255] Train Loss:    tensor(0.8582, device='cuda:0', grad_fn=<NllLossBackward>)
[4256] Train Loss:    tensor(0.8332, device='cuda:0', grad_fn=<NllLossBackward>)
[4257] Train Loss:    tensor(0.8285, device='cuda:0', grad_fn=<NllLossBackward>)
[4258] Train Loss:    tensor(0.8311, device='cuda:0', grad_fn=<NllLossBackward>)
[4259] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[4260] Train Loss:    tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[4261] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[4262] Train Loss:    tensor(0.8145, device='cuda:0', grad_fn=<NllLossBackward>)
[4263] Train Loss:    tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[4264] Train Loss:    tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[4265] Train Loss:    tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[4266] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[4267] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[4268] Train Loss:    tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[4269] Train Loss:    tensor(0.8467, device='cuda:0', grad_fn=<NllLossBackward>)
[4270] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[4271] Train Loss:    tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[4272] Train Loss:    tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[4273] Train Loss:    tensor(0.8290, device='cuda:0', grad_fn=<NllLossBackward>)
[4274] Train Loss:    tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[4275] Train Loss:    tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[4276] Train Loss:    tensor(0.8261, device='cuda:0', grad_fn=<NllLossBackward>)
[4277] Train Loss:    tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[4278] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[4279] Train Loss:    tensor(0.8042, device='cuda:0', grad_fn=<NllLossBackward>)
[4280] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[4281] Train Loss:    tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[4282] Train Loss:    tensor(0.8253, device='cuda:0', grad_fn=<NllLossBackward>)
[4283] Train Loss:    tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[4284] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4285] Train Loss:   tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[4286] Train Loss:   tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[4287] Train Loss:   tensor(0.7901, device='cuda:0', grad_fn=<NllLossBackward>)
[4288] Train Loss:   tensor(0.8359, device='cuda:0', grad_fn=<NllLossBackward>)
[4289] Train Loss:   tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[4290] Train Loss:   tensor(0.8327, device='cuda:0', grad_fn=<NllLossBackward>)
[4291] Train Loss:   tensor(0.8012, device='cuda:0', grad_fn=<NllLossBackward>)
[4292] Train Loss:   tensor(0.8408, device='cuda:0', grad_fn=<NllLossBackward>)
[4293] Train Loss:   tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[4294] Train Loss:   tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[4295] Train Loss:   tensor(0.8389, device='cuda:0', grad_fn=<NllLossBackward>)
[4296] Train Loss:   tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[4297] Train Loss:   tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[4298] Train Loss:   tensor(0.8056, device='cuda:0', grad_fn=<NllLossBackward>)
[4299] Train Loss:   tensor(0.8192, device='cuda:0', grad_fn=<NllLossBackward>)
[4300] Train Loss:   tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[4301] Train Loss:   tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[4302] Train Loss:   tensor(0.8109, device='cuda:0', grad_fn=<NllLossBackward>)
[4303] Train Loss:   tensor(0.8020, device='cuda:0', grad_fn=<NllLossBackward>)
[4304] Train Loss:   tensor(0.8218, device='cuda:0', grad_fn=<NllLossBackward>)
[4305] Train Loss:   tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[4306] Train Loss:   tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[4307] Train Loss:   tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[4308] Train Loss:   tensor(0.8080, device='cuda:0', grad_fn=<NllLossBackward>)
[4309] Train Loss:   tensor(0.8266, device='cuda:0', grad_fn=<NllLossBackward>)
[4310] Train Loss:   tensor(0.8427, device='cuda:0', grad_fn=<NllLossBackward>)
[4311] Train Loss:   tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[4312] Train Loss:   tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[4313] Train Loss:   tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[4314] Train Loss:   tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[4315] Train Loss:   tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[4316] Train Loss:   tensor(0.8457, device='cuda:0', grad_fn=<NllLossBackward>)
[4317] Train Loss:   tensor(0.8351, device='cuda:0', grad_fn=<NllLossBackward>)
[4318] Train Loss:   tensor(0.8303, device='cuda:0', grad_fn=<NllLossBackward>)
[4319] Train Loss:   tensor(0.8119, device='cuda:0', grad_fn=<NllLossBackward>)
[4320] Train Loss:   tensor(0.8312, device='cuda:0', grad_fn=<NllLossBackward>)
[4321] Train Loss:   tensor(0.8150, device='cuda:0', grad_fn=<NllLossBackward>)
[4322] Train Loss:   tensor(0.8406, device='cuda:0', grad_fn=<NllLossBackward>)
[4323] Train Loss:   tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[4324] Train Loss:   tensor(0.8372, device='cuda:0', grad_fn=<NllLossBackward>)
[4325] Train Loss:   tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[4326] Train Loss:   tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[4327] Train Loss:   tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
[4328] Train Loss:   tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[4329] Train Loss:   tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[4330] Train Loss:   tensor(0.7784, device='cuda:0', grad_fn=<NllLossBackward>)
[4331] Train Loss:   tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[4332] Train Loss:   tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4333] Train Loss:   tensor(0.8168, device='cuda:0', grad_fn=<NllLossBackward>)
[4334] Train Loss:   tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[4335] Train Loss:   tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[4336] Train Loss:   tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[4337] Train Loss:   tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[4338] Train Loss:   tensor(0.8248, device='cuda:0', grad_fn=<NllLossBackward>)
[4339] Train Loss:   tensor(0.8025, device='cuda:0', grad_fn=<NllLossBackward>)
[4340] Train Loss:   tensor(0.7945, device='cuda:0', grad_fn=<NllLossBackward>)
[4341] Train Loss:   tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[4342] Train Loss:   tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[4343] Train Loss:   tensor(0.8620, device='cuda:0', grad_fn=<NllLossBackward>)
[4344] Train Loss:   tensor(0.8385, device='cuda:0', grad_fn=<NllLossBackward>)
[4345] Train Loss:   tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[4346] Train Loss:   tensor(0.8327, device='cuda:0', grad_fn=<NllLossBackward>)
[4347] Train Loss:   tensor(0.8497, device='cuda:0', grad_fn=<NllLossBackward>)
[4348] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[4349] Train Loss:   tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[4350] Train Loss:   tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[4351] Train Loss:   tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[4352] Train Loss:   tensor(0.8435, device='cuda:0', grad_fn=<NllLossBackward>)
[4353] Train Loss:   tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
[4354] Train Loss:   tensor(0.8524, device='cuda:0', grad_fn=<NllLossBackward>)
[4355] Train Loss:   tensor(0.8139, device='cuda:0', grad_fn=<NllLossBackward>)
[4356] Train Loss:   tensor(0.8337, device='cuda:0', grad_fn=<NllLossBackward>)
[4357] Train Loss:   tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[4358] Train Loss:   tensor(0.7909, device='cuda:0', grad_fn=<NllLossBackward>)
[4359] Train Loss:   tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[4360] Train Loss:   tensor(0.8193, device='cuda:0', grad_fn=<NllLossBackward>)
[4361] Train Loss:   tensor(0.8432, device='cuda:0', grad_fn=<NllLossBackward>)
[4362] Train Loss:   tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[4363] Train Loss:   tensor(0.8228, device='cuda:0', grad_fn=<NllLossBackward>)
[4364] Train Loss:   tensor(0.8301, device='cuda:0', grad_fn=<NllLossBackward>)
[4365] Train Loss:   tensor(0.8143, device='cuda:0', grad_fn=<NllLossBackward>)
[4366] Train Loss:   tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[4367] Train Loss:   tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[4368] Train Loss:   tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[4369] Train Loss:   tensor(0.8171, device='cuda:0', grad_fn=<NllLossBackward>)
[4370] Train Loss:   tensor(0.8155, device='cuda:0', grad_fn=<NllLossBackward>)
[4371] Train Loss:   tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[4372] Train Loss:   tensor(0.8468, device='cuda:0', grad_fn=<NllLossBackward>)
[4373] Train Loss:   tensor(0.8227, device='cuda:0', grad_fn=<NllLossBackward>)
[4374] Train Loss:   tensor(0.8425, device='cuda:0', grad_fn=<NllLossBackward>)
[4375] Train Loss:   tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[4376] Train Loss:   tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[4377] Train Loss:   tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[4378] Train Loss:   tensor(0.8104, device='cuda:0', grad_fn=<NllLossBackward>)
[4379] Train Loss:   tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[4380] Train Loss:   tensor(0.8301, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4381] Train Loss:    tensor(0.8056, device='cuda:0', grad_fn=<NllLossBackward>)
[4382] Train Loss:    tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[4383] Train Loss:    tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
[4384] Train Loss:    tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
[4385] Train Loss:    tensor(0.8071, device='cuda:0', grad_fn=<NllLossBackward>)
[4386] Train Loss:    tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[4387] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[4388] Train Loss:    tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[4389] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[4390] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[4391] Train Loss:    tensor(0.7969, device='cuda:0', grad_fn=<NllLossBackward>)
[4392] Train Loss:    tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[4393] Train Loss:    tensor(0.8308, device='cuda:0', grad_fn=<NllLossBackward>)
[4394] Train Loss:    tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[4395] Train Loss:    tensor(0.8414, device='cuda:0', grad_fn=<NllLossBackward>)
[4396] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[4397] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[4398] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[4399] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[4400] Train Loss:    tensor(0.8622, device='cuda:0', grad_fn=<NllLossBackward>)
[4401] Train Loss:    tensor(0.8241, device='cuda:0', grad_fn=<NllLossBackward>)
[4402] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[4403] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[4404] Train Loss:    tensor(0.7955, device='cuda:0', grad_fn=<NllLossBackward>)
[4405] Train Loss:    tensor(0.8209, device='cuda:0', grad_fn=<NllLossBackward>)
[4406] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[4407] Train Loss:    tensor(0.8004, device='cuda:0', grad_fn=<NllLossBackward>)
[4408] Train Loss:    tensor(0.8147, device='cuda:0', grad_fn=<NllLossBackward>)
[4409] Train Loss:    tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[4410] Train Loss:    tensor(0.8502, device='cuda:0', grad_fn=<NllLossBackward>)
[4411] Train Loss:    tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[4412] Train Loss:    tensor(0.8161, device='cuda:0', grad_fn=<NllLossBackward>)
[4413] Train Loss:    tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[4414] Train Loss:    tensor(0.8345, device='cuda:0', grad_fn=<NllLossBackward>)
[4415] Train Loss:    tensor(0.8514, device='cuda:0', grad_fn=<NllLossBackward>)
[4416] Train Loss:    tensor(0.8109, device='cuda:0', grad_fn=<NllLossBackward>)
[4417] Train Loss:    tensor(0.8095, device='cuda:0', grad_fn=<NllLossBackward>)
[4418] Train Loss:    tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[4419] Train Loss:    tensor(0.8199, device='cuda:0', grad_fn=<NllLossBackward>)
[4420] Train Loss:    tensor(0.8632, device='cuda:0', grad_fn=<NllLossBackward>)
[4421] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[4422] Train Loss:    tensor(0.8662, device='cuda:0', grad_fn=<NllLossBackward>)
[4423] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[4424] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[4425] Train Loss:    tensor(0.8203, device='cuda:0', grad_fn=<NllLossBackward>)
[4426] Train Loss:    tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[4427] Train Loss:    tensor(0.8199, device='cuda:0', grad_fn=<NllLossBackward>)
[4428] Train Loss:    tensor(0.8349, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4429] Train Loss:    tensor(0.7930, device='cuda:0', grad_fn=<NllLossBackward>)
[4430] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[4431] Train Loss:    tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[4432] Train Loss:    tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[4433] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[4434] Train Loss:    tensor(0.8641, device='cuda:0', grad_fn=<NllLossBackward>)
[4435] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[4436] Train Loss:    tensor(0.8442, device='cuda:0', grad_fn=<NllLossBackward>)
[4437] Train Loss:    tensor(0.8165, device='cuda:0', grad_fn=<NllLossBackward>)
[4438] Train Loss:    tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[4439] Train Loss:    tensor(0.8145, device='cuda:0', grad_fn=<NllLossBackward>)
[4440] Train Loss:    tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[4441] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[4442] Train Loss:    tensor(0.8169, device='cuda:0', grad_fn=<NllLossBackward>)
[4443] Train Loss:    tensor(0.8359, device='cuda:0', grad_fn=<NllLossBackward>)
[4444] Train Loss:    tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[4445] Train Loss:    tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[4446] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[4447] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[4448] Train Loss:    tensor(0.8088, device='cuda:0', grad_fn=<NllLossBackward>)
[4449] Train Loss:    tensor(0.8384, device='cuda:0', grad_fn=<NllLossBackward>)
[4450] Train Loss:    tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[4451] Train Loss:    tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[4452] Train Loss:    tensor(0.8383, device='cuda:0', grad_fn=<NllLossBackward>)
[4453] Train Loss:    tensor(0.8339, device='cuda:0', grad_fn=<NllLossBackward>)
[4454] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[4455] Train Loss:    tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[4456] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[4457] Train Loss:    tensor(0.8318, device='cuda:0', grad_fn=<NllLossBackward>)
[4458] Train Loss:    tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[4459] Train Loss:    tensor(0.8223, device='cuda:0', grad_fn=<NllLossBackward>)
[4460] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[4461] Train Loss:    tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
[4462] Train Loss:    tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[4463] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[4464] Train Loss:    tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[4465] Train Loss:    tensor(0.8183, device='cuda:0', grad_fn=<NllLossBackward>)
[4466] Train Loss:    tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[4467] Train Loss:    tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[4468] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[4469] Train Loss:    tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[4470] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[4471] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[4472] Train Loss:    tensor(0.8212, device='cuda:0', grad_fn=<NllLossBackward>)
[4473] Train Loss:    tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[4474] Train Loss:    tensor(0.8381, device='cuda:0', grad_fn=<NllLossBackward>)
[4475] Train Loss:    tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[4476] Train Loss:    tensor(0.8101, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4477] Train Loss:  tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[4478] Train Loss:  tensor(0.8341, device='cuda:0', grad_fn=<NllLossBackward>)
[4479] Train Loss:  tensor(0.8375, device='cuda:0', grad_fn=<NllLossBackward>)
[4480] Train Loss:  tensor(0.8162, device='cuda:0', grad_fn=<NllLossBackward>)
[4481] Train Loss:  tensor(0.8137, device='cuda:0', grad_fn=<NllLossBackward>)
[4482] Train Loss:  tensor(0.8109, device='cuda:0', grad_fn=<NllLossBackward>)
[4483] Train Loss:  tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[4484] Train Loss:  tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[4485] Train Loss:  tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[4486] Train Loss:  tensor(0.8162, device='cuda:0', grad_fn=<NllLossBackward>)
[4487] Train Loss:  tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[4488] Train Loss:  tensor(0.7909, device='cuda:0', grad_fn=<NllLossBackward>)
[4489] Train Loss:  tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[4490] Train Loss:  tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[4491] Train Loss:  tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[4492] Train Loss:  tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[4493] Train Loss:  tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[4494] Train Loss:  tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[4495] Train Loss:  tensor(0.8373, device='cuda:0', grad_fn=<NllLossBackward>)
[4496] Train Loss:  tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[4497] Train Loss:  tensor(0.8076, device='cuda:0', grad_fn=<NllLossBackward>)
[4498] Train Loss:  tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
[4499] Train Loss:  tensor(0.8430, device='cuda:0', grad_fn=<NllLossBackward>)
[4500] Train Loss:  tensor(0.7955, device='cuda:0', grad_fn=<NllLossBackward>)
[4501] Train Loss:  tensor(0.8499, device='cuda:0', grad_fn=<NllLossBackward>)
[4502] Train Loss:  tensor(0.8171, device='cuda:0', grad_fn=<NllLossBackward>)
[4503] Train Loss:  tensor(0.8303, device='cuda:0', grad_fn=<NllLossBackward>)
[4504] Train Loss:  tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[4505] Train Loss:  tensor(0.8394, device='cuda:0', grad_fn=<NllLossBackward>)
[4506] Train Loss:  tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[4507] Train Loss:  tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[4508] Train Loss:  tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[4509] Train Loss:  tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[4510] Train Loss:  tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[4511] Train Loss:  tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
Training Loss:  0.82780744941205 Time:  2520.960553884506 s
Training Accuaracy:  74.78449936454177 %
[0] Valid Loss:  tensor(0.7629, device='cuda:0')
[1] Valid Loss:  tensor(0.9257, device='cuda:0')
[2] Valid Loss:  tensor(0.7415, device='cuda:0')
[3] Valid Loss:  tensor(0.6687, device='cuda:0')
[4] Valid Loss:  tensor(0.7882, device='cuda:0')
[5] Valid Loss:  tensor(0.7834, device='cuda:0')
[6] Valid Loss:  tensor(0.7807, device='cuda:0')
[7] Valid Loss:  tensor(0.8840, device='cuda:0')
[8] Valid Loss:  tensor(0.8427, device='cuda:0')
[9] Valid Loss:  tensor(0.8767, device='cuda:0')
[10] Valid Loss:  tensor(0.7345, device='cuda:0')
```

```
[11] Valid Loss:  tensor(0.7144, device='cuda:0')
[12] Valid Loss:  tensor(1.1644, device='cuda:0')
[13] Valid Loss:  tensor(0.9973, device='cuda:0')
[14] Valid Loss:  tensor(0.6694, device='cuda:0')
[15] Valid Loss:  tensor(0.7954, device='cuda:0')
[16] Valid Loss:  tensor(0.6846, device='cuda:0')
[17] Valid Loss:  tensor(0.8328, device='cuda:0')
[18] Valid Loss:  tensor(0.7236, device='cuda:0')
[19] Valid Loss:  tensor(0.6793, device='cuda:0')
[20] Valid Loss:  tensor(0.7294, device='cuda:0')
[21] Valid Loss:  tensor(0.6516, device='cuda:0')
[22] Valid Loss:  tensor(0.6396, device='cuda:0')
[23] Valid Loss:  tensor(0.8780, device='cuda:0')
[24] Valid Loss:  tensor(0.7844, device='cuda:0')
[25] Valid Loss:  tensor(0.9940, device='cuda:0')
[26] Valid Loss:  tensor(0.5716, device='cuda:0')
[27] Valid Loss:  tensor(0.9234, device='cuda:0')
[28] Valid Loss:  tensor(0.7433, device='cuda:0')
[29] Valid Loss:  tensor(0.7678, device='cuda:0')
[30] Valid Loss:  tensor(0.6725, device='cuda:0')
[31] Valid Loss:  tensor(0.7936, device='cuda:0')
[32] Valid Loss:  tensor(0.9023, device='cuda:0')
[33] Valid Loss:  tensor(0.8536, device='cuda:0')
[34] Valid Loss:  tensor(0.8261, device='cuda:0')
[35] Valid Loss:  tensor(0.9114, device='cuda:0')
[36] Valid Loss:  tensor(0.7418, device='cuda:0')
[37] Valid Loss:  tensor(1.0632, device='cuda:0')
[38] Valid Loss:  tensor(0.8757, device='cuda:0')
[39] Valid Loss:  tensor(0.7287, device='cuda:0')
[40] Valid Loss:  tensor(0.8349, device='cuda:0')
[41] Valid Loss:  tensor(0.9137, device='cuda:0')
[42] Valid Loss:  tensor(0.8300, device='cuda:0')
[43] Valid Loss:  tensor(0.6177, device='cuda:0')
[44] Valid Loss:  tensor(0.7333, device='cuda:0')
[45] Valid Loss:  tensor(0.6322, device='cuda:0')
[46] Valid Loss:  tensor(0.9337, device='cuda:0')
[47] Valid Loss:  tensor(0.8555, device='cuda:0')
[48] Valid Loss:  tensor(0.7304, device='cuda:0')
[49] Valid Loss:  tensor(0.7886, device='cuda:0')
[50] Valid Loss:  tensor(0.7747, device='cuda:0')
[51] Valid Loss:  tensor(0.7068, device='cuda:0')
[52] Valid Loss:  tensor(0.6817, device='cuda:0')
[53] Valid Loss:  tensor(0.9305, device='cuda:0')
[54] Valid Loss:  tensor(0.6716, device='cuda:0')
[55] Valid Loss:  tensor(0.6226, device='cuda:0')
[56] Valid Loss:  tensor(0.9220, device='cuda:0')
[57] Valid Loss:  tensor(0.8642, device='cuda:0')
[58] Valid Loss:  tensor(0.8188, device='cuda:0')
```

```
[59] Valid Loss:  tensor(0.7241, device='cuda:0')
[60] Valid Loss:  tensor(1.0560, device='cuda:0')
[61] Valid Loss:  tensor(0.7541, device='cuda:0')
[62] Valid Loss:  tensor(0.8981, device='cuda:0')
[63] Valid Loss:  tensor(0.7357, device='cuda:0')
[64] Valid Loss:  tensor(0.7966, device='cuda:0')
[65] Valid Loss:  tensor(0.9079, device='cuda:0')
[66] Valid Loss:  tensor(0.7688, device='cuda:0')
[67] Valid Loss:  tensor(0.7906, device='cuda:0')
[68] Valid Loss:  tensor(0.7771, device='cuda:0')
[69] Valid Loss:  tensor(0.6243, device='cuda:0')
[70] Valid Loss:  tensor(0.6436, device='cuda:0')
[71] Valid Loss:  tensor(0.8592, device='cuda:0')
[72] Valid Loss:  tensor(0.8020, device='cuda:0')
[73] Valid Loss:  tensor(0.7461, device='cuda:0')
[74] Valid Loss:  tensor(0.8055, device='cuda:0')
[75] Valid Loss:  tensor(0.9166, device='cuda:0')
[76] Valid Loss:  tensor(0.8537, device='cuda:0')
[77] Valid Loss:  tensor(0.7472, device='cuda:0')
[78] Valid Loss:  tensor(0.8547, device='cuda:0')
[79] Valid Loss:  tensor(0.6915, device='cuda:0')
[80] Valid Loss:  tensor(0.7924, device='cuda:0')
[81] Valid Loss:  tensor(0.6689, device='cuda:0')
[82] Valid Loss:  tensor(0.9303, device='cuda:0')
[83] Valid Loss:  tensor(0.8637, device='cuda:0')
[84] Valid Loss:  tensor(0.6813, device='cuda:0')
[85] Valid Loss:  tensor(0.7911, device='cuda:0')
[86] Valid Loss:  tensor(0.8129, device='cuda:0')
[87] Valid Loss:  tensor(0.8759, device='cuda:0')
[88] Valid Loss:  tensor(0.6268, device='cuda:0')
[89] Valid Loss:  tensor(0.8279, device='cuda:0')
[90] Valid Loss:  tensor(0.9683, device='cuda:0')
[91] Valid Loss:  tensor(0.8092, device='cuda:0')
[92] Valid Loss:  tensor(0.8291, device='cuda:0')
[93] Valid Loss:  tensor(0.8404, device='cuda:0')
[94] Valid Loss:  tensor(0.8834, device='cuda:0')
[95] Valid Loss:  tensor(0.7477, device='cuda:0')
[96] Valid Loss:  tensor(0.9122, device='cuda:0')
[97] Valid Loss:  tensor(0.8496, device='cuda:0')
[98] Valid Loss:  tensor(0.8662, device='cuda:0')
[99] Valid Loss:  tensor(0.6864, device='cuda:0')
[100] Valid Loss:  tensor(0.8236, device='cuda:0')
[101] Valid Loss:  tensor(0.7733, device='cuda:0')
[102] Valid Loss:  tensor(0.9317, device='cuda:0')
[103] Valid Loss:  tensor(0.9119, device='cuda:0')
[104] Valid Loss:  tensor(0.7134, device='cuda:0')
[105] Valid Loss:  tensor(0.7085, device='cuda:0')
[106] Valid Loss:  tensor(0.6377, device='cuda:0')
```

```
[107] Valid Loss:   tensor(0.8292, device='cuda:0')
[108] Valid Loss:   tensor(0.8931, device='cuda:0')
[109] Valid Loss:   tensor(0.7991, device='cuda:0')
[110] Valid Loss:   tensor(0.8566, device='cuda:0')
[111] Valid Loss:   tensor(0.8055, device='cuda:0')
[112] Valid Loss:   tensor(0.8527, device='cuda:0')
[113] Valid Loss:   tensor(0.8165, device='cuda:0')
[114] Valid Loss:   tensor(0.7865, device='cuda:0')
[115] Valid Loss:   tensor(0.8869, device='cuda:0')
[116] Valid Loss:   tensor(0.8693, device='cuda:0')
[117] Valid Loss:   tensor(0.9012, device='cuda:0')
[118] Valid Loss:   tensor(0.7279, device='cuda:0')
[119] Valid Loss:   tensor(0.6482, device='cuda:0')
[120] Valid Loss:   tensor(1.0270, device='cuda:0')
[121] Valid Loss:   tensor(0.7292, device='cuda:0')
[122] Valid Loss:   tensor(0.8049, device='cuda:0')
[123] Valid Loss:   tensor(0.7179, device='cuda:0')
[124] Valid Loss:   tensor(0.9424, device='cuda:0')
[125] Valid Loss:   tensor(1.0041, device='cuda:0')
[126] Valid Loss:   tensor(0.6398, device='cuda:0')
[127] Valid Loss:   tensor(0.6792, device='cuda:0')
[128] Valid Loss:   tensor(0.9032, device='cuda:0')
[129] Valid Loss:   tensor(0.7635, device='cuda:0')
[130] Valid Loss:   tensor(0.8963, device='cuda:0')
[131] Valid Loss:   tensor(0.7319, device='cuda:0')
[132] Valid Loss:   tensor(0.7495, device='cuda:0')
[133] Valid Loss:   tensor(0.8360, device='cuda:0')
[134] Valid Loss:   tensor(0.7310, device='cuda:0')
[135] Valid Loss:   tensor(0.6552, device='cuda:0')
[136] Valid Loss:   tensor(0.6734, device='cuda:0')
[137] Valid Loss:   tensor(1.0277, device='cuda:0')
[138] Valid Loss:   tensor(0.8674, device='cuda:0')
[139] Valid Loss:   tensor(1.0959, device='cuda:0')
[140] Valid Loss:   tensor(0.8540, device='cuda:0')
[141] Valid Loss:   tensor(0.6850, device='cuda:0')
[142] Valid Loss:   tensor(0.9839, device='cuda:0')
[143] Valid Loss:   tensor(0.6940, device='cuda:0')
[144] Valid Loss:   tensor(0.8214, device='cuda:0')
[145] Valid Loss:   tensor(1.0036, device='cuda:0')
[146] Valid Loss:   tensor(0.7426, device='cuda:0')
[147] Valid Loss:   tensor(0.8500, device='cuda:0')
[148] Valid Loss:   tensor(0.7519, device='cuda:0')
[149] Valid Loss:   tensor(0.8081, device='cuda:0')
[150] Valid Loss:   tensor(0.8110, device='cuda:0')
[151] Valid Loss:   tensor(0.8018, device='cuda:0')
[152] Valid Loss:   tensor(0.7538, device='cuda:0')
[153] Valid Loss:   tensor(0.7432, device='cuda:0')
[154] Valid Loss:   tensor(0.7556, device='cuda:0')
```

```
[155] Valid Loss:  tensor(0.8062, device='cuda:0')
[156] Valid Loss:  tensor(0.8567, device='cuda:0')
[157] Valid Loss:  tensor(0.7155, device='cuda:0')
[158] Valid Loss:  tensor(0.7576, device='cuda:0')
[159] Valid Loss:  tensor(0.7825, device='cuda:0')
[160] Valid Loss:  tensor(0.6544, device='cuda:0')
[161] Valid Loss:  tensor(0.8741, device='cuda:0')
[162] Valid Loss:  tensor(0.8140, device='cuda:0')
[163] Valid Loss:  tensor(0.7117, device='cuda:0')
[164] Valid Loss:  tensor(1.0057, device='cuda:0')
[165] Valid Loss:  tensor(0.7781, device='cuda:0')
[166] Valid Loss:  tensor(0.7883, device='cuda:0')
[167] Valid Loss:  tensor(0.6932, device='cuda:0')
[168] Valid Loss:  tensor(0.7346, device='cuda:0')
[169] Valid Loss:  tensor(0.7306, device='cuda:0')
[170] Valid Loss:  tensor(0.7304, device='cuda:0')
[171] Valid Loss:  tensor(0.6193, device='cuda:0')
[172] Valid Loss:  tensor(0.8477, device='cuda:0')
[173] Valid Loss:  tensor(0.6950, device='cuda:0')
[174] Valid Loss:  tensor(0.8173, device='cuda:0')
[175] Valid Loss:  tensor(0.6941, device='cuda:0')
[176] Valid Loss:  tensor(0.6275, device='cuda:0')
[177] Valid Loss:  tensor(0.8146, device='cuda:0')
[178] Valid Loss:  tensor(0.8443, device='cuda:0')
[179] Valid Loss:  tensor(0.8023, device='cuda:0')
[180] Valid Loss:  tensor(0.8736, device='cuda:0')
[181] Valid Loss:  tensor(0.7910, device='cuda:0')
[182] Valid Loss:  tensor(0.8012, device='cuda:0')
[183] Valid Loss:  tensor(0.8040, device='cuda:0')
[184] Valid Loss:  tensor(0.9520, device='cuda:0')
[185] Valid Loss:  tensor(0.8910, device='cuda:0')
[186] Valid Loss:  tensor(0.7532, device='cuda:0')
[187] Valid Loss:  tensor(0.8792, device='cuda:0')
[188] Valid Loss:  tensor(0.9347, device='cuda:0')
[189] Valid Loss:  tensor(0.7830, device='cuda:0')
[190] Valid Loss:  tensor(0.7934, device='cuda:0')
[191] Valid Loss:  tensor(0.6622, device='cuda:0')
[192] Valid Loss:  tensor(1.0384, device='cuda:0')
[193] Valid Loss:  tensor(0.8223, device='cuda:0')
[194] Valid Loss:  tensor(0.8890, device='cuda:0')
[195] Valid Loss:  tensor(0.8888, device='cuda:0')
[196] Valid Loss:  tensor(0.6727, device='cuda:0')
[197] Valid Loss:  tensor(0.8239, device='cuda:0')
[198] Valid Loss:  tensor(0.7912, device='cuda:0')
[199] Valid Loss:  tensor(0.8473, device='cuda:0')
[200] Valid Loss:  tensor(0.8194, device='cuda:0')
[201] Valid Loss:  tensor(0.7642, device='cuda:0')
[202] Valid Loss:  tensor(0.7315, device='cuda:0')
```

```
[203] Valid Loss:  tensor(0.9584, device='cuda:0')
[204] Valid Loss:  tensor(0.9102, device='cuda:0')
[205] Valid Loss:  tensor(0.8958, device='cuda:0')
[206] Valid Loss:  tensor(0.9703, device='cuda:0')
[207] Valid Loss:  tensor(0.8134, device='cuda:0')
[208] Valid Loss:  tensor(0.6883, device='cuda:0')
[209] Valid Loss:  tensor(0.7738, device='cuda:0')
[210] Valid Loss:  tensor(1.1507, device='cuda:0')
[211] Valid Loss:  tensor(0.8019, device='cuda:0')
[212] Valid Loss:  tensor(0.6751, device='cuda:0')
[213] Valid Loss:  tensor(0.7687, device='cuda:0')
[214] Valid Loss:  tensor(0.7435, device='cuda:0')
[215] Valid Loss:  tensor(0.8508, device='cuda:0')
[216] Valid Loss:  tensor(0.7898, device='cuda:0')
[217] Valid Loss:  tensor(0.8147, device='cuda:0')
[218] Valid Loss:  tensor(0.6324, device='cuda:0')
[219] Valid Loss:  tensor(0.7433, device='cuda:0')
[220] Valid Loss:  tensor(0.7202, device='cuda:0')
[221] Valid Loss:  tensor(0.8288, device='cuda:0')
[222] Valid Loss:  tensor(0.6331, device='cuda:0')
[223] Valid Loss:  tensor(0.7457, device='cuda:0')
[224] Valid Loss:  tensor(0.8733, device='cuda:0')
[225] Valid Loss:  tensor(0.6747, device='cuda:0')
[226] Valid Loss:  tensor(0.7810, device='cuda:0')
[227] Valid Loss:  tensor(0.6725, device='cuda:0')
[228] Valid Loss:  tensor(0.8749, device='cuda:0')
[229] Valid Loss:  tensor(0.7298, device='cuda:0')
[230] Valid Loss:  tensor(0.8360, device='cuda:0')
[231] Valid Loss:  tensor(0.7376, device='cuda:0')
[232] Valid Loss:  tensor(0.6395, device='cuda:0')
[233] Valid Loss:  tensor(0.7422, device='cuda:0')
[234] Valid Loss:  tensor(0.6784, device='cuda:0')
[235] Valid Loss:  tensor(0.6534, device='cuda:0')
[236] Valid Loss:  tensor(0.8102, device='cuda:0')
[237] Valid Loss:  tensor(0.8459, device='cuda:0')
[238] Valid Loss:  tensor(0.5957, device='cuda:0')
[239] Valid Loss:  tensor(0.8329, device='cuda:0')
[240] Valid Loss:  tensor(0.7569, device='cuda:0')
[241] Valid Loss:  tensor(0.6900, device='cuda:0')
[242] Valid Loss:  tensor(0.5907, device='cuda:0')
[243] Valid Loss:  tensor(1.2206, device='cuda:0')
[244] Valid Loss:  tensor(0.8524, device='cuda:0')
[245] Valid Loss:  tensor(0.7507, device='cuda:0')
[246] Valid Loss:  tensor(0.7323, device='cuda:0')
[247] Valid Loss:  tensor(0.7565, device='cuda:0')
[248] Valid Loss:  tensor(0.9837, device='cuda:0')
[249] Valid Loss:  tensor(0.7711, device='cuda:0')
[250] Valid Loss:  tensor(0.8242, device='cuda:0')
```

```
[251] Valid Loss:  tensor(0.8833, device='cuda:0')
[252] Valid Loss:  tensor(0.6781, device='cuda:0')
[253] Valid Loss:  tensor(0.6891, device='cuda:0')
[254] Valid Loss:  tensor(0.7198, device='cuda:0')
[255] Valid Loss:  tensor(0.8404, device='cuda:0')
[256] Valid Loss:  tensor(0.8615, device='cuda:0')
[257] Valid Loss:  tensor(0.7102, device='cuda:0')
[258] Valid Loss:  tensor(0.8886, device='cuda:0')
[259] Valid Loss:  tensor(0.7798, device='cuda:0')
[260] Valid Loss:  tensor(0.9505, device='cuda:0')
[261] Valid Loss:  tensor(0.8321, device='cuda:0')
[262] Valid Loss:  tensor(0.9619, device='cuda:0')
[263] Valid Loss:  tensor(0.9375, device='cuda:0')
[264] Valid Loss:  tensor(0.8767, device='cuda:0')
[265] Valid Loss:  tensor(0.8549, device='cuda:0')
[266] Valid Loss:  tensor(0.7178, device='cuda:0')
[267] Valid Loss:  tensor(0.7807, device='cuda:0')
[268] Valid Loss:  tensor(0.7572, device='cuda:0')
[269] Valid Loss:  tensor(0.7497, device='cuda:0')
[270] Valid Loss:  tensor(0.6914, device='cuda:0')
[271] Valid Loss:  tensor(1.0171, device='cuda:0')
[272] Valid Loss:  tensor(0.7768, device='cuda:0')
[273] Valid Loss:  tensor(0.9041, device='cuda:0')
[274] Valid Loss:  tensor(0.7914, device='cuda:0')
[275] Valid Loss:  tensor(0.8135, device='cuda:0')
[276] Valid Loss:  tensor(0.7083, device='cuda:0')
[277] Valid Loss:  tensor(0.6856, device='cuda:0')
[278] Valid Loss:  tensor(0.9568, device='cuda:0')
[279] Valid Loss:  tensor(0.8554, device='cuda:0')
[280] Valid Loss:  tensor(0.9413, device='cuda:0')
[281] Valid Loss:  tensor(0.7154, device='cuda:0')
[282] Valid Loss:  tensor(0.8453, device='cuda:0')
[283] Valid Loss:  tensor(0.9404, device='cuda:0')
[284] Valid Loss:  tensor(0.7293, device='cuda:0')
[285] Valid Loss:  tensor(0.6167, device='cuda:0')
[286] Valid Loss:  tensor(0.8631, device='cuda:0')
[287] Valid Loss:  tensor(0.9913, device='cuda:0')
[288] Valid Loss:  tensor(0.6115, device='cuda:0')
[289] Valid Loss:  tensor(0.8466, device='cuda:0')
[290] Valid Loss:  tensor(0.6524, device='cuda:0')
[291] Valid Loss:  tensor(0.7609, device='cuda:0')
[292] Valid Loss:  tensor(0.8929, device='cuda:0')
[293] Valid Loss:  tensor(0.7079, device='cuda:0')
[294] Valid Loss:  tensor(0.9051, device='cuda:0')
[295] Valid Loss:  tensor(0.7029, device='cuda:0')
[296] Valid Loss:  tensor(0.8162, device='cuda:0')
[297] Valid Loss:  tensor(0.8410, device='cuda:0')
[298] Valid Loss:  tensor(0.7274, device='cuda:0')
```

```
[299] Valid Loss:   tensor(0.8512, device='cuda:0')
[300] Valid Loss:   tensor(0.9480, device='cuda:0')
[301] Valid Loss:   tensor(0.6327, device='cuda:0')
[302] Valid Loss:   tensor(0.6202, device='cuda:0')
[303] Valid Loss:   tensor(0.7522, device='cuda:0')
[304] Valid Loss:   tensor(0.7346, device='cuda:0')
[305] Valid Loss:   tensor(0.7953, device='cuda:0')
[306] Valid Loss:   tensor(0.8274, device='cuda:0')
[307] Valid Loss:   tensor(0.6514, device='cuda:0')
[308] Valid Loss:   tensor(0.6875, device='cuda:0')
[309] Valid Loss:   tensor(0.7642, device='cuda:0')
[310] Valid Loss:   tensor(0.7741, device='cuda:0')
[311] Valid Loss:   tensor(0.6910, device='cuda:0')
[312] Valid Loss:   tensor(0.6813, device='cuda:0')
[313] Valid Loss:   tensor(0.7658, device='cuda:0')
[314] Valid Loss:   tensor(0.7599, device='cuda:0')
[315] Valid Loss:   tensor(1.0553, device='cuda:0')
[316] Valid Loss:   tensor(0.8825, device='cuda:0')
[317] Valid Loss:   tensor(0.7496, device='cuda:0')
[318] Valid Loss:   tensor(0.7424, device='cuda:0')
[319] Valid Loss:   tensor(0.7049, device='cuda:0')
[320] Valid Loss:   tensor(0.7161, device='cuda:0')
[321] Valid Loss:   tensor(0.7171, device='cuda:0')
[322] Valid Loss:   tensor(0.7138, device='cuda:0')
[323] Valid Loss:   tensor(0.8613, device='cuda:0')
[324] Valid Loss:   tensor(0.7910, device='cuda:0')
[325] Valid Loss:   tensor(0.6971, device='cuda:0')
[326] Valid Loss:   tensor(0.7795, device='cuda:0')
[327] Valid Loss:   tensor(0.5259, device='cuda:0')
[328] Valid Loss:   tensor(0.7432, device='cuda:0')
[329] Valid Loss:   tensor(0.8436, device='cuda:0')
[330] Valid Loss:   tensor(0.6474, device='cuda:0')
[331] Valid Loss:   tensor(0.9786, device='cuda:0')
[332] Valid Loss:   tensor(0.6708, device='cuda:0')
[333] Valid Loss:   tensor(0.6457, device='cuda:0')
[334] Valid Loss:   tensor(0.6693, device='cuda:0')
[335] Valid Loss:   tensor(0.7489, device='cuda:0')
[336] Valid Loss:   tensor(0.8396, device='cuda:0')
[337] Valid Loss:   tensor(0.9371, device='cuda:0')
[338] Valid Loss:   tensor(0.8236, device='cuda:0')
[339] Valid Loss:   tensor(0.7532, device='cuda:0')
[340] Valid Loss:   tensor(0.8186, device='cuda:0')
[341] Valid Loss:   tensor(0.8812, device='cuda:0')
[342] Valid Loss:   tensor(0.9539, device='cuda:0')
[343] Valid Loss:   tensor(0.8067, device='cuda:0')
[344] Valid Loss:   tensor(0.6951, device='cuda:0')
[345] Valid Loss:   tensor(0.6148, device='cuda:0')
[346] Valid Loss:   tensor(0.7654, device='cuda:0')
```

```
[347] Valid Loss:  tensor(0.8888, device='cuda:0')
[348] Valid Loss:  tensor(0.7537, device='cuda:0')
[349] Valid Loss:  tensor(0.7202, device='cuda:0')
[350] Valid Loss:  tensor(0.7742, device='cuda:0')
[351] Valid Loss:  tensor(0.8846, device='cuda:0')
[352] Valid Loss:  tensor(0.6835, device='cuda:0')
[353] Valid Loss:  tensor(0.7978, device='cuda:0')
[354] Valid Loss:  tensor(0.9479, device='cuda:0')
[355] Valid Loss:  tensor(0.6028, device='cuda:0')
[356] Valid Loss:  tensor(0.7203, device='cuda:0')
[357] Valid Loss:  tensor(0.8573, device='cuda:0')
[358] Valid Loss:  tensor(0.7440, device='cuda:0')
[359] Valid Loss:  tensor(0.8428, device='cuda:0')
[360] Valid Loss:  tensor(0.7602, device='cuda:0')
[361] Valid Loss:  tensor(0.7954, device='cuda:0')
[362] Valid Loss:  tensor(0.7722, device='cuda:0')
[363] Valid Loss:  tensor(0.8652, device='cuda:0')
[364] Valid Loss:  tensor(0.9656, device='cuda:0')
[365] Valid Loss:  tensor(0.7420, device='cuda:0')
[366] Valid Loss:  tensor(0.7461, device='cuda:0')
[367] Valid Loss:  tensor(0.8335, device='cuda:0')
[368] Valid Loss:  tensor(0.8945, device='cuda:0')
[369] Valid Loss:  tensor(0.7799, device='cuda:0')
[370] Valid Loss:  tensor(0.6571, device='cuda:0')
[371] Valid Loss:  tensor(0.6794, device='cuda:0')
[372] Valid Loss:  tensor(0.8797, device='cuda:0')
[373] Valid Loss:  tensor(0.6042, device='cuda:0')
[374] Valid Loss:  tensor(0.7637, device='cuda:0')
[375] Valid Loss:  tensor(0.8645, device='cuda:0')
[376] Valid Loss:  tensor(0.7030, device='cuda:0')
[377] Valid Loss:  tensor(0.9690, device='cuda:0')
[378] Valid Loss:  tensor(0.6725, device='cuda:0')
[379] Valid Loss:  tensor(0.8609, device='cuda:0')
[380] Valid Loss:  tensor(1.0919, device='cuda:0')
Testing Loss:  0.796104325397121
Testing Accuaracy:  75.39057263461181 %
Epoch Number:  3
[0] Train Loss:  tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[1] Train Loss:  tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[2] Train Loss:  tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[3] Train Loss:  tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[4] Train Loss:  tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[5] Train Loss:  tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[6] Train Loss:  tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[7] Train Loss:  tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[8] Train Loss:  tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[9] Train Loss:  tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[10] Train Loss:  tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[11] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[12] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[13] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[14] Train Loss:    tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[15] Train Loss:    tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[16] Train Loss:    tensor(0.8149, device='cuda:0', grad_fn=<NllLossBackward>)
[17] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[18] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[19] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[20] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[21] Train Loss:    tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[22] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[23] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[24] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[25] Train Loss:    tensor(0.7975, device='cuda:0', grad_fn=<NllLossBackward>)
[26] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[27] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[28] Train Loss:    tensor(0.8021, device='cuda:0', grad_fn=<NllLossBackward>)
[29] Train Loss:    tensor(0.8080, device='cuda:0', grad_fn=<NllLossBackward>)
[30] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[31] Train Loss:    tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[32] Train Loss:    tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[33] Train Loss:    tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[34] Train Loss:    tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[35] Train Loss:    tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[36] Train Loss:    tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[37] Train Loss:    tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[38] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[39] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[40] Train Loss:    tensor(0.8235, device='cuda:0', grad_fn=<NllLossBackward>)
[41] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[42] Train Loss:    tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[43] Train Loss:    tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[44] Train Loss:    tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[45] Train Loss:    tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[46] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[47] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[48] Train Loss:    tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[49] Train Loss:    tensor(0.7975, device='cuda:0', grad_fn=<NllLossBackward>)
[50] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[51] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[52] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[53] Train Loss:    tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[54] Train Loss:    tensor(0.8266, device='cuda:0', grad_fn=<NllLossBackward>)
[55] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[56] Train Loss:    tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[57] Train Loss:    tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[58] Train Loss:    tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[59] Train Loss:   tensor(0.8179, device='cuda:0', grad_fn=<NllLossBackward>)
[60] Train Loss:   tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[61] Train Loss:   tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[62] Train Loss:   tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[63] Train Loss:   tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[64] Train Loss:   tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[65] Train Loss:   tensor(0.8100, device='cuda:0', grad_fn=<NllLossBackward>)
[66] Train Loss:   tensor(0.8553, device='cuda:0', grad_fn=<NllLossBackward>)
[67] Train Loss:   tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[68] Train Loss:   tensor(0.8025, device='cuda:0', grad_fn=<NllLossBackward>)
[69] Train Loss:   tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[70] Train Loss:   tensor(0.8425, device='cuda:0', grad_fn=<NllLossBackward>)
[71] Train Loss:   tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[72] Train Loss:   tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[73] Train Loss:   tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[74] Train Loss:   tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[75] Train Loss:   tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[76] Train Loss:   tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[77] Train Loss:   tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[78] Train Loss:   tensor(0.8350, device='cuda:0', grad_fn=<NllLossBackward>)
[79] Train Loss:   tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[80] Train Loss:   tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[81] Train Loss:   tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[82] Train Loss:   tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[83] Train Loss:   tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[84] Train Loss:   tensor(0.8171, device='cuda:0', grad_fn=<NllLossBackward>)
[85] Train Loss:   tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[86] Train Loss:   tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[87] Train Loss:   tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[88] Train Loss:   tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[89] Train Loss:   tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[90] Train Loss:   tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[91] Train Loss:   tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[92] Train Loss:   tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[93] Train Loss:   tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[94] Train Loss:   tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[95] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[96] Train Loss:   tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[97] Train Loss:   tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[98] Train Loss:   tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[99] Train Loss:   tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[100] Train Loss:  tensor(0.8178, device='cuda:0', grad_fn=<NllLossBackward>)
[101] Train Loss:  tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[102] Train Loss:  tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[103] Train Loss:  tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[104] Train Loss:  tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[105] Train Loss:  tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[106] Train Loss:  tensor(0.8326, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[107] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[108] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[109] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[110] Train Loss:    tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[111] Train Loss:    tensor(0.8075, device='cuda:0', grad_fn=<NllLossBackward>)
[112] Train Loss:    tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[113] Train Loss:    tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[114] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[115] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[116] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[117] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[118] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[119] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[120] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[121] Train Loss:    tensor(0.8020, device='cuda:0', grad_fn=<NllLossBackward>)
[122] Train Loss:    tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[123] Train Loss:    tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[124] Train Loss:    tensor(0.8487, device='cuda:0', grad_fn=<NllLossBackward>)
[125] Train Loss:    tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[126] Train Loss:    tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[127] Train Loss:    tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[128] Train Loss:    tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[129] Train Loss:    tensor(0.8010, device='cuda:0', grad_fn=<NllLossBackward>)
[130] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[131] Train Loss:    tensor(0.8128, device='cuda:0', grad_fn=<NllLossBackward>)
[132] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[133] Train Loss:    tensor(0.8109, device='cuda:0', grad_fn=<NllLossBackward>)
[134] Train Loss:    tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[135] Train Loss:    tensor(0.7858, device='cuda:0', grad_fn=<NllLossBackward>)
[136] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[137] Train Loss:    tensor(0.8021, device='cuda:0', grad_fn=<NllLossBackward>)
[138] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[139] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[140] Train Loss:    tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[141] Train Loss:    tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[142] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[143] Train Loss:    tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[144] Train Loss:    tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[145] Train Loss:    tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[146] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[147] Train Loss:    tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[148] Train Loss:    tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[149] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[150] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[151] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[152] Train Loss:    tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[153] Train Loss:    tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[154] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[155] Train Loss:   tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[156] Train Loss:   tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[157] Train Loss:   tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[158] Train Loss:   tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[159] Train Loss:   tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[160] Train Loss:   tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[161] Train Loss:   tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[162] Train Loss:   tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[163] Train Loss:   tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[164] Train Loss:   tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[165] Train Loss:   tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[166] Train Loss:   tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[167] Train Loss:   tensor(0.8239, device='cuda:0', grad_fn=<NllLossBackward>)
[168] Train Loss:   tensor(0.8340, device='cuda:0', grad_fn=<NllLossBackward>)
[169] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[170] Train Loss:   tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[171] Train Loss:   tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[172] Train Loss:   tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[173] Train Loss:   tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[174] Train Loss:   tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[175] Train Loss:   tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[176] Train Loss:   tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[177] Train Loss:   tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[178] Train Loss:   tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[179] Train Loss:   tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[180] Train Loss:   tensor(0.8186, device='cuda:0', grad_fn=<NllLossBackward>)
[181] Train Loss:   tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[182] Train Loss:   tensor(0.8137, device='cuda:0', grad_fn=<NllLossBackward>)
[183] Train Loss:   tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[184] Train Loss:   tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[185] Train Loss:   tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[186] Train Loss:   tensor(0.8181, device='cuda:0', grad_fn=<NllLossBackward>)
[187] Train Loss:   tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[188] Train Loss:   tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[189] Train Loss:   tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[190] Train Loss:   tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[191] Train Loss:   tensor(0.8327, device='cuda:0', grad_fn=<NllLossBackward>)
[192] Train Loss:   tensor(0.8075, device='cuda:0', grad_fn=<NllLossBackward>)
[193] Train Loss:   tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[194] Train Loss:   tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[195] Train Loss:   tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[196] Train Loss:   tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[197] Train Loss:   tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[198] Train Loss:   tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[199] Train Loss:   tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[200] Train Loss:   tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[201] Train Loss:   tensor(0.8024, device='cuda:0', grad_fn=<NllLossBackward>)
[202] Train Loss:   tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[203] Train Loss:  tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[204] Train Loss:  tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[205] Train Loss:  tensor(0.8055, device='cuda:0', grad_fn=<NllLossBackward>)
[206] Train Loss:  tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[207] Train Loss:  tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[208] Train Loss:  tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[209] Train Loss:  tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[210] Train Loss:  tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[211] Train Loss:  tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[212] Train Loss:  tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[213] Train Loss:  tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[214] Train Loss:  tensor(0.8306, device='cuda:0', grad_fn=<NllLossBackward>)
[215] Train Loss:  tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
[216] Train Loss:  tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[217] Train Loss:  tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[218] Train Loss:  tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[219] Train Loss:  tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[220] Train Loss:  tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[221] Train Loss:  tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
[222] Train Loss:  tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[223] Train Loss:  tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[224] Train Loss:  tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[225] Train Loss:  tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[226] Train Loss:  tensor(0.7991, device='cuda:0', grad_fn=<NllLossBackward>)
[227] Train Loss:  tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[228] Train Loss:  tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[229] Train Loss:  tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[230] Train Loss:  tensor(0.8090, device='cuda:0', grad_fn=<NllLossBackward>)
[231] Train Loss:  tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[232] Train Loss:  tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[233] Train Loss:  tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[234] Train Loss:  tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[235] Train Loss:  tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[236] Train Loss:  tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[237] Train Loss:  tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[238] Train Loss:  tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[239] Train Loss:  tensor(0.8250, device='cuda:0', grad_fn=<NllLossBackward>)
[240] Train Loss:  tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[241] Train Loss:  tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[242] Train Loss:  tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[243] Train Loss:  tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[244] Train Loss:  tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[245] Train Loss:  tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[246] Train Loss:  tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[247] Train Loss:  tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[248] Train Loss:  tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[249] Train Loss:  tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[250] Train Loss:  tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[251] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[252] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[253] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[254] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[255] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[256] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[257] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[258] Train Loss:    tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[259] Train Loss:    tensor(0.8065, device='cuda:0', grad_fn=<NllLossBackward>)
[260] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[261] Train Loss:    tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[262] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[263] Train Loss:    tensor(0.8282, device='cuda:0', grad_fn=<NllLossBackward>)
[264] Train Loss:    tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[265] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[266] Train Loss:    tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[267] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[268] Train Loss:    tensor(0.8221, device='cuda:0', grad_fn=<NllLossBackward>)
[269] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[270] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[271] Train Loss:    tensor(0.7930, device='cuda:0', grad_fn=<NllLossBackward>)
[272] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[273] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[274] Train Loss:    tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[275] Train Loss:    tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[276] Train Loss:    tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[277] Train Loss:    tensor(0.8104, device='cuda:0', grad_fn=<NllLossBackward>)
[278] Train Loss:    tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
[279] Train Loss:    tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[280] Train Loss:    tensor(0.8236, device='cuda:0', grad_fn=<NllLossBackward>)
[281] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[282] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[283] Train Loss:    tensor(0.7975, device='cuda:0', grad_fn=<NllLossBackward>)
[284] Train Loss:    tensor(0.8353, device='cuda:0', grad_fn=<NllLossBackward>)
[285] Train Loss:    tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[286] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[287] Train Loss:    tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[288] Train Loss:    tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[289] Train Loss:    tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[290] Train Loss:    tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[291] Train Loss:    tensor(0.8161, device='cuda:0', grad_fn=<NllLossBackward>)
[292] Train Loss:    tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[293] Train Loss:    tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[294] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[295] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[296] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[297] Train Loss:    tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[298] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[299] Train Loss:   tensor(0.8143, device='cuda:0', grad_fn=<NllLossBackward>)
[300] Train Loss:   tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[301] Train Loss:   tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[302] Train Loss:   tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[303] Train Loss:   tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[304] Train Loss:   tensor(0.7996, device='cuda:0', grad_fn=<NllLossBackward>)
[305] Train Loss:   tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[306] Train Loss:   tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[307] Train Loss:   tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[308] Train Loss:   tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[309] Train Loss:   tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[310] Train Loss:   tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[311] Train Loss:   tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[312] Train Loss:   tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[313] Train Loss:   tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[314] Train Loss:   tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[315] Train Loss:   tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[316] Train Loss:   tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[317] Train Loss:   tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[318] Train Loss:   tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[319] Train Loss:   tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[320] Train Loss:   tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[321] Train Loss:   tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[322] Train Loss:   tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[323] Train Loss:   tensor(0.7705, device='cuda:0', grad_fn=<NllLossBackward>)
[324] Train Loss:   tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[325] Train Loss:   tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[326] Train Loss:   tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[327] Train Loss:   tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[328] Train Loss:   tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[329] Train Loss:   tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[330] Train Loss:   tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[331] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[332] Train Loss:   tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[333] Train Loss:   tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[334] Train Loss:   tensor(0.8251, device='cuda:0', grad_fn=<NllLossBackward>)
[335] Train Loss:   tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[336] Train Loss:   tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[337] Train Loss:   tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[338] Train Loss:   tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[339] Train Loss:   tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[340] Train Loss:   tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[341] Train Loss:   tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[342] Train Loss:   tensor(0.8086, device='cuda:0', grad_fn=<NllLossBackward>)
[343] Train Loss:   tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[344] Train Loss:   tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[345] Train Loss:   tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[346] Train Loss:   tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[347] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[348] Train Loss:    tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[349] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[350] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[351] Train Loss:    tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[352] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[353] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[354] Train Loss:    tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[355] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[356] Train Loss:    tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[357] Train Loss:    tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[358] Train Loss:    tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[359] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[360] Train Loss:    tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[361] Train Loss:    tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[362] Train Loss:    tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[363] Train Loss:    tensor(0.8311, device='cuda:0', grad_fn=<NllLossBackward>)
[364] Train Loss:    tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[365] Train Loss:    tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[366] Train Loss:    tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[367] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[368] Train Loss:    tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[369] Train Loss:    tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[370] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[371] Train Loss:    tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[372] Train Loss:    tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[373] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[374] Train Loss:    tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[375] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[376] Train Loss:    tensor(0.8395, device='cuda:0', grad_fn=<NllLossBackward>)
[377] Train Loss:    tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[378] Train Loss:    tensor(0.7901, device='cuda:0', grad_fn=<NllLossBackward>)
[379] Train Loss:    tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
[380] Train Loss:    tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[381] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[382] Train Loss:    tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[383] Train Loss:    tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[384] Train Loss:    tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[385] Train Loss:    tensor(0.8221, device='cuda:0', grad_fn=<NllLossBackward>)
[386] Train Loss:    tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[387] Train Loss:    tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[388] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[389] Train Loss:    tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[390] Train Loss:    tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[391] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[392] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[393] Train Loss:    tensor(0.8191, device='cuda:0', grad_fn=<NllLossBackward>)
[394] Train Loss:    tensor(0.8150, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[395] Train Loss:   tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[396] Train Loss:   tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[397] Train Loss:   tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[398] Train Loss:   tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
[399] Train Loss:   tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[400] Train Loss:   tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[401] Train Loss:   tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[402] Train Loss:   tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[403] Train Loss:   tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[404] Train Loss:   tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[405] Train Loss:   tensor(0.8276, device='cuda:0', grad_fn=<NllLossBackward>)
[406] Train Loss:   tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[407] Train Loss:   tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
[408] Train Loss:   tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
[409] Train Loss:   tensor(0.8056, device='cuda:0', grad_fn=<NllLossBackward>)
[410] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[411] Train Loss:   tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[412] Train Loss:   tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[413] Train Loss:   tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[414] Train Loss:   tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[415] Train Loss:   tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[416] Train Loss:   tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[417] Train Loss:   tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[418] Train Loss:   tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[419] Train Loss:   tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[420] Train Loss:   tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[421] Train Loss:   tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[422] Train Loss:   tensor(0.8174, device='cuda:0', grad_fn=<NllLossBackward>)
[423] Train Loss:   tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[424] Train Loss:   tensor(0.8010, device='cuda:0', grad_fn=<NllLossBackward>)
[425] Train Loss:   tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[426] Train Loss:   tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[427] Train Loss:   tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[428] Train Loss:   tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[429] Train Loss:   tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[430] Train Loss:   tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[431] Train Loss:   tensor(0.7895, device='cuda:0', grad_fn=<NllLossBackward>)
[432] Train Loss:   tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[433] Train Loss:   tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[434] Train Loss:   tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[435] Train Loss:   tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[436] Train Loss:   tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[437] Train Loss:   tensor(0.8121, device='cuda:0', grad_fn=<NllLossBackward>)
[438] Train Loss:   tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[439] Train Loss:   tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[440] Train Loss:   tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[441] Train Loss:   tensor(0.8140, device='cuda:0', grad_fn=<NllLossBackward>)
[442] Train Loss:   tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[443] Train Loss:    tensor(0.8416, device='cuda:0', grad_fn=<NllLossBackward>)
[444] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[445] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[446] Train Loss:    tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[447] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[448] Train Loss:    tensor(0.8390, device='cuda:0', grad_fn=<NllLossBackward>)
[449] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[450] Train Loss:    tensor(0.8097, device='cuda:0', grad_fn=<NllLossBackward>)
[451] Train Loss:    tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[452] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[453] Train Loss:    tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[454] Train Loss:    tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[455] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[456] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[457] Train Loss:    tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
[458] Train Loss:    tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[459] Train Loss:    tensor(0.8119, device='cuda:0', grad_fn=<NllLossBackward>)
[460] Train Loss:    tensor(0.8249, device='cuda:0', grad_fn=<NllLossBackward>)
[461] Train Loss:    tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[462] Train Loss:    tensor(0.8095, device='cuda:0', grad_fn=<NllLossBackward>)
[463] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[464] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[465] Train Loss:    tensor(0.8301, device='cuda:0', grad_fn=<NllLossBackward>)
[466] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[467] Train Loss:    tensor(0.8345, device='cuda:0', grad_fn=<NllLossBackward>)
[468] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[469] Train Loss:    tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[470] Train Loss:    tensor(0.8004, device='cuda:0', grad_fn=<NllLossBackward>)
[471] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[472] Train Loss:    tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[473] Train Loss:    tensor(0.7856, device='cuda:0', grad_fn=<NllLossBackward>)
[474] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[475] Train Loss:    tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[476] Train Loss:    tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[477] Train Loss:    tensor(0.8293, device='cuda:0', grad_fn=<NllLossBackward>)
[478] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[479] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[480] Train Loss:    tensor(0.8139, device='cuda:0', grad_fn=<NllLossBackward>)
[481] Train Loss:    tensor(0.8399, device='cuda:0', grad_fn=<NllLossBackward>)
[482] Train Loss:    tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[483] Train Loss:    tensor(0.8201, device='cuda:0', grad_fn=<NllLossBackward>)
[484] Train Loss:    tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[485] Train Loss:    tensor(0.8059, device='cuda:0', grad_fn=<NllLossBackward>)
[486] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[487] Train Loss:    tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[488] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[489] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[490] Train Loss:    tensor(0.7875, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[491] Train Loss:   tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[492] Train Loss:   tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[493] Train Loss:   tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[494] Train Loss:   tensor(0.8088, device='cuda:0', grad_fn=<NllLossBackward>)
[495] Train Loss:   tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[496] Train Loss:   tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[497] Train Loss:   tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[498] Train Loss:   tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[499] Train Loss:   tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[500] Train Loss:   tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[501] Train Loss:   tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[502] Train Loss:   tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[503] Train Loss:   tensor(0.8104, device='cuda:0', grad_fn=<NllLossBackward>)
[504] Train Loss:   tensor(0.7859, device='cuda:0', grad_fn=<NllLossBackward>)
[505] Train Loss:   tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[506] Train Loss:   tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[507] Train Loss:   tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[508] Train Loss:   tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[509] Train Loss:   tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[510] Train Loss:   tensor(0.7975, device='cuda:0', grad_fn=<NllLossBackward>)
[511] Train Loss:   tensor(0.8171, device='cuda:0', grad_fn=<NllLossBackward>)
[512] Train Loss:   tensor(0.8055, device='cuda:0', grad_fn=<NllLossBackward>)
[513] Train Loss:   tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[514] Train Loss:   tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[515] Train Loss:   tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[516] Train Loss:   tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[517] Train Loss:   tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[518] Train Loss:   tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[519] Train Loss:   tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[520] Train Loss:   tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[521] Train Loss:   tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[522] Train Loss:   tensor(0.8276, device='cuda:0', grad_fn=<NllLossBackward>)
[523] Train Loss:   tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[524] Train Loss:   tensor(0.8009, device='cuda:0', grad_fn=<NllLossBackward>)
[525] Train Loss:   tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[526] Train Loss:   tensor(0.7895, device='cuda:0', grad_fn=<NllLossBackward>)
[527] Train Loss:   tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[528] Train Loss:   tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[529] Train Loss:   tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[530] Train Loss:   tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[531] Train Loss:   tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[532] Train Loss:   tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[533] Train Loss:   tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[534] Train Loss:   tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[535] Train Loss:   tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[536] Train Loss:   tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[537] Train Loss:   tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[538] Train Loss:   tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[539] Train Loss:    tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[540] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[541] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[542] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[543] Train Loss:    tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[544] Train Loss:    tensor(0.8121, device='cuda:0', grad_fn=<NllLossBackward>)
[545] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[546] Train Loss:    tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[547] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[548] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[549] Train Loss:    tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[550] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[551] Train Loss:    tensor(0.8576, device='cuda:0', grad_fn=<NllLossBackward>)
[552] Train Loss:    tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[553] Train Loss:    tensor(0.8109, device='cuda:0', grad_fn=<NllLossBackward>)
[554] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[555] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[556] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[557] Train Loss:    tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[558] Train Loss:    tensor(0.8570, device='cuda:0', grad_fn=<NllLossBackward>)
[559] Train Loss:    tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[560] Train Loss:    tensor(0.8085, device='cuda:0', grad_fn=<NllLossBackward>)
[561] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[562] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[563] Train Loss:    tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[564] Train Loss:    tensor(0.8244, device='cuda:0', grad_fn=<NllLossBackward>)
[565] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[566] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[567] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[568] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[569] Train Loss:    tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
[570] Train Loss:    tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[571] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[572] Train Loss:    tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[573] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[574] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[575] Train Loss:    tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[576] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[577] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[578] Train Loss:    tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[579] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[580] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[581] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[582] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[583] Train Loss:    tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[584] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[585] Train Loss:    tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[586] Train Loss:    tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[587] Train Loss:   tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[588] Train Loss:   tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[589] Train Loss:   tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[590] Train Loss:   tensor(0.8155, device='cuda:0', grad_fn=<NllLossBackward>)
[591] Train Loss:   tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[592] Train Loss:   tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[593] Train Loss:   tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[594] Train Loss:   tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[595] Train Loss:   tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[596] Train Loss:   tensor(0.8100, device='cuda:0', grad_fn=<NllLossBackward>)
[597] Train Loss:   tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[598] Train Loss:   tensor(0.8336, device='cuda:0', grad_fn=<NllLossBackward>)
[599] Train Loss:   tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[600] Train Loss:   tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[601] Train Loss:   tensor(0.8178, device='cuda:0', grad_fn=<NllLossBackward>)
[602] Train Loss:   tensor(0.7732, device='cuda:0', grad_fn=<NllLossBackward>)
[603] Train Loss:   tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[604] Train Loss:   tensor(0.8163, device='cuda:0', grad_fn=<NllLossBackward>)
[605] Train Loss:   tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[606] Train Loss:   tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[607] Train Loss:   tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[608] Train Loss:   tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[609] Train Loss:   tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[610] Train Loss:   tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[611] Train Loss:   tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[612] Train Loss:   tensor(0.8190, device='cuda:0', grad_fn=<NllLossBackward>)
[613] Train Loss:   tensor(0.7730, device='cuda:0', grad_fn=<NllLossBackward>)
[614] Train Loss:   tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[615] Train Loss:   tensor(0.8152, device='cuda:0', grad_fn=<NllLossBackward>)
[616] Train Loss:   tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[617] Train Loss:   tensor(0.8291, device='cuda:0', grad_fn=<NllLossBackward>)
[618] Train Loss:   tensor(0.8334, device='cuda:0', grad_fn=<NllLossBackward>)
[619] Train Loss:   tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[620] Train Loss:   tensor(0.8140, device='cuda:0', grad_fn=<NllLossBackward>)
[621] Train Loss:   tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[622] Train Loss:   tensor(0.8028, device='cuda:0', grad_fn=<NllLossBackward>)
[623] Train Loss:   tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[624] Train Loss:   tensor(0.8129, device='cuda:0', grad_fn=<NllLossBackward>)
[625] Train Loss:   tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[626] Train Loss:   tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[627] Train Loss:   tensor(0.8186, device='cuda:0', grad_fn=<NllLossBackward>)
[628] Train Loss:   tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[629] Train Loss:   tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[630] Train Loss:   tensor(0.8121, device='cuda:0', grad_fn=<NllLossBackward>)
[631] Train Loss:   tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[632] Train Loss:   tensor(0.8071, device='cuda:0', grad_fn=<NllLossBackward>)
[633] Train Loss:   tensor(0.8191, device='cuda:0', grad_fn=<NllLossBackward>)
[634] Train Loss:   tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[635] Train Loss:   tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[636] Train Loss:   tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[637] Train Loss:   tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[638] Train Loss:   tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[639] Train Loss:   tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[640] Train Loss:   tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[641] Train Loss:   tensor(0.8509, device='cuda:0', grad_fn=<NllLossBackward>)
[642] Train Loss:   tensor(0.8317, device='cuda:0', grad_fn=<NllLossBackward>)
[643] Train Loss:   tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[644] Train Loss:   tensor(0.8253, device='cuda:0', grad_fn=<NllLossBackward>)
[645] Train Loss:   tensor(0.8168, device='cuda:0', grad_fn=<NllLossBackward>)
[646] Train Loss:   tensor(0.8159, device='cuda:0', grad_fn=<NllLossBackward>)
[647] Train Loss:   tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[648] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[649] Train Loss:   tensor(0.8075, device='cuda:0', grad_fn=<NllLossBackward>)
[650] Train Loss:   tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[651] Train Loss:   tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[652] Train Loss:   tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[653] Train Loss:   tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[654] Train Loss:   tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[655] Train Loss:   tensor(0.8393, device='cuda:0', grad_fn=<NllLossBackward>)
[656] Train Loss:   tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[657] Train Loss:   tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[658] Train Loss:   tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[659] Train Loss:   tensor(0.7844, device='cuda:0', grad_fn=<NllLossBackward>)
[660] Train Loss:   tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[661] Train Loss:   tensor(0.7982, device='cuda:0', grad_fn=<NllLossBackward>)
[662] Train Loss:   tensor(0.7888, device='cuda:0', grad_fn=<NllLossBackward>)
[663] Train Loss:   tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[664] Train Loss:   tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[665] Train Loss:   tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[666] Train Loss:   tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[667] Train Loss:   tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[668] Train Loss:   tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[669] Train Loss:   tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[670] Train Loss:   tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[671] Train Loss:   tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[672] Train Loss:   tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[673] Train Loss:   tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[674] Train Loss:   tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[675] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[676] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[677] Train Loss:   tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[678] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[679] Train Loss:   tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[680] Train Loss:   tensor(0.8180, device='cuda:0', grad_fn=<NllLossBackward>)
[681] Train Loss:   tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[682] Train Loss:   tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[683] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[684] Train Loss:    tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[685] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[686] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[687] Train Loss:    tensor(0.8250, device='cuda:0', grad_fn=<NllLossBackward>)
[688] Train Loss:    tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[689] Train Loss:    tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[690] Train Loss:    tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[691] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[692] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[693] Train Loss:    tensor(0.8399, device='cuda:0', grad_fn=<NllLossBackward>)
[694] Train Loss:    tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[695] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[696] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[697] Train Loss:    tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[698] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[699] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[700] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[701] Train Loss:    tensor(0.8020, device='cuda:0', grad_fn=<NllLossBackward>)
[702] Train Loss:    tensor(0.8077, device='cuda:0', grad_fn=<NllLossBackward>)
[703] Train Loss:    tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[704] Train Loss:    tensor(0.8025, device='cuda:0', grad_fn=<NllLossBackward>)
[705] Train Loss:    tensor(0.8169, device='cuda:0', grad_fn=<NllLossBackward>)
[706] Train Loss:    tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[707] Train Loss:    tensor(0.8334, device='cuda:0', grad_fn=<NllLossBackward>)
[708] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[709] Train Loss:    tensor(0.8076, device='cuda:0', grad_fn=<NllLossBackward>)
[710] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[711] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[712] Train Loss:    tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[713] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[714] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[715] Train Loss:    tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[716] Train Loss:    tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[717] Train Loss:    tensor(0.7996, device='cuda:0', grad_fn=<NllLossBackward>)
[718] Train Loss:    tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[719] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[720] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[721] Train Loss:    tensor(0.8329, device='cuda:0', grad_fn=<NllLossBackward>)
[722] Train Loss:    tensor(0.7986, device='cuda:0', grad_fn=<NllLossBackward>)
[723] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[724] Train Loss:    tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[725] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[726] Train Loss:    tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[727] Train Loss:    tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[728] Train Loss:    tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[729] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[730] Train Loss:    tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[731] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[732] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[733] Train Loss:    tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[734] Train Loss:    tensor(0.8344, device='cuda:0', grad_fn=<NllLossBackward>)
[735] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[736] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[737] Train Loss:    tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[738] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[739] Train Loss:    tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[740] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[741] Train Loss:    tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[742] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[743] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[744] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[745] Train Loss:    tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[746] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[747] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[748] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[749] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[750] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[751] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[752] Train Loss:    tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[753] Train Loss:    tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[754] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[755] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[756] Train Loss:    tensor(0.8040, device='cuda:0', grad_fn=<NllLossBackward>)
[757] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[758] Train Loss:    tensor(0.8101, device='cuda:0', grad_fn=<NllLossBackward>)
[759] Train Loss:    tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[760] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[761] Train Loss:    tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[762] Train Loss:    tensor(0.8264, device='cuda:0', grad_fn=<NllLossBackward>)
[763] Train Loss:    tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[764] Train Loss:    tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[765] Train Loss:    tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[766] Train Loss:    tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[767] Train Loss:    tensor(0.8153, device='cuda:0', grad_fn=<NllLossBackward>)
[768] Train Loss:    tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[769] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[770] Train Loss:    tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[771] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[772] Train Loss:    tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[773] Train Loss:    tensor(0.8015, device='cuda:0', grad_fn=<NllLossBackward>)
[774] Train Loss:    tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[775] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[776] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[777] Train Loss:    tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[778] Train Loss:    tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[779] Train Loss:   tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[780] Train Loss:   tensor(0.8212, device='cuda:0', grad_fn=<NllLossBackward>)
[781] Train Loss:   tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[782] Train Loss:   tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[783] Train Loss:   tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[784] Train Loss:   tensor(0.8278, device='cuda:0', grad_fn=<NllLossBackward>)
[785] Train Loss:   tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[786] Train Loss:   tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[787] Train Loss:   tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[788] Train Loss:   tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[789] Train Loss:   tensor(0.7895, device='cuda:0', grad_fn=<NllLossBackward>)
[790] Train Loss:   tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[791] Train Loss:   tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[792] Train Loss:   tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[793] Train Loss:   tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[794] Train Loss:   tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[795] Train Loss:   tensor(0.8227, device='cuda:0', grad_fn=<NllLossBackward>)
[796] Train Loss:   tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[797] Train Loss:   tensor(0.7875, device='cuda:0', grad_fn=<NllLossBackward>)
[798] Train Loss:   tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
[799] Train Loss:   tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[800] Train Loss:   tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[801] Train Loss:   tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[802] Train Loss:   tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[803] Train Loss:   tensor(0.8090, device='cuda:0', grad_fn=<NllLossBackward>)
[804] Train Loss:   tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[805] Train Loss:   tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[806] Train Loss:   tensor(0.7909, device='cuda:0', grad_fn=<NllLossBackward>)
[807] Train Loss:   tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[808] Train Loss:   tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[809] Train Loss:   tensor(0.8162, device='cuda:0', grad_fn=<NllLossBackward>)
[810] Train Loss:   tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[811] Train Loss:   tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[812] Train Loss:   tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[813] Train Loss:   tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[814] Train Loss:   tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[815] Train Loss:   tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[816] Train Loss:   tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[817] Train Loss:   tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[818] Train Loss:   tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[819] Train Loss:   tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[820] Train Loss:   tensor(0.8076, device='cuda:0', grad_fn=<NllLossBackward>)
[821] Train Loss:   tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[822] Train Loss:   tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[823] Train Loss:   tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
[824] Train Loss:   tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[825] Train Loss:   tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[826] Train Loss:   tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[827] Train Loss:   tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[828] Train Loss:   tensor(0.8157, device='cuda:0', grad_fn=<NllLossBackward>)
[829] Train Loss:   tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[830] Train Loss:   tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[831] Train Loss:   tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[832] Train Loss:   tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[833] Train Loss:   tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[834] Train Loss:   tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[835] Train Loss:   tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[836] Train Loss:   tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[837] Train Loss:   tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[838] Train Loss:   tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[839] Train Loss:   tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[840] Train Loss:   tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[841] Train Loss:   tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[842] Train Loss:   tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[843] Train Loss:   tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[844] Train Loss:   tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[845] Train Loss:   tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[846] Train Loss:   tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[847] Train Loss:   tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[848] Train Loss:   tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[849] Train Loss:   tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[850] Train Loss:   tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[851] Train Loss:   tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[852] Train Loss:   tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[853] Train Loss:   tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[854] Train Loss:   tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[855] Train Loss:   tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[856] Train Loss:   tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[857] Train Loss:   tensor(0.8425, device='cuda:0', grad_fn=<NllLossBackward>)
[858] Train Loss:   tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[859] Train Loss:   tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[860] Train Loss:   tensor(0.8280, device='cuda:0', grad_fn=<NllLossBackward>)
[861] Train Loss:   tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
[862] Train Loss:   tensor(0.8030, device='cuda:0', grad_fn=<NllLossBackward>)
[863] Train Loss:   tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[864] Train Loss:   tensor(0.7801, device='cuda:0', grad_fn=<NllLossBackward>)
[865] Train Loss:   tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[866] Train Loss:   tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[867] Train Loss:   tensor(0.8071, device='cuda:0', grad_fn=<NllLossBackward>)
[868] Train Loss:   tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[869] Train Loss:   tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[870] Train Loss:   tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[871] Train Loss:   tensor(0.8146, device='cuda:0', grad_fn=<NllLossBackward>)
[872] Train Loss:   tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[873] Train Loss:   tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[874] Train Loss:   tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[875] Train Loss:   tensor(0.8075, device='cuda:0', grad_fn=<NllLossBackward>)
[876] Train Loss:   tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[877] Train Loss:   tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[878] Train Loss:   tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[879] Train Loss:   tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[880] Train Loss:   tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[881] Train Loss:   tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[882] Train Loss:   tensor(0.8046, device='cuda:0', grad_fn=<NllLossBackward>)
[883] Train Loss:   tensor(0.8119, device='cuda:0', grad_fn=<NllLossBackward>)
[884] Train Loss:   tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[885] Train Loss:   tensor(0.7705, device='cuda:0', grad_fn=<NllLossBackward>)
[886] Train Loss:   tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[887] Train Loss:   tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[888] Train Loss:   tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[889] Train Loss:   tensor(0.8286, device='cuda:0', grad_fn=<NllLossBackward>)
[890] Train Loss:   tensor(0.8100, device='cuda:0', grad_fn=<NllLossBackward>)
[891] Train Loss:   tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[892] Train Loss:   tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[893] Train Loss:   tensor(0.8428, device='cuda:0', grad_fn=<NllLossBackward>)
[894] Train Loss:   tensor(0.8319, device='cuda:0', grad_fn=<NllLossBackward>)
[895] Train Loss:   tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[896] Train Loss:   tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[897] Train Loss:   tensor(0.8021, device='cuda:0', grad_fn=<NllLossBackward>)
[898] Train Loss:   tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[899] Train Loss:   tensor(0.8044, device='cuda:0', grad_fn=<NllLossBackward>)
[900] Train Loss:   tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[901] Train Loss:   tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[902] Train Loss:   tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[903] Train Loss:   tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[904] Train Loss:   tensor(0.8489, device='cuda:0', grad_fn=<NllLossBackward>)
[905] Train Loss:   tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[906] Train Loss:   tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[907] Train Loss:   tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[908] Train Loss:   tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[909] Train Loss:   tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[910] Train Loss:   tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[911] Train Loss:   tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[912] Train Loss:   tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[913] Train Loss:   tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[914] Train Loss:   tensor(0.8050, device='cuda:0', grad_fn=<NllLossBackward>)
[915] Train Loss:   tensor(0.8009, device='cuda:0', grad_fn=<NllLossBackward>)
[916] Train Loss:   tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[917] Train Loss:   tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[918] Train Loss:   tensor(0.8071, device='cuda:0', grad_fn=<NllLossBackward>)
[919] Train Loss:   tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
[920] Train Loss:   tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[921] Train Loss:   tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[922] Train Loss:   tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[923] Train Loss:    tensor(0.8380, device='cuda:0', grad_fn=<NllLossBackward>)
[924] Train Loss:    tensor(0.7898, device='cuda:0', grad_fn=<NllLossBackward>)
[925] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[926] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[927] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[928] Train Loss:    tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[929] Train Loss:    tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[930] Train Loss:    tensor(0.8157, device='cuda:0', grad_fn=<NllLossBackward>)
[931] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[932] Train Loss:    tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[933] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[934] Train Loss:    tensor(0.8163, device='cuda:0', grad_fn=<NllLossBackward>)
[935] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[936] Train Loss:    tensor(0.8326, device='cuda:0', grad_fn=<NllLossBackward>)
[937] Train Loss:    tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[938] Train Loss:    tensor(0.8169, device='cuda:0', grad_fn=<NllLossBackward>)
[939] Train Loss:    tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[940] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[941] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[942] Train Loss:    tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[943] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[944] Train Loss:    tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[945] Train Loss:    tensor(0.8162, device='cuda:0', grad_fn=<NllLossBackward>)
[946] Train Loss:    tensor(0.7726, device='cuda:0', grad_fn=<NllLossBackward>)
[947] Train Loss:    tensor(0.8169, device='cuda:0', grad_fn=<NllLossBackward>)
[948] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[949] Train Loss:    tensor(0.8020, device='cuda:0', grad_fn=<NllLossBackward>)
[950] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[951] Train Loss:    tensor(0.8028, device='cuda:0', grad_fn=<NllLossBackward>)
[952] Train Loss:    tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[953] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[954] Train Loss:    tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[955] Train Loss:    tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[956] Train Loss:    tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[957] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[958] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[959] Train Loss:    tensor(0.8158, device='cuda:0', grad_fn=<NllLossBackward>)
[960] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[961] Train Loss:    tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[962] Train Loss:    tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[963] Train Loss:    tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[964] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[965] Train Loss:    tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[966] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[967] Train Loss:    tensor(0.8030, device='cuda:0', grad_fn=<NllLossBackward>)
[968] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[969] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[970] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[971] Train Loss:  tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[972] Train Loss:  tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[973] Train Loss:  tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[974] Train Loss:  tensor(0.7970, device='cuda:0', grad_fn=<NllLossBackward>)
[975] Train Loss:  tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[976] Train Loss:  tensor(0.8044, device='cuda:0', grad_fn=<NllLossBackward>)
[977] Train Loss:  tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[978] Train Loss:  tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[979] Train Loss:  tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[980] Train Loss:  tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[981] Train Loss:  tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[982] Train Loss:  tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[983] Train Loss:  tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[984] Train Loss:  tensor(0.8651, device='cuda:0', grad_fn=<NllLossBackward>)
[985] Train Loss:  tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[986] Train Loss:  tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[987] Train Loss:  tensor(0.8191, device='cuda:0', grad_fn=<NllLossBackward>)
[988] Train Loss:  tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[989] Train Loss:  tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[990] Train Loss:  tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[991] Train Loss:  tensor(0.8262, device='cuda:0', grad_fn=<NllLossBackward>)
[992] Train Loss:  tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[993] Train Loss:  tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[994] Train Loss:  tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[995] Train Loss:  tensor(0.8310, device='cuda:0', grad_fn=<NllLossBackward>)
[996] Train Loss:  tensor(0.8323, device='cuda:0', grad_fn=<NllLossBackward>)
[997] Train Loss:  tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[998] Train Loss:  tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[999] Train Loss:  tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[1000] Train Loss:  tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[1001] Train Loss:  tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[1002] Train Loss:  tensor(0.8552, device='cuda:0', grad_fn=<NllLossBackward>)
[1003] Train Loss:  tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[1004] Train Loss:  tensor(0.7854, device='cuda:0', grad_fn=<NllLossBackward>)
[1005] Train Loss:  tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[1006] Train Loss:  tensor(0.8087, device='cuda:0', grad_fn=<NllLossBackward>)
[1007] Train Loss:  tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[1008] Train Loss:  tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[1009] Train Loss:  tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[1010] Train Loss:  tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[1011] Train Loss:  tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[1012] Train Loss:  tensor(0.8206, device='cuda:0', grad_fn=<NllLossBackward>)
[1013] Train Loss:  tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[1014] Train Loss:  tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[1015] Train Loss:  tensor(0.7801, device='cuda:0', grad_fn=<NllLossBackward>)
[1016] Train Loss:  tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[1017] Train Loss:  tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[1018] Train Loss:  tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1019] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[1020] Train Loss:    tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[1021] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[1022] Train Loss:    tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[1023] Train Loss:    tensor(0.8100, device='cuda:0', grad_fn=<NllLossBackward>)
[1024] Train Loss:    tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[1025] Train Loss:    tensor(0.8044, device='cuda:0', grad_fn=<NllLossBackward>)
[1026] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[1027] Train Loss:    tensor(0.8149, device='cuda:0', grad_fn=<NllLossBackward>)
[1028] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[1029] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[1030] Train Loss:    tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[1031] Train Loss:    tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
[1032] Train Loss:    tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[1033] Train Loss:    tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[1034] Train Loss:    tensor(0.8238, device='cuda:0', grad_fn=<NllLossBackward>)
[1035] Train Loss:    tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
[1036] Train Loss:    tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[1037] Train Loss:    tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[1038] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[1039] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[1040] Train Loss:    tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[1041] Train Loss:    tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[1042] Train Loss:    tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[1043] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[1044] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[1045] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[1046] Train Loss:    tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[1047] Train Loss:    tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[1048] Train Loss:    tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[1049] Train Loss:    tensor(0.8237, device='cuda:0', grad_fn=<NllLossBackward>)
[1050] Train Loss:    tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[1051] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[1052] Train Loss:    tensor(0.7909, device='cuda:0', grad_fn=<NllLossBackward>)
[1053] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[1054] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[1055] Train Loss:    tensor(0.8098, device='cuda:0', grad_fn=<NllLossBackward>)
[1056] Train Loss:    tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[1057] Train Loss:    tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[1058] Train Loss:    tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[1059] Train Loss:    tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[1060] Train Loss:    tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[1061] Train Loss:    tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[1062] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[1063] Train Loss:    tensor(0.8176, device='cuda:0', grad_fn=<NllLossBackward>)
[1064] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[1065] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[1066] Train Loss:    tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1067] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[1068] Train Loss:    tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[1069] Train Loss:    tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[1070] Train Loss:    tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[1071] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[1072] Train Loss:    tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[1073] Train Loss:    tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[1074] Train Loss:    tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[1075] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[1076] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[1077] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[1078] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[1079] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[1080] Train Loss:    tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[1081] Train Loss:    tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[1082] Train Loss:    tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[1083] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[1084] Train Loss:    tensor(0.8080, device='cuda:0', grad_fn=<NllLossBackward>)
[1085] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[1086] Train Loss:    tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[1087] Train Loss:    tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[1088] Train Loss:    tensor(0.7991, device='cuda:0', grad_fn=<NllLossBackward>)
[1089] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[1090] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[1091] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[1092] Train Loss:    tensor(0.8316, device='cuda:0', grad_fn=<NllLossBackward>)
[1093] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[1094] Train Loss:    tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[1095] Train Loss:    tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[1096] Train Loss:    tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[1097] Train Loss:    tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[1098] Train Loss:    tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[1099] Train Loss:    tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[1100] Train Loss:    tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[1101] Train Loss:    tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[1102] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[1103] Train Loss:    tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[1104] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[1105] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[1106] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[1107] Train Loss:    tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[1108] Train Loss:    tensor(0.8119, device='cuda:0', grad_fn=<NllLossBackward>)
[1109] Train Loss:    tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[1110] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[1111] Train Loss:    tensor(0.7980, device='cuda:0', grad_fn=<NllLossBackward>)
[1112] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[1113] Train Loss:    tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[1114] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1115] Train Loss:    tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[1116] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[1117] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[1118] Train Loss:    tensor(0.8042, device='cuda:0', grad_fn=<NllLossBackward>)
[1119] Train Loss:    tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[1120] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[1121] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[1122] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[1123] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[1124] Train Loss:    tensor(0.8036, device='cuda:0', grad_fn=<NllLossBackward>)
[1125] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[1126] Train Loss:    tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[1127] Train Loss:    tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[1128] Train Loss:    tensor(0.8586, device='cuda:0', grad_fn=<NllLossBackward>)
[1129] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[1130] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[1131] Train Loss:    tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[1132] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[1133] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[1134] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[1135] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[1136] Train Loss:    tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[1137] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[1138] Train Loss:    tensor(0.8161, device='cuda:0', grad_fn=<NllLossBackward>)
[1139] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[1140] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[1141] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[1142] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[1143] Train Loss:    tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[1144] Train Loss:    tensor(0.8231, device='cuda:0', grad_fn=<NllLossBackward>)
[1145] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[1146] Train Loss:    tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[1147] Train Loss:    tensor(0.7942, device='cuda:0', grad_fn=<NllLossBackward>)
[1148] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[1149] Train Loss:    tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[1150] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[1151] Train Loss:    tensor(0.8159, device='cuda:0', grad_fn=<NllLossBackward>)
[1152] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[1153] Train Loss:    tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[1154] Train Loss:    tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[1155] Train Loss:    tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[1156] Train Loss:    tensor(0.8185, device='cuda:0', grad_fn=<NllLossBackward>)
[1157] Train Loss:    tensor(0.8118, device='cuda:0', grad_fn=<NllLossBackward>)
[1158] Train Loss:    tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[1159] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[1160] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[1161] Train Loss:    tensor(0.7980, device='cuda:0', grad_fn=<NllLossBackward>)
[1162] Train Loss:    tensor(0.7930, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1163] Train Loss:    tensor(0.8279, device='cuda:0', grad_fn=<NllLossBackward>)
[1164] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[1165] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[1166] Train Loss:    tensor(0.7942, device='cuda:0', grad_fn=<NllLossBackward>)
[1167] Train Loss:    tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[1168] Train Loss:    tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[1169] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[1170] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[1171] Train Loss:    tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[1172] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[1173] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[1174] Train Loss:    tensor(0.8131, device='cuda:0', grad_fn=<NllLossBackward>)
[1175] Train Loss:    tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[1176] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[1177] Train Loss:    tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[1178] Train Loss:    tensor(0.8028, device='cuda:0', grad_fn=<NllLossBackward>)
[1179] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[1180] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[1181] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[1182] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[1183] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[1184] Train Loss:    tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[1185] Train Loss:    tensor(0.7991, device='cuda:0', grad_fn=<NllLossBackward>)
[1186] Train Loss:    tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[1187] Train Loss:    tensor(0.8080, device='cuda:0', grad_fn=<NllLossBackward>)
[1188] Train Loss:    tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[1189] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[1190] Train Loss:    tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[1191] Train Loss:    tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[1192] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[1193] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[1194] Train Loss:    tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[1195] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[1196] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[1197] Train Loss:    tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[1198] Train Loss:    tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
[1199] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[1200] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[1201] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[1202] Train Loss:    tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[1203] Train Loss:    tensor(0.8127, device='cuda:0', grad_fn=<NllLossBackward>)
[1204] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[1205] Train Loss:    tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[1206] Train Loss:    tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[1207] Train Loss:    tensor(0.7996, device='cuda:0', grad_fn=<NllLossBackward>)
[1208] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[1209] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[1210] Train Loss:    tensor(0.8281, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1211] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[1212] Train Loss:    tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[1213] Train Loss:    tensor(0.8104, device='cuda:0', grad_fn=<NllLossBackward>)
[1214] Train Loss:    tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[1215] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[1216] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[1217] Train Loss:    tensor(0.8294, device='cuda:0', grad_fn=<NllLossBackward>)
[1218] Train Loss:    tensor(0.7582, device='cuda:0', grad_fn=<NllLossBackward>)
[1219] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[1220] Train Loss:    tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[1221] Train Loss:    tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[1222] Train Loss:    tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[1223] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[1224] Train Loss:    tensor(0.8470, device='cuda:0', grad_fn=<NllLossBackward>)
[1225] Train Loss:    tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[1226] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[1227] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[1228] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[1229] Train Loss:    tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[1230] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[1231] Train Loss:    tensor(0.8046, device='cuda:0', grad_fn=<NllLossBackward>)
[1232] Train Loss:    tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[1233] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[1234] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[1235] Train Loss:    tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[1236] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[1237] Train Loss:    tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
[1238] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[1239] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[1240] Train Loss:    tensor(0.7726, device='cuda:0', grad_fn=<NllLossBackward>)
[1241] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[1242] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[1243] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[1244] Train Loss:    tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[1245] Train Loss:    tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[1246] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[1247] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[1248] Train Loss:    tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[1249] Train Loss:    tensor(0.8235, device='cuda:0', grad_fn=<NllLossBackward>)
[1250] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[1251] Train Loss:    tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[1252] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[1253] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[1254] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[1255] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[1256] Train Loss:    tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[1257] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[1258] Train Loss:    tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1259] Train Loss:    tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[1260] Train Loss:    tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[1261] Train Loss:    tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[1262] Train Loss:    tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[1263] Train Loss:    tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[1264] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[1265] Train Loss:    tensor(0.8420, device='cuda:0', grad_fn=<NllLossBackward>)
[1266] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[1267] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[1268] Train Loss:    tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[1269] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[1270] Train Loss:    tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[1271] Train Loss:    tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[1272] Train Loss:    tensor(0.8076, device='cuda:0', grad_fn=<NllLossBackward>)
[1273] Train Loss:    tensor(0.8156, device='cuda:0', grad_fn=<NllLossBackward>)
[1274] Train Loss:    tensor(0.8337, device='cuda:0', grad_fn=<NllLossBackward>)
[1275] Train Loss:    tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[1276] Train Loss:    tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[1277] Train Loss:    tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[1278] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[1279] Train Loss:    tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[1280] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[1281] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[1282] Train Loss:    tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[1283] Train Loss:    tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[1284] Train Loss:    tensor(0.8024, device='cuda:0', grad_fn=<NllLossBackward>)
[1285] Train Loss:    tensor(0.8092, device='cuda:0', grad_fn=<NllLossBackward>)
[1286] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[1287] Train Loss:    tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[1288] Train Loss:    tensor(0.7705, device='cuda:0', grad_fn=<NllLossBackward>)
[1289] Train Loss:    tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[1290] Train Loss:    tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[1291] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[1292] Train Loss:    tensor(0.8036, device='cuda:0', grad_fn=<NllLossBackward>)
[1293] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[1294] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[1295] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[1296] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[1297] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[1298] Train Loss:    tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
[1299] Train Loss:    tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[1300] Train Loss:    tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[1301] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[1302] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[1303] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[1304] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[1305] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[1306] Train Loss:    tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1307] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[1308] Train Loss:    tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[1309] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[1310] Train Loss:    tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[1311] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[1312] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[1313] Train Loss:    tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[1314] Train Loss:    tensor(0.8176, device='cuda:0', grad_fn=<NllLossBackward>)
[1315] Train Loss:    tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[1316] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[1317] Train Loss:    tensor(0.8236, device='cuda:0', grad_fn=<NllLossBackward>)
[1318] Train Loss:    tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[1319] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[1320] Train Loss:    tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[1321] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[1322] Train Loss:    tensor(0.8216, device='cuda:0', grad_fn=<NllLossBackward>)
[1323] Train Loss:    tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[1324] Train Loss:    tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[1325] Train Loss:    tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[1326] Train Loss:    tensor(0.8335, device='cuda:0', grad_fn=<NllLossBackward>)
[1327] Train Loss:    tensor(0.8252, device='cuda:0', grad_fn=<NllLossBackward>)
[1328] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[1329] Train Loss:    tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[1330] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[1331] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[1332] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[1333] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[1334] Train Loss:    tensor(0.8077, device='cuda:0', grad_fn=<NllLossBackward>)
[1335] Train Loss:    tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[1336] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[1337] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[1338] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[1339] Train Loss:    tensor(0.7865, device='cuda:0', grad_fn=<NllLossBackward>)
[1340] Train Loss:    tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[1341] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[1342] Train Loss:    tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[1343] Train Loss:    tensor(0.8456, device='cuda:0', grad_fn=<NllLossBackward>)
[1344] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1345] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[1346] Train Loss:    tensor(0.8291, device='cuda:0', grad_fn=<NllLossBackward>)
[1347] Train Loss:    tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[1348] Train Loss:    tensor(0.8121, device='cuda:0', grad_fn=<NllLossBackward>)
[1349] Train Loss:    tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[1350] Train Loss:    tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[1351] Train Loss:    tensor(0.8212, device='cuda:0', grad_fn=<NllLossBackward>)
[1352] Train Loss:    tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[1353] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[1354] Train Loss:    tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1355] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[1356] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[1357] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[1358] Train Loss:    tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[1359] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[1360] Train Loss:    tensor(0.8338, device='cuda:0', grad_fn=<NllLossBackward>)
[1361] Train Loss:    tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[1362] Train Loss:    tensor(0.8218, device='cuda:0', grad_fn=<NllLossBackward>)
[1363] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[1364] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[1365] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[1366] Train Loss:    tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[1367] Train Loss:    tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[1368] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[1369] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[1370] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[1371] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[1372] Train Loss:    tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[1373] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[1374] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[1375] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[1376] Train Loss:    tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[1377] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[1378] Train Loss:    tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[1379] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[1380] Train Loss:    tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[1381] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[1382] Train Loss:    tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[1383] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[1384] Train Loss:    tensor(0.7898, device='cuda:0', grad_fn=<NllLossBackward>)
[1385] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[1386] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[1387] Train Loss:    tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[1388] Train Loss:    tensor(0.8015, device='cuda:0', grad_fn=<NllLossBackward>)
[1389] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[1390] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[1391] Train Loss:    tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[1392] Train Loss:    tensor(0.8226, device='cuda:0', grad_fn=<NllLossBackward>)
[1393] Train Loss:    tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[1394] Train Loss:    tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[1395] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[1396] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[1397] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[1398] Train Loss:    tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[1399] Train Loss:    tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[1400] Train Loss:    tensor(0.7991, device='cuda:0', grad_fn=<NllLossBackward>)
[1401] Train Loss:    tensor(0.8118, device='cuda:0', grad_fn=<NllLossBackward>)
[1402] Train Loss:    tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1403] Train Loss:   tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[1404] Train Loss:   tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[1405] Train Loss:   tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[1406] Train Loss:   tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[1407] Train Loss:   tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[1408] Train Loss:   tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[1409] Train Loss:   tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[1410] Train Loss:   tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[1411] Train Loss:   tensor(0.8271, device='cuda:0', grad_fn=<NllLossBackward>)
[1412] Train Loss:   tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[1413] Train Loss:   tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[1414] Train Loss:   tensor(0.8276, device='cuda:0', grad_fn=<NllLossBackward>)
[1415] Train Loss:   tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[1416] Train Loss:   tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[1417] Train Loss:   tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[1418] Train Loss:   tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[1419] Train Loss:   tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[1420] Train Loss:   tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[1421] Train Loss:   tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[1422] Train Loss:   tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[1423] Train Loss:   tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[1424] Train Loss:   tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[1425] Train Loss:   tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[1426] Train Loss:   tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[1427] Train Loss:   tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[1428] Train Loss:   tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[1429] Train Loss:   tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[1430] Train Loss:   tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[1431] Train Loss:   tensor(0.8373, device='cuda:0', grad_fn=<NllLossBackward>)
[1432] Train Loss:   tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[1433] Train Loss:   tensor(0.7945, device='cuda:0', grad_fn=<NllLossBackward>)
[1434] Train Loss:   tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[1435] Train Loss:   tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[1436] Train Loss:   tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[1437] Train Loss:   tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[1438] Train Loss:   tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[1439] Train Loss:   tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[1440] Train Loss:   tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[1441] Train Loss:   tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[1442] Train Loss:   tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[1443] Train Loss:   tensor(0.7823, device='cuda:0', grad_fn=<NllLossBackward>)
[1444] Train Loss:   tensor(0.8251, device='cuda:0', grad_fn=<NllLossBackward>)
[1445] Train Loss:   tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[1446] Train Loss:   tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[1447] Train Loss:   tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[1448] Train Loss:   tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[1449] Train Loss:   tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[1450] Train Loss:   tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1451] Train Loss:   tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[1452] Train Loss:   tensor(0.8126, device='cuda:0', grad_fn=<NllLossBackward>)
[1453] Train Loss:   tensor(0.8081, device='cuda:0', grad_fn=<NllLossBackward>)
[1454] Train Loss:   tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[1455] Train Loss:   tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[1456] Train Loss:   tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[1457] Train Loss:   tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[1458] Train Loss:   tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[1459] Train Loss:   tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[1460] Train Loss:   tensor(0.7955, device='cuda:0', grad_fn=<NllLossBackward>)
[1461] Train Loss:   tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[1462] Train Loss:   tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[1463] Train Loss:   tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[1464] Train Loss:   tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[1465] Train Loss:   tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[1466] Train Loss:   tensor(0.8110, device='cuda:0', grad_fn=<NllLossBackward>)
[1467] Train Loss:   tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[1468] Train Loss:   tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[1469] Train Loss:   tensor(0.8354, device='cuda:0', grad_fn=<NllLossBackward>)
[1470] Train Loss:   tensor(0.8325, device='cuda:0', grad_fn=<NllLossBackward>)
[1471] Train Loss:   tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[1472] Train Loss:   tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[1473] Train Loss:   tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[1474] Train Loss:   tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[1475] Train Loss:   tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[1476] Train Loss:   tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[1477] Train Loss:   tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[1478] Train Loss:   tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[1479] Train Loss:   tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[1480] Train Loss:   tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[1481] Train Loss:   tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[1482] Train Loss:   tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[1483] Train Loss:   tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[1484] Train Loss:   tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[1485] Train Loss:   tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[1486] Train Loss:   tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[1487] Train Loss:   tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[1488] Train Loss:   tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[1489] Train Loss:   tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[1490] Train Loss:   tensor(0.8178, device='cuda:0', grad_fn=<NllLossBackward>)
[1491] Train Loss:   tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[1492] Train Loss:   tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[1493] Train Loss:   tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[1494] Train Loss:   tensor(0.8183, device='cuda:0', grad_fn=<NllLossBackward>)
[1495] Train Loss:   tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[1496] Train Loss:   tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[1497] Train Loss:   tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[1498] Train Loss:   tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1499] Train Loss:    tensor(0.8080, device='cuda:0', grad_fn=<NllLossBackward>)
[1500] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[1501] Train Loss:    tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[1502] Train Loss:    tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[1503] Train Loss:    tensor(0.8171, device='cuda:0', grad_fn=<NllLossBackward>)
[1504] Train Loss:    tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[1505] Train Loss:    tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[1506] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[1507] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[1508] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[1509] Train Loss:    tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[1510] Train Loss:    tensor(0.8065, device='cuda:0', grad_fn=<NllLossBackward>)
[1511] Train Loss:    tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[1512] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[1513] Train Loss:    tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[1514] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[1515] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[1516] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[1517] Train Loss:    tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[1518] Train Loss:    tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[1519] Train Loss:    tensor(0.8153, device='cuda:0', grad_fn=<NllLossBackward>)
[1520] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[1521] Train Loss:    tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[1522] Train Loss:    tensor(0.8157, device='cuda:0', grad_fn=<NllLossBackward>)
[1523] Train Loss:    tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[1524] Train Loss:    tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[1525] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[1526] Train Loss:    tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[1527] Train Loss:    tensor(0.8409, device='cuda:0', grad_fn=<NllLossBackward>)
[1528] Train Loss:    tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[1529] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[1530] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[1531] Train Loss:    tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[1532] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[1533] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[1534] Train Loss:    tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[1535] Train Loss:    tensor(0.8128, device='cuda:0', grad_fn=<NllLossBackward>)
[1536] Train Loss:    tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[1537] Train Loss:    tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[1538] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[1539] Train Loss:    tensor(0.7793, device='cuda:0', grad_fn=<NllLossBackward>)
[1540] Train Loss:    tensor(0.8077, device='cuda:0', grad_fn=<NllLossBackward>)
[1541] Train Loss:    tensor(0.8426, device='cuda:0', grad_fn=<NllLossBackward>)
[1542] Train Loss:    tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[1543] Train Loss:    tensor(0.8162, device='cuda:0', grad_fn=<NllLossBackward>)
[1544] Train Loss:    tensor(0.8143, device='cuda:0', grad_fn=<NllLossBackward>)
[1545] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[1546] Train Loss:    tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1547] Train Loss:    tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[1548] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[1549] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[1550] Train Loss:    tensor(0.7865, device='cuda:0', grad_fn=<NllLossBackward>)
[1551] Train Loss:    tensor(0.7858, device='cuda:0', grad_fn=<NllLossBackward>)
[1552] Train Loss:    tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[1553] Train Loss:    tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[1554] Train Loss:    tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[1555] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[1556] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[1557] Train Loss:    tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[1558] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[1559] Train Loss:    tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[1560] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[1561] Train Loss:    tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[1562] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[1563] Train Loss:    tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[1564] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[1565] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[1566] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[1567] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[1568] Train Loss:    tensor(0.8334, device='cuda:0', grad_fn=<NllLossBackward>)
[1569] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[1570] Train Loss:    tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[1571] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[1572] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[1573] Train Loss:    tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[1574] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[1575] Train Loss:    tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[1576] Train Loss:    tensor(0.8373, device='cuda:0', grad_fn=<NllLossBackward>)
[1577] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[1578] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[1579] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[1580] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[1581] Train Loss:    tensor(0.7982, device='cuda:0', grad_fn=<NllLossBackward>)
[1582] Train Loss:    tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[1583] Train Loss:    tensor(0.8154, device='cuda:0', grad_fn=<NllLossBackward>)
[1584] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[1585] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[1586] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[1587] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[1588] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[1589] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[1590] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[1591] Train Loss:    tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[1592] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[1593] Train Loss:    tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[1594] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1595] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[1596] Train Loss:    tensor(0.8071, device='cuda:0', grad_fn=<NllLossBackward>)
[1597] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[1598] Train Loss:    tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[1599] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[1600] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[1601] Train Loss:    tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[1602] Train Loss:    tensor(0.8055, device='cuda:0', grad_fn=<NllLossBackward>)
[1603] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[1604] Train Loss:    tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[1605] Train Loss:    tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[1606] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[1607] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[1608] Train Loss:    tensor(0.7955, device='cuda:0', grad_fn=<NllLossBackward>)
[1609] Train Loss:    tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[1610] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[1611] Train Loss:    tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[1612] Train Loss:    tensor(0.8109, device='cuda:0', grad_fn=<NllLossBackward>)
[1613] Train Loss:    tensor(0.8221, device='cuda:0', grad_fn=<NllLossBackward>)
[1614] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[1615] Train Loss:    tensor(0.8004, device='cuda:0', grad_fn=<NllLossBackward>)
[1616] Train Loss:    tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[1617] Train Loss:    tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[1618] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[1619] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[1620] Train Loss:    tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[1621] Train Loss:    tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[1622] Train Loss:    tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[1623] Train Loss:    tensor(0.8140, device='cuda:0', grad_fn=<NllLossBackward>)
[1624] Train Loss:    tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[1625] Train Loss:    tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[1626] Train Loss:    tensor(0.8383, device='cuda:0', grad_fn=<NllLossBackward>)
[1627] Train Loss:    tensor(0.7955, device='cuda:0', grad_fn=<NllLossBackward>)
[1628] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[1629] Train Loss:    tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[1630] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[1631] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[1632] Train Loss:    tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
[1633] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[1634] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[1635] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[1636] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[1637] Train Loss:    tensor(0.8389, device='cuda:0', grad_fn=<NllLossBackward>)
[1638] Train Loss:    tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[1639] Train Loss:    tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[1640] Train Loss:    tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[1641] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[1642] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1643] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[1644] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[1645] Train Loss:    tensor(0.8059, device='cuda:0', grad_fn=<NllLossBackward>)
[1646] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[1647] Train Loss:    tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[1648] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[1649] Train Loss:    tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[1650] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[1651] Train Loss:    tensor(0.8249, device='cuda:0', grad_fn=<NllLossBackward>)
[1652] Train Loss:    tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[1653] Train Loss:    tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[1654] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[1655] Train Loss:    tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[1656] Train Loss:    tensor(0.7705, device='cuda:0', grad_fn=<NllLossBackward>)
[1657] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[1658] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[1659] Train Loss:    tensor(0.8049, device='cuda:0', grad_fn=<NllLossBackward>)
[1660] Train Loss:    tensor(0.8029, device='cuda:0', grad_fn=<NllLossBackward>)
[1661] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[1662] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[1663] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[1664] Train Loss:    tensor(0.7816, device='cuda:0', grad_fn=<NllLossBackward>)
[1665] Train Loss:    tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[1666] Train Loss:    tensor(0.8174, device='cuda:0', grad_fn=<NllLossBackward>)
[1667] Train Loss:    tensor(0.7989, device='cuda:0', grad_fn=<NllLossBackward>)
[1668] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[1669] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[1670] Train Loss:    tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[1671] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[1672] Train Loss:    tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[1673] Train Loss:    tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[1674] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[1675] Train Loss:    tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[1676] Train Loss:    tensor(0.8055, device='cuda:0', grad_fn=<NllLossBackward>)
[1677] Train Loss:    tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[1678] Train Loss:    tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[1679] Train Loss:    tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[1680] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[1681] Train Loss:    tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[1682] Train Loss:    tensor(0.8170, device='cuda:0', grad_fn=<NllLossBackward>)
[1683] Train Loss:    tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[1684] Train Loss:    tensor(0.8012, device='cuda:0', grad_fn=<NllLossBackward>)
[1685] Train Loss:    tensor(0.7816, device='cuda:0', grad_fn=<NllLossBackward>)
[1686] Train Loss:    tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
[1687] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[1688] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[1689] Train Loss:    tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[1690] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1691] Train Loss:    tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[1692] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[1693] Train Loss:    tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[1694] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[1695] Train Loss:    tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[1696] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[1697] Train Loss:    tensor(0.7982, device='cuda:0', grad_fn=<NllLossBackward>)
[1698] Train Loss:    tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[1699] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[1700] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[1701] Train Loss:    tensor(0.8341, device='cuda:0', grad_fn=<NllLossBackward>)
[1702] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[1703] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[1704] Train Loss:    tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[1705] Train Loss:    tensor(0.8055, device='cuda:0', grad_fn=<NllLossBackward>)
[1706] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[1707] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[1708] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[1709] Train Loss:    tensor(0.8077, device='cuda:0', grad_fn=<NllLossBackward>)
[1710] Train Loss:    tensor(0.8224, device='cuda:0', grad_fn=<NllLossBackward>)
[1711] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[1712] Train Loss:    tensor(0.7895, device='cuda:0', grad_fn=<NllLossBackward>)
[1713] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[1714] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[1715] Train Loss:    tensor(0.8100, device='cuda:0', grad_fn=<NllLossBackward>)
[1716] Train Loss:    tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[1717] Train Loss:    tensor(0.8156, device='cuda:0', grad_fn=<NllLossBackward>)
[1718] Train Loss:    tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[1719] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[1720] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[1721] Train Loss:    tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[1722] Train Loss:    tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[1723] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[1724] Train Loss:    tensor(0.8336, device='cuda:0', grad_fn=<NllLossBackward>)
[1725] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[1726] Train Loss:    tensor(0.7942, device='cuda:0', grad_fn=<NllLossBackward>)
[1727] Train Loss:    tensor(0.8090, device='cuda:0', grad_fn=<NllLossBackward>)
[1728] Train Loss:    tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[1729] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[1730] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[1731] Train Loss:    tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[1732] Train Loss:    tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[1733] Train Loss:    tensor(0.8167, device='cuda:0', grad_fn=<NllLossBackward>)
[1734] Train Loss:    tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[1735] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[1736] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[1737] Train Loss:    tensor(0.8137, device='cuda:0', grad_fn=<NllLossBackward>)
[1738] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1739] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1740] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[1741] Train Loss:    tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
[1742] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[1743] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[1744] Train Loss:    tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[1745] Train Loss:    tensor(0.8159, device='cuda:0', grad_fn=<NllLossBackward>)
[1746] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[1747] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[1748] Train Loss:    tensor(0.7903, device='cuda:0', grad_fn=<NllLossBackward>)
[1749] Train Loss:    tensor(0.8097, device='cuda:0', grad_fn=<NllLossBackward>)
[1750] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[1751] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[1752] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[1753] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1754] Train Loss:    tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[1755] Train Loss:    tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
[1756] Train Loss:    tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[1757] Train Loss:    tensor(0.8012, device='cuda:0', grad_fn=<NllLossBackward>)
[1758] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[1759] Train Loss:    tensor(0.8310, device='cuda:0', grad_fn=<NllLossBackward>)
[1760] Train Loss:    tensor(0.8149, device='cuda:0', grad_fn=<NllLossBackward>)
[1761] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[1762] Train Loss:    tensor(0.8065, device='cuda:0', grad_fn=<NllLossBackward>)
[1763] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[1764] Train Loss:    tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[1765] Train Loss:    tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[1766] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[1767] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[1768] Train Loss:    tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[1769] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[1770] Train Loss:    tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[1771] Train Loss:    tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[1772] Train Loss:    tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[1773] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[1774] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1775] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[1776] Train Loss:    tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[1777] Train Loss:    tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
[1778] Train Loss:    tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[1779] Train Loss:    tensor(0.8087, device='cuda:0', grad_fn=<NllLossBackward>)
[1780] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[1781] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[1782] Train Loss:    tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[1783] Train Loss:    tensor(0.8438, device='cuda:0', grad_fn=<NllLossBackward>)
[1784] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[1785] Train Loss:    tensor(0.8015, device='cuda:0', grad_fn=<NllLossBackward>)
[1786] Train Loss:    tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1787] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[1788] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[1789] Train Loss:    tensor(0.8128, device='cuda:0', grad_fn=<NllLossBackward>)
[1790] Train Loss:    tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[1791] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[1792] Train Loss:    tensor(0.8150, device='cuda:0', grad_fn=<NllLossBackward>)
[1793] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[1794] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[1795] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[1796] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[1797] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[1798] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[1799] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[1800] Train Loss:    tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[1801] Train Loss:    tensor(0.8161, device='cuda:0', grad_fn=<NllLossBackward>)
[1802] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[1803] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[1804] Train Loss:    tensor(0.7784, device='cuda:0', grad_fn=<NllLossBackward>)
[1805] Train Loss:    tensor(0.8033, device='cuda:0', grad_fn=<NllLossBackward>)
[1806] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[1807] Train Loss:    tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[1808] Train Loss:    tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[1809] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[1810] Train Loss:    tensor(0.7982, device='cuda:0', grad_fn=<NllLossBackward>)
[1811] Train Loss:    tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[1812] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[1813] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[1814] Train Loss:    tensor(0.7895, device='cuda:0', grad_fn=<NllLossBackward>)
[1815] Train Loss:    tensor(0.7854, device='cuda:0', grad_fn=<NllLossBackward>)
[1816] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[1817] Train Loss:    tensor(0.8009, device='cuda:0', grad_fn=<NllLossBackward>)
[1818] Train Loss:    tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[1819] Train Loss:    tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
[1820] Train Loss:    tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[1821] Train Loss:    tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[1822] Train Loss:    tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[1823] Train Loss:    tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[1824] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[1825] Train Loss:    tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[1826] Train Loss:    tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[1827] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[1828] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[1829] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[1830] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[1831] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[1832] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[1833] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[1834] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1835] Train Loss:    tensor(0.7844, device='cuda:0', grad_fn=<NllLossBackward>)
[1836] Train Loss:    tensor(0.8128, device='cuda:0', grad_fn=<NllLossBackward>)
[1837] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[1838] Train Loss:    tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[1839] Train Loss:    tensor(0.7970, device='cuda:0', grad_fn=<NllLossBackward>)
[1840] Train Loss:    tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[1841] Train Loss:    tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[1842] Train Loss:    tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[1843] Train Loss:    tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[1844] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[1845] Train Loss:    tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
[1846] Train Loss:    tensor(0.8406, device='cuda:0', grad_fn=<NllLossBackward>)
[1847] Train Loss:    tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[1848] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[1849] Train Loss:    tensor(0.8160, device='cuda:0', grad_fn=<NllLossBackward>)
[1850] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[1851] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1852] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[1853] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[1854] Train Loss:    tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[1855] Train Loss:    tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[1856] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[1857] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[1858] Train Loss:    tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[1859] Train Loss:    tensor(0.7801, device='cuda:0', grad_fn=<NllLossBackward>)
[1860] Train Loss:    tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[1861] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[1862] Train Loss:    tensor(0.7992, device='cuda:0', grad_fn=<NllLossBackward>)
[1863] Train Loss:    tensor(0.7726, device='cuda:0', grad_fn=<NllLossBackward>)
[1864] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[1865] Train Loss:    tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[1866] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[1867] Train Loss:    tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[1868] Train Loss:    tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[1869] Train Loss:    tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[1870] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[1871] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[1872] Train Loss:    tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[1873] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[1874] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[1875] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[1876] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[1877] Train Loss:    tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[1878] Train Loss:    tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[1879] Train Loss:    tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[1880] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[1881] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[1882] Train Loss:    tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1883] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[1884] Train Loss:    tensor(0.8165, device='cuda:0', grad_fn=<NllLossBackward>)
[1885] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[1886] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[1887] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[1888] Train Loss:    tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[1889] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[1890] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[1891] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[1892] Train Loss:    tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[1893] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[1894] Train Loss:    tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[1895] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[1896] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[1897] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[1898] Train Loss:    tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[1899] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[1900] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[1901] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[1902] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[1903] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[1904] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[1905] Train Loss:    tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[1906] Train Loss:    tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[1907] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[1908] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[1909] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[1910] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[1911] Train Loss:    tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[1912] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[1913] Train Loss:    tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[1914] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[1915] Train Loss:    tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[1916] Train Loss:    tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[1917] Train Loss:    tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[1918] Train Loss:    tensor(0.8270, device='cuda:0', grad_fn=<NllLossBackward>)
[1919] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[1920] Train Loss:    tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[1921] Train Loss:    tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[1922] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[1923] Train Loss:    tensor(0.8137, device='cuda:0', grad_fn=<NllLossBackward>)
[1924] Train Loss:    tensor(0.8183, device='cuda:0', grad_fn=<NllLossBackward>)
[1925] Train Loss:    tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[1926] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[1927] Train Loss:    tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[1928] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[1929] Train Loss:    tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[1930] Train Loss:    tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1931] Train Loss:   tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[1932] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[1933] Train Loss:   tensor(0.7823, device='cuda:0', grad_fn=<NllLossBackward>)
[1934] Train Loss:   tensor(0.8265, device='cuda:0', grad_fn=<NllLossBackward>)
[1935] Train Loss:   tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[1936] Train Loss:   tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[1937] Train Loss:   tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[1938] Train Loss:   tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[1939] Train Loss:   tensor(0.7854, device='cuda:0', grad_fn=<NllLossBackward>)
[1940] Train Loss:   tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[1941] Train Loss:   tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[1942] Train Loss:   tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[1943] Train Loss:   tensor(0.8472, device='cuda:0', grad_fn=<NllLossBackward>)
[1944] Train Loss:   tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[1945] Train Loss:   tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[1946] Train Loss:   tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[1947] Train Loss:   tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[1948] Train Loss:   tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[1949] Train Loss:   tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[1950] Train Loss:   tensor(0.7942, device='cuda:0', grad_fn=<NllLossBackward>)
[1951] Train Loss:   tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[1952] Train Loss:   tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[1953] Train Loss:   tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[1954] Train Loss:   tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[1955] Train Loss:   tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[1956] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[1957] Train Loss:   tensor(0.8396, device='cuda:0', grad_fn=<NllLossBackward>)
[1958] Train Loss:   tensor(0.8147, device='cuda:0', grad_fn=<NllLossBackward>)
[1959] Train Loss:   tensor(0.8085, device='cuda:0', grad_fn=<NllLossBackward>)
[1960] Train Loss:   tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[1961] Train Loss:   tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[1962] Train Loss:   tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[1963] Train Loss:   tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[1964] Train Loss:   tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[1965] Train Loss:   tensor(0.8171, device='cuda:0', grad_fn=<NllLossBackward>)
[1966] Train Loss:   tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[1967] Train Loss:   tensor(0.8252, device='cuda:0', grad_fn=<NllLossBackward>)
[1968] Train Loss:   tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[1969] Train Loss:   tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[1970] Train Loss:   tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[1971] Train Loss:   tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[1972] Train Loss:   tensor(0.7793, device='cuda:0', grad_fn=<NllLossBackward>)
[1973] Train Loss:   tensor(0.8059, device='cuda:0', grad_fn=<NllLossBackward>)
[1974] Train Loss:   tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[1975] Train Loss:   tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
[1976] Train Loss:   tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[1977] Train Loss:   tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[1978] Train Loss:   tensor(0.8143, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1979] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[1980] Train Loss:    tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[1981] Train Loss:    tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[1982] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[1983] Train Loss:    tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[1984] Train Loss:    tensor(0.8283, device='cuda:0', grad_fn=<NllLossBackward>)
[1985] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[1986] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[1987] Train Loss:    tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[1988] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[1989] Train Loss:    tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[1990] Train Loss:    tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
[1991] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[1992] Train Loss:    tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[1993] Train Loss:    tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[1994] Train Loss:    tensor(0.8191, device='cuda:0', grad_fn=<NllLossBackward>)
[1995] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[1996] Train Loss:    tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[1997] Train Loss:    tensor(0.7816, device='cuda:0', grad_fn=<NllLossBackward>)
[1998] Train Loss:    tensor(0.8085, device='cuda:0', grad_fn=<NllLossBackward>)
[1999] Train Loss:    tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[2000] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[2001] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[2002] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[2003] Train Loss:    tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[2004] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[2005] Train Loss:    tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[2006] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[2007] Train Loss:    tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[2008] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[2009] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[2010] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[2011] Train Loss:    tensor(0.7975, device='cuda:0', grad_fn=<NllLossBackward>)
[2012] Train Loss:    tensor(0.8192, device='cuda:0', grad_fn=<NllLossBackward>)
[2013] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[2014] Train Loss:    tensor(0.7909, device='cuda:0', grad_fn=<NllLossBackward>)
[2015] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[2016] Train Loss:    tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[2017] Train Loss:    tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[2018] Train Loss:    tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[2019] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[2020] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[2021] Train Loss:    tensor(0.8228, device='cuda:0', grad_fn=<NllLossBackward>)
[2022] Train Loss:    tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[2023] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[2024] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[2025] Train Loss:    tensor(0.7858, device='cuda:0', grad_fn=<NllLossBackward>)
[2026] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2027] Train Loss:   tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[2028] Train Loss:   tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[2029] Train Loss:   tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[2030] Train Loss:   tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[2031] Train Loss:   tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[2032] Train Loss:   tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[2033] Train Loss:   tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[2034] Train Loss:   tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[2035] Train Loss:   tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[2036] Train Loss:   tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[2037] Train Loss:   tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[2038] Train Loss:   tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[2039] Train Loss:   tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[2040] Train Loss:   tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[2041] Train Loss:   tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[2042] Train Loss:   tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[2043] Train Loss:   tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[2044] Train Loss:   tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[2045] Train Loss:   tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[2046] Train Loss:   tensor(0.8376, device='cuda:0', grad_fn=<NllLossBackward>)
[2047] Train Loss:   tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[2048] Train Loss:   tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[2049] Train Loss:   tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[2050] Train Loss:   tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[2051] Train Loss:   tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[2052] Train Loss:   tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[2053] Train Loss:   tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[2054] Train Loss:   tensor(0.8269, device='cuda:0', grad_fn=<NllLossBackward>)
[2055] Train Loss:   tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[2056] Train Loss:   tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[2057] Train Loss:   tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[2058] Train Loss:   tensor(0.8266, device='cuda:0', grad_fn=<NllLossBackward>)
[2059] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[2060] Train Loss:   tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[2061] Train Loss:   tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[2062] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[2063] Train Loss:   tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[2064] Train Loss:   tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[2065] Train Loss:   tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[2066] Train Loss:   tensor(0.8147, device='cuda:0', grad_fn=<NllLossBackward>)
[2067] Train Loss:   tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[2068] Train Loss:   tensor(0.8152, device='cuda:0', grad_fn=<NllLossBackward>)
[2069] Train Loss:   tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[2070] Train Loss:   tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[2071] Train Loss:   tensor(0.8012, device='cuda:0', grad_fn=<NllLossBackward>)
[2072] Train Loss:   tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[2073] Train Loss:   tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[2074] Train Loss:   tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2075] Train Loss:   tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[2076] Train Loss:   tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[2077] Train Loss:   tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[2078] Train Loss:   tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[2079] Train Loss:   tensor(0.7816, device='cuda:0', grad_fn=<NllLossBackward>)
[2080] Train Loss:   tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[2081] Train Loss:   tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[2082] Train Loss:   tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[2083] Train Loss:   tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[2084] Train Loss:   tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[2085] Train Loss:   tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[2086] Train Loss:   tensor(0.8310, device='cuda:0', grad_fn=<NllLossBackward>)
[2087] Train Loss:   tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[2088] Train Loss:   tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[2089] Train Loss:   tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[2090] Train Loss:   tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[2091] Train Loss:   tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[2092] Train Loss:   tensor(0.8274, device='cuda:0', grad_fn=<NllLossBackward>)
[2093] Train Loss:   tensor(0.8175, device='cuda:0', grad_fn=<NllLossBackward>)
[2094] Train Loss:   tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[2095] Train Loss:   tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[2096] Train Loss:   tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[2097] Train Loss:   tensor(0.8178, device='cuda:0', grad_fn=<NllLossBackward>)
[2098] Train Loss:   tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[2099] Train Loss:   tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[2100] Train Loss:   tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[2101] Train Loss:   tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[2102] Train Loss:   tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
[2103] Train Loss:   tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[2104] Train Loss:   tensor(0.8181, device='cuda:0', grad_fn=<NllLossBackward>)
[2105] Train Loss:   tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[2106] Train Loss:   tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[2107] Train Loss:   tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[2108] Train Loss:   tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[2109] Train Loss:   tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[2110] Train Loss:   tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[2111] Train Loss:   tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[2112] Train Loss:   tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[2113] Train Loss:   tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[2114] Train Loss:   tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[2115] Train Loss:   tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[2116] Train Loss:   tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[2117] Train Loss:   tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
[2118] Train Loss:   tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[2119] Train Loss:   tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[2120] Train Loss:   tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[2121] Train Loss:   tensor(0.8380, device='cuda:0', grad_fn=<NllLossBackward>)
[2122] Train Loss:   tensor(0.8296, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2123] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[2124] Train Loss:    tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[2125] Train Loss:    tensor(0.7816, device='cuda:0', grad_fn=<NllLossBackward>)
[2126] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[2127] Train Loss:    tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[2128] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[2129] Train Loss:    tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[2130] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[2131] Train Loss:    tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[2132] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[2133] Train Loss:    tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[2134] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[2135] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[2136] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[2137] Train Loss:    tensor(0.8139, device='cuda:0', grad_fn=<NllLossBackward>)
[2138] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[2139] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[2140] Train Loss:    tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[2141] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[2142] Train Loss:    tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[2143] Train Loss:    tensor(0.8198, device='cuda:0', grad_fn=<NllLossBackward>)
[2144] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[2145] Train Loss:    tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[2146] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[2147] Train Loss:    tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[2148] Train Loss:    tensor(0.8092, device='cuda:0', grad_fn=<NllLossBackward>)
[2149] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[2150] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[2151] Train Loss:    tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[2152] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[2153] Train Loss:    tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[2154] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[2155] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[2156] Train Loss:    tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[2157] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[2158] Train Loss:    tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[2159] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[2160] Train Loss:    tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[2161] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[2162] Train Loss:    tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[2163] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[2164] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[2165] Train Loss:    tensor(0.8095, device='cuda:0', grad_fn=<NllLossBackward>)
[2166] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[2167] Train Loss:    tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[2168] Train Loss:    tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[2169] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[2170] Train Loss:    tensor(0.8075, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2171] Train Loss:    tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[2172] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[2173] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[2174] Train Loss:    tensor(0.7859, device='cuda:0', grad_fn=<NllLossBackward>)
[2175] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[2176] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[2177] Train Loss:    tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[2178] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[2179] Train Loss:    tensor(0.7730, device='cuda:0', grad_fn=<NllLossBackward>)
[2180] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[2181] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[2182] Train Loss:    tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[2183] Train Loss:    tensor(0.8433, device='cuda:0', grad_fn=<NllLossBackward>)
[2184] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2185] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[2186] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[2187] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[2188] Train Loss:    tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[2189] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[2190] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[2191] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[2192] Train Loss:    tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[2193] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[2194] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[2195] Train Loss:    tensor(0.8089, device='cuda:0', grad_fn=<NllLossBackward>)
[2196] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[2197] Train Loss:    tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[2198] Train Loss:    tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[2199] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2200] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[2201] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[2202] Train Loss:    tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[2203] Train Loss:    tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[2204] Train Loss:    tensor(0.8522, device='cuda:0', grad_fn=<NllLossBackward>)
[2205] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[2206] Train Loss:    tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[2207] Train Loss:    tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
[2208] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[2209] Train Loss:    tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[2210] Train Loss:    tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[2211] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[2212] Train Loss:    tensor(0.7898, device='cuda:0', grad_fn=<NllLossBackward>)
[2213] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[2214] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[2215] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[2216] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[2217] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[2218] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2219] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[2220] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[2221] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[2222] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[2223] Train Loss:    tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[2224] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[2225] Train Loss:    tensor(0.8128, device='cuda:0', grad_fn=<NllLossBackward>)
[2226] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[2227] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[2228] Train Loss:    tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[2229] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[2230] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[2231] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[2232] Train Loss:    tensor(0.7838, device='cuda:0', grad_fn=<NllLossBackward>)
[2233] Train Loss:    tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[2234] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[2235] Train Loss:    tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[2236] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[2237] Train Loss:    tensor(0.8244, device='cuda:0', grad_fn=<NllLossBackward>)
[2238] Train Loss:    tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[2239] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[2240] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[2241] Train Loss:    tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[2242] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[2243] Train Loss:    tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
[2244] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[2245] Train Loss:    tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[2246] Train Loss:    tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[2247] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[2248] Train Loss:    tensor(0.8209, device='cuda:0', grad_fn=<NllLossBackward>)
[2249] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[2250] Train Loss:    tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[2251] Train Loss:    tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[2252] Train Loss:    tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[2253] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[2254] Train Loss:    tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[2255] Train Loss:    tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[2256] Train Loss:    tensor(0.8289, device='cuda:0', grad_fn=<NllLossBackward>)
[2257] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[2258] Train Loss:    tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[2259] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[2260] Train Loss:    tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[2261] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[2262] Train Loss:    tensor(0.8028, device='cuda:0', grad_fn=<NllLossBackward>)
[2263] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[2264] Train Loss:    tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[2265] Train Loss:    tensor(0.7980, device='cuda:0', grad_fn=<NllLossBackward>)
[2266] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2267] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[2268] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[2269] Train Loss:    tensor(0.8279, device='cuda:0', grad_fn=<NllLossBackward>)
[2270] Train Loss:    tensor(0.8172, device='cuda:0', grad_fn=<NllLossBackward>)
[2271] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[2272] Train Loss:    tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[2273] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[2274] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[2275] Train Loss:    tensor(0.8042, device='cuda:0', grad_fn=<NllLossBackward>)
[2276] Train Loss:    tensor(0.7985, device='cuda:0', grad_fn=<NllLossBackward>)
[2277] Train Loss:    tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[2278] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[2279] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[2280] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[2281] Train Loss:    tensor(0.7838, device='cuda:0', grad_fn=<NllLossBackward>)
[2282] Train Loss:    tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[2283] Train Loss:    tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[2284] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[2285] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[2286] Train Loss:    tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[2287] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[2288] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[2289] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[2290] Train Loss:    tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[2291] Train Loss:    tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[2292] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[2293] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[2294] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[2295] Train Loss:    tensor(0.8036, device='cuda:0', grad_fn=<NllLossBackward>)
[2296] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[2297] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[2298] Train Loss:    tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[2299] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[2300] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[2301] Train Loss:    tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[2302] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[2303] Train Loss:    tensor(0.8176, device='cuda:0', grad_fn=<NllLossBackward>)
[2304] Train Loss:    tensor(0.8134, device='cuda:0', grad_fn=<NllLossBackward>)
[2305] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[2306] Train Loss:    tensor(0.7726, device='cuda:0', grad_fn=<NllLossBackward>)
[2307] Train Loss:    tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[2308] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[2309] Train Loss:    tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[2310] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[2311] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[2312] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[2313] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[2314] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2315] Train Loss:   tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[2316] Train Loss:   tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[2317] Train Loss:   tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[2318] Train Loss:   tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[2319] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[2320] Train Loss:   tensor(0.7888, device='cuda:0', grad_fn=<NllLossBackward>)
[2321] Train Loss:   tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[2322] Train Loss:   tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[2323] Train Loss:   tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[2324] Train Loss:   tensor(0.8208, device='cuda:0', grad_fn=<NllLossBackward>)
[2325] Train Loss:   tensor(0.7969, device='cuda:0', grad_fn=<NllLossBackward>)
[2326] Train Loss:   tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[2327] Train Loss:   tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[2328] Train Loss:   tensor(0.8009, device='cuda:0', grad_fn=<NllLossBackward>)
[2329] Train Loss:   tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[2330] Train Loss:   tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[2331] Train Loss:   tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[2332] Train Loss:   tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[2333] Train Loss:   tensor(0.8021, device='cuda:0', grad_fn=<NllLossBackward>)
[2334] Train Loss:   tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[2335] Train Loss:   tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[2336] Train Loss:   tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[2337] Train Loss:   tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[2338] Train Loss:   tensor(0.8036, device='cuda:0', grad_fn=<NllLossBackward>)
[2339] Train Loss:   tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[2340] Train Loss:   tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[2341] Train Loss:   tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[2342] Train Loss:   tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[2343] Train Loss:   tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[2344] Train Loss:   tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[2345] Train Loss:   tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[2346] Train Loss:   tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[2347] Train Loss:   tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[2348] Train Loss:   tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[2349] Train Loss:   tensor(0.8015, device='cuda:0', grad_fn=<NllLossBackward>)
[2350] Train Loss:   tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[2351] Train Loss:   tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[2352] Train Loss:   tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[2353] Train Loss:   tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[2354] Train Loss:   tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[2355] Train Loss:   tensor(0.7948, device='cuda:0', grad_fn=<NllLossBackward>)
[2356] Train Loss:   tensor(0.8123, device='cuda:0', grad_fn=<NllLossBackward>)
[2357] Train Loss:   tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[2358] Train Loss:   tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[2359] Train Loss:   tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[2360] Train Loss:   tensor(0.8059, device='cuda:0', grad_fn=<NllLossBackward>)
[2361] Train Loss:   tensor(0.7793, device='cuda:0', grad_fn=<NllLossBackward>)
[2362] Train Loss:   tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2363] Train Loss:    tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[2364] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[2365] Train Loss:    tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[2366] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[2367] Train Loss:    tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[2368] Train Loss:    tensor(0.8444, device='cuda:0', grad_fn=<NllLossBackward>)
[2369] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[2370] Train Loss:    tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[2371] Train Loss:    tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[2372] Train Loss:    tensor(0.8162, device='cuda:0', grad_fn=<NllLossBackward>)
[2373] Train Loss:    tensor(0.7858, device='cuda:0', grad_fn=<NllLossBackward>)
[2374] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[2375] Train Loss:    tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[2376] Train Loss:    tensor(0.8215, device='cuda:0', grad_fn=<NllLossBackward>)
[2377] Train Loss:    tensor(0.7732, device='cuda:0', grad_fn=<NllLossBackward>)
[2378] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[2379] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[2380] Train Loss:    tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[2381] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[2382] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[2383] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[2384] Train Loss:    tensor(0.8145, device='cuda:0', grad_fn=<NllLossBackward>)
[2385] Train Loss:    tensor(0.8157, device='cuda:0', grad_fn=<NllLossBackward>)
[2386] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[2387] Train Loss:    tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[2388] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[2389] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[2390] Train Loss:    tensor(0.8255, device='cuda:0', grad_fn=<NllLossBackward>)
[2391] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[2392] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[2393] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[2394] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[2395] Train Loss:    tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[2396] Train Loss:    tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[2397] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[2398] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[2399] Train Loss:    tensor(0.8366, device='cuda:0', grad_fn=<NllLossBackward>)
[2400] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[2401] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[2402] Train Loss:    tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[2403] Train Loss:    tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[2404] Train Loss:    tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[2405] Train Loss:    tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[2406] Train Loss:    tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[2407] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[2408] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[2409] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[2410] Train Loss:    tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2411] Train Loss:    tensor(0.7582, device='cuda:0', grad_fn=<NllLossBackward>)
[2412] Train Loss:    tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[2413] Train Loss:    tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[2414] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[2415] Train Loss:    tensor(0.8044, device='cuda:0', grad_fn=<NllLossBackward>)
[2416] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[2417] Train Loss:    tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[2418] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[2419] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[2420] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[2421] Train Loss:    tensor(0.7969, device='cuda:0', grad_fn=<NllLossBackward>)
[2422] Train Loss:    tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[2423] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[2424] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[2425] Train Loss:    tensor(0.8321, device='cuda:0', grad_fn=<NllLossBackward>)
[2426] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[2427] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[2428] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[2429] Train Loss:    tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[2430] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[2431] Train Loss:    tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[2432] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[2433] Train Loss:    tensor(0.7955, device='cuda:0', grad_fn=<NllLossBackward>)
[2434] Train Loss:    tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[2435] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[2436] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[2437] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[2438] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[2439] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[2440] Train Loss:    tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[2441] Train Loss:    tensor(0.8301, device='cuda:0', grad_fn=<NllLossBackward>)
[2442] Train Loss:    tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[2443] Train Loss:    tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[2444] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[2445] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2446] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[2447] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[2448] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[2449] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[2450] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[2451] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[2452] Train Loss:    tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[2453] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[2454] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2455] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[2456] Train Loss:    tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[2457] Train Loss:    tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[2458] Train Loss:    tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2459] Train Loss:   tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[2460] Train Loss:   tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[2461] Train Loss:   tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[2462] Train Loss:   tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
[2463] Train Loss:   tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[2464] Train Loss:   tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[2465] Train Loss:   tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[2466] Train Loss:   tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[2467] Train Loss:   tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
[2468] Train Loss:   tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[2469] Train Loss:   tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[2470] Train Loss:   tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2471] Train Loss:   tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[2472] Train Loss:   tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[2473] Train Loss:   tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[2474] Train Loss:   tensor(0.8087, device='cuda:0', grad_fn=<NllLossBackward>)
[2475] Train Loss:   tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[2476] Train Loss:   tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[2477] Train Loss:   tensor(0.7705, device='cuda:0', grad_fn=<NllLossBackward>)
[2478] Train Loss:   tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[2479] Train Loss:   tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[2480] Train Loss:   tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[2481] Train Loss:   tensor(0.8280, device='cuda:0', grad_fn=<NllLossBackward>)
[2482] Train Loss:   tensor(0.8158, device='cuda:0', grad_fn=<NllLossBackward>)
[2483] Train Loss:   tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[2484] Train Loss:   tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[2485] Train Loss:   tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[2486] Train Loss:   tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[2487] Train Loss:   tensor(0.8233, device='cuda:0', grad_fn=<NllLossBackward>)
[2488] Train Loss:   tensor(0.8097, device='cuda:0', grad_fn=<NllLossBackward>)
[2489] Train Loss:   tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[2490] Train Loss:   tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[2491] Train Loss:   tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[2492] Train Loss:   tensor(0.8090, device='cuda:0', grad_fn=<NllLossBackward>)
[2493] Train Loss:   tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[2494] Train Loss:   tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[2495] Train Loss:   tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[2496] Train Loss:   tensor(0.7985, device='cuda:0', grad_fn=<NllLossBackward>)
[2497] Train Loss:   tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[2498] Train Loss:   tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[2499] Train Loss:   tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[2500] Train Loss:   tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[2501] Train Loss:   tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[2502] Train Loss:   tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[2503] Train Loss:   tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[2504] Train Loss:   tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[2505] Train Loss:   tensor(0.8358, device='cuda:0', grad_fn=<NllLossBackward>)
[2506] Train Loss:   tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2507] Train Loss:    tensor(0.8195, device='cuda:0', grad_fn=<NllLossBackward>)
[2508] Train Loss:    tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[2509] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[2510] Train Loss:    tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[2511] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[2512] Train Loss:    tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
[2513] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[2514] Train Loss:    tensor(0.8268, device='cuda:0', grad_fn=<NllLossBackward>)
[2515] Train Loss:    tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[2516] Train Loss:    tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[2517] Train Loss:    tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[2518] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[2519] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[2520] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[2521] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[2522] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[2523] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[2524] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[2525] Train Loss:    tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[2526] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[2527] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[2528] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[2529] Train Loss:    tensor(0.7898, device='cuda:0', grad_fn=<NllLossBackward>)
[2530] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[2531] Train Loss:    tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[2532] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[2533] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[2534] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[2535] Train Loss:    tensor(0.8156, device='cuda:0', grad_fn=<NllLossBackward>)
[2536] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[2537] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[2538] Train Loss:    tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
[2539] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[2540] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[2541] Train Loss:    tensor(0.8086, device='cuda:0', grad_fn=<NllLossBackward>)
[2542] Train Loss:    tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[2543] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[2544] Train Loss:    tensor(0.8101, device='cuda:0', grad_fn=<NllLossBackward>)
[2545] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[2546] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[2547] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[2548] Train Loss:    tensor(0.8176, device='cuda:0', grad_fn=<NllLossBackward>)
[2549] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[2550] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[2551] Train Loss:    tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[2552] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[2553] Train Loss:    tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[2554] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2555] Train Loss:    tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[2556] Train Loss:    tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[2557] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[2558] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[2559] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[2560] Train Loss:    tensor(0.7942, device='cuda:0', grad_fn=<NllLossBackward>)
[2561] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[2562] Train Loss:    tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[2563] Train Loss:    tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[2564] Train Loss:    tensor(0.8033, device='cuda:0', grad_fn=<NllLossBackward>)
[2565] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[2566] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[2567] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[2568] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[2569] Train Loss:    tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[2570] Train Loss:    tensor(0.8218, device='cuda:0', grad_fn=<NllLossBackward>)
[2571] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[2572] Train Loss:    tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[2573] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[2574] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[2575] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[2576] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[2577] Train Loss:    tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[2578] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[2579] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[2580] Train Loss:    tensor(0.8178, device='cuda:0', grad_fn=<NllLossBackward>)
[2581] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[2582] Train Loss:    tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[2583] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[2584] Train Loss:    tensor(0.7793, device='cuda:0', grad_fn=<NllLossBackward>)
[2585] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[2586] Train Loss:    tensor(0.8221, device='cuda:0', grad_fn=<NllLossBackward>)
[2587] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[2588] Train Loss:    tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[2589] Train Loss:    tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[2590] Train Loss:    tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[2591] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[2592] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[2593] Train Loss:    tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[2594] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[2595] Train Loss:    tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[2596] Train Loss:    tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[2597] Train Loss:    tensor(0.8219, device='cuda:0', grad_fn=<NllLossBackward>)
[2598] Train Loss:    tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[2599] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[2600] Train Loss:    tensor(0.7784, device='cuda:0', grad_fn=<NllLossBackward>)
[2601] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[2602] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2603] Train Loss:    tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[2604] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[2605] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[2606] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[2607] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[2608] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2609] Train Loss:    tensor(0.8009, device='cuda:0', grad_fn=<NllLossBackward>)
[2610] Train Loss:    tensor(0.8238, device='cuda:0', grad_fn=<NllLossBackward>)
[2611] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[2612] Train Loss:    tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[2613] Train Loss:    tensor(0.8163, device='cuda:0', grad_fn=<NllLossBackward>)
[2614] Train Loss:    tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[2615] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[2616] Train Loss:    tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[2617] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[2618] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[2619] Train Loss:    tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[2620] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[2621] Train Loss:    tensor(0.8184, device='cuda:0', grad_fn=<NllLossBackward>)
[2622] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[2623] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[2624] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[2625] Train Loss:    tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[2626] Train Loss:    tensor(0.8177, device='cuda:0', grad_fn=<NllLossBackward>)
[2627] Train Loss:    tensor(0.8042, device='cuda:0', grad_fn=<NllLossBackward>)
[2628] Train Loss:    tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
[2629] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[2630] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[2631] Train Loss:    tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[2632] Train Loss:    tensor(0.7901, device='cuda:0', grad_fn=<NllLossBackward>)
[2633] Train Loss:    tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[2634] Train Loss:    tensor(0.8218, device='cuda:0', grad_fn=<NllLossBackward>)
[2635] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[2636] Train Loss:    tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[2637] Train Loss:    tensor(0.8268, device='cuda:0', grad_fn=<NllLossBackward>)
[2638] Train Loss:    tensor(0.7854, device='cuda:0', grad_fn=<NllLossBackward>)
[2639] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[2640] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[2641] Train Loss:    tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[2642] Train Loss:    tensor(0.8238, device='cuda:0', grad_fn=<NllLossBackward>)
[2643] Train Loss:    tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[2644] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[2645] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[2646] Train Loss:    tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[2647] Train Loss:    tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[2648] Train Loss:    tensor(0.8119, device='cuda:0', grad_fn=<NllLossBackward>)
[2649] Train Loss:    tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[2650] Train Loss:    tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2651] Train Loss:    tensor(0.8166, device='cuda:0', grad_fn=<NllLossBackward>)
[2652] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[2653] Train Loss:    tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[2654] Train Loss:    tensor(0.8147, device='cuda:0', grad_fn=<NllLossBackward>)
[2655] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[2656] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[2657] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[2658] Train Loss:    tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[2659] Train Loss:    tensor(0.8357, device='cuda:0', grad_fn=<NllLossBackward>)
[2660] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[2661] Train Loss:    tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[2662] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[2663] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[2664] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[2665] Train Loss:    tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[2666] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[2667] Train Loss:    tensor(0.8276, device='cuda:0', grad_fn=<NllLossBackward>)
[2668] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[2669] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[2670] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[2671] Train Loss:    tensor(0.8106, device='cuda:0', grad_fn=<NllLossBackward>)
[2672] Train Loss:    tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[2673] Train Loss:    tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[2674] Train Loss:    tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[2675] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[2676] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[2677] Train Loss:    tensor(0.8170, device='cuda:0', grad_fn=<NllLossBackward>)
[2678] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[2679] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[2680] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[2681] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[2682] Train Loss:    tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[2683] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[2684] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[2685] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[2686] Train Loss:    tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[2687] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[2688] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[2689] Train Loss:    tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[2690] Train Loss:    tensor(0.8246, device='cuda:0', grad_fn=<NllLossBackward>)
[2691] Train Loss:    tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[2692] Train Loss:    tensor(0.8071, device='cuda:0', grad_fn=<NllLossBackward>)
[2693] Train Loss:    tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[2694] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[2695] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[2696] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[2697] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[2698] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2699] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[2700] Train Loss:    tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[2701] Train Loss:    tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[2702] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[2703] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[2704] Train Loss:    tensor(0.7898, device='cuda:0', grad_fn=<NllLossBackward>)
[2705] Train Loss:    tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[2706] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[2707] Train Loss:    tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[2708] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[2709] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[2710] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[2711] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2712] Train Loss:    tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[2713] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[2714] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[2715] Train Loss:    tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[2716] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[2717] Train Loss:    tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
[2718] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[2719] Train Loss:    tensor(0.8503, device='cuda:0', grad_fn=<NllLossBackward>)
[2720] Train Loss:    tensor(0.8087, device='cuda:0', grad_fn=<NllLossBackward>)
[2721] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[2722] Train Loss:    tensor(0.8252, device='cuda:0', grad_fn=<NllLossBackward>)
[2723] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[2724] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[2725] Train Loss:    tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[2726] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[2727] Train Loss:    tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[2728] Train Loss:    tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[2729] Train Loss:    tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[2730] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[2731] Train Loss:    tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[2732] Train Loss:    tensor(0.8020, device='cuda:0', grad_fn=<NllLossBackward>)
[2733] Train Loss:    tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[2734] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[2735] Train Loss:    tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[2736] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[2737] Train Loss:    tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[2738] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[2739] Train Loss:    tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[2740] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[2741] Train Loss:    tensor(0.7816, device='cuda:0', grad_fn=<NllLossBackward>)
[2742] Train Loss:    tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[2743] Train Loss:    tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[2744] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[2745] Train Loss:    tensor(0.8331, device='cuda:0', grad_fn=<NllLossBackward>)
[2746] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2747] Train Loss:   tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[2748] Train Loss:   tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[2749] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[2750] Train Loss:   tensor(0.8464, device='cuda:0', grad_fn=<NllLossBackward>)
[2751] Train Loss:   tensor(0.7844, device='cuda:0', grad_fn=<NllLossBackward>)
[2752] Train Loss:   tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[2753] Train Loss:   tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[2754] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[2755] Train Loss:   tensor(0.8164, device='cuda:0', grad_fn=<NllLossBackward>)
[2756] Train Loss:   tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[2757] Train Loss:   tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[2758] Train Loss:   tensor(0.8192, device='cuda:0', grad_fn=<NllLossBackward>)
[2759] Train Loss:   tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[2760] Train Loss:   tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[2761] Train Loss:   tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[2762] Train Loss:   tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[2763] Train Loss:   tensor(0.8330, device='cuda:0', grad_fn=<NllLossBackward>)
[2764] Train Loss:   tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[2765] Train Loss:   tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[2766] Train Loss:   tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[2767] Train Loss:   tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
[2768] Train Loss:   tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[2769] Train Loss:   tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[2770] Train Loss:   tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[2771] Train Loss:   tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[2772] Train Loss:   tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[2773] Train Loss:   tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[2774] Train Loss:   tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[2775] Train Loss:   tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
[2776] Train Loss:   tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[2777] Train Loss:   tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[2778] Train Loss:   tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[2779] Train Loss:   tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[2780] Train Loss:   tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[2781] Train Loss:   tensor(0.7865, device='cuda:0', grad_fn=<NllLossBackward>)
[2782] Train Loss:   tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
[2783] Train Loss:   tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[2784] Train Loss:   tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[2785] Train Loss:   tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[2786] Train Loss:   tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[2787] Train Loss:   tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[2788] Train Loss:   tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[2789] Train Loss:   tensor(0.8044, device='cuda:0', grad_fn=<NllLossBackward>)
[2790] Train Loss:   tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[2791] Train Loss:   tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[2792] Train Loss:   tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[2793] Train Loss:   tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[2794] Train Loss:   tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2795] Train Loss:    tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[2796] Train Loss:    tensor(0.8044, device='cuda:0', grad_fn=<NllLossBackward>)
[2797] Train Loss:    tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[2798] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[2799] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[2800] Train Loss:    tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[2801] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[2802] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[2803] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[2804] Train Loss:    tensor(0.8030, device='cuda:0', grad_fn=<NllLossBackward>)
[2805] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[2806] Train Loss:    tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[2807] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[2808] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[2809] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[2810] Train Loss:    tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[2811] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[2812] Train Loss:    tensor(0.8070, device='cuda:0', grad_fn=<NllLossBackward>)
[2813] Train Loss:    tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[2814] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[2815] Train Loss:    tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[2816] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[2817] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[2818] Train Loss:    tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
[2819] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[2820] Train Loss:    tensor(0.7970, device='cuda:0', grad_fn=<NllLossBackward>)
[2821] Train Loss:    tensor(0.8267, device='cuda:0', grad_fn=<NllLossBackward>)
[2822] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[2823] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[2824] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[2825] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[2826] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[2827] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[2828] Train Loss:    tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[2829] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[2830] Train Loss:    tensor(0.8324, device='cuda:0', grad_fn=<NllLossBackward>)
[2831] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[2832] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[2833] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[2834] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[2835] Train Loss:    tensor(0.8028, device='cuda:0', grad_fn=<NllLossBackward>)
[2836] Train Loss:    tensor(0.8197, device='cuda:0', grad_fn=<NllLossBackward>)
[2837] Train Loss:    tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[2838] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[2839] Train Loss:    tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[2840] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[2841] Train Loss:    tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[2842] Train Loss:    tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2843] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[2844] Train Loss:    tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[2845] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[2846] Train Loss:    tensor(0.8298, device='cuda:0', grad_fn=<NllLossBackward>)
[2847] Train Loss:    tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[2848] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[2849] Train Loss:    tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[2850] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[2851] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[2852] Train Loss:    tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[2853] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[2854] Train Loss:    tensor(0.8020, device='cuda:0', grad_fn=<NllLossBackward>)
[2855] Train Loss:    tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[2856] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[2857] Train Loss:    tensor(0.8055, device='cuda:0', grad_fn=<NllLossBackward>)
[2858] Train Loss:    tensor(0.8191, device='cuda:0', grad_fn=<NllLossBackward>)
[2859] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[2860] Train Loss:    tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[2861] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[2862] Train Loss:    tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[2863] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[2864] Train Loss:    tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[2865] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[2866] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[2867] Train Loss:    tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[2868] Train Loss:    tensor(0.8157, device='cuda:0', grad_fn=<NllLossBackward>)
[2869] Train Loss:    tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[2870] Train Loss:    tensor(0.8236, device='cuda:0', grad_fn=<NllLossBackward>)
[2871] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[2872] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[2873] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[2874] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[2875] Train Loss:    tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[2876] Train Loss:    tensor(0.8332, device='cuda:0', grad_fn=<NllLossBackward>)
[2877] Train Loss:    tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[2878] Train Loss:    tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[2879] Train Loss:    tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[2880] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[2881] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[2882] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[2883] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[2884] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[2885] Train Loss:    tensor(0.8012, device='cuda:0', grad_fn=<NllLossBackward>)
[2886] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[2887] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[2888] Train Loss:    tensor(0.8088, device='cuda:0', grad_fn=<NllLossBackward>)
[2889] Train Loss:    tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[2890] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2891] Train Loss:    tensor(0.8179, device='cuda:0', grad_fn=<NllLossBackward>)
[2892] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[2893] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[2894] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[2895] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[2896] Train Loss:    tensor(0.8033, device='cuda:0', grad_fn=<NllLossBackward>)
[2897] Train Loss:    tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[2898] Train Loss:    tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[2899] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[2900] Train Loss:    tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[2901] Train Loss:    tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[2902] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[2903] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[2904] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[2905] Train Loss:    tensor(0.8356, device='cuda:0', grad_fn=<NllLossBackward>)
[2906] Train Loss:    tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[2907] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[2908] Train Loss:    tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[2909] Train Loss:    tensor(0.8227, device='cuda:0', grad_fn=<NllLossBackward>)
[2910] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[2911] Train Loss:    tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[2912] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[2913] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[2914] Train Loss:    tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[2915] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[2916] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[2917] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[2918] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[2919] Train Loss:    tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[2920] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[2921] Train Loss:    tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[2922] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[2923] Train Loss:    tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[2924] Train Loss:    tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[2925] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[2926] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[2927] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[2928] Train Loss:    tensor(0.8163, device='cuda:0', grad_fn=<NllLossBackward>)
[2929] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[2930] Train Loss:    tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[2931] Train Loss:    tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[2932] Train Loss:    tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[2933] Train Loss:    tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[2934] Train Loss:    tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[2935] Train Loss:    tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[2936] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[2937] Train Loss:    tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[2938] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2939] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[2940] Train Loss:    tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[2941] Train Loss:    tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[2942] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[2943] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[2944] Train Loss:    tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[2945] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[2946] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[2947] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[2948] Train Loss:    tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[2949] Train Loss:    tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[2950] Train Loss:    tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[2951] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[2952] Train Loss:    tensor(0.7970, device='cuda:0', grad_fn=<NllLossBackward>)
[2953] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[2954] Train Loss:    tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[2955] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[2956] Train Loss:    tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[2957] Train Loss:    tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[2958] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[2959] Train Loss:    tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[2960] Train Loss:    tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[2961] Train Loss:    tensor(0.7859, device='cuda:0', grad_fn=<NllLossBackward>)
[2962] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[2963] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[2964] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[2965] Train Loss:    tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[2966] Train Loss:    tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[2967] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[2968] Train Loss:    tensor(0.7784, device='cuda:0', grad_fn=<NllLossBackward>)
[2969] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[2970] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[2971] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[2972] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[2973] Train Loss:    tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[2974] Train Loss:    tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[2975] Train Loss:    tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[2976] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[2977] Train Loss:    tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[2978] Train Loss:    tensor(0.8060, device='cuda:0', grad_fn=<NllLossBackward>)
[2979] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[2980] Train Loss:    tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[2981] Train Loss:    tensor(0.8025, device='cuda:0', grad_fn=<NllLossBackward>)
[2982] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[2983] Train Loss:    tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[2984] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[2985] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[2986] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2987] Train Loss:  tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[2988] Train Loss:  tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[2989] Train Loss:  tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[2990] Train Loss:  tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[2991] Train Loss:  tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[2992] Train Loss:  tensor(0.8168, device='cuda:0', grad_fn=<NllLossBackward>)
[2993] Train Loss:  tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[2994] Train Loss:  tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[2995] Train Loss:  tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[2996] Train Loss:  tensor(0.8379, device='cuda:0', grad_fn=<NllLossBackward>)
[2997] Train Loss:  tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[2998] Train Loss:  tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[2999] Train Loss:  tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[3000] Train Loss:  tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3001] Train Loss:  tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[3002] Train Loss:  tensor(0.8154, device='cuda:0', grad_fn=<NllLossBackward>)
[3003] Train Loss:  tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[3004] Train Loss:  tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3005] Train Loss:  tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[3006] Train Loss:  tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[3007] Train Loss:  tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[3008] Train Loss:  tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[3009] Train Loss:  tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[3010] Train Loss:  tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[3011] Train Loss:  tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[3012] Train Loss:  tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[3013] Train Loss:  tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[3014] Train Loss:  tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[3015] Train Loss:  tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[3016] Train Loss:  tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[3017] Train Loss:  tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[3018] Train Loss:  tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[3019] Train Loss:  tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[3020] Train Loss:  tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[3021] Train Loss:  tensor(0.8273, device='cuda:0', grad_fn=<NllLossBackward>)
[3022] Train Loss:  tensor(0.7875, device='cuda:0', grad_fn=<NllLossBackward>)
[3023] Train Loss:  tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3024] Train Loss:  tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3025] Train Loss:  tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[3026] Train Loss:  tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[3027] Train Loss:  tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[3028] Train Loss:  tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[3029] Train Loss:  tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[3030] Train Loss:  tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[3031] Train Loss:  tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[3032] Train Loss:  tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[3033] Train Loss:  tensor(0.7980, device='cuda:0', grad_fn=<NllLossBackward>)
[3034] Train Loss:  tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3035] Train Loss:   tensor(0.8156, device='cuda:0', grad_fn=<NllLossBackward>)
[3036] Train Loss:   tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[3037] Train Loss:   tensor(0.7859, device='cuda:0', grad_fn=<NllLossBackward>)
[3038] Train Loss:   tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[3039] Train Loss:   tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[3040] Train Loss:   tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[3041] Train Loss:   tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[3042] Train Loss:   tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[3043] Train Loss:   tensor(0.8021, device='cuda:0', grad_fn=<NllLossBackward>)
[3044] Train Loss:   tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[3045] Train Loss:   tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[3046] Train Loss:   tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[3047] Train Loss:   tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[3048] Train Loss:   tensor(0.7909, device='cuda:0', grad_fn=<NllLossBackward>)
[3049] Train Loss:   tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[3050] Train Loss:   tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[3051] Train Loss:   tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[3052] Train Loss:   tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[3053] Train Loss:   tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[3054] Train Loss:   tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[3055] Train Loss:   tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[3056] Train Loss:   tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[3057] Train Loss:   tensor(0.8118, device='cuda:0', grad_fn=<NllLossBackward>)
[3058] Train Loss:   tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[3059] Train Loss:   tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[3060] Train Loss:   tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[3061] Train Loss:   tensor(0.8249, device='cuda:0', grad_fn=<NllLossBackward>)
[3062] Train Loss:   tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[3063] Train Loss:   tensor(0.8143, device='cuda:0', grad_fn=<NllLossBackward>)
[3064] Train Loss:   tensor(0.7969, device='cuda:0', grad_fn=<NllLossBackward>)
[3065] Train Loss:   tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[3066] Train Loss:   tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[3067] Train Loss:   tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[3068] Train Loss:   tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[3069] Train Loss:   tensor(0.8077, device='cuda:0', grad_fn=<NllLossBackward>)
[3070] Train Loss:   tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[3071] Train Loss:   tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[3072] Train Loss:   tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[3073] Train Loss:   tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[3074] Train Loss:   tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[3075] Train Loss:   tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[3076] Train Loss:   tensor(0.8075, device='cuda:0', grad_fn=<NllLossBackward>)
[3077] Train Loss:   tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[3078] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[3079] Train Loss:   tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[3080] Train Loss:   tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[3081] Train Loss:   tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[3082] Train Loss:   tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3083] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[3084] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[3085] Train Loss:    tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[3086] Train Loss:    tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[3087] Train Loss:    tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[3088] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[3089] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[3090] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[3091] Train Loss:    tensor(0.8305, device='cuda:0', grad_fn=<NllLossBackward>)
[3092] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[3093] Train Loss:    tensor(0.8083, device='cuda:0', grad_fn=<NllLossBackward>)
[3094] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[3095] Train Loss:    tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[3096] Train Loss:    tensor(0.8408, device='cuda:0', grad_fn=<NllLossBackward>)
[3097] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[3098] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[3099] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[3100] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3101] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[3102] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[3103] Train Loss:    tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[3104] Train Loss:    tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[3105] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[3106] Train Loss:    tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[3107] Train Loss:    tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[3108] Train Loss:    tensor(0.8275, device='cuda:0', grad_fn=<NllLossBackward>)
[3109] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3110] Train Loss:    tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[3111] Train Loss:    tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[3112] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[3113] Train Loss:    tensor(0.8451, device='cuda:0', grad_fn=<NllLossBackward>)
[3114] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[3115] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[3116] Train Loss:    tensor(0.8195, device='cuda:0', grad_fn=<NllLossBackward>)
[3117] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[3118] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[3119] Train Loss:    tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[3120] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[3121] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[3122] Train Loss:    tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[3123] Train Loss:    tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[3124] Train Loss:    tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[3125] Train Loss:    tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
[3126] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[3127] Train Loss:    tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[3128] Train Loss:    tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[3129] Train Loss:    tensor(0.8179, device='cuda:0', grad_fn=<NllLossBackward>)
[3130] Train Loss:    tensor(0.8086, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3131] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[3132] Train Loss:    tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[3133] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[3134] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[3135] Train Loss:    tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[3136] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[3137] Train Loss:    tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[3138] Train Loss:    tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[3139] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[3140] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[3141] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[3142] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[3143] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[3144] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[3145] Train Loss:    tensor(0.8085, device='cuda:0', grad_fn=<NllLossBackward>)
[3146] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3147] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[3148] Train Loss:    tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[3149] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3150] Train Loss:    tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[3151] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[3152] Train Loss:    tensor(0.7895, device='cuda:0', grad_fn=<NllLossBackward>)
[3153] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[3154] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[3155] Train Loss:    tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[3156] Train Loss:    tensor(0.8073, device='cuda:0', grad_fn=<NllLossBackward>)
[3157] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[3158] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[3159] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[3160] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[3161] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[3162] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[3163] Train Loss:    tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[3164] Train Loss:    tensor(0.7859, device='cuda:0', grad_fn=<NllLossBackward>)
[3165] Train Loss:    tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[3166] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[3167] Train Loss:    tensor(0.8393, device='cuda:0', grad_fn=<NllLossBackward>)
[3168] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[3169] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[3170] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[3171] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[3172] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[3173] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[3174] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[3175] Train Loss:    tensor(0.8174, device='cuda:0', grad_fn=<NllLossBackward>)
[3176] Train Loss:    tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[3177] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[3178] Train Loss:    tensor(0.8131, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3179] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[3180] Train Loss:    tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[3181] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[3182] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[3183] Train Loss:    tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[3184] Train Loss:    tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[3185] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[3186] Train Loss:    tensor(0.8107, device='cuda:0', grad_fn=<NllLossBackward>)
[3187] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[3188] Train Loss:    tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[3189] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[3190] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[3191] Train Loss:    tensor(0.7793, device='cuda:0', grad_fn=<NllLossBackward>)
[3192] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[3193] Train Loss:    tensor(0.7901, device='cuda:0', grad_fn=<NllLossBackward>)
[3194] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[3195] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[3196] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[3197] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[3198] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[3199] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[3200] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[3201] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[3202] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[3203] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[3204] Train Loss:    tensor(0.8065, device='cuda:0', grad_fn=<NllLossBackward>)
[3205] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[3206] Train Loss:    tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[3207] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[3208] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[3209] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[3210] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[3211] Train Loss:    tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[3212] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[3213] Train Loss:    tensor(0.7667, device='cuda:0', grad_fn=<NllLossBackward>)
[3214] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[3215] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3216] Train Loss:    tensor(0.7985, device='cuda:0', grad_fn=<NllLossBackward>)
[3217] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[3218] Train Loss:    tensor(0.8056, device='cuda:0', grad_fn=<NllLossBackward>)
[3219] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[3220] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[3221] Train Loss:    tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[3222] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[3223] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[3224] Train Loss:    tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
[3225] Train Loss:    tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[3226] Train Loss:    tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3227] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[3228] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[3229] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[3230] Train Loss:    tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[3231] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[3232] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[3233] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[3234] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[3235] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[3236] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[3237] Train Loss:    tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[3238] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[3239] Train Loss:    tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[3240] Train Loss:    tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[3241] Train Loss:    tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[3242] Train Loss:    tensor(0.8234, device='cuda:0', grad_fn=<NllLossBackward>)
[3243] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[3244] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[3245] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[3246] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[3247] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3248] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[3249] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[3250] Train Loss:    tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[3251] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[3252] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[3253] Train Loss:    tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[3254] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[3255] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[3256] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[3257] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[3258] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[3259] Train Loss:    tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[3260] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[3261] Train Loss:    tensor(0.8188, device='cuda:0', grad_fn=<NllLossBackward>)
[3262] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[3263] Train Loss:    tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[3264] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[3265] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[3266] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[3267] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[3268] Train Loss:    tensor(0.7909, device='cuda:0', grad_fn=<NllLossBackward>)
[3269] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[3270] Train Loss:    tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[3271] Train Loss:    tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[3272] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[3273] Train Loss:    tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[3274] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3275] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[3276] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[3277] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[3278] Train Loss:    tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[3279] Train Loss:    tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[3280] Train Loss:    tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[3281] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[3282] Train Loss:    tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[3283] Train Loss:    tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[3284] Train Loss:    tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[3285] Train Loss:    tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[3286] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[3287] Train Loss:    tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[3288] Train Loss:    tensor(0.8207, device='cuda:0', grad_fn=<NllLossBackward>)
[3289] Train Loss:    tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[3290] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[3291] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[3292] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[3293] Train Loss:    tensor(0.7969, device='cuda:0', grad_fn=<NllLossBackward>)
[3294] Train Loss:    tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[3295] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[3296] Train Loss:    tensor(0.7942, device='cuda:0', grad_fn=<NllLossBackward>)
[3297] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[3298] Train Loss:    tensor(0.7793, device='cuda:0', grad_fn=<NllLossBackward>)
[3299] Train Loss:    tensor(0.8295, device='cuda:0', grad_fn=<NllLossBackward>)
[3300] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[3301] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[3302] Train Loss:    tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[3303] Train Loss:    tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[3304] Train Loss:    tensor(0.8373, device='cuda:0', grad_fn=<NllLossBackward>)
[3305] Train Loss:    tensor(0.7934, device='cuda:0', grad_fn=<NllLossBackward>)
[3306] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[3307] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[3308] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[3309] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[3310] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[3311] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[3312] Train Loss:    tensor(0.8203, device='cuda:0', grad_fn=<NllLossBackward>)
[3313] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[3314] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[3315] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3316] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[3317] Train Loss:    tensor(0.8214, device='cuda:0', grad_fn=<NllLossBackward>)
[3318] Train Loss:    tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[3319] Train Loss:    tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[3320] Train Loss:    tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[3321] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3322] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3323] Train Loss:   tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[3324] Train Loss:   tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[3325] Train Loss:   tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[3326] Train Loss:   tensor(0.8103, device='cuda:0', grad_fn=<NllLossBackward>)
[3327] Train Loss:   tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[3328] Train Loss:   tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[3329] Train Loss:   tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[3330] Train Loss:   tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[3331] Train Loss:   tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[3332] Train Loss:   tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[3333] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[3334] Train Loss:   tensor(0.7801, device='cuda:0', grad_fn=<NllLossBackward>)
[3335] Train Loss:   tensor(0.8134, device='cuda:0', grad_fn=<NllLossBackward>)
[3336] Train Loss:   tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
[3337] Train Loss:   tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[3338] Train Loss:   tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[3339] Train Loss:   tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[3340] Train Loss:   tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[3341] Train Loss:   tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[3342] Train Loss:   tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[3343] Train Loss:   tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[3344] Train Loss:   tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[3345] Train Loss:   tensor(0.8004, device='cuda:0', grad_fn=<NllLossBackward>)
[3346] Train Loss:   tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[3347] Train Loss:   tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[3348] Train Loss:   tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[3349] Train Loss:   tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[3350] Train Loss:   tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[3351] Train Loss:   tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[3352] Train Loss:   tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[3353] Train Loss:   tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[3354] Train Loss:   tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[3355] Train Loss:   tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[3356] Train Loss:   tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[3357] Train Loss:   tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[3358] Train Loss:   tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[3359] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3360] Train Loss:   tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[3361] Train Loss:   tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[3362] Train Loss:   tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[3363] Train Loss:   tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[3364] Train Loss:   tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[3365] Train Loss:   tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3366] Train Loss:   tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[3367] Train Loss:   tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[3368] Train Loss:   tensor(0.8142, device='cuda:0', grad_fn=<NllLossBackward>)
[3369] Train Loss:   tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[3370] Train Loss:   tensor(0.8005, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3371] Train Loss:    tensor(0.8084, device='cuda:0', grad_fn=<NllLossBackward>)
[3372] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[3373] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[3374] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[3375] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[3376] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[3377] Train Loss:    tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[3378] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[3379] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[3380] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[3381] Train Loss:    tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[3382] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[3383] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[3384] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[3385] Train Loss:    tensor(0.8025, device='cuda:0', grad_fn=<NllLossBackward>)
[3386] Train Loss:    tensor(0.7838, device='cuda:0', grad_fn=<NllLossBackward>)
[3387] Train Loss:    tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[3388] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[3389] Train Loss:    tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[3390] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[3391] Train Loss:    tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[3392] Train Loss:    tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[3393] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[3394] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[3395] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[3396] Train Loss:    tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
[3397] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[3398] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[3399] Train Loss:    tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[3400] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[3401] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[3402] Train Loss:    tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[3403] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[3404] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[3405] Train Loss:    tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[3406] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[3407] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[3408] Train Loss:    tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[3409] Train Loss:    tensor(0.8155, device='cuda:0', grad_fn=<NllLossBackward>)
[3410] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[3411] Train Loss:    tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[3412] Train Loss:    tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[3413] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[3414] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[3415] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[3416] Train Loss:    tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[3417] Train Loss:    tensor(0.7865, device='cuda:0', grad_fn=<NllLossBackward>)
[3418] Train Loss:    tensor(0.7816, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3419] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[3420] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[3421] Train Loss:    tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[3422] Train Loss:    tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[3423] Train Loss:    tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[3424] Train Loss:    tensor(0.8087, device='cuda:0', grad_fn=<NllLossBackward>)
[3425] Train Loss:    tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[3426] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[3427] Train Loss:    tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[3428] Train Loss:    tensor(0.8050, device='cuda:0', grad_fn=<NllLossBackward>)
[3429] Train Loss:    tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[3430] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[3431] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[3432] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[3433] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3434] Train Loss:    tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[3435] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[3436] Train Loss:    tensor(0.8230, device='cuda:0', grad_fn=<NllLossBackward>)
[3437] Train Loss:    tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[3438] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[3439] Train Loss:    tensor(0.7931, device='cuda:0', grad_fn=<NllLossBackward>)
[3440] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[3441] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[3442] Train Loss:    tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[3443] Train Loss:    tensor(0.8132, device='cuda:0', grad_fn=<NllLossBackward>)
[3444] Train Loss:    tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[3445] Train Loss:    tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[3446] Train Loss:    tensor(0.8257, device='cuda:0', grad_fn=<NllLossBackward>)
[3447] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[3448] Train Loss:    tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[3449] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[3450] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[3451] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[3452] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[3453] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[3454] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3455] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[3456] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[3457] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[3458] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[3459] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[3460] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[3461] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[3462] Train Loss:    tensor(0.8161, device='cuda:0', grad_fn=<NllLossBackward>)
[3463] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[3464] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[3465] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[3466] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3467] Train Loss:    tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[3468] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[3469] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[3470] Train Loss:    tensor(0.8101, device='cuda:0', grad_fn=<NllLossBackward>)
[3471] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[3472] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[3473] Train Loss:    tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[3474] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[3475] Train Loss:    tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[3476] Train Loss:    tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[3477] Train Loss:    tensor(0.8152, device='cuda:0', grad_fn=<NllLossBackward>)
[3478] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[3479] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[3480] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3481] Train Loss:    tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[3482] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[3483] Train Loss:    tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[3484] Train Loss:    tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[3485] Train Loss:    tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[3486] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[3487] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[3488] Train Loss:    tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[3489] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[3490] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[3491] Train Loss:    tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[3492] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[3493] Train Loss:    tensor(0.8174, device='cuda:0', grad_fn=<NllLossBackward>)
[3494] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[3495] Train Loss:    tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[3496] Train Loss:    tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[3497] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[3498] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3499] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[3500] Train Loss:    tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[3501] Train Loss:    tensor(0.7948, device='cuda:0', grad_fn=<NllLossBackward>)
[3502] Train Loss:    tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[3503] Train Loss:    tensor(0.7888, device='cuda:0', grad_fn=<NllLossBackward>)
[3504] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[3505] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[3506] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[3507] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[3508] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[3509] Train Loss:    tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[3510] Train Loss:    tensor(0.7823, device='cuda:0', grad_fn=<NllLossBackward>)
[3511] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[3512] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[3513] Train Loss:    tensor(0.8139, device='cuda:0', grad_fn=<NllLossBackward>)
[3514] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3515] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[3516] Train Loss:    tensor(0.8166, device='cuda:0', grad_fn=<NllLossBackward>)
[3517] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[3518] Train Loss:    tensor(0.8030, device='cuda:0', grad_fn=<NllLossBackward>)
[3519] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[3520] Train Loss:    tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[3521] Train Loss:    tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[3522] Train Loss:    tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[3523] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[3524] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[3525] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[3526] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[3527] Train Loss:    tensor(0.8136, device='cuda:0', grad_fn=<NllLossBackward>)
[3528] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[3529] Train Loss:    tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[3530] Train Loss:    tensor(0.8157, device='cuda:0', grad_fn=<NllLossBackward>)
[3531] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[3532] Train Loss:    tensor(0.8037, device='cuda:0', grad_fn=<NllLossBackward>)
[3533] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[3534] Train Loss:    tensor(0.8138, device='cuda:0', grad_fn=<NllLossBackward>)
[3535] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[3536] Train Loss:    tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[3537] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[3538] Train Loss:    tensor(0.7986, device='cuda:0', grad_fn=<NllLossBackward>)
[3539] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[3540] Train Loss:    tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[3541] Train Loss:    tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[3542] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[3543] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[3544] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[3545] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[3546] Train Loss:    tensor(0.8205, device='cuda:0', grad_fn=<NllLossBackward>)
[3547] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[3548] Train Loss:    tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[3549] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[3550] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[3551] Train Loss:    tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[3552] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[3553] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[3554] Train Loss:    tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[3555] Train Loss:    tensor(0.8112, device='cuda:0', grad_fn=<NllLossBackward>)
[3556] Train Loss:    tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[3557] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[3558] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[3559] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[3560] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[3561] Train Loss:    tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[3562] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3563] Train Loss:   tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[3564] Train Loss:   tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[3565] Train Loss:   tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[3566] Train Loss:   tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[3567] Train Loss:   tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[3568] Train Loss:   tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[3569] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[3570] Train Loss:   tensor(0.8030, device='cuda:0', grad_fn=<NllLossBackward>)
[3571] Train Loss:   tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[3572] Train Loss:   tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[3573] Train Loss:   tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[3574] Train Loss:   tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[3575] Train Loss:   tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[3576] Train Loss:   tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[3577] Train Loss:   tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[3578] Train Loss:   tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[3579] Train Loss:   tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[3580] Train Loss:   tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[3581] Train Loss:   tensor(0.8211, device='cuda:0', grad_fn=<NllLossBackward>)
[3582] Train Loss:   tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[3583] Train Loss:   tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[3584] Train Loss:   tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[3585] Train Loss:   tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[3586] Train Loss:   tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[3587] Train Loss:   tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[3588] Train Loss:   tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[3589] Train Loss:   tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[3590] Train Loss:   tensor(0.8339, device='cuda:0', grad_fn=<NllLossBackward>)
[3591] Train Loss:   tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[3592] Train Loss:   tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3593] Train Loss:   tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[3594] Train Loss:   tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[3595] Train Loss:   tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[3596] Train Loss:   tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[3597] Train Loss:   tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[3598] Train Loss:   tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[3599] Train Loss:   tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[3600] Train Loss:   tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[3601] Train Loss:   tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[3602] Train Loss:   tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[3603] Train Loss:   tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3604] Train Loss:   tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[3605] Train Loss:   tensor(0.8162, device='cuda:0', grad_fn=<NllLossBackward>)
[3606] Train Loss:   tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[3607] Train Loss:   tensor(0.8302, device='cuda:0', grad_fn=<NllLossBackward>)
[3608] Train Loss:   tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[3609] Train Loss:   tensor(0.8072, device='cuda:0', grad_fn=<NllLossBackward>)
[3610] Train Loss:   tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3611] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[3612] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[3613] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[3614] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[3615] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[3616] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[3617] Train Loss:    tensor(0.8260, device='cuda:0', grad_fn=<NllLossBackward>)
[3618] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[3619] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[3620] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[3621] Train Loss:    tensor(0.8148, device='cuda:0', grad_fn=<NllLossBackward>)
[3622] Train Loss:    tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[3623] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[3624] Train Loss:    tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
[3625] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[3626] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[3627] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[3628] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[3629] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[3630] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[3631] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[3632] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[3633] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[3634] Train Loss:    tensor(0.8147, device='cuda:0', grad_fn=<NllLossBackward>)
[3635] Train Loss:    tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[3636] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[3637] Train Loss:    tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[3638] Train Loss:    tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[3639] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[3640] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[3641] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[3642] Train Loss:    tensor(0.7731, device='cuda:0', grad_fn=<NllLossBackward>)
[3643] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[3644] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[3645] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[3646] Train Loss:    tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[3647] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[3648] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[3649] Train Loss:    tensor(0.8004, device='cuda:0', grad_fn=<NllLossBackward>)
[3650] Train Loss:    tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[3651] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[3652] Train Loss:    tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[3653] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[3654] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[3655] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[3656] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[3657] Train Loss:    tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[3658] Train Loss:    tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3659] Train Loss:    tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[3660] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[3661] Train Loss:    tensor(0.8240, device='cuda:0', grad_fn=<NllLossBackward>)
[3662] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[3663] Train Loss:    tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[3664] Train Loss:    tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[3665] Train Loss:    tensor(0.8220, device='cuda:0', grad_fn=<NllLossBackward>)
[3666] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[3667] Train Loss:    tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[3668] Train Loss:    tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[3669] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[3670] Train Loss:    tensor(0.8139, device='cuda:0', grad_fn=<NllLossBackward>)
[3671] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[3672] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[3673] Train Loss:    tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[3674] Train Loss:    tensor(0.8277, device='cuda:0', grad_fn=<NllLossBackward>)
[3675] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[3676] Train Loss:    tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[3677] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[3678] Train Loss:    tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[3679] Train Loss:    tensor(0.8139, device='cuda:0', grad_fn=<NllLossBackward>)
[3680] Train Loss:    tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[3681] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[3682] Train Loss:    tensor(0.8150, device='cuda:0', grad_fn=<NllLossBackward>)
[3683] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[3684] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[3685] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[3686] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[3687] Train Loss:    tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[3688] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[3689] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[3690] Train Loss:    tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[3691] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[3692] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[3693] Train Loss:    tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[3694] Train Loss:    tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[3695] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[3696] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[3697] Train Loss:    tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[3698] Train Loss:    tensor(0.7996, device='cuda:0', grad_fn=<NllLossBackward>)
[3699] Train Loss:    tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[3700] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[3701] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[3702] Train Loss:    tensor(0.8071, device='cuda:0', grad_fn=<NllLossBackward>)
[3703] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[3704] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[3705] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[3706] Train Loss:    tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3707] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[3708] Train Loss:    tensor(0.8135, device='cuda:0', grad_fn=<NllLossBackward>)
[3709] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[3710] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[3711] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3712] Train Loss:    tensor(0.8259, device='cuda:0', grad_fn=<NllLossBackward>)
[3713] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[3714] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[3715] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[3716] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[3717] Train Loss:    tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[3718] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[3719] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[3720] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[3721] Train Loss:    tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[3722] Train Loss:    tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[3723] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[3724] Train Loss:    tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[3725] Train Loss:    tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[3726] Train Loss:    tensor(0.8160, device='cuda:0', grad_fn=<NllLossBackward>)
[3727] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[3728] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[3729] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[3730] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[3731] Train Loss:    tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[3732] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[3733] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[3734] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[3735] Train Loss:    tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[3736] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3737] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[3738] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[3739] Train Loss:    tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[3740] Train Loss:    tensor(0.8341, device='cuda:0', grad_fn=<NllLossBackward>)
[3741] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[3742] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[3743] Train Loss:    tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[3744] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[3745] Train Loss:    tensor(0.7793, device='cuda:0', grad_fn=<NllLossBackward>)
[3746] Train Loss:    tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[3747] Train Loss:    tensor(0.7909, device='cuda:0', grad_fn=<NllLossBackward>)
[3748] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[3749] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[3750] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3751] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[3752] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[3753] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[3754] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3755] Train Loss:   tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[3756] Train Loss:   tensor(0.8133, device='cuda:0', grad_fn=<NllLossBackward>)
[3757] Train Loss:   tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3758] Train Loss:   tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[3759] Train Loss:   tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[3760] Train Loss:   tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[3761] Train Loss:   tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[3762] Train Loss:   tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[3763] Train Loss:   tensor(0.7838, device='cuda:0', grad_fn=<NllLossBackward>)
[3764] Train Loss:   tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[3765] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3766] Train Loss:   tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[3767] Train Loss:   tensor(0.8009, device='cuda:0', grad_fn=<NllLossBackward>)
[3768] Train Loss:   tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[3769] Train Loss:   tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[3770] Train Loss:   tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[3771] Train Loss:   tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[3772] Train Loss:   tensor(0.7784, device='cuda:0', grad_fn=<NllLossBackward>)
[3773] Train Loss:   tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[3774] Train Loss:   tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[3775] Train Loss:   tensor(0.7971, device='cuda:0', grad_fn=<NllLossBackward>)
[3776] Train Loss:   tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[3777] Train Loss:   tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[3778] Train Loss:   tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[3779] Train Loss:   tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[3780] Train Loss:   tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[3781] Train Loss:   tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3782] Train Loss:   tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[3783] Train Loss:   tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[3784] Train Loss:   tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[3785] Train Loss:   tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[3786] Train Loss:   tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[3787] Train Loss:   tensor(0.7996, device='cuda:0', grad_fn=<NllLossBackward>)
[3788] Train Loss:   tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[3789] Train Loss:   tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[3790] Train Loss:   tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[3791] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[3792] Train Loss:   tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3793] Train Loss:   tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[3794] Train Loss:   tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[3795] Train Loss:   tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[3796] Train Loss:   tensor(0.7854, device='cuda:0', grad_fn=<NllLossBackward>)
[3797] Train Loss:   tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[3798] Train Loss:   tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[3799] Train Loss:   tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[3800] Train Loss:   tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[3801] Train Loss:   tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[3802] Train Loss:   tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3803] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3804] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[3805] Train Loss:    tensor(0.8128, device='cuda:0', grad_fn=<NllLossBackward>)
[3806] Train Loss:    tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[3807] Train Loss:    tensor(0.7985, device='cuda:0', grad_fn=<NllLossBackward>)
[3808] Train Loss:    tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[3809] Train Loss:    tensor(0.8247, device='cuda:0', grad_fn=<NllLossBackward>)
[3810] Train Loss:    tensor(0.7730, device='cuda:0', grad_fn=<NllLossBackward>)
[3811] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[3812] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[3813] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[3814] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[3815] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[3816] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[3817] Train Loss:    tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[3818] Train Loss:    tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[3819] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[3820] Train Loss:    tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[3821] Train Loss:    tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[3822] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[3823] Train Loss:    tensor(0.8222, device='cuda:0', grad_fn=<NllLossBackward>)
[3824] Train Loss:    tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[3825] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[3826] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[3827] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[3828] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[3829] Train Loss:    tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[3830] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[3831] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[3832] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[3833] Train Loss:    tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[3834] Train Loss:    tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[3835] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[3836] Train Loss:    tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[3837] Train Loss:    tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[3838] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[3839] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[3840] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[3841] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[3842] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[3843] Train Loss:    tensor(0.8167, device='cuda:0', grad_fn=<NllLossBackward>)
[3844] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[3845] Train Loss:    tensor(0.8030, device='cuda:0', grad_fn=<NllLossBackward>)
[3846] Train Loss:    tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[3847] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[3848] Train Loss:    tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[3849] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[3850] Train Loss:    tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3851] Train Loss:   tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[3852] Train Loss:   tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[3853] Train Loss:   tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[3854] Train Loss:   tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[3855] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[3856] Train Loss:   tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
[3857] Train Loss:   tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[3858] Train Loss:   tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[3859] Train Loss:   tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[3860] Train Loss:   tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[3861] Train Loss:   tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[3862] Train Loss:   tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[3863] Train Loss:   tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[3864] Train Loss:   tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[3865] Train Loss:   tensor(0.8115, device='cuda:0', grad_fn=<NllLossBackward>)
[3866] Train Loss:   tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[3867] Train Loss:   tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[3868] Train Loss:   tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[3869] Train Loss:   tensor(0.7985, device='cuda:0', grad_fn=<NllLossBackward>)
[3870] Train Loss:   tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[3871] Train Loss:   tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[3872] Train Loss:   tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[3873] Train Loss:   tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[3874] Train Loss:   tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[3875] Train Loss:   tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[3876] Train Loss:   tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[3877] Train Loss:   tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[3878] Train Loss:   tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[3879] Train Loss:   tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[3880] Train Loss:   tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[3881] Train Loss:   tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[3882] Train Loss:   tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[3883] Train Loss:   tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[3884] Train Loss:   tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[3885] Train Loss:   tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[3886] Train Loss:   tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[3887] Train Loss:   tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[3888] Train Loss:   tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[3889] Train Loss:   tensor(0.8196, device='cuda:0', grad_fn=<NllLossBackward>)
[3890] Train Loss:   tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[3891] Train Loss:   tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[3892] Train Loss:   tensor(0.7986, device='cuda:0', grad_fn=<NllLossBackward>)
[3893] Train Loss:   tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[3894] Train Loss:   tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[3895] Train Loss:   tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[3896] Train Loss:   tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[3897] Train Loss:   tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[3898] Train Loss:   tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3899] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[3900] Train Loss:    tensor(0.8624, device='cuda:0', grad_fn=<NllLossBackward>)
[3901] Train Loss:    tensor(0.7934, device='cuda:0', grad_fn=<NllLossBackward>)
[3902] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3903] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[3904] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3905] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[3906] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[3907] Train Loss:    tensor(0.7854, device='cuda:0', grad_fn=<NllLossBackward>)
[3908] Train Loss:    tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[3909] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[3910] Train Loss:    tensor(0.7980, device='cuda:0', grad_fn=<NllLossBackward>)
[3911] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[3912] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[3913] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[3914] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[3915] Train Loss:    tensor(0.8187, device='cuda:0', grad_fn=<NllLossBackward>)
[3916] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3917] Train Loss:    tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[3918] Train Loss:    tensor(0.7956, device='cuda:0', grad_fn=<NllLossBackward>)
[3919] Train Loss:    tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
[3920] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[3921] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[3922] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[3923] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[3924] Train Loss:    tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[3925] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[3926] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[3927] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[3928] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[3929] Train Loss:    tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[3930] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[3931] Train Loss:    tensor(0.8218, device='cuda:0', grad_fn=<NllLossBackward>)
[3932] Train Loss:    tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[3933] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[3934] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[3935] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3936] Train Loss:    tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
[3937] Train Loss:    tensor(0.8004, device='cuda:0', grad_fn=<NllLossBackward>)
[3938] Train Loss:    tensor(0.7969, device='cuda:0', grad_fn=<NllLossBackward>)
[3939] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3940] Train Loss:    tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[3941] Train Loss:    tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[3942] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[3943] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[3944] Train Loss:    tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[3945] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[3946] Train Loss:    tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3947] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[3948] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[3949] Train Loss:    tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[3950] Train Loss:    tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[3951] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[3952] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[3953] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[3954] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[3955] Train Loss:    tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[3956] Train Loss:    tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[3957] Train Loss:    tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[3958] Train Loss:    tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[3959] Train Loss:    tensor(0.8042, device='cuda:0', grad_fn=<NllLossBackward>)
[3960] Train Loss:    tensor(0.8104, device='cuda:0', grad_fn=<NllLossBackward>)
[3961] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[3962] Train Loss:    tensor(0.8327, device='cuda:0', grad_fn=<NllLossBackward>)
[3963] Train Loss:    tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[3964] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[3965] Train Loss:    tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[3966] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[3967] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[3968] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[3969] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[3970] Train Loss:    tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[3971] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[3972] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[3973] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[3974] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[3975] Train Loss:    tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[3976] Train Loss:    tensor(0.8242, device='cuda:0', grad_fn=<NllLossBackward>)
[3977] Train Loss:    tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[3978] Train Loss:    tensor(0.8011, device='cuda:0', grad_fn=<NllLossBackward>)
[3979] Train Loss:    tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[3980] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[3981] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[3982] Train Loss:    tensor(0.8104, device='cuda:0', grad_fn=<NllLossBackward>)
[3983] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[3984] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[3985] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[3986] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3987] Train Loss:    tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[3988] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[3989] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[3990] Train Loss:    tensor(0.8054, device='cuda:0', grad_fn=<NllLossBackward>)
[3991] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[3992] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[3993] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[3994] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3995] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[3996] Train Loss:    tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[3997] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[3998] Train Loss:    tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[3999] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[4000] Train Loss:    tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[4001] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[4002] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[4003] Train Loss:    tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[4004] Train Loss:    tensor(0.7898, device='cuda:0', grad_fn=<NllLossBackward>)
[4005] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[4006] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[4007] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[4008] Train Loss:    tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[4009] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[4010] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[4011] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[4012] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[4013] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[4014] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[4015] Train Loss:    tensor(0.7996, device='cuda:0', grad_fn=<NllLossBackward>)
[4016] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[4017] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[4018] Train Loss:    tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[4019] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[4020] Train Loss:    tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[4021] Train Loss:    tensor(0.8018, device='cuda:0', grad_fn=<NllLossBackward>)
[4022] Train Loss:    tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[4023] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[4024] Train Loss:    tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[4025] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[4026] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[4027] Train Loss:    tensor(0.8052, device='cuda:0', grad_fn=<NllLossBackward>)
[4028] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[4029] Train Loss:    tensor(0.7854, device='cuda:0', grad_fn=<NllLossBackward>)
[4030] Train Loss:    tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[4031] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[4032] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[4033] Train Loss:    tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[4034] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[4035] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[4036] Train Loss:    tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[4037] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[4038] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[4039] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[4040] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[4041] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[4042] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4043] Train Loss:    tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[4044] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[4045] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[4046] Train Loss:    tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[4047] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[4048] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[4049] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[4050] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[4051] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[4052] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[4053] Train Loss:    tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[4054] Train Loss:    tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[4055] Train Loss:    tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[4056] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[4057] Train Loss:    tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[4058] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[4059] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[4060] Train Loss:    tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[4061] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[4062] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[4063] Train Loss:    tensor(0.8226, device='cuda:0', grad_fn=<NllLossBackward>)
[4064] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[4065] Train Loss:    tensor(0.8162, device='cuda:0', grad_fn=<NllLossBackward>)
[4066] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[4067] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[4068] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[4069] Train Loss:    tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[4070] Train Loss:    tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[4071] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[4072] Train Loss:    tensor(0.8369, device='cuda:0', grad_fn=<NllLossBackward>)
[4073] Train Loss:    tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[4074] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[4075] Train Loss:    tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[4076] Train Loss:    tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[4077] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[4078] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[4079] Train Loss:    tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[4080] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[4081] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[4082] Train Loss:    tensor(0.8074, device='cuda:0', grad_fn=<NllLossBackward>)
[4083] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[4084] Train Loss:    tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[4085] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[4086] Train Loss:    tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[4087] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[4088] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[4089] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[4090] Train Loss:    tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4091] Train Loss:   tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[4092] Train Loss:   tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[4093] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[4094] Train Loss:   tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[4095] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[4096] Train Loss:   tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[4097] Train Loss:   tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[4098] Train Loss:   tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[4099] Train Loss:   tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[4100] Train Loss:   tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[4101] Train Loss:   tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[4102] Train Loss:   tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[4103] Train Loss:   tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[4104] Train Loss:   tensor(0.7945, device='cuda:0', grad_fn=<NllLossBackward>)
[4105] Train Loss:   tensor(0.8065, device='cuda:0', grad_fn=<NllLossBackward>)
[4106] Train Loss:   tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[4107] Train Loss:   tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[4108] Train Loss:   tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[4109] Train Loss:   tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[4110] Train Loss:   tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[4111] Train Loss:   tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[4112] Train Loss:   tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[4113] Train Loss:   tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[4114] Train Loss:   tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[4115] Train Loss:   tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[4116] Train Loss:   tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[4117] Train Loss:   tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[4118] Train Loss:   tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[4119] Train Loss:   tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[4120] Train Loss:   tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[4121] Train Loss:   tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[4122] Train Loss:   tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[4123] Train Loss:   tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[4124] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[4125] Train Loss:   tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[4126] Train Loss:   tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[4127] Train Loss:   tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[4128] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[4129] Train Loss:   tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[4130] Train Loss:   tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
[4131] Train Loss:   tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[4132] Train Loss:   tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[4133] Train Loss:   tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[4134] Train Loss:   tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[4135] Train Loss:   tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[4136] Train Loss:   tensor(0.7895, device='cuda:0', grad_fn=<NllLossBackward>)
[4137] Train Loss:   tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[4138] Train Loss:   tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4139] Train Loss:    tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[4140] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[4141] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[4142] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[4143] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[4144] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[4145] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[4146] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[4147] Train Loss:    tensor(0.7858, device='cuda:0', grad_fn=<NllLossBackward>)
[4148] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[4149] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[4150] Train Loss:    tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[4151] Train Loss:    tensor(0.7801, device='cuda:0', grad_fn=<NllLossBackward>)
[4152] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[4153] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[4154] Train Loss:    tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[4155] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[4156] Train Loss:    tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[4157] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[4158] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[4159] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[4160] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[4161] Train Loss:    tensor(0.7944, device='cuda:0', grad_fn=<NllLossBackward>)
[4162] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[4163] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[4164] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[4165] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[4166] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[4167] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[4168] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[4169] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[4170] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[4171] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[4172] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[4173] Train Loss:    tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[4174] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[4175] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[4176] Train Loss:    tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[4177] Train Loss:    tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[4178] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[4179] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[4180] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[4181] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[4182] Train Loss:    tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[4183] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[4184] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[4185] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[4186] Train Loss:    tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4187] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[4188] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[4189] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[4190] Train Loss:    tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[4191] Train Loss:    tensor(0.8091, device='cuda:0', grad_fn=<NllLossBackward>)
[4192] Train Loss:    tensor(0.8130, device='cuda:0', grad_fn=<NllLossBackward>)
[4193] Train Loss:    tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[4194] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[4195] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[4196] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[4197] Train Loss:    tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[4198] Train Loss:    tensor(0.8232, device='cuda:0', grad_fn=<NllLossBackward>)
[4199] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[4200] Train Loss:    tensor(0.8048, device='cuda:0', grad_fn=<NllLossBackward>)
[4201] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[4202] Train Loss:    tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[4203] Train Loss:    tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[4204] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[4205] Train Loss:    tensor(0.8284, device='cuda:0', grad_fn=<NllLossBackward>)
[4206] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[4207] Train Loss:    tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[4208] Train Loss:    tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[4209] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[4210] Train Loss:    tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[4211] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[4212] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[4213] Train Loss:    tensor(0.8064, device='cuda:0', grad_fn=<NllLossBackward>)
[4214] Train Loss:    tensor(0.7667, device='cuda:0', grad_fn=<NllLossBackward>)
[4215] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[4216] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[4217] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[4218] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[4219] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[4220] Train Loss:    tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[4221] Train Loss:    tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[4222] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[4223] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[4224] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[4225] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[4226] Train Loss:    tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[4227] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[4228] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[4229] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[4230] Train Loss:    tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[4231] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[4232] Train Loss:    tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[4233] Train Loss:    tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[4234] Train Loss:    tensor(0.7975, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4235] Train Loss:    tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[4236] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[4237] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[4238] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[4239] Train Loss:    tensor(0.8210, device='cuda:0', grad_fn=<NllLossBackward>)
[4240] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[4241] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[4242] Train Loss:    tensor(0.8097, device='cuda:0', grad_fn=<NllLossBackward>)
[4243] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[4244] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[4245] Train Loss:    tensor(0.8299, device='cuda:0', grad_fn=<NllLossBackward>)
[4246] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[4247] Train Loss:    tensor(0.8141, device='cuda:0', grad_fn=<NllLossBackward>)
[4248] Train Loss:    tensor(0.7982, device='cuda:0', grad_fn=<NllLossBackward>)
[4249] Train Loss:    tensor(0.8102, device='cuda:0', grad_fn=<NllLossBackward>)
[4250] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[4251] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[4252] Train Loss:    tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[4253] Train Loss:    tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[4254] Train Loss:    tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[4255] Train Loss:    tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[4256] Train Loss:    tensor(0.7793, device='cuda:0', grad_fn=<NllLossBackward>)
[4257] Train Loss:    tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[4258] Train Loss:    tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
[4259] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[4260] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[4261] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[4262] Train Loss:    tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[4263] Train Loss:    tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[4264] Train Loss:    tensor(0.7964, device='cuda:0', grad_fn=<NllLossBackward>)
[4265] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[4266] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[4267] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[4268] Train Loss:    tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[4269] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[4270] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[4271] Train Loss:    tensor(0.7875, device='cuda:0', grad_fn=<NllLossBackward>)
[4272] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[4273] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[4274] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[4275] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[4276] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[4277] Train Loss:    tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[4278] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[4279] Train Loss:    tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[4280] Train Loss:    tensor(0.7667, device='cuda:0', grad_fn=<NllLossBackward>)
[4281] Train Loss:    tensor(0.8166, device='cuda:0', grad_fn=<NllLossBackward>)
[4282] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4283] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[4284] Train Loss:    tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[4285] Train Loss:    tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[4286] Train Loss:    tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[4287] Train Loss:    tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[4288] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[4289] Train Loss:    tensor(0.7942, device='cuda:0', grad_fn=<NllLossBackward>)
[4290] Train Loss:    tensor(0.7978, device='cuda:0', grad_fn=<NllLossBackward>)
[4291] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[4292] Train Loss:    tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[4293] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[4294] Train Loss:    tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[4295] Train Loss:    tensor(0.8092, device='cuda:0', grad_fn=<NllLossBackward>)
[4296] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[4297] Train Loss:    tensor(0.8179, device='cuda:0', grad_fn=<NllLossBackward>)
[4298] Train Loss:    tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[4299] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[4300] Train Loss:    tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[4301] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[4302] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[4303] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[4304] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[4305] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[4306] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[4307] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[4308] Train Loss:    tensor(0.7895, device='cuda:0', grad_fn=<NllLossBackward>)
[4309] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[4310] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[4311] Train Loss:    tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[4312] Train Loss:    tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[4313] Train Loss:    tensor(0.8121, device='cuda:0', grad_fn=<NllLossBackward>)
[4314] Train Loss:    tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
[4315] Train Loss:    tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[4316] Train Loss:    tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[4317] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[4318] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[4319] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[4320] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[4321] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[4322] Train Loss:    tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[4323] Train Loss:    tensor(0.8125, device='cuda:0', grad_fn=<NllLossBackward>)
[4324] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[4325] Train Loss:    tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[4326] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[4327] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[4328] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[4329] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[4330] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4331] Train Loss:    tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[4332] Train Loss:    tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[4333] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[4334] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[4335] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[4336] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[4337] Train Loss:    tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[4338] Train Loss:    tensor(0.8044, device='cuda:0', grad_fn=<NllLossBackward>)
[4339] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[4340] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[4341] Train Loss:    tensor(0.8024, device='cuda:0', grad_fn=<NllLossBackward>)
[4342] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[4343] Train Loss:    tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[4344] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[4345] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[4346] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[4347] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[4348] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[4349] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[4350] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[4351] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[4352] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[4353] Train Loss:    tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[4354] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[4355] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[4356] Train Loss:    tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[4357] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[4358] Train Loss:    tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[4359] Train Loss:    tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[4360] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[4361] Train Loss:    tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[4362] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[4363] Train Loss:    tensor(0.7859, device='cuda:0', grad_fn=<NllLossBackward>)
[4364] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[4365] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[4366] Train Loss:    tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[4367] Train Loss:    tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[4368] Train Loss:    tensor(0.7582, device='cuda:0', grad_fn=<NllLossBackward>)
[4369] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[4370] Train Loss:    tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[4371] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[4372] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[4373] Train Loss:    tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[4374] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[4375] Train Loss:    tensor(0.8014, device='cuda:0', grad_fn=<NllLossBackward>)
[4376] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[4377] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[4378] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4379] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[4380] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[4381] Train Loss:    tensor(0.7980, device='cuda:0', grad_fn=<NllLossBackward>)
[4382] Train Loss:    tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[4383] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[4384] Train Loss:    tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[4385] Train Loss:    tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
[4386] Train Loss:    tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[4387] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[4388] Train Loss:    tensor(0.8122, device='cuda:0', grad_fn=<NllLossBackward>)
[4389] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[4390] Train Loss:    tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[4391] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[4392] Train Loss:    tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[4393] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[4394] Train Loss:    tensor(0.8194, device='cuda:0', grad_fn=<NllLossBackward>)
[4395] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[4396] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[4397] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[4398] Train Loss:    tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[4399] Train Loss:    tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[4400] Train Loss:    tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[4401] Train Loss:    tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[4402] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[4403] Train Loss:    tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[4404] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[4405] Train Loss:    tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[4406] Train Loss:    tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[4407] Train Loss:    tensor(0.8015, device='cuda:0', grad_fn=<NllLossBackward>)
[4408] Train Loss:    tensor(0.8053, device='cuda:0', grad_fn=<NllLossBackward>)
[4409] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[4410] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[4411] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[4412] Train Loss:    tensor(0.8272, device='cuda:0', grad_fn=<NllLossBackward>)
[4413] Train Loss:    tensor(0.7823, device='cuda:0', grad_fn=<NllLossBackward>)
[4414] Train Loss:    tensor(0.7982, device='cuda:0', grad_fn=<NllLossBackward>)
[4415] Train Loss:    tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[4416] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[4417] Train Loss:    tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[4418] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[4419] Train Loss:    tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[4420] Train Loss:    tensor(0.7732, device='cuda:0', grad_fn=<NllLossBackward>)
[4421] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[4422] Train Loss:    tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[4423] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[4424] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[4425] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[4426] Train Loss:    tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4427] Train Loss:    tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[4428] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[4429] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[4430] Train Loss:    tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[4431] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[4432] Train Loss:    tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[4433] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[4434] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[4435] Train Loss:    tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[4436] Train Loss:    tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[4437] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[4438] Train Loss:    tensor(0.8045, device='cuda:0', grad_fn=<NllLossBackward>)
[4439] Train Loss:    tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[4440] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[4441] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[4442] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[4443] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[4444] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[4445] Train Loss:    tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[4446] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[4447] Train Loss:    tensor(0.7896, device='cuda:0', grad_fn=<NllLossBackward>)
[4448] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[4449] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[4450] Train Loss:    tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[4451] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[4452] Train Loss:    tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[4453] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[4454] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[4455] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[4456] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[4457] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[4458] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[4459] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[4460] Train Loss:    tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[4461] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[4462] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[4463] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[4464] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[4465] Train Loss:    tensor(0.7999, device='cuda:0', grad_fn=<NllLossBackward>)
[4466] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[4467] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[4468] Train Loss:    tensor(0.7903, device='cuda:0', grad_fn=<NllLossBackward>)
[4469] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[4470] Train Loss:    tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[4471] Train Loss:    tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[4472] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[4473] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[4474] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4475] Train Loss:  tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[4476] Train Loss:  tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[4477] Train Loss:  tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[4478] Train Loss:  tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[4479] Train Loss:  tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[4480] Train Loss:  tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[4481] Train Loss:  tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[4482] Train Loss:  tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[4483] Train Loss:  tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[4484] Train Loss:  tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[4485] Train Loss:  tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[4486] Train Loss:  tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[4487] Train Loss:  tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[4488] Train Loss:  tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[4489] Train Loss:  tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[4490] Train Loss:  tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[4491] Train Loss:  tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[4492] Train Loss:  tensor(0.8026, device='cuda:0', grad_fn=<NllLossBackward>)
[4493] Train Loss:  tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[4494] Train Loss:  tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[4495] Train Loss:  tensor(0.8258, device='cuda:0', grad_fn=<NllLossBackward>)
[4496] Train Loss:  tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[4497] Train Loss:  tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[4498] Train Loss:  tensor(0.8087, device='cuda:0', grad_fn=<NllLossBackward>)
[4499] Train Loss:  tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[4500] Train Loss:  tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[4501] Train Loss:  tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[4502] Train Loss:  tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[4503] Train Loss:  tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[4504] Train Loss:  tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[4505] Train Loss:  tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[4506] Train Loss:  tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[4507] Train Loss:  tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[4508] Train Loss:  tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[4509] Train Loss:  tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[4510] Train Loss:  tensor(0.8003, device='cuda:0', grad_fn=<NllLossBackward>)
[4511] Train Loss:  tensor(0.7955, device='cuda:0', grad_fn=<NllLossBackward>)
Training Loss:  0.787917179801891 Time:  2449.2736966609955 s
Training Accuaracy:  75.91425036707669 %
[0] Valid Loss:  tensor(0.7497, device='cuda:0')
[1] Valid Loss:  tensor(0.9172, device='cuda:0')
[2] Valid Loss:  tensor(0.7370, device='cuda:0')
[3] Valid Loss:  tensor(0.6501, device='cuda:0')
[4] Valid Loss:  tensor(0.7505, device='cuda:0')
[5] Valid Loss:  tensor(0.7616, device='cuda:0')
[6] Valid Loss:  tensor(0.7708, device='cuda:0')
[7] Valid Loss:  tensor(0.8945, device='cuda:0')
[8] Valid Loss:  tensor(0.8245, device='cuda:0')
```

```
[9] Valid Loss:  tensor(0.8503, device='cuda:0')
[10] Valid Loss:  tensor(0.7246, device='cuda:0')
[11] Valid Loss:  tensor(0.7190, device='cuda:0')
[12] Valid Loss:  tensor(1.1533, device='cuda:0')
[13] Valid Loss:  tensor(1.0036, device='cuda:0')
[14] Valid Loss:  tensor(0.6557, device='cuda:0')
[15] Valid Loss:  tensor(0.7805, device='cuda:0')
[16] Valid Loss:  tensor(0.6658, device='cuda:0')
[17] Valid Loss:  tensor(0.8140, device='cuda:0')
[18] Valid Loss:  tensor(0.7173, device='cuda:0')
[19] Valid Loss:  tensor(0.6707, device='cuda:0')
[20] Valid Loss:  tensor(0.7240, device='cuda:0')
[21] Valid Loss:  tensor(0.6196, device='cuda:0')
[22] Valid Loss:  tensor(0.6095, device='cuda:0')
[23] Valid Loss:  tensor(0.8564, device='cuda:0')
[24] Valid Loss:  tensor(0.7631, device='cuda:0')
[25] Valid Loss:  tensor(0.9776, device='cuda:0')
[26] Valid Loss:  tensor(0.5680, device='cuda:0')
[27] Valid Loss:  tensor(0.9430, device='cuda:0')
[28] Valid Loss:  tensor(0.7121, device='cuda:0')
[29] Valid Loss:  tensor(0.7447, device='cuda:0')
[30] Valid Loss:  tensor(0.6657, device='cuda:0')
[31] Valid Loss:  tensor(0.7659, device='cuda:0')
[32] Valid Loss:  tensor(0.8687, device='cuda:0')
[33] Valid Loss:  tensor(0.8582, device='cuda:0')
[34] Valid Loss:  tensor(0.8126, device='cuda:0')
[35] Valid Loss:  tensor(0.8902, device='cuda:0')
[36] Valid Loss:  tensor(0.7288, device='cuda:0')
[37] Valid Loss:  tensor(1.0364, device='cuda:0')
[38] Valid Loss:  tensor(0.8737, device='cuda:0')
[39] Valid Loss:  tensor(0.7025, device='cuda:0')
[40] Valid Loss:  tensor(0.8352, device='cuda:0')
[41] Valid Loss:  tensor(0.9040, device='cuda:0')
[42] Valid Loss:  tensor(0.8037, device='cuda:0')
[43] Valid Loss:  tensor(0.5987, device='cuda:0')
[44] Valid Loss:  tensor(0.7257, device='cuda:0')
[45] Valid Loss:  tensor(0.6110, device='cuda:0')
[46] Valid Loss:  tensor(0.8964, device='cuda:0')
[47] Valid Loss:  tensor(0.8464, device='cuda:0')
[48] Valid Loss:  tensor(0.7077, device='cuda:0')
[49] Valid Loss:  tensor(0.7850, device='cuda:0')
[50] Valid Loss:  tensor(0.7493, device='cuda:0')
[51] Valid Loss:  tensor(0.7096, device='cuda:0')
[52] Valid Loss:  tensor(0.6613, device='cuda:0')
[53] Valid Loss:  tensor(0.9351, device='cuda:0')
[54] Valid Loss:  tensor(0.6604, device='cuda:0')
[55] Valid Loss:  tensor(0.5898, device='cuda:0')
[56] Valid Loss:  tensor(0.9097, device='cuda:0')
```

```
[57] Valid Loss:  tensor(0.8602, device='cuda:0')
[58] Valid Loss:  tensor(0.8057, device='cuda:0')
[59] Valid Loss:  tensor(0.7019, device='cuda:0')
[60] Valid Loss:  tensor(1.0707, device='cuda:0')
[61] Valid Loss:  tensor(0.7261, device='cuda:0')
[62] Valid Loss:  tensor(0.8863, device='cuda:0')
[63] Valid Loss:  tensor(0.7248, device='cuda:0')
[64] Valid Loss:  tensor(0.7962, device='cuda:0')
[65] Valid Loss:  tensor(0.8938, device='cuda:0')
[66] Valid Loss:  tensor(0.7527, device='cuda:0')
[67] Valid Loss:  tensor(0.7726, device='cuda:0')
[68] Valid Loss:  tensor(0.7567, device='cuda:0')
[69] Valid Loss:  tensor(0.6097, device='cuda:0')
[70] Valid Loss:  tensor(0.6239, device='cuda:0')
[71] Valid Loss:  tensor(0.8914, device='cuda:0')
[72] Valid Loss:  tensor(0.7715, device='cuda:0')
[73] Valid Loss:  tensor(0.7370, device='cuda:0')
[74] Valid Loss:  tensor(0.8077, device='cuda:0')
[75] Valid Loss:  tensor(0.8964, device='cuda:0')
[76] Valid Loss:  tensor(0.8285, device='cuda:0')
[77] Valid Loss:  tensor(0.7246, device='cuda:0')
[78] Valid Loss:  tensor(0.8100, device='cuda:0')
[79] Valid Loss:  tensor(0.6691, device='cuda:0')
[80] Valid Loss:  tensor(0.7662, device='cuda:0')
[81] Valid Loss:  tensor(0.6528, device='cuda:0')
[82] Valid Loss:  tensor(0.9086, device='cuda:0')
[83] Valid Loss:  tensor(0.8446, device='cuda:0')
[84] Valid Loss:  tensor(0.6595, device='cuda:0')
[85] Valid Loss:  tensor(0.7756, device='cuda:0')
[86] Valid Loss:  tensor(0.7877, device='cuda:0')
[87] Valid Loss:  tensor(0.8445, device='cuda:0')
[88] Valid Loss:  tensor(0.6269, device='cuda:0')
[89] Valid Loss:  tensor(0.8317, device='cuda:0')
[90] Valid Loss:  tensor(0.9619, device='cuda:0')
[91] Valid Loss:  tensor(0.7769, device='cuda:0')
[92] Valid Loss:  tensor(0.8379, device='cuda:0')
[93] Valid Loss:  tensor(0.8286, device='cuda:0')
[94] Valid Loss:  tensor(0.8629, device='cuda:0')
[95] Valid Loss:  tensor(0.7348, device='cuda:0')
[96] Valid Loss:  tensor(0.8708, device='cuda:0')
[97] Valid Loss:  tensor(0.8328, device='cuda:0')
[98] Valid Loss:  tensor(0.8414, device='cuda:0')
[99] Valid Loss:  tensor(0.6475, device='cuda:0')
[100] Valid Loss:  tensor(0.7781, device='cuda:0')
[101] Valid Loss:  tensor(0.7609, device='cuda:0')
[102] Valid Loss:  tensor(0.9285, device='cuda:0')
[103] Valid Loss:  tensor(0.9089, device='cuda:0')
[104] Valid Loss:  tensor(0.6999, device='cuda:0')
```

```
[105] Valid Loss:   tensor(0.6865, device='cuda:0')
[106] Valid Loss:   tensor(0.6293, device='cuda:0')
[107] Valid Loss:   tensor(0.8194, device='cuda:0')
[108] Valid Loss:   tensor(0.8875, device='cuda:0')
[109] Valid Loss:   tensor(0.7822, device='cuda:0')
[110] Valid Loss:   tensor(0.8267, device='cuda:0')
[111] Valid Loss:   tensor(0.8159, device='cuda:0')
[112] Valid Loss:   tensor(0.8506, device='cuda:0')
[113] Valid Loss:   tensor(0.8127, device='cuda:0')
[114] Valid Loss:   tensor(0.7613, device='cuda:0')
[115] Valid Loss:   tensor(0.8670, device='cuda:0')
[116] Valid Loss:   tensor(0.8639, device='cuda:0')
[117] Valid Loss:   tensor(0.8804, device='cuda:0')
[118] Valid Loss:   tensor(0.6958, device='cuda:0')
[119] Valid Loss:   tensor(0.6528, device='cuda:0')
[120] Valid Loss:   tensor(0.9955, device='cuda:0')
[121] Valid Loss:   tensor(0.7331, device='cuda:0')
[122] Valid Loss:   tensor(0.7624, device='cuda:0')
[123] Valid Loss:   tensor(0.7146, device='cuda:0')
[124] Valid Loss:   tensor(0.9287, device='cuda:0')
[125] Valid Loss:   tensor(0.9990, device='cuda:0')
[126] Valid Loss:   tensor(0.6060, device='cuda:0')
[127] Valid Loss:   tensor(0.6642, device='cuda:0')
[128] Valid Loss:   tensor(0.9037, device='cuda:0')
[129] Valid Loss:   tensor(0.7167, device='cuda:0')
[130] Valid Loss:   tensor(0.8634, device='cuda:0')
[131] Valid Loss:   tensor(0.7358, device='cuda:0')
[132] Valid Loss:   tensor(0.7559, device='cuda:0')
[133] Valid Loss:   tensor(0.7690, device='cuda:0')
[134] Valid Loss:   tensor(0.6891, device='cuda:0')
[135] Valid Loss:   tensor(0.6239, device='cuda:0')
[136] Valid Loss:   tensor(0.6505, device='cuda:0')
[137] Valid Loss:   tensor(0.9909, device='cuda:0')
[138] Valid Loss:   tensor(0.8601, device='cuda:0')
[139] Valid Loss:   tensor(1.0651, device='cuda:0')
[140] Valid Loss:   tensor(0.8493, device='cuda:0')
[141] Valid Loss:   tensor(0.6710, device='cuda:0')
[142] Valid Loss:   tensor(0.9577, device='cuda:0')
[143] Valid Loss:   tensor(0.6795, device='cuda:0')
[144] Valid Loss:   tensor(0.8032, device='cuda:0')
[145] Valid Loss:   tensor(0.9903, device='cuda:0')
[146] Valid Loss:   tensor(0.7165, device='cuda:0')
[147] Valid Loss:   tensor(0.8258, device='cuda:0')
[148] Valid Loss:   tensor(0.7400, device='cuda:0')
[149] Valid Loss:   tensor(0.7946, device='cuda:0')
[150] Valid Loss:   tensor(0.8081, device='cuda:0')
[151] Valid Loss:   tensor(0.7831, device='cuda:0')
[152] Valid Loss:   tensor(0.7050, device='cuda:0')
```

```
[153] Valid Loss:   tensor(0.7154, device='cuda:0')
[154] Valid Loss:   tensor(0.7401, device='cuda:0')
[155] Valid Loss:   tensor(0.7776, device='cuda:0')
[156] Valid Loss:   tensor(0.8687, device='cuda:0')
[157] Valid Loss:   tensor(0.6926, device='cuda:0')
[158] Valid Loss:   tensor(0.7613, device='cuda:0')
[159] Valid Loss:   tensor(0.7588, device='cuda:0')
[160] Valid Loss:   tensor(0.6420, device='cuda:0')
[161] Valid Loss:   tensor(0.8643, device='cuda:0')
[162] Valid Loss:   tensor(0.8020, device='cuda:0')
[163] Valid Loss:   tensor(0.6836, device='cuda:0')
[164] Valid Loss:   tensor(0.9623, device='cuda:0')
[165] Valid Loss:   tensor(0.7498, device='cuda:0')
[166] Valid Loss:   tensor(0.7520, device='cuda:0')
[167] Valid Loss:   tensor(0.6586, device='cuda:0')
[168] Valid Loss:   tensor(0.7310, device='cuda:0')
[169] Valid Loss:   tensor(0.6984, device='cuda:0')
[170] Valid Loss:   tensor(0.7302, device='cuda:0')
[171] Valid Loss:   tensor(0.6059, device='cuda:0')
[172] Valid Loss:   tensor(0.8078, device='cuda:0')
[173] Valid Loss:   tensor(0.6757, device='cuda:0')
[174] Valid Loss:   tensor(0.7993, device='cuda:0')
[175] Valid Loss:   tensor(0.6759, device='cuda:0')
[176] Valid Loss:   tensor(0.6127, device='cuda:0')
[177] Valid Loss:   tensor(0.8019, device='cuda:0')
[178] Valid Loss:   tensor(0.8460, device='cuda:0')
[179] Valid Loss:   tensor(0.7952, device='cuda:0')
[180] Valid Loss:   tensor(0.8379, device='cuda:0')
[181] Valid Loss:   tensor(0.7800, device='cuda:0')
[182] Valid Loss:   tensor(0.7699, device='cuda:0')
[183] Valid Loss:   tensor(0.7730, device='cuda:0')
[184] Valid Loss:   tensor(0.9305, device='cuda:0')
[185] Valid Loss:   tensor(0.8472, device='cuda:0')
[186] Valid Loss:   tensor(0.7452, device='cuda:0')
[187] Valid Loss:   tensor(0.8531, device='cuda:0')
[188] Valid Loss:   tensor(0.9255, device='cuda:0')
[189] Valid Loss:   tensor(0.7477, device='cuda:0')
[190] Valid Loss:   tensor(0.7752, device='cuda:0')
[191] Valid Loss:   tensor(0.6529, device='cuda:0')
[192] Valid Loss:   tensor(1.0238, device='cuda:0')
[193] Valid Loss:   tensor(0.7844, device='cuda:0')
[194] Valid Loss:   tensor(0.8681, device='cuda:0')
[195] Valid Loss:   tensor(0.9133, device='cuda:0')
[196] Valid Loss:   tensor(0.6420, device='cuda:0')
[197] Valid Loss:   tensor(0.8199, device='cuda:0')
[198] Valid Loss:   tensor(0.7779, device='cuda:0')
[199] Valid Loss:   tensor(0.8360, device='cuda:0')
[200] Valid Loss:   tensor(0.7810, device='cuda:0')
```

```
[201] Valid Loss:   tensor(0.7430, device='cuda:0')
[202] Valid Loss:   tensor(0.6888, device='cuda:0')
[203] Valid Loss:   tensor(0.9209, device='cuda:0')
[204] Valid Loss:   tensor(0.8902, device='cuda:0')
[205] Valid Loss:   tensor(0.8862, device='cuda:0')
[206] Valid Loss:   tensor(0.9740, device='cuda:0')
[207] Valid Loss:   tensor(0.8263, device='cuda:0')
[208] Valid Loss:   tensor(0.6438, device='cuda:0')
[209] Valid Loss:   tensor(0.7836, device='cuda:0')
[210] Valid Loss:   tensor(1.1525, device='cuda:0')
[211] Valid Loss:   tensor(0.7686, device='cuda:0')
[212] Valid Loss:   tensor(0.6701, device='cuda:0')
[213] Valid Loss:   tensor(0.7433, device='cuda:0')
[214] Valid Loss:   tensor(0.7295, device='cuda:0')
[215] Valid Loss:   tensor(0.8386, device='cuda:0')
[216] Valid Loss:   tensor(0.7897, device='cuda:0')
[217] Valid Loss:   tensor(0.7775, device='cuda:0')
[218] Valid Loss:   tensor(0.6186, device='cuda:0')
[219] Valid Loss:   tensor(0.7305, device='cuda:0')
[220] Valid Loss:   tensor(0.6941, device='cuda:0')
[221] Valid Loss:   tensor(0.7993, device='cuda:0')
[222] Valid Loss:   tensor(0.6125, device='cuda:0')
[223] Valid Loss:   tensor(0.7164, device='cuda:0')
[224] Valid Loss:   tensor(0.8413, device='cuda:0')
[225] Valid Loss:   tensor(0.6336, device='cuda:0')
[226] Valid Loss:   tensor(0.7740, device='cuda:0')
[227] Valid Loss:   tensor(0.6325, device='cuda:0')
[228] Valid Loss:   tensor(0.8624, device='cuda:0')
[229] Valid Loss:   tensor(0.7155, device='cuda:0')
[230] Valid Loss:   tensor(0.8264, device='cuda:0')
[231] Valid Loss:   tensor(0.7258, device='cuda:0')
[232] Valid Loss:   tensor(0.6361, device='cuda:0')
[233] Valid Loss:   tensor(0.7299, device='cuda:0')
[234] Valid Loss:   tensor(0.6344, device='cuda:0')
[235] Valid Loss:   tensor(0.6558, device='cuda:0')
[236] Valid Loss:   tensor(0.7920, device='cuda:0')
[237] Valid Loss:   tensor(0.8427, device='cuda:0')
[238] Valid Loss:   tensor(0.5664, device='cuda:0')
[239] Valid Loss:   tensor(0.8225, device='cuda:0')
[240] Valid Loss:   tensor(0.7369, device='cuda:0')
[241] Valid Loss:   tensor(0.6725, device='cuda:0')
[242] Valid Loss:   tensor(0.5617, device='cuda:0')
[243] Valid Loss:   tensor(1.1815, device='cuda:0')
[244] Valid Loss:   tensor(0.7975, device='cuda:0')
[245] Valid Loss:   tensor(0.7405, device='cuda:0')
[246] Valid Loss:   tensor(0.7031, device='cuda:0')
[247] Valid Loss:   tensor(0.7607, device='cuda:0')
[248] Valid Loss:   tensor(0.9633, device='cuda:0')
```

```
[249] Valid Loss:  tensor(0.7507, device='cuda:0')
[250] Valid Loss:  tensor(0.8112, device='cuda:0')
[251] Valid Loss:  tensor(0.8838, device='cuda:0')
[252] Valid Loss:  tensor(0.6607, device='cuda:0')
[253] Valid Loss:  tensor(0.6809, device='cuda:0')
[254] Valid Loss:  tensor(0.6674, device='cuda:0')
[255] Valid Loss:  tensor(0.8186, device='cuda:0')
[256] Valid Loss:  tensor(0.8496, device='cuda:0')
[257] Valid Loss:  tensor(0.6978, device='cuda:0')
[258] Valid Loss:  tensor(0.8912, device='cuda:0')
[259] Valid Loss:  tensor(0.7517, device='cuda:0')
[260] Valid Loss:  tensor(0.9392, device='cuda:0')
[261] Valid Loss:  tensor(0.7915, device='cuda:0')
[262] Valid Loss:  tensor(0.9392, device='cuda:0')
[263] Valid Loss:  tensor(0.9013, device='cuda:0')
[264] Valid Loss:  tensor(0.8896, device='cuda:0')
[265] Valid Loss:  tensor(0.8439, device='cuda:0')
[266] Valid Loss:  tensor(0.7255, device='cuda:0')
[267] Valid Loss:  tensor(0.7622, device='cuda:0')
[268] Valid Loss:  tensor(0.7473, device='cuda:0')
[269] Valid Loss:  tensor(0.7080, device='cuda:0')
[270] Valid Loss:  tensor(0.6649, device='cuda:0')
[271] Valid Loss:  tensor(1.0073, device='cuda:0')
[272] Valid Loss:  tensor(0.7799, device='cuda:0')
[273] Valid Loss:  tensor(0.9162, device='cuda:0')
[274] Valid Loss:  tensor(0.7312, device='cuda:0')
[275] Valid Loss:  tensor(0.7750, device='cuda:0')
[276] Valid Loss:  tensor(0.6833, device='cuda:0')
[277] Valid Loss:  tensor(0.6533, device='cuda:0')
[278] Valid Loss:  tensor(0.9313, device='cuda:0')
[279] Valid Loss:  tensor(0.8151, device='cuda:0')
[280] Valid Loss:  tensor(0.9299, device='cuda:0')
[281] Valid Loss:  tensor(0.6991, device='cuda:0')
[282] Valid Loss:  tensor(0.8102, device='cuda:0')
[283] Valid Loss:  tensor(0.9488, device='cuda:0')
[284] Valid Loss:  tensor(0.7033, device='cuda:0')
[285] Valid Loss:  tensor(0.6089, device='cuda:0')
[286] Valid Loss:  tensor(0.8374, device='cuda:0')
[287] Valid Loss:  tensor(0.9829, device='cuda:0')
[288] Valid Loss:  tensor(0.5959, device='cuda:0')
[289] Valid Loss:  tensor(0.8179, device='cuda:0')
[290] Valid Loss:  tensor(0.6286, device='cuda:0')
[291] Valid Loss:  tensor(0.7454, device='cuda:0')
[292] Valid Loss:  tensor(0.8826, device='cuda:0')
[293] Valid Loss:  tensor(0.6894, device='cuda:0')
[294] Valid Loss:  tensor(0.8931, device='cuda:0')
[295] Valid Loss:  tensor(0.6824, device='cuda:0')
[296] Valid Loss:  tensor(0.8036, device='cuda:0')
```

```
[297] Valid Loss:  tensor(0.8400, device='cuda:0')
[298] Valid Loss:  tensor(0.6980, device='cuda:0')
[299] Valid Loss:  tensor(0.8308, device='cuda:0')
[300] Valid Loss:  tensor(0.9660, device='cuda:0')
[301] Valid Loss:  tensor(0.6265, device='cuda:0')
[302] Valid Loss:  tensor(0.6194, device='cuda:0')
[303] Valid Loss:  tensor(0.7154, device='cuda:0')
[304] Valid Loss:  tensor(0.7122, device='cuda:0')
[305] Valid Loss:  tensor(0.7968, device='cuda:0')
[306] Valid Loss:  tensor(0.8139, device='cuda:0')
[307] Valid Loss:  tensor(0.6152, device='cuda:0')
[308] Valid Loss:  tensor(0.6842, device='cuda:0')
[309] Valid Loss:  tensor(0.7426, device='cuda:0')
[310] Valid Loss:  tensor(0.7427, device='cuda:0')
[311] Valid Loss:  tensor(0.6651, device='cuda:0')
[312] Valid Loss:  tensor(0.6685, device='cuda:0')
[313] Valid Loss:  tensor(0.7432, device='cuda:0')
[314] Valid Loss:  tensor(0.7389, device='cuda:0')
[315] Valid Loss:  tensor(1.0494, device='cuda:0')
[316] Valid Loss:  tensor(0.8728, device='cuda:0')
[317] Valid Loss:  tensor(0.7386, device='cuda:0')
[318] Valid Loss:  tensor(0.7266, device='cuda:0')
[319] Valid Loss:  tensor(0.6817, device='cuda:0')
[320] Valid Loss:  tensor(0.7057, device='cuda:0')
[321] Valid Loss:  tensor(0.7252, device='cuda:0')
[322] Valid Loss:  tensor(0.6704, device='cuda:0')
[323] Valid Loss:  tensor(0.8491, device='cuda:0')
[324] Valid Loss:  tensor(0.8018, device='cuda:0')
[325] Valid Loss:  tensor(0.6628, device='cuda:0')
[326] Valid Loss:  tensor(0.7576, device='cuda:0')
[327] Valid Loss:  tensor(0.4981, device='cuda:0')
[328] Valid Loss:  tensor(0.7237, device='cuda:0')
[329] Valid Loss:  tensor(0.8331, device='cuda:0')
[330] Valid Loss:  tensor(0.6251, device='cuda:0')
[331] Valid Loss:  tensor(0.9502, device='cuda:0')
[332] Valid Loss:  tensor(0.6554, device='cuda:0')
[333] Valid Loss:  tensor(0.6124, device='cuda:0')
[334] Valid Loss:  tensor(0.6550, device='cuda:0')
[335] Valid Loss:  tensor(0.7448, device='cuda:0')
[336] Valid Loss:  tensor(0.8122, device='cuda:0')
[337] Valid Loss:  tensor(0.9407, device='cuda:0')
[338] Valid Loss:  tensor(0.7811, device='cuda:0')
[339] Valid Loss:  tensor(0.7507, device='cuda:0')
[340] Valid Loss:  tensor(0.7667, device='cuda:0')
[341] Valid Loss:  tensor(0.8550, device='cuda:0')
[342] Valid Loss:  tensor(0.9495, device='cuda:0')
[343] Valid Loss:  tensor(0.8126, device='cuda:0')
[344] Valid Loss:  tensor(0.6858, device='cuda:0')
```

```
[345] Valid Loss:  tensor(0.5977, device='cuda:0')
[346] Valid Loss:  tensor(0.7449, device='cuda:0')
[347] Valid Loss:  tensor(0.8741, device='cuda:0')
[348] Valid Loss:  tensor(0.7394, device='cuda:0')
[349] Valid Loss:  tensor(0.7363, device='cuda:0')
[350] Valid Loss:  tensor(0.7653, device='cuda:0')
[351] Valid Loss:  tensor(0.8838, device='cuda:0')
[352] Valid Loss:  tensor(0.6605, device='cuda:0')
[353] Valid Loss:  tensor(0.7875, device='cuda:0')
[354] Valid Loss:  tensor(0.9628, device='cuda:0')
[355] Valid Loss:  tensor(0.5571, device='cuda:0')
[356] Valid Loss:  tensor(0.7140, device='cuda:0')
[357] Valid Loss:  tensor(0.8446, device='cuda:0')
[358] Valid Loss:  tensor(0.7066, device='cuda:0')
[359] Valid Loss:  tensor(0.8179, device='cuda:0')
[360] Valid Loss:  tensor(0.7471, device='cuda:0')
[361] Valid Loss:  tensor(0.7714, device='cuda:0')
[362] Valid Loss:  tensor(0.7409, device='cuda:0')
[363] Valid Loss:  tensor(0.8036, device='cuda:0')
[364] Valid Loss:  tensor(0.9337, device='cuda:0')
[365] Valid Loss:  tensor(0.7191, device='cuda:0')
[366] Valid Loss:  tensor(0.7156, device='cuda:0')
[367] Valid Loss:  tensor(0.8202, device='cuda:0')
[368] Valid Loss:  tensor(0.8796, device='cuda:0')
[369] Valid Loss:  tensor(0.7604, device='cuda:0')
[370] Valid Loss:  tensor(0.6504, device='cuda:0')
[371] Valid Loss:  tensor(0.6564, device='cuda:0')
[372] Valid Loss:  tensor(0.8384, device='cuda:0')
[373] Valid Loss:  tensor(0.5738, device='cuda:0')
[374] Valid Loss:  tensor(0.7391, device='cuda:0')
[375] Valid Loss:  tensor(0.8433, device='cuda:0')
[376] Valid Loss:  tensor(0.6741, device='cuda:0')
[377] Valid Loss:  tensor(0.9777, device='cuda:0')
[378] Valid Loss:  tensor(0.6837, device='cuda:0')
[379] Valid Loss:  tensor(0.8507, device='cuda:0')
[380] Valid Loss:  tensor(1.0793, device='cuda:0')
Testing Loss:  0.7788032918308038
Testing Accuaracy:  75.91730129174238 %
Epoch Number:  4
[0] Train Loss:  tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[1] Train Loss:  tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[2] Train Loss:  tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[3] Train Loss:  tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[4] Train Loss:  tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[5] Train Loss:  tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[6] Train Loss:  tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[7] Train Loss:  tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[8] Train Loss:  tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[9] Train Loss:  tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[10] Train Loss:  tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[11] Train Loss:  tensor(0.7726, device='cuda:0', grad_fn=<NllLossBackward>)
[12] Train Loss:  tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[13] Train Loss:  tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[14] Train Loss:  tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[15] Train Loss:  tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[16] Train Loss:  tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[17] Train Loss:  tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[18] Train Loss:  tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[19] Train Loss:  tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[20] Train Loss:  tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[21] Train Loss:  tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[22] Train Loss:  tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[23] Train Loss:  tensor(0.7726, device='cuda:0', grad_fn=<NllLossBackward>)
[24] Train Loss:  tensor(0.8082, device='cuda:0', grad_fn=<NllLossBackward>)
[25] Train Loss:  tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[26] Train Loss:  tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[27] Train Loss:  tensor(0.7667, device='cuda:0', grad_fn=<NllLossBackward>)
[28] Train Loss:  tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[29] Train Loss:  tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[30] Train Loss:  tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[31] Train Loss:  tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[32] Train Loss:  tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[33] Train Loss:  tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[34] Train Loss:  tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[35] Train Loss:  tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[36] Train Loss:  tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[37] Train Loss:  tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[38] Train Loss:  tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[39] Train Loss:  tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[40] Train Loss:  tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[41] Train Loss:  tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[42] Train Loss:  tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[43] Train Loss:  tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[44] Train Loss:  tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[45] Train Loss:  tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[46] Train Loss:  tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[47] Train Loss:  tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[48] Train Loss:  tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[49] Train Loss:  tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[50] Train Loss:  tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[51] Train Loss:  tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[52] Train Loss:  tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[53] Train Loss:  tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[54] Train Loss:  tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[55] Train Loss:  tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[56] Train Loss:  tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[57] Train Loss:   tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[58] Train Loss:   tensor(0.8117, device='cuda:0', grad_fn=<NllLossBackward>)
[59] Train Loss:   tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[60] Train Loss:   tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[61] Train Loss:   tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[62] Train Loss:   tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[63] Train Loss:   tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[64] Train Loss:   tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[65] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[66] Train Loss:   tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[67] Train Loss:   tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[68] Train Loss:   tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[69] Train Loss:   tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[70] Train Loss:   tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[71] Train Loss:   tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[72] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[73] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[74] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[75] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[76] Train Loss:   tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[77] Train Loss:   tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[78] Train Loss:   tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[79] Train Loss:   tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[80] Train Loss:   tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[81] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[82] Train Loss:   tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[83] Train Loss:   tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[84] Train Loss:   tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[85] Train Loss:   tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[86] Train Loss:   tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[87] Train Loss:   tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[88] Train Loss:   tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[89] Train Loss:   tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[90] Train Loss:   tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[91] Train Loss:   tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[92] Train Loss:   tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[93] Train Loss:   tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[94] Train Loss:   tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[95] Train Loss:   tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[96] Train Loss:   tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[97] Train Loss:   tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[98] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[99] Train Loss:   tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[100] Train Loss:  tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[101] Train Loss:  tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[102] Train Loss:  tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[103] Train Loss:  tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[104] Train Loss:  tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[105] Train Loss:   tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[106] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[107] Train Loss:   tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[108] Train Loss:   tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[109] Train Loss:   tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[110] Train Loss:   tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[111] Train Loss:   tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[112] Train Loss:   tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[113] Train Loss:   tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[114] Train Loss:   tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[115] Train Loss:   tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[116] Train Loss:   tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[117] Train Loss:   tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[118] Train Loss:   tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[119] Train Loss:   tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[120] Train Loss:   tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[121] Train Loss:   tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[122] Train Loss:   tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[123] Train Loss:   tensor(0.8057, device='cuda:0', grad_fn=<NllLossBackward>)
[124] Train Loss:   tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[125] Train Loss:   tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[126] Train Loss:   tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[127] Train Loss:   tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[128] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[129] Train Loss:   tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[130] Train Loss:   tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[131] Train Loss:   tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[132] Train Loss:   tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[133] Train Loss:   tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[134] Train Loss:   tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[135] Train Loss:   tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[136] Train Loss:   tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[137] Train Loss:   tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[138] Train Loss:   tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[139] Train Loss:   tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[140] Train Loss:   tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[141] Train Loss:   tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[142] Train Loss:   tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[143] Train Loss:   tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[144] Train Loss:   tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[145] Train Loss:   tensor(0.7875, device='cuda:0', grad_fn=<NllLossBackward>)
[146] Train Loss:   tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[147] Train Loss:   tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[148] Train Loss:   tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[149] Train Loss:   tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[150] Train Loss:   tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[151] Train Loss:   tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[152] Train Loss:   tensor(0.8186, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[153] Train Loss:    tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[154] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[155] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[156] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[157] Train Loss:    tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[158] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[159] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[160] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[161] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[162] Train Loss:    tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[163] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[164] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[165] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[166] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[167] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[168] Train Loss:    tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[169] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[170] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[171] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[172] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[173] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[174] Train Loss:    tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[175] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[176] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[177] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[178] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[179] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[180] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[181] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[182] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[183] Train Loss:    tensor(0.8041, device='cuda:0', grad_fn=<NllLossBackward>)
[184] Train Loss:    tensor(0.7947, device='cuda:0', grad_fn=<NllLossBackward>)
[185] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[186] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[187] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[188] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[189] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[190] Train Loss:    tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[191] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[192] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[193] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[194] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[195] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[196] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[197] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[198] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[199] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[200] Train Loss:    tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[201] Train Loss:   tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[202] Train Loss:   tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[203] Train Loss:   tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[204] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[205] Train Loss:   tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[206] Train Loss:   tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[207] Train Loss:   tensor(0.7854, device='cuda:0', grad_fn=<NllLossBackward>)
[208] Train Loss:   tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[209] Train Loss:   tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[210] Train Loss:   tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[211] Train Loss:   tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[212] Train Loss:   tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[213] Train Loss:   tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[214] Train Loss:   tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[215] Train Loss:   tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[216] Train Loss:   tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[217] Train Loss:   tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[218] Train Loss:   tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[219] Train Loss:   tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[220] Train Loss:   tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[221] Train Loss:   tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[222] Train Loss:   tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[223] Train Loss:   tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[224] Train Loss:   tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[225] Train Loss:   tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[226] Train Loss:   tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[227] Train Loss:   tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[228] Train Loss:   tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[229] Train Loss:   tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[230] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[231] Train Loss:   tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[232] Train Loss:   tensor(0.7732, device='cuda:0', grad_fn=<NllLossBackward>)
[233] Train Loss:   tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[234] Train Loss:   tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[235] Train Loss:   tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[236] Train Loss:   tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[237] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[238] Train Loss:   tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[239] Train Loss:   tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[240] Train Loss:   tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[241] Train Loss:   tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[242] Train Loss:   tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[243] Train Loss:   tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[244] Train Loss:   tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[245] Train Loss:   tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
[246] Train Loss:   tensor(0.7801, device='cuda:0', grad_fn=<NllLossBackward>)
[247] Train Loss:   tensor(0.8173, device='cuda:0', grad_fn=<NllLossBackward>)
[248] Train Loss:   tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[249] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[250] Train Loss:   tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[251] Train Loss:   tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[252] Train Loss:   tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[253] Train Loss:   tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[254] Train Loss:   tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[255] Train Loss:   tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[256] Train Loss:   tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[257] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[258] Train Loss:   tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[259] Train Loss:   tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[260] Train Loss:   tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[261] Train Loss:   tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[262] Train Loss:   tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[263] Train Loss:   tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[264] Train Loss:   tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[265] Train Loss:   tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[266] Train Loss:   tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[267] Train Loss:   tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[268] Train Loss:   tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[269] Train Loss:   tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[270] Train Loss:   tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[271] Train Loss:   tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[272] Train Loss:   tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[273] Train Loss:   tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[274] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[275] Train Loss:   tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[276] Train Loss:   tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[277] Train Loss:   tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[278] Train Loss:   tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[279] Train Loss:   tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[280] Train Loss:   tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[281] Train Loss:   tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[282] Train Loss:   tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[283] Train Loss:   tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[284] Train Loss:   tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[285] Train Loss:   tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[286] Train Loss:   tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[287] Train Loss:   tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[288] Train Loss:   tensor(0.8111, device='cuda:0', grad_fn=<NllLossBackward>)
[289] Train Loss:   tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[290] Train Loss:   tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[291] Train Loss:   tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[292] Train Loss:   tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[293] Train Loss:   tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[294] Train Loss:   tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[295] Train Loss:   tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[296] Train Loss:   tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[297] Train Loss:   tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[298] Train Loss:   tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[299] Train Loss:   tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[300] Train Loss:   tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[301] Train Loss:   tensor(0.7970, device='cuda:0', grad_fn=<NllLossBackward>)
[302] Train Loss:   tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[303] Train Loss:   tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[304] Train Loss:   tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[305] Train Loss:   tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[306] Train Loss:   tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[307] Train Loss:   tensor(0.7843, device='cuda:0', grad_fn=<NllLossBackward>)
[308] Train Loss:   tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[309] Train Loss:   tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[310] Train Loss:   tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[311] Train Loss:   tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[312] Train Loss:   tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[313] Train Loss:   tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[314] Train Loss:   tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[315] Train Loss:   tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[316] Train Loss:   tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[317] Train Loss:   tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[318] Train Loss:   tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[319] Train Loss:   tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[320] Train Loss:   tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[321] Train Loss:   tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[322] Train Loss:   tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[323] Train Loss:   tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[324] Train Loss:   tensor(0.8031, device='cuda:0', grad_fn=<NllLossBackward>)
[325] Train Loss:   tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[326] Train Loss:   tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[327] Train Loss:   tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[328] Train Loss:   tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[329] Train Loss:   tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[330] Train Loss:   tensor(0.7927, device='cuda:0', grad_fn=<NllLossBackward>)
[331] Train Loss:   tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[332] Train Loss:   tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[333] Train Loss:   tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[334] Train Loss:   tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[335] Train Loss:   tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[336] Train Loss:   tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[337] Train Loss:   tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[338] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[339] Train Loss:   tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[340] Train Loss:   tensor(0.7945, device='cuda:0', grad_fn=<NllLossBackward>)
[341] Train Loss:   tensor(0.7875, device='cuda:0', grad_fn=<NllLossBackward>)
[342] Train Loss:   tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[343] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[344] Train Loss:   tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[345] Train Loss:   tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[346] Train Loss:   tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[347] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[348] Train Loss:   tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[349] Train Loss:   tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[350] Train Loss:   tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[351] Train Loss:   tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[352] Train Loss:   tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[353] Train Loss:   tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[354] Train Loss:   tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[355] Train Loss:   tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[356] Train Loss:   tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[357] Train Loss:   tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[358] Train Loss:   tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[359] Train Loss:   tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[360] Train Loss:   tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[361] Train Loss:   tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[362] Train Loss:   tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[363] Train Loss:   tensor(0.7823, device='cuda:0', grad_fn=<NllLossBackward>)
[364] Train Loss:   tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[365] Train Loss:   tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[366] Train Loss:   tensor(0.7731, device='cuda:0', grad_fn=<NllLossBackward>)
[367] Train Loss:   tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[368] Train Loss:   tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[369] Train Loss:   tensor(0.8147, device='cuda:0', grad_fn=<NllLossBackward>)
[370] Train Loss:   tensor(0.7930, device='cuda:0', grad_fn=<NllLossBackward>)
[371] Train Loss:   tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[372] Train Loss:   tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[373] Train Loss:   tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[374] Train Loss:   tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[375] Train Loss:   tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[376] Train Loss:   tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[377] Train Loss:   tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[378] Train Loss:   tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[379] Train Loss:   tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[380] Train Loss:   tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[381] Train Loss:   tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[382] Train Loss:   tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[383] Train Loss:   tensor(0.8050, device='cuda:0', grad_fn=<NllLossBackward>)
[384] Train Loss:   tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[385] Train Loss:   tensor(0.8114, device='cuda:0', grad_fn=<NllLossBackward>)
[386] Train Loss:   tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[387] Train Loss:   tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[388] Train Loss:   tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[389] Train Loss:   tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[390] Train Loss:   tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[391] Train Loss:   tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[392] Train Loss:   tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[393] Train Loss:   tensor(0.8063, device='cuda:0', grad_fn=<NllLossBackward>)
[394] Train Loss:   tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[395] Train Loss:   tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[396] Train Loss:   tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[397] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[398] Train Loss:   tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[399] Train Loss:   tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[400] Train Loss:   tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[401] Train Loss:   tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[402] Train Loss:   tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[403] Train Loss:   tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[404] Train Loss:   tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[405] Train Loss:   tensor(0.7898, device='cuda:0', grad_fn=<NllLossBackward>)
[406] Train Loss:   tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[407] Train Loss:   tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[408] Train Loss:   tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[409] Train Loss:   tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[410] Train Loss:   tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[411] Train Loss:   tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[412] Train Loss:   tensor(0.8043, device='cuda:0', grad_fn=<NllLossBackward>)
[413] Train Loss:   tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[414] Train Loss:   tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[415] Train Loss:   tensor(0.8004, device='cuda:0', grad_fn=<NllLossBackward>)
[416] Train Loss:   tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[417] Train Loss:   tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[418] Train Loss:   tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[419] Train Loss:   tensor(0.7907, device='cuda:0', grad_fn=<NllLossBackward>)
[420] Train Loss:   tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[421] Train Loss:   tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[422] Train Loss:   tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[423] Train Loss:   tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[424] Train Loss:   tensor(0.8029, device='cuda:0', grad_fn=<NllLossBackward>)
[425] Train Loss:   tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[426] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[427] Train Loss:   tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[428] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[429] Train Loss:   tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[430] Train Loss:   tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[431] Train Loss:   tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[432] Train Loss:   tensor(0.8225, device='cuda:0', grad_fn=<NllLossBackward>)
[433] Train Loss:   tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[434] Train Loss:   tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[435] Train Loss:   tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[436] Train Loss:   tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[437] Train Loss:   tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[438] Train Loss:   tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[439] Train Loss:   tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[440] Train Loss:   tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[441] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[442] Train Loss:   tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[443] Train Loss:   tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[444] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[445] Train Loss:   tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[446] Train Loss:   tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[447] Train Loss:   tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[448] Train Loss:   tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[449] Train Loss:   tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[450] Train Loss:   tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[451] Train Loss:   tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[452] Train Loss:   tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[453] Train Loss:   tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[454] Train Loss:   tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[455] Train Loss:   tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[456] Train Loss:   tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[457] Train Loss:   tensor(0.7898, device='cuda:0', grad_fn=<NllLossBackward>)
[458] Train Loss:   tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[459] Train Loss:   tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[460] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[461] Train Loss:   tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[462] Train Loss:   tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[463] Train Loss:   tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[464] Train Loss:   tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[465] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[466] Train Loss:   tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[467] Train Loss:   tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[468] Train Loss:   tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[469] Train Loss:   tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[470] Train Loss:   tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[471] Train Loss:   tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[472] Train Loss:   tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[473] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[474] Train Loss:   tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[475] Train Loss:   tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[476] Train Loss:   tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[477] Train Loss:   tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[478] Train Loss:   tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[479] Train Loss:   tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[480] Train Loss:   tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[481] Train Loss:   tensor(0.8129, device='cuda:0', grad_fn=<NllLossBackward>)
[482] Train Loss:   tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[483] Train Loss:   tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[484] Train Loss:   tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[485] Train Loss:   tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[486] Train Loss:   tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[487] Train Loss:   tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[488] Train Loss:   tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[489] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[490] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[491] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[492] Train Loss:    tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[493] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[494] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[495] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[496] Train Loss:    tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[497] Train Loss:    tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[498] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[499] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[500] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[501] Train Loss:    tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[502] Train Loss:    tensor(0.8217, device='cuda:0', grad_fn=<NllLossBackward>)
[503] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[504] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[505] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[506] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[507] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[508] Train Loss:    tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[509] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[510] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[511] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[512] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[513] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[514] Train Loss:    tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[515] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[516] Train Loss:    tensor(0.8044, device='cuda:0', grad_fn=<NllLossBackward>)
[517] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[518] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[519] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[520] Train Loss:    tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[521] Train Loss:    tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[522] Train Loss:    tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[523] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[524] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[525] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[526] Train Loss:    tensor(0.7858, device='cuda:0', grad_fn=<NllLossBackward>)
[527] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[528] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[529] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[530] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[531] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[532] Train Loss:    tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[533] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[534] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[535] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[536] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[537] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[538] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[539] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[540] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[541] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[542] Train Loss:    tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[543] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[544] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[545] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[546] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[547] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[548] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[549] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[550] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[551] Train Loss:    tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[552] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[553] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[554] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[555] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[556] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[557] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[558] Train Loss:    tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[559] Train Loss:    tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[560] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[561] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[562] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[563] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[564] Train Loss:    tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[565] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[566] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[567] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[568] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[569] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[570] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[571] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[572] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[573] Train Loss:    tensor(0.7972, device='cuda:0', grad_fn=<NllLossBackward>)
[574] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[575] Train Loss:    tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[576] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[577] Train Loss:    tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[578] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[579] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[580] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[581] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[582] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[583] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[584] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[585] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[586] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[587] Train Loss:    tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[588] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[589] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[590] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[591] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[592] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[593] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[594] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[595] Train Loss:    tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[596] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[597] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[598] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[599] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[600] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[601] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[602] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[603] Train Loss:    tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[604] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[605] Train Loss:    tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[606] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[607] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[608] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[609] Train Loss:    tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[610] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[611] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[612] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[613] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[614] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[615] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[616] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[617] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[618] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[619] Train Loss:    tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[620] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[621] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[622] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[623] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[624] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[625] Train Loss:    tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[626] Train Loss:    tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[627] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[628] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[629] Train Loss:    tensor(0.8254, device='cuda:0', grad_fn=<NllLossBackward>)
[630] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[631] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[632] Train Loss:    tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[633] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[634] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[635] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[636] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[637] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[638] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[639] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[640] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[641] Train Loss:    tensor(0.7705, device='cuda:0', grad_fn=<NllLossBackward>)
[642] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[643] Train Loss:    tensor(0.7784, device='cuda:0', grad_fn=<NllLossBackward>)
[644] Train Loss:    tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[645] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[646] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[647] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[648] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[649] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[650] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[651] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[652] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[653] Train Loss:    tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[654] Train Loss:    tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[655] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[656] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[657] Train Loss:    tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[658] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[659] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[660] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[661] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[662] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[663] Train Loss:    tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[664] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[665] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[666] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[667] Train Loss:    tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[668] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[669] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[670] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[671] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[672] Train Loss:    tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[673] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[674] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[675] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[676] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[677] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[678] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[679] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[680] Train Loss:    tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[681] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[682] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[683] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[684] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[685] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[686] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[687] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[688] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[689] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[690] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[691] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[692] Train Loss:    tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[693] Train Loss:    tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[694] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[695] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[696] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[697] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[698] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[699] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[700] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[701] Train Loss:    tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[702] Train Loss:    tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[703] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[704] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[705] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[706] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[707] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[708] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[709] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[710] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[711] Train Loss:    tensor(0.8034, device='cuda:0', grad_fn=<NllLossBackward>)
[712] Train Loss:    tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
[713] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[714] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[715] Train Loss:    tensor(0.8085, device='cuda:0', grad_fn=<NllLossBackward>)
[716] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[717] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[718] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[719] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[720] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[721] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[722] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[723] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[724] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[725] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[726] Train Loss:    tensor(0.8109, device='cuda:0', grad_fn=<NllLossBackward>)
[727] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[728] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[729] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[730] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[731] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[732] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[733] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[734] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[735] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[736] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[737] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[738] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[739] Train Loss:    tensor(0.7859, device='cuda:0', grad_fn=<NllLossBackward>)
[740] Train Loss:    tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[741] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[742] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[743] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[744] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[745] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[746] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[747] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[748] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[749] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[750] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[751] Train Loss:    tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[752] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[753] Train Loss:    tensor(0.8025, device='cuda:0', grad_fn=<NllLossBackward>)
[754] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[755] Train Loss:    tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[756] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[757] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[758] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[759] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[760] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[761] Train Loss:    tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[762] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[763] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[764] Train Loss:    tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[765] Train Loss:    tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[766] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[767] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[768] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[769] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[770] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[771] Train Loss:    tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[772] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[773] Train Loss:    tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[774] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[775] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[776] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[777] Train Loss:   tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[778] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[779] Train Loss:   tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[780] Train Loss:   tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[781] Train Loss:   tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[782] Train Loss:   tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[783] Train Loss:   tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[784] Train Loss:   tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[785] Train Loss:   tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[786] Train Loss:   tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[787] Train Loss:   tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[788] Train Loss:   tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[789] Train Loss:   tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[790] Train Loss:   tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[791] Train Loss:   tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[792] Train Loss:   tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[793] Train Loss:   tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[794] Train Loss:   tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[795] Train Loss:   tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[796] Train Loss:   tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[797] Train Loss:   tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[798] Train Loss:   tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[799] Train Loss:   tensor(0.7980, device='cuda:0', grad_fn=<NllLossBackward>)
[800] Train Loss:   tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[801] Train Loss:   tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[802] Train Loss:   tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[803] Train Loss:   tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[804] Train Loss:   tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[805] Train Loss:   tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[806] Train Loss:   tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[807] Train Loss:   tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[808] Train Loss:   tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[809] Train Loss:   tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[810] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[811] Train Loss:   tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[812] Train Loss:   tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[813] Train Loss:   tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[814] Train Loss:   tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[815] Train Loss:   tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[816] Train Loss:   tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[817] Train Loss:   tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[818] Train Loss:   tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[819] Train Loss:   tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[820] Train Loss:   tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[821] Train Loss:   tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[822] Train Loss:   tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[823] Train Loss:   tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[824] Train Loss:   tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[825] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[826] Train Loss:    tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[827] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[828] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[829] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[830] Train Loss:    tensor(0.7844, device='cuda:0', grad_fn=<NllLossBackward>)
[831] Train Loss:    tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[832] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[833] Train Loss:    tensor(0.7985, device='cuda:0', grad_fn=<NllLossBackward>)
[834] Train Loss:    tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[835] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[836] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[837] Train Loss:    tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[838] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[839] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[840] Train Loss:    tensor(0.8021, device='cuda:0', grad_fn=<NllLossBackward>)
[841] Train Loss:    tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[842] Train Loss:    tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[843] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[844] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[845] Train Loss:    tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[846] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[847] Train Loss:    tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[848] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[849] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[850] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[851] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[852] Train Loss:    tensor(0.8061, device='cuda:0', grad_fn=<NllLossBackward>)
[853] Train Loss:    tensor(0.7838, device='cuda:0', grad_fn=<NllLossBackward>)
[854] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[855] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[856] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[857] Train Loss:    tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[858] Train Loss:    tensor(0.7934, device='cuda:0', grad_fn=<NllLossBackward>)
[859] Train Loss:    tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[860] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[861] Train Loss:    tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[862] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[863] Train Loss:    tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[864] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[865] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[866] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[867] Train Loss:    tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[868] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[869] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[870] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[871] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[872] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[873] Train Loss:   tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[874] Train Loss:   tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[875] Train Loss:   tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[876] Train Loss:   tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[877] Train Loss:   tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[878] Train Loss:   tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[879] Train Loss:   tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[880] Train Loss:   tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[881] Train Loss:   tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[882] Train Loss:   tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[883] Train Loss:   tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[884] Train Loss:   tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[885] Train Loss:   tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[886] Train Loss:   tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[887] Train Loss:   tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[888] Train Loss:   tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[889] Train Loss:   tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[890] Train Loss:   tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[891] Train Loss:   tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[892] Train Loss:   tensor(0.8015, device='cuda:0', grad_fn=<NllLossBackward>)
[893] Train Loss:   tensor(0.8017, device='cuda:0', grad_fn=<NllLossBackward>)
[894] Train Loss:   tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[895] Train Loss:   tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[896] Train Loss:   tensor(0.7903, device='cuda:0', grad_fn=<NllLossBackward>)
[897] Train Loss:   tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[898] Train Loss:   tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[899] Train Loss:   tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[900] Train Loss:   tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[901] Train Loss:   tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[902] Train Loss:   tensor(0.8120, device='cuda:0', grad_fn=<NllLossBackward>)
[903] Train Loss:   tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[904] Train Loss:   tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[905] Train Loss:   tensor(0.7726, device='cuda:0', grad_fn=<NllLossBackward>)
[906] Train Loss:   tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[907] Train Loss:   tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[908] Train Loss:   tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[909] Train Loss:   tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[910] Train Loss:   tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[911] Train Loss:   tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[912] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[913] Train Loss:   tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[914] Train Loss:   tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[915] Train Loss:   tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[916] Train Loss:   tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[917] Train Loss:   tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[918] Train Loss:   tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[919] Train Loss:   tensor(0.8062, device='cuda:0', grad_fn=<NllLossBackward>)
[920] Train Loss:   tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[921] Train Loss:   tensor(0.8166, device='cuda:0', grad_fn=<NllLossBackward>)
[922] Train Loss:   tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[923] Train Loss:   tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[924] Train Loss:   tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[925] Train Loss:   tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[926] Train Loss:   tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[927] Train Loss:   tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[928] Train Loss:   tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[929] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[930] Train Loss:   tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[931] Train Loss:   tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[932] Train Loss:   tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[933] Train Loss:   tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[934] Train Loss:   tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[935] Train Loss:   tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[936] Train Loss:   tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[937] Train Loss:   tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[938] Train Loss:   tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[939] Train Loss:   tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[940] Train Loss:   tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[941] Train Loss:   tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[942] Train Loss:   tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[943] Train Loss:   tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[944] Train Loss:   tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[945] Train Loss:   tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[946] Train Loss:   tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[947] Train Loss:   tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[948] Train Loss:   tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[949] Train Loss:   tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[950] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[951] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[952] Train Loss:   tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[953] Train Loss:   tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[954] Train Loss:   tensor(0.8144, device='cuda:0', grad_fn=<NllLossBackward>)
[955] Train Loss:   tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[956] Train Loss:   tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[957] Train Loss:   tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[958] Train Loss:   tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[959] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[960] Train Loss:   tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[961] Train Loss:   tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[962] Train Loss:   tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[963] Train Loss:   tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[964] Train Loss:   tensor(0.8108, device='cuda:0', grad_fn=<NllLossBackward>)
[965] Train Loss:   tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
[966] Train Loss:   tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[967] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[968] Train Loss:   tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[969] Train Loss:  tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[970] Train Loss:  tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[971] Train Loss:  tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[972] Train Loss:  tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[973] Train Loss:  tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[974] Train Loss:  tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[975] Train Loss:  tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[976] Train Loss:  tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[977] Train Loss:  tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[978] Train Loss:  tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[979] Train Loss:  tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[980] Train Loss:  tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[981] Train Loss:  tensor(0.7961, device='cuda:0', grad_fn=<NllLossBackward>)
[982] Train Loss:  tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[983] Train Loss:  tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[984] Train Loss:  tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[985] Train Loss:  tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[986] Train Loss:  tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[987] Train Loss:  tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[988] Train Loss:  tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[989] Train Loss:  tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[990] Train Loss:  tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[991] Train Loss:  tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[992] Train Loss:  tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[993] Train Loss:  tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[994] Train Loss:  tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[995] Train Loss:  tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[996] Train Loss:  tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[997] Train Loss:  tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[998] Train Loss:  tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[999] Train Loss:  tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[1000] Train Loss:  tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[1001] Train Loss:  tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[1002] Train Loss:  tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[1003] Train Loss:  tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[1004] Train Loss:  tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[1005] Train Loss:  tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[1006] Train Loss:  tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[1007] Train Loss:  tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[1008] Train Loss:  tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[1009] Train Loss:  tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[1010] Train Loss:  tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[1011] Train Loss:  tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[1012] Train Loss:  tensor(0.7667, device='cuda:0', grad_fn=<NllLossBackward>)
[1013] Train Loss:  tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[1014] Train Loss:  tensor(0.7823, device='cuda:0', grad_fn=<NllLossBackward>)
[1015] Train Loss:  tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[1016] Train Loss:  tensor(0.7934, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1017] Train Loss:    tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[1018] Train Loss:    tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[1019] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[1020] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[1021] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[1022] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[1023] Train Loss:    tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[1024] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[1025] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[1026] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[1027] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[1028] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[1029] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[1030] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[1031] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[1032] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[1033] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[1034] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[1035] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[1036] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[1037] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[1038] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[1039] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[1040] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[1041] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[1042] Train Loss:    tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[1043] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[1044] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[1045] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[1046] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[1047] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[1048] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[1049] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[1050] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[1051] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[1052] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[1053] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[1054] Train Loss:    tensor(0.7955, device='cuda:0', grad_fn=<NllLossBackward>)
[1055] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[1056] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[1057] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[1058] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[1059] Train Loss:    tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[1060] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[1061] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[1062] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[1063] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[1064] Train Loss:    tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1065] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[1066] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[1067] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[1068] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[1069] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[1070] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[1071] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[1072] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[1073] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[1074] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[1075] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[1076] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[1077] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[1078] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[1079] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[1080] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[1081] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[1082] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[1083] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[1084] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[1085] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[1086] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[1087] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[1088] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[1089] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[1090] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[1091] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[1092] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[1093] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[1094] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[1095] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[1096] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[1097] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[1098] Train Loss:    tensor(0.7856, device='cuda:0', grad_fn=<NllLossBackward>)
[1099] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[1100] Train Loss:    tensor(0.8213, device='cuda:0', grad_fn=<NllLossBackward>)
[1101] Train Loss:    tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[1102] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[1103] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[1104] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[1105] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[1106] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[1107] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[1108] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[1109] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[1110] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[1111] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[1112] Train Loss:    tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1113] Train Loss:    tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[1114] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[1115] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[1116] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[1117] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[1118] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[1119] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[1120] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[1121] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[1122] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[1123] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[1124] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[1125] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[1126] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1127] Train Loss:    tensor(0.8013, device='cuda:0', grad_fn=<NllLossBackward>)
[1128] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[1129] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[1130] Train Loss:    tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[1131] Train Loss:    tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[1132] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[1133] Train Loss:    tensor(0.7955, device='cuda:0', grad_fn=<NllLossBackward>)
[1134] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[1135] Train Loss:    tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[1136] Train Loss:    tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[1137] Train Loss:    tensor(0.7967, device='cuda:0', grad_fn=<NllLossBackward>)
[1138] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1139] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[1140] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1141] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[1142] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[1143] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[1144] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[1145] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[1146] Train Loss:    tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[1147] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[1148] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[1149] Train Loss:    tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[1150] Train Loss:    tensor(0.7991, device='cuda:0', grad_fn=<NllLossBackward>)
[1151] Train Loss:    tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[1152] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[1153] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[1154] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[1155] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[1156] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[1157] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[1158] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[1159] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[1160] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1161] Train Loss:   tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[1162] Train Loss:   tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[1163] Train Loss:   tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[1164] Train Loss:   tensor(0.8069, device='cuda:0', grad_fn=<NllLossBackward>)
[1165] Train Loss:   tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[1166] Train Loss:   tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[1167] Train Loss:   tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[1168] Train Loss:   tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[1169] Train Loss:   tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[1170] Train Loss:   tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[1171] Train Loss:   tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[1172] Train Loss:   tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[1173] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[1174] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[1175] Train Loss:   tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[1176] Train Loss:   tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[1177] Train Loss:   tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[1178] Train Loss:   tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[1179] Train Loss:   tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[1180] Train Loss:   tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[1181] Train Loss:   tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[1182] Train Loss:   tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[1183] Train Loss:   tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[1184] Train Loss:   tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[1185] Train Loss:   tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[1186] Train Loss:   tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[1187] Train Loss:   tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[1188] Train Loss:   tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[1189] Train Loss:   tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[1190] Train Loss:   tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[1191] Train Loss:   tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[1192] Train Loss:   tensor(0.7998, device='cuda:0', grad_fn=<NllLossBackward>)
[1193] Train Loss:   tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[1194] Train Loss:   tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[1195] Train Loss:   tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[1196] Train Loss:   tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[1197] Train Loss:   tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[1198] Train Loss:   tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[1199] Train Loss:   tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[1200] Train Loss:   tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[1201] Train Loss:   tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1202] Train Loss:   tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[1203] Train Loss:   tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[1204] Train Loss:   tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[1205] Train Loss:   tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[1206] Train Loss:   tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[1207] Train Loss:   tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[1208] Train Loss:   tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1209] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[1210] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1211] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[1212] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[1213] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[1214] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[1215] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[1216] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[1217] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[1218] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[1219] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[1220] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[1221] Train Loss:    tensor(0.8006, device='cuda:0', grad_fn=<NllLossBackward>)
[1222] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[1223] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[1224] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[1225] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[1226] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[1227] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[1228] Train Loss:    tensor(0.8168, device='cuda:0', grad_fn=<NllLossBackward>)
[1229] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[1230] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[1231] Train Loss:    tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[1232] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[1233] Train Loss:    tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[1234] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[1235] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[1236] Train Loss:    tensor(0.7985, device='cuda:0', grad_fn=<NllLossBackward>)
[1237] Train Loss:    tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[1238] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[1239] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[1240] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[1241] Train Loss:    tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[1242] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[1243] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[1244] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[1245] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[1246] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[1247] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[1248] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[1249] Train Loss:    tensor(0.7979, device='cuda:0', grad_fn=<NllLossBackward>)
[1250] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[1251] Train Loss:    tensor(0.7859, device='cuda:0', grad_fn=<NllLossBackward>)
[1252] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[1253] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[1254] Train Loss:    tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[1255] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[1256] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1257] Train Loss:   tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[1258] Train Loss:   tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[1259] Train Loss:   tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[1260] Train Loss:   tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[1261] Train Loss:   tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[1262] Train Loss:   tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[1263] Train Loss:   tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[1264] Train Loss:   tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[1265] Train Loss:   tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[1266] Train Loss:   tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[1267] Train Loss:   tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[1268] Train Loss:   tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[1269] Train Loss:   tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[1270] Train Loss:   tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[1271] Train Loss:   tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[1272] Train Loss:   tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[1273] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[1274] Train Loss:   tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[1275] Train Loss:   tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[1276] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[1277] Train Loss:   tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[1278] Train Loss:   tensor(0.7957, device='cuda:0', grad_fn=<NllLossBackward>)
[1279] Train Loss:   tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[1280] Train Loss:   tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[1281] Train Loss:   tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[1282] Train Loss:   tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[1283] Train Loss:   tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[1284] Train Loss:   tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[1285] Train Loss:   tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[1286] Train Loss:   tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[1287] Train Loss:   tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[1288] Train Loss:   tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[1289] Train Loss:   tensor(0.7730, device='cuda:0', grad_fn=<NllLossBackward>)
[1290] Train Loss:   tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[1291] Train Loss:   tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[1292] Train Loss:   tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[1293] Train Loss:   tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[1294] Train Loss:   tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[1295] Train Loss:   tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[1296] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[1297] Train Loss:   tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[1298] Train Loss:   tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[1299] Train Loss:   tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[1300] Train Loss:   tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[1301] Train Loss:   tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[1302] Train Loss:   tensor(0.7888, device='cuda:0', grad_fn=<NllLossBackward>)
[1303] Train Loss:   tensor(0.7875, device='cuda:0', grad_fn=<NllLossBackward>)
[1304] Train Loss:   tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1305] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[1306] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[1307] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[1308] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[1309] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[1310] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[1311] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[1312] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[1313] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[1314] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[1315] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[1316] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[1317] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[1318] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[1319] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[1320] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[1321] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[1322] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[1323] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[1324] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[1325] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[1326] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[1327] Train Loss:    tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[1328] Train Loss:    tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[1329] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[1330] Train Loss:    tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[1331] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[1332] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[1333] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[1334] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[1335] Train Loss:    tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[1336] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[1337] Train Loss:    tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[1338] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[1339] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[1340] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[1341] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[1342] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[1343] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[1344] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[1345] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1346] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[1347] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[1348] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[1349] Train Loss:    tensor(0.8039, device='cuda:0', grad_fn=<NllLossBackward>)
[1350] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[1351] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[1352] Train Loss:    tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1353] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[1354] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[1355] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[1356] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1357] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[1358] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[1359] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[1360] Train Loss:    tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[1361] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[1362] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[1363] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[1364] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[1365] Train Loss:    tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[1366] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[1367] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[1368] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[1369] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[1370] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[1371] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[1372] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[1373] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[1374] Train Loss:    tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[1375] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[1376] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[1377] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1378] Train Loss:    tensor(0.7986, device='cuda:0', grad_fn=<NllLossBackward>)
[1379] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[1380] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[1381] Train Loss:    tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[1382] Train Loss:    tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[1383] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[1384] Train Loss:    tensor(0.7823, device='cuda:0', grad_fn=<NllLossBackward>)
[1385] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[1386] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[1387] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[1388] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1389] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[1390] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[1391] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[1392] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[1393] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[1394] Train Loss:    tensor(0.7976, device='cuda:0', grad_fn=<NllLossBackward>)
[1395] Train Loss:    tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[1396] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[1397] Train Loss:    tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[1398] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[1399] Train Loss:    tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[1400] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1401] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[1402] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[1403] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[1404] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[1405] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[1406] Train Loss:    tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[1407] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[1408] Train Loss:    tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[1409] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[1410] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[1411] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[1412] Train Loss:    tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[1413] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[1414] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[1415] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[1416] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[1417] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[1418] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[1419] Train Loss:    tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[1420] Train Loss:    tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[1421] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[1422] Train Loss:    tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[1423] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[1424] Train Loss:    tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[1425] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[1426] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[1427] Train Loss:    tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[1428] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[1429] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[1430] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[1431] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[1432] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[1433] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[1434] Train Loss:    tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[1435] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[1436] Train Loss:    tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[1437] Train Loss:    tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[1438] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[1439] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[1440] Train Loss:    tensor(0.8078, device='cuda:0', grad_fn=<NllLossBackward>)
[1441] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[1442] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[1443] Train Loss:    tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[1444] Train Loss:    tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[1445] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[1446] Train Loss:    tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[1447] Train Loss:    tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[1448] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1449] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[1450] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[1451] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[1452] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[1453] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[1454] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[1455] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[1456] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[1457] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[1458] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[1459] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[1460] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[1461] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[1462] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[1463] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[1464] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[1465] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[1466] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[1467] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[1468] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[1469] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[1470] Train Loss:    tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[1471] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[1472] Train Loss:    tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[1473] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[1474] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[1475] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[1476] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[1477] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[1478] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[1479] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[1480] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[1481] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[1482] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[1483] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[1484] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[1485] Train Loss:    tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[1486] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1487] Train Loss:    tensor(0.7838, device='cuda:0', grad_fn=<NllLossBackward>)
[1488] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[1489] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[1490] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[1491] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[1492] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[1493] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[1494] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[1495] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[1496] Train Loss:    tensor(0.7731, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1497] Train Loss:    tensor(0.8100, device='cuda:0', grad_fn=<NllLossBackward>)
[1498] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[1499] Train Loss:    tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[1500] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[1501] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[1502] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[1503] Train Loss:    tensor(0.7970, device='cuda:0', grad_fn=<NllLossBackward>)
[1504] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[1505] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[1506] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[1507] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[1508] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[1509] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[1510] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[1511] Train Loss:    tensor(0.8124, device='cuda:0', grad_fn=<NllLossBackward>)
[1512] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[1513] Train Loss:    tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[1514] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[1515] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[1516] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[1517] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[1518] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[1519] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[1520] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[1521] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[1522] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[1523] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[1524] Train Loss:    tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[1525] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[1526] Train Loss:    tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[1527] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[1528] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[1529] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[1530] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[1531] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[1532] Train Loss:    tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[1533] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[1534] Train Loss:    tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[1535] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[1536] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[1537] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[1538] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[1539] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[1540] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[1541] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[1542] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[1543] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[1544] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1545] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[1546] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[1547] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[1548] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[1549] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[1550] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[1551] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[1552] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[1553] Train Loss:    tensor(0.7856, device='cuda:0', grad_fn=<NllLossBackward>)
[1554] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[1555] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[1556] Train Loss:    tensor(0.8134, device='cuda:0', grad_fn=<NllLossBackward>)
[1557] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[1558] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[1559] Train Loss:    tensor(0.8204, device='cuda:0', grad_fn=<NllLossBackward>)
[1560] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[1561] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[1562] Train Loss:    tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[1563] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[1564] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[1565] Train Loss:    tensor(0.7731, device='cuda:0', grad_fn=<NllLossBackward>)
[1566] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1567] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1568] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[1569] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[1570] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[1571] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[1572] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[1573] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[1574] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[1575] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[1576] Train Loss:    tensor(0.7730, device='cuda:0', grad_fn=<NllLossBackward>)
[1577] Train Loss:    tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[1578] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[1579] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[1580] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[1581] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[1582] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[1583] Train Loss:    tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[1584] Train Loss:    tensor(0.7888, device='cuda:0', grad_fn=<NllLossBackward>)
[1585] Train Loss:    tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[1586] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[1587] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1588] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[1589] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[1590] Train Loss:    tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[1591] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[1592] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1593] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[1594] Train Loss:   tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[1595] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[1596] Train Loss:   tensor(0.7965, device='cuda:0', grad_fn=<NllLossBackward>)
[1597] Train Loss:   tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[1598] Train Loss:   tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[1599] Train Loss:   tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[1600] Train Loss:   tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[1601] Train Loss:   tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[1602] Train Loss:   tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[1603] Train Loss:   tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[1604] Train Loss:   tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[1605] Train Loss:   tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[1606] Train Loss:   tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[1607] Train Loss:   tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[1608] Train Loss:   tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[1609] Train Loss:   tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[1610] Train Loss:   tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[1611] Train Loss:   tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[1612] Train Loss:   tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[1613] Train Loss:   tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[1614] Train Loss:   tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[1615] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[1616] Train Loss:   tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[1617] Train Loss:   tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[1618] Train Loss:   tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[1619] Train Loss:   tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[1620] Train Loss:   tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[1621] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[1622] Train Loss:   tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[1623] Train Loss:   tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[1624] Train Loss:   tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[1625] Train Loss:   tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[1626] Train Loss:   tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[1627] Train Loss:   tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[1628] Train Loss:   tensor(0.7934, device='cuda:0', grad_fn=<NllLossBackward>)
[1629] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[1630] Train Loss:   tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[1631] Train Loss:   tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[1632] Train Loss:   tensor(0.7943, device='cuda:0', grad_fn=<NllLossBackward>)
[1633] Train Loss:   tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[1634] Train Loss:   tensor(0.7987, device='cuda:0', grad_fn=<NllLossBackward>)
[1635] Train Loss:   tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[1636] Train Loss:   tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[1637] Train Loss:   tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[1638] Train Loss:   tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[1639] Train Loss:   tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[1640] Train Loss:   tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1641] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[1642] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[1643] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[1644] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[1645] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[1646] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[1647] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[1648] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[1649] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[1650] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[1651] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[1652] Train Loss:    tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[1653] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[1654] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[1655] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[1656] Train Loss:    tensor(0.7942, device='cuda:0', grad_fn=<NllLossBackward>)
[1657] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[1658] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[1659] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[1660] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[1661] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[1662] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[1663] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[1664] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[1665] Train Loss:    tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[1666] Train Loss:    tensor(0.7875, device='cuda:0', grad_fn=<NllLossBackward>)
[1667] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[1668] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[1669] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[1670] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[1671] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[1672] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[1673] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[1674] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[1675] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[1676] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[1677] Train Loss:    tensor(0.7844, device='cuda:0', grad_fn=<NllLossBackward>)
[1678] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[1679] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[1680] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[1681] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[1682] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[1683] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[1684] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[1685] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[1686] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[1687] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[1688] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1689] Train Loss:    tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[1690] Train Loss:    tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[1691] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[1692] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[1693] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[1694] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[1695] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[1696] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[1697] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[1698] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[1699] Train Loss:    tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[1700] Train Loss:    tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[1701] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[1702] Train Loss:    tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[1703] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[1704] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[1705] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[1706] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[1707] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[1708] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[1709] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[1710] Train Loss:    tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[1711] Train Loss:    tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[1712] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[1713] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[1714] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[1715] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[1716] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[1717] Train Loss:    tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[1718] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[1719] Train Loss:    tensor(0.7745, device='cuda:0', grad_fn=<NllLossBackward>)
[1720] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[1721] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[1722] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[1723] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[1724] Train Loss:    tensor(0.8079, device='cuda:0', grad_fn=<NllLossBackward>)
[1725] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[1726] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[1727] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[1728] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[1729] Train Loss:    tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[1730] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[1731] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[1732] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[1733] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[1734] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[1735] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[1736] Train Loss:    tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1737] Train Loss:   tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[1738] Train Loss:   tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[1739] Train Loss:   tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[1740] Train Loss:   tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[1741] Train Loss:   tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[1742] Train Loss:   tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[1743] Train Loss:   tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[1744] Train Loss:   tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[1745] Train Loss:   tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[1746] Train Loss:   tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[1747] Train Loss:   tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[1748] Train Loss:   tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[1749] Train Loss:   tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[1750] Train Loss:   tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[1751] Train Loss:   tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[1752] Train Loss:   tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[1753] Train Loss:   tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[1754] Train Loss:   tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[1755] Train Loss:   tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[1756] Train Loss:   tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[1757] Train Loss:   tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[1758] Train Loss:   tensor(0.7954, device='cuda:0', grad_fn=<NllLossBackward>)
[1759] Train Loss:   tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[1760] Train Loss:   tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[1761] Train Loss:   tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[1762] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[1763] Train Loss:   tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[1764] Train Loss:   tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[1765] Train Loss:   tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[1766] Train Loss:   tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[1767] Train Loss:   tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[1768] Train Loss:   tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[1769] Train Loss:   tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[1770] Train Loss:   tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[1771] Train Loss:   tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[1772] Train Loss:   tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[1773] Train Loss:   tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[1774] Train Loss:   tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[1775] Train Loss:   tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[1776] Train Loss:   tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[1777] Train Loss:   tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[1778] Train Loss:   tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[1779] Train Loss:   tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[1780] Train Loss:   tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[1781] Train Loss:   tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[1782] Train Loss:   tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[1783] Train Loss:   tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[1784] Train Loss:   tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1785] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[1786] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[1787] Train Loss:    tensor(0.8027, device='cuda:0', grad_fn=<NllLossBackward>)
[1788] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[1789] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[1790] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[1791] Train Loss:    tensor(0.8166, device='cuda:0', grad_fn=<NllLossBackward>)
[1792] Train Loss:    tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[1793] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[1794] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[1795] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[1796] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[1797] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[1798] Train Loss:    tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[1799] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[1800] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[1801] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[1802] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[1803] Train Loss:    tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[1804] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[1805] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[1806] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[1807] Train Loss:    tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[1808] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[1809] Train Loss:    tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[1810] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[1811] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[1812] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[1813] Train Loss:    tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[1814] Train Loss:    tensor(0.7667, device='cuda:0', grad_fn=<NllLossBackward>)
[1815] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[1816] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[1817] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[1818] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[1819] Train Loss:    tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[1820] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[1821] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[1822] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[1823] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[1824] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[1825] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[1826] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[1827] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[1828] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[1829] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[1830] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[1831] Train Loss:    tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[1832] Train Loss:    tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1833] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[1834] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[1835] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[1836] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[1837] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1838] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[1839] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[1840] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[1841] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[1842] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[1843] Train Loss:    tensor(0.8099, device='cuda:0', grad_fn=<NllLossBackward>)
[1844] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[1845] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[1846] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[1847] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[1848] Train Loss:    tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[1849] Train Loss:    tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[1850] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[1851] Train Loss:    tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[1852] Train Loss:    tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[1853] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[1854] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[1855] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[1856] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1857] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[1858] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[1859] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[1860] Train Loss:    tensor(0.7937, device='cuda:0', grad_fn=<NllLossBackward>)
[1861] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[1862] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1863] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[1864] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[1865] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[1866] Train Loss:    tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[1867] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[1868] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[1869] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[1870] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[1871] Train Loss:    tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[1872] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[1873] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[1874] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[1875] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[1876] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[1877] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1878] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[1879] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[1880] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1881] Train Loss:    tensor(0.8058, device='cuda:0', grad_fn=<NllLossBackward>)
[1882] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[1883] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[1884] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[1885] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[1886] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[1887] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[1888] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[1889] Train Loss:    tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[1890] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[1891] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[1892] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[1893] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[1894] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[1895] Train Loss:    tensor(0.7902, device='cuda:0', grad_fn=<NllLossBackward>)
[1896] Train Loss:    tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[1897] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[1898] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[1899] Train Loss:    tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[1900] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[1901] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[1902] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[1903] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[1904] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[1905] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[1906] Train Loss:    tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[1907] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[1908] Train Loss:    tensor(0.7844, device='cuda:0', grad_fn=<NllLossBackward>)
[1909] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[1910] Train Loss:    tensor(0.8055, device='cuda:0', grad_fn=<NllLossBackward>)
[1911] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[1912] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[1913] Train Loss:    tensor(0.7910, device='cuda:0', grad_fn=<NllLossBackward>)
[1914] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[1915] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[1916] Train Loss:    tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[1917] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[1918] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[1919] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[1920] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[1921] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[1922] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[1923] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[1924] Train Loss:    tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[1925] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[1926] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[1927] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[1928] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1929] Train Loss:    tensor(0.7914, device='cuda:0', grad_fn=<NllLossBackward>)
[1930] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[1931] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[1932] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[1933] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[1934] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[1935] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[1936] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[1937] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[1938] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[1939] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[1940] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[1941] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[1942] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[1943] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[1944] Train Loss:    tensor(0.8046, device='cuda:0', grad_fn=<NllLossBackward>)
[1945] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[1946] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[1947] Train Loss:    tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[1948] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[1949] Train Loss:    tensor(0.7731, device='cuda:0', grad_fn=<NllLossBackward>)
[1950] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[1951] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[1952] Train Loss:    tensor(0.8140, device='cuda:0', grad_fn=<NllLossBackward>)
[1953] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[1954] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[1955] Train Loss:    tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[1956] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[1957] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[1958] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[1959] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1960] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[1961] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[1962] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[1963] Train Loss:    tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[1964] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[1965] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[1966] Train Loss:    tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[1967] Train Loss:    tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[1968] Train Loss:    tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[1969] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[1970] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[1971] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[1972] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[1973] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[1974] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[1975] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[1976] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1977] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[1978] Train Loss:    tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[1979] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[1980] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[1981] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[1982] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[1983] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[1984] Train Loss:    tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[1985] Train Loss:    tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[1986] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[1987] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[1988] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[1989] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[1990] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[1991] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[1992] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[1993] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[1994] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[1995] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[1996] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[1997] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[1998] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[1999] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[2000] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[2001] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[2002] Train Loss:    tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[2003] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[2004] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[2005] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
[2006] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[2007] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[2008] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[2009] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[2010] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[2011] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[2012] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[2013] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[2014] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[2015] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[2016] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[2017] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[2018] Train Loss:    tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[2019] Train Loss:    tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[2020] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[2021] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[2022] Train Loss:    tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[2023] Train Loss:    tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[2024] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2025] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[2026] Train Loss:    tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[2027] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[2028] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[2029] Train Loss:    tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[2030] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[2031] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[2032] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[2033] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2034] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[2035] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[2036] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[2037] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[2038] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[2039] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[2040] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[2041] Train Loss:    tensor(0.8080, device='cuda:0', grad_fn=<NllLossBackward>)
[2042] Train Loss:    tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[2043] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[2044] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[2045] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[2046] Train Loss:    tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[2047] Train Loss:    tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
[2048] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[2049] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[2050] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[2051] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[2052] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[2053] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[2054] Train Loss:    tensor(0.7995, device='cuda:0', grad_fn=<NllLossBackward>)
[2055] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[2056] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[2057] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[2058] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[2059] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[2060] Train Loss:    tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
[2061] Train Loss:    tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[2062] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[2063] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[2064] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[2065] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[2066] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[2067] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[2068] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[2069] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[2070] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[2071] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[2072] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2073] Train Loss:   tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[2074] Train Loss:   tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[2075] Train Loss:   tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[2076] Train Loss:   tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[2077] Train Loss:   tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[2078] Train Loss:   tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[2079] Train Loss:   tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[2080] Train Loss:   tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[2081] Train Loss:   tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[2082] Train Loss:   tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[2083] Train Loss:   tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[2084] Train Loss:   tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[2085] Train Loss:   tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[2086] Train Loss:   tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[2087] Train Loss:   tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[2088] Train Loss:   tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[2089] Train Loss:   tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[2090] Train Loss:   tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[2091] Train Loss:   tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[2092] Train Loss:   tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[2093] Train Loss:   tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[2094] Train Loss:   tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[2095] Train Loss:   tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[2096] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[2097] Train Loss:   tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[2098] Train Loss:   tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[2099] Train Loss:   tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[2100] Train Loss:   tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[2101] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[2102] Train Loss:   tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[2103] Train Loss:   tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[2104] Train Loss:   tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[2105] Train Loss:   tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[2106] Train Loss:   tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[2107] Train Loss:   tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[2108] Train Loss:   tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[2109] Train Loss:   tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[2110] Train Loss:   tensor(0.8002, device='cuda:0', grad_fn=<NllLossBackward>)
[2111] Train Loss:   tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[2112] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[2113] Train Loss:   tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[2114] Train Loss:   tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[2115] Train Loss:   tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[2116] Train Loss:   tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[2117] Train Loss:   tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[2118] Train Loss:   tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[2119] Train Loss:   tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[2120] Train Loss:   tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2121] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[2122] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[2123] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[2124] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[2125] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[2126] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[2127] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2128] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[2129] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[2130] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[2131] Train Loss:    tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
[2132] Train Loss:    tensor(0.8029, device='cuda:0', grad_fn=<NllLossBackward>)
[2133] Train Loss:    tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[2134] Train Loss:    tensor(0.7939, device='cuda:0', grad_fn=<NllLossBackward>)
[2135] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[2136] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[2137] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[2138] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[2139] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[2140] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[2141] Train Loss:    tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[2142] Train Loss:    tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[2143] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[2144] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[2145] Train Loss:    tensor(0.7667, device='cuda:0', grad_fn=<NllLossBackward>)
[2146] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[2147] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[2148] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[2149] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[2150] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[2151] Train Loss:    tensor(0.8086, device='cuda:0', grad_fn=<NllLossBackward>)
[2152] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[2153] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[2154] Train Loss:    tensor(0.7985, device='cuda:0', grad_fn=<NllLossBackward>)
[2155] Train Loss:    tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[2156] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[2157] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[2158] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[2159] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[2160] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[2161] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[2162] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[2163] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[2164] Train Loss:    tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[2165] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[2166] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[2167] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[2168] Train Loss:    tensor(0.8001, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2169] Train Loss:   tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[2170] Train Loss:   tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[2171] Train Loss:   tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[2172] Train Loss:   tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[2173] Train Loss:   tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[2174] Train Loss:   tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[2175] Train Loss:   tensor(0.8038, device='cuda:0', grad_fn=<NllLossBackward>)
[2176] Train Loss:   tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[2177] Train Loss:   tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[2178] Train Loss:   tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[2179] Train Loss:   tensor(0.8051, device='cuda:0', grad_fn=<NllLossBackward>)
[2180] Train Loss:   tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[2181] Train Loss:   tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[2182] Train Loss:   tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[2183] Train Loss:   tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[2184] Train Loss:   tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[2185] Train Loss:   tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[2186] Train Loss:   tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[2187] Train Loss:   tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[2188] Train Loss:   tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[2189] Train Loss:   tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[2190] Train Loss:   tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[2191] Train Loss:   tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[2192] Train Loss:   tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
[2193] Train Loss:   tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[2194] Train Loss:   tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[2195] Train Loss:   tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[2196] Train Loss:   tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2197] Train Loss:   tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[2198] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[2199] Train Loss:   tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[2200] Train Loss:   tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[2201] Train Loss:   tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[2202] Train Loss:   tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[2203] Train Loss:   tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[2204] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[2205] Train Loss:   tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[2206] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[2207] Train Loss:   tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[2208] Train Loss:   tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[2209] Train Loss:   tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[2210] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[2211] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2212] Train Loss:   tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[2213] Train Loss:   tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[2214] Train Loss:   tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[2215] Train Loss:   tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[2216] Train Loss:   tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2217] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[2218] Train Loss:    tensor(0.7877, device='cuda:0', grad_fn=<NllLossBackward>)
[2219] Train Loss:    tensor(0.8019, device='cuda:0', grad_fn=<NllLossBackward>)
[2220] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[2221] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[2222] Train Loss:    tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[2223] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[2224] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[2225] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[2226] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2227] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[2228] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[2229] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[2230] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[2231] Train Loss:    tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[2232] Train Loss:    tensor(0.7986, device='cuda:0', grad_fn=<NllLossBackward>)
[2233] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[2234] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2235] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[2236] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[2237] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[2238] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[2239] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[2240] Train Loss:    tensor(0.7881, device='cuda:0', grad_fn=<NllLossBackward>)
[2241] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[2242] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[2243] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[2244] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[2245] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[2246] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[2247] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[2248] Train Loss:    tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[2249] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[2250] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[2251] Train Loss:    tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[2252] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[2253] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[2254] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[2255] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[2256] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[2257] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2258] Train Loss:    tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
[2259] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[2260] Train Loss:    tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[2261] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[2262] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[2263] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[2264] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2265] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[2266] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2267] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[2268] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[2269] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[2270] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[2271] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[2272] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[2273] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[2274] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[2275] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[2276] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[2277] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[2278] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[2279] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[2280] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[2281] Train Loss:    tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[2282] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[2283] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[2284] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[2285] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[2286] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[2287] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[2288] Train Loss:    tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[2289] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[2290] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[2291] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[2292] Train Loss:    tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[2293] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[2294] Train Loss:    tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[2295] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[2296] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[2297] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[2298] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[2299] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[2300] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[2301] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[2302] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[2303] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[2304] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[2305] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[2306] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[2307] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[2308] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[2309] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[2310] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[2311] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[2312] Train Loss:    tensor(0.7946, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2313] Train Loss:   tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[2314] Train Loss:   tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[2315] Train Loss:   tensor(0.7866, device='cuda:0', grad_fn=<NllLossBackward>)
[2316] Train Loss:   tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[2317] Train Loss:   tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[2318] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[2319] Train Loss:   tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
[2320] Train Loss:   tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[2321] Train Loss:   tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[2322] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[2323] Train Loss:   tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[2324] Train Loss:   tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[2325] Train Loss:   tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[2326] Train Loss:   tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[2327] Train Loss:   tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[2328] Train Loss:   tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[2329] Train Loss:   tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[2330] Train Loss:   tensor(0.8022, device='cuda:0', grad_fn=<NllLossBackward>)
[2331] Train Loss:   tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[2332] Train Loss:   tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[2333] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[2334] Train Loss:   tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[2335] Train Loss:   tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[2336] Train Loss:   tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[2337] Train Loss:   tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[2338] Train Loss:   tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[2339] Train Loss:   tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[2340] Train Loss:   tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[2341] Train Loss:   tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[2342] Train Loss:   tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[2343] Train Loss:   tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[2344] Train Loss:   tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[2345] Train Loss:   tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[2346] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[2347] Train Loss:   tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[2348] Train Loss:   tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[2349] Train Loss:   tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[2350] Train Loss:   tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[2351] Train Loss:   tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[2352] Train Loss:   tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2353] Train Loss:   tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[2354] Train Loss:   tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[2355] Train Loss:   tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[2356] Train Loss:   tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[2357] Train Loss:   tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[2358] Train Loss:   tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[2359] Train Loss:   tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[2360] Train Loss:   tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2361] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[2362] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[2363] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[2364] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[2365] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[2366] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[2367] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[2368] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[2369] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[2370] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[2371] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[2372] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[2373] Train Loss:    tensor(0.7935, device='cuda:0', grad_fn=<NllLossBackward>)
[2374] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[2375] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[2376] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2377] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[2378] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[2379] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[2380] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[2381] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[2382] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[2383] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[2384] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[2385] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2386] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[2387] Train Loss:    tensor(0.7968, device='cuda:0', grad_fn=<NllLossBackward>)
[2388] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2389] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[2390] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[2391] Train Loss:    tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[2392] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[2393] Train Loss:    tensor(0.8097, device='cuda:0', grad_fn=<NllLossBackward>)
[2394] Train Loss:    tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[2395] Train Loss:    tensor(0.7793, device='cuda:0', grad_fn=<NllLossBackward>)
[2396] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[2397] Train Loss:    tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[2398] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[2399] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[2400] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[2401] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[2402] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[2403] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2404] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[2405] Train Loss:    tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[2406] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[2407] Train Loss:    tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[2408] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2409] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[2410] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[2411] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[2412] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[2413] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[2414] Train Loss:    tensor(0.7837, device='cuda:0', grad_fn=<NllLossBackward>)
[2415] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[2416] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[2417] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[2418] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[2419] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[2420] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[2421] Train Loss:    tensor(0.7888, device='cuda:0', grad_fn=<NllLossBackward>)
[2422] Train Loss:    tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[2423] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[2424] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[2425] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[2426] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[2427] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[2428] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[2429] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[2430] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[2431] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[2432] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[2433] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[2434] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[2435] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[2436] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[2437] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[2438] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[2439] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[2440] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[2441] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[2442] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[2443] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[2444] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[2445] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[2446] Train Loss:    tensor(0.8101, device='cuda:0', grad_fn=<NllLossBackward>)
[2447] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[2448] Train Loss:    tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[2449] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[2450] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2451] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[2452] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[2453] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[2454] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[2455] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[2456] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2457] Train Loss:    tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[2458] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[2459] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[2460] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[2461] Train Loss:    tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[2462] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[2463] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[2464] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[2465] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[2466] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[2467] Train Loss:    tensor(0.8024, device='cuda:0', grad_fn=<NllLossBackward>)
[2468] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2469] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[2470] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[2471] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[2472] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[2473] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[2474] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[2475] Train Loss:    tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[2476] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2477] Train Loss:    tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[2478] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[2479] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[2480] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[2481] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[2482] Train Loss:    tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[2483] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[2484] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[2485] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[2486] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[2487] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[2488] Train Loss:    tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[2489] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[2490] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2491] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[2492] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[2493] Train Loss:    tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[2494] Train Loss:    tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[2495] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[2496] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[2497] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[2498] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[2499] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[2500] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[2501] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[2502] Train Loss:    tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[2503] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[2504] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2505] Train Loss:   tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[2506] Train Loss:   tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2507] Train Loss:   tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[2508] Train Loss:   tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[2509] Train Loss:   tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[2510] Train Loss:   tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[2511] Train Loss:   tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[2512] Train Loss:   tensor(0.7888, device='cuda:0', grad_fn=<NllLossBackward>)
[2513] Train Loss:   tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[2514] Train Loss:   tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[2515] Train Loss:   tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[2516] Train Loss:   tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[2517] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[2518] Train Loss:   tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[2519] Train Loss:   tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[2520] Train Loss:   tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[2521] Train Loss:   tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[2522] Train Loss:   tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[2523] Train Loss:   tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[2524] Train Loss:   tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[2525] Train Loss:   tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[2526] Train Loss:   tensor(0.7858, device='cuda:0', grad_fn=<NllLossBackward>)
[2527] Train Loss:   tensor(0.7959, device='cuda:0', grad_fn=<NllLossBackward>)
[2528] Train Loss:   tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[2529] Train Loss:   tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[2530] Train Loss:   tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[2531] Train Loss:   tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[2532] Train Loss:   tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[2533] Train Loss:   tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[2534] Train Loss:   tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[2535] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[2536] Train Loss:   tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[2537] Train Loss:   tensor(0.7844, device='cuda:0', grad_fn=<NllLossBackward>)
[2538] Train Loss:   tensor(0.8010, device='cuda:0', grad_fn=<NllLossBackward>)
[2539] Train Loss:   tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[2540] Train Loss:   tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[2541] Train Loss:   tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[2542] Train Loss:   tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[2543] Train Loss:   tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[2544] Train Loss:   tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[2545] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2546] Train Loss:   tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[2547] Train Loss:   tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[2548] Train Loss:   tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[2549] Train Loss:   tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[2550] Train Loss:   tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[2551] Train Loss:   tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[2552] Train Loss:   tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2553] Train Loss:   tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[2554] Train Loss:   tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[2555] Train Loss:   tensor(0.7810, device='cuda:0', grad_fn=<NllLossBackward>)
[2556] Train Loss:   tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[2557] Train Loss:   tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[2558] Train Loss:   tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[2559] Train Loss:   tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[2560] Train Loss:   tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[2561] Train Loss:   tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[2562] Train Loss:   tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[2563] Train Loss:   tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[2564] Train Loss:   tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[2565] Train Loss:   tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[2566] Train Loss:   tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[2567] Train Loss:   tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[2568] Train Loss:   tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[2569] Train Loss:   tensor(0.8081, device='cuda:0', grad_fn=<NllLossBackward>)
[2570] Train Loss:   tensor(0.7919, device='cuda:0', grad_fn=<NllLossBackward>)
[2571] Train Loss:   tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[2572] Train Loss:   tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[2573] Train Loss:   tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[2574] Train Loss:   tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[2575] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2576] Train Loss:   tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[2577] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[2578] Train Loss:   tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[2579] Train Loss:   tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[2580] Train Loss:   tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[2581] Train Loss:   tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[2582] Train Loss:   tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[2583] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[2584] Train Loss:   tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[2585] Train Loss:   tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[2586] Train Loss:   tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[2587] Train Loss:   tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2588] Train Loss:   tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[2589] Train Loss:   tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[2590] Train Loss:   tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[2591] Train Loss:   tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[2592] Train Loss:   tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[2593] Train Loss:   tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[2594] Train Loss:   tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[2595] Train Loss:   tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[2596] Train Loss:   tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[2597] Train Loss:   tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[2598] Train Loss:   tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[2599] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[2600] Train Loss:   tensor(0.7801, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2601] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[2602] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[2603] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[2604] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[2605] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[2606] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[2607] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[2608] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[2609] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[2610] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[2611] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[2612] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[2613] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[2614] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[2615] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[2616] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[2617] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[2618] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[2619] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[2620] Train Loss:    tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[2621] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[2622] Train Loss:    tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[2623] Train Loss:    tensor(0.7958, device='cuda:0', grad_fn=<NllLossBackward>)
[2624] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[2625] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[2626] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[2627] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[2628] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[2629] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[2630] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[2631] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[2632] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[2633] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[2634] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[2635] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[2636] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[2637] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[2638] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[2639] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[2640] Train Loss:    tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[2641] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[2642] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[2643] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[2644] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[2645] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[2646] Train Loss:    tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[2647] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[2648] Train Loss:    tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2649] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[2650] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[2651] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[2652] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[2653] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[2654] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[2655] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[2656] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[2657] Train Loss:    tensor(0.8200, device='cuda:0', grad_fn=<NllLossBackward>)
[2658] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[2659] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[2660] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[2661] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[2662] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[2663] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[2664] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[2665] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[2666] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[2667] Train Loss:    tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[2668] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2669] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[2670] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[2671] Train Loss:    tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[2672] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[2673] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[2674] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[2675] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[2676] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[2677] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[2678] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[2679] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[2680] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[2681] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[2682] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[2683] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2684] Train Loss:    tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[2685] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[2686] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[2687] Train Loss:    tensor(0.7953, device='cuda:0', grad_fn=<NllLossBackward>)
[2688] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[2689] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[2690] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[2691] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[2692] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[2693] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[2694] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[2695] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[2696] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2697] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2698] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[2699] Train Loss:    tensor(0.7732, device='cuda:0', grad_fn=<NllLossBackward>)
[2700] Train Loss:    tensor(0.7864, device='cuda:0', grad_fn=<NllLossBackward>)
[2701] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[2702] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[2703] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[2704] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[2705] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[2706] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[2707] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[2708] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[2709] Train Loss:    tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[2710] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[2711] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[2712] Train Loss:    tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[2713] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[2714] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[2715] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[2716] Train Loss:    tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[2717] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[2718] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[2719] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[2720] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[2721] Train Loss:    tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[2722] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[2723] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2724] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[2725] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[2726] Train Loss:    tensor(0.8182, device='cuda:0', grad_fn=<NllLossBackward>)
[2727] Train Loss:    tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[2728] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[2729] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[2730] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[2731] Train Loss:    tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[2732] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[2733] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[2734] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[2735] Train Loss:    tensor(0.7731, device='cuda:0', grad_fn=<NllLossBackward>)
[2736] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[2737] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[2738] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[2739] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[2740] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[2741] Train Loss:    tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[2742] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[2743] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[2744] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2745] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[2746] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[2747] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[2748] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[2749] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[2750] Train Loss:    tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[2751] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[2752] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[2753] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2754] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[2755] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[2756] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[2757] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[2758] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[2759] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[2760] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[2761] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[2762] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[2763] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[2764] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[2765] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[2766] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2767] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[2768] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[2769] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[2770] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[2771] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[2772] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[2773] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[2774] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[2775] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[2776] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[2777] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[2778] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[2779] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[2780] Train Loss:    tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[2781] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[2782] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[2783] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[2784] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[2785] Train Loss:    tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[2786] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[2787] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[2788] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[2789] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[2790] Train Loss:    tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[2791] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[2792] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2793] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[2794] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[2795] Train Loss:    tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[2796] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[2797] Train Loss:    tensor(0.7793, device='cuda:0', grad_fn=<NllLossBackward>)
[2798] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[2799] Train Loss:    tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[2800] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[2801] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[2802] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[2803] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[2804] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[2805] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[2806] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[2807] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[2808] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[2809] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[2810] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[2811] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[2812] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[2813] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[2814] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[2815] Train Loss:    tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[2816] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[2817] Train Loss:    tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[2818] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[2819] Train Loss:    tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[2820] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[2821] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[2822] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[2823] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[2824] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2825] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[2826] Train Loss:    tensor(0.8055, device='cuda:0', grad_fn=<NllLossBackward>)
[2827] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[2828] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[2829] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[2830] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[2831] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[2832] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[2833] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[2834] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[2835] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[2836] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[2837] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[2838] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[2839] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[2840] Train Loss:    tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2841] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[2842] Train Loss:    tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[2843] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[2844] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[2845] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[2846] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[2847] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[2848] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[2849] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[2850] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[2851] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[2852] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2853] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[2854] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[2855] Train Loss:    tensor(0.8050, device='cuda:0', grad_fn=<NllLossBackward>)
[2856] Train Loss:    tensor(0.7858, device='cuda:0', grad_fn=<NllLossBackward>)
[2857] Train Loss:    tensor(0.7835, device='cuda:0', grad_fn=<NllLossBackward>)
[2858] Train Loss:    tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[2859] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[2860] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[2861] Train Loss:    tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[2862] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[2863] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[2864] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[2865] Train Loss:    tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[2866] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[2867] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[2868] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[2869] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[2870] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[2871] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[2872] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[2873] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[2874] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[2875] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[2876] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[2877] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[2878] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[2879] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[2880] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[2881] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[2882] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[2883] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[2884] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[2885] Train Loss:    tensor(0.7991, device='cuda:0', grad_fn=<NllLossBackward>)
[2886] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[2887] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[2888] Train Loss:    tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2889] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[2890] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[2891] Train Loss:    tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[2892] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[2893] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[2894] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[2895] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[2896] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[2897] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[2898] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[2899] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2900] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[2901] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[2902] Train Loss:    tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[2903] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[2904] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[2905] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[2906] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[2907] Train Loss:    tensor(0.8050, device='cuda:0', grad_fn=<NllLossBackward>)
[2908] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[2909] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[2910] Train Loss:    tensor(0.7879, device='cuda:0', grad_fn=<NllLossBackward>)
[2911] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[2912] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[2913] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[2914] Train Loss:    tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[2915] Train Loss:    tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[2916] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[2917] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[2918] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[2919] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[2920] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[2921] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[2922] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[2923] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[2924] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[2925] Train Loss:    tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[2926] Train Loss:    tensor(0.8143, device='cuda:0', grad_fn=<NllLossBackward>)
[2927] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[2928] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[2929] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[2930] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[2931] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[2932] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[2933] Train Loss:    tensor(0.7942, device='cuda:0', grad_fn=<NllLossBackward>)
[2934] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[2935] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[2936] Train Loss:    tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2937] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2938] Train Loss:    tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[2939] Train Loss:    tensor(0.7853, device='cuda:0', grad_fn=<NllLossBackward>)
[2940] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[2941] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[2942] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[2943] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[2944] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[2945] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[2946] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[2947] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[2948] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[2949] Train Loss:    tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[2950] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[2951] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[2952] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[2953] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[2954] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[2955] Train Loss:    tensor(0.7858, device='cuda:0', grad_fn=<NllLossBackward>)
[2956] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[2957] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[2958] Train Loss:    tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[2959] Train Loss:    tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[2960] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[2961] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[2962] Train Loss:    tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[2963] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[2964] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[2965] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[2966] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[2967] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[2968] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[2969] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[2970] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[2971] Train Loss:    tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[2972] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[2973] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[2974] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[2975] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[2976] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[2977] Train Loss:    tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[2978] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[2979] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[2980] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[2981] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[2982] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[2983] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[2984] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2985] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[2986] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[2987] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[2988] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[2989] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[2990] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[2991] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[2992] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[2993] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[2994] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[2995] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[2996] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[2997] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[2998] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[2999] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[3000] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3001] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[3002] Train Loss:    tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[3003] Train Loss:    tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[3004] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[3005] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[3006] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[3007] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[3008] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[3009] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[3010] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[3011] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[3012] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[3013] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[3014] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[3015] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[3016] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[3017] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[3018] Train Loss:    tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[3019] Train Loss:    tensor(0.8116, device='cuda:0', grad_fn=<NllLossBackward>)
[3020] Train Loss:    tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[3021] Train Loss:    tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[3022] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[3023] Train Loss:    tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[3024] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[3025] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[3026] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[3027] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[3028] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[3029] Train Loss:    tensor(0.7731, device='cuda:0', grad_fn=<NllLossBackward>)
[3030] Train Loss:    tensor(0.8113, device='cuda:0', grad_fn=<NllLossBackward>)
[3031] Train Loss:    tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[3032] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3033] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[3034] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[3035] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[3036] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[3037] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[3038] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[3039] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[3040] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[3041] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[3042] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[3043] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[3044] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[3045] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[3046] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[3047] Train Loss:    tensor(0.7789, device='cuda:0', grad_fn=<NllLossBackward>)
[3048] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[3049] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[3050] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[3051] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[3052] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[3053] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[3054] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3055] Train Loss:    tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[3056] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[3057] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[3058] Train Loss:    tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[3059] Train Loss:    tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[3060] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[3061] Train Loss:    tensor(0.8023, device='cuda:0', grad_fn=<NllLossBackward>)
[3062] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3063] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[3064] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[3065] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[3066] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[3067] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[3068] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[3069] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[3070] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[3071] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[3072] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[3073] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[3074] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[3075] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[3076] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[3077] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[3078] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[3079] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[3080] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3081] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[3082] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[3083] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[3084] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[3085] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[3086] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[3087] Train Loss:    tensor(0.7838, device='cuda:0', grad_fn=<NllLossBackward>)
[3088] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3089] Train Loss:    tensor(0.7962, device='cuda:0', grad_fn=<NllLossBackward>)
[3090] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[3091] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[3092] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[3093] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[3094] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[3095] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[3096] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[3097] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[3098] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[3099] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[3100] Train Loss:    tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[3101] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[3102] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[3103] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[3104] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[3105] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[3106] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[3107] Train Loss:    tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[3108] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3109] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[3110] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3111] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[3112] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[3113] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3114] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[3115] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[3116] Train Loss:    tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[3117] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[3118] Train Loss:    tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[3119] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[3120] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[3121] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[3122] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[3123] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[3124] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[3125] Train Loss:    tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[3126] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[3127] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[3128] Train Loss:    tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3129] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[3130] Train Loss:    tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[3131] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[3132] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[3133] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[3134] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[3135] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3136] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[3137] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[3138] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3139] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[3140] Train Loss:    tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[3141] Train Loss:    tensor(0.7980, device='cuda:0', grad_fn=<NllLossBackward>)
[3142] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[3143] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[3144] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[3145] Train Loss:    tensor(0.7925, device='cuda:0', grad_fn=<NllLossBackward>)
[3146] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[3147] Train Loss:    tensor(0.8008, device='cuda:0', grad_fn=<NllLossBackward>)
[3148] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[3149] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[3150] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[3151] Train Loss:    tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[3152] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[3153] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[3154] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[3155] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[3156] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[3157] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[3158] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3159] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[3160] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[3161] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[3162] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[3163] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[3164] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[3165] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[3166] Train Loss:    tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[3167] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[3168] Train Loss:    tensor(0.7951, device='cuda:0', grad_fn=<NllLossBackward>)
[3169] Train Loss:    tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[3170] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[3171] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[3172] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[3173] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[3174] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[3175] Train Loss:    tensor(0.7869, device='cuda:0', grad_fn=<NllLossBackward>)
[3176] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3177] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[3178] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[3179] Train Loss:    tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[3180] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[3181] Train Loss:    tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[3182] Train Loss:    tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[3183] Train Loss:    tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[3184] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[3185] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[3186] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[3187] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[3188] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[3189] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[3190] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[3191] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3192] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[3193] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[3194] Train Loss:    tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[3195] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[3196] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[3197] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[3198] Train Loss:    tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[3199] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[3200] Train Loss:    tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[3201] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[3202] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3203] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[3204] Train Loss:    tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[3205] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[3206] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[3207] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3208] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[3209] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[3210] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[3211] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3212] Train Loss:    tensor(0.7838, device='cuda:0', grad_fn=<NllLossBackward>)
[3213] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[3214] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[3215] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3216] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[3217] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[3218] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[3219] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[3220] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[3221] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[3222] Train Loss:    tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[3223] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[3224] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3225] Train Loss:   tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[3226] Train Loss:   tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[3227] Train Loss:   tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[3228] Train Loss:   tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[3229] Train Loss:   tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[3230] Train Loss:   tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[3231] Train Loss:   tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[3232] Train Loss:   tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[3233] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[3234] Train Loss:   tensor(0.7732, device='cuda:0', grad_fn=<NllLossBackward>)
[3235] Train Loss:   tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[3236] Train Loss:   tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[3237] Train Loss:   tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[3238] Train Loss:   tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[3239] Train Loss:   tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[3240] Train Loss:   tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[3241] Train Loss:   tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[3242] Train Loss:   tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[3243] Train Loss:   tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[3244] Train Loss:   tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[3245] Train Loss:   tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[3246] Train Loss:   tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[3247] Train Loss:   tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[3248] Train Loss:   tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[3249] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[3250] Train Loss:   tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[3251] Train Loss:   tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[3252] Train Loss:   tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[3253] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3254] Train Loss:   tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[3255] Train Loss:   tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[3256] Train Loss:   tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[3257] Train Loss:   tensor(0.7818, device='cuda:0', grad_fn=<NllLossBackward>)
[3258] Train Loss:   tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[3259] Train Loss:   tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[3260] Train Loss:   tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[3261] Train Loss:   tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[3262] Train Loss:   tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[3263] Train Loss:   tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[3264] Train Loss:   tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[3265] Train Loss:   tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[3266] Train Loss:   tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[3267] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[3268] Train Loss:   tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[3269] Train Loss:   tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[3270] Train Loss:   tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[3271] Train Loss:   tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[3272] Train Loss:   tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3273] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[3274] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[3275] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[3276] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[3277] Train Loss:    tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
[3278] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[3279] Train Loss:    tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[3280] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[3281] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3282] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[3283] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3284] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[3285] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[3286] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3287] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[3288] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[3289] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[3290] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[3291] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[3292] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3293] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[3294] Train Loss:    tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[3295] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[3296] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[3297] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[3298] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[3299] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3300] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[3301] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[3302] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[3303] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3304] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[3305] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[3306] Train Loss:    tensor(0.8032, device='cuda:0', grad_fn=<NllLossBackward>)
[3307] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[3308] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[3309] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[3310] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[3311] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[3312] Train Loss:    tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[3313] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[3314] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[3315] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[3316] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[3317] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[3318] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[3319] Train Loss:    tensor(0.8042, device='cuda:0', grad_fn=<NllLossBackward>)
[3320] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3321] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[3322] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3323] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[3324] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[3325] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[3326] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[3327] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[3328] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[3329] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[3330] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[3331] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[3332] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[3333] Train Loss:    tensor(0.7732, device='cuda:0', grad_fn=<NllLossBackward>)
[3334] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[3335] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3336] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[3337] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[3338] Train Loss:    tensor(0.8355, device='cuda:0', grad_fn=<NllLossBackward>)
[3339] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[3340] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[3341] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[3342] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[3343] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[3344] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[3345] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[3346] Train Loss:    tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[3347] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[3348] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[3349] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[3350] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[3351] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[3352] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[3353] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[3354] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[3355] Train Loss:    tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[3356] Train Loss:    tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[3357] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[3358] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[3359] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[3360] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[3361] Train Loss:    tensor(0.7940, device='cuda:0', grad_fn=<NllLossBackward>)
[3362] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[3363] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[3364] Train Loss:    tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[3365] Train Loss:    tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[3366] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[3367] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[3368] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3369] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[3370] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[3371] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[3372] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[3373] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[3374] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[3375] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[3376] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[3377] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[3378] Train Loss:    tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[3379] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[3380] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[3381] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[3382] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[3383] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3384] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[3385] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[3386] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[3387] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[3388] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[3389] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[3390] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[3391] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[3392] Train Loss:    tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[3393] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[3394] Train Loss:    tensor(0.8059, device='cuda:0', grad_fn=<NllLossBackward>)
[3395] Train Loss:    tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[3396] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[3397] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[3398] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[3399] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[3400] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[3401] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[3402] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[3403] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[3404] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[3405] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[3406] Train Loss:    tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[3407] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[3408] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[3409] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[3410] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[3411] Train Loss:    tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[3412] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[3413] Train Loss:    tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[3414] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[3415] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[3416] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3417] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[3418] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[3419] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[3420] Train Loss:    tensor(0.7923, device='cuda:0', grad_fn=<NllLossBackward>)
[3421] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[3422] Train Loss:    tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[3423] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[3424] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[3425] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[3426] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[3427] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[3428] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[3429] Train Loss:    tensor(0.7966, device='cuda:0', grad_fn=<NllLossBackward>)
[3430] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[3431] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[3432] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[3433] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[3434] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[3435] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[3436] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[3437] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[3438] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[3439] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[3440] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[3441] Train Loss:    tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[3442] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[3443] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[3444] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[3445] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[3446] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[3447] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[3448] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[3449] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[3450] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[3451] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[3452] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[3453] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[3454] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[3455] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[3456] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[3457] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[3458] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[3459] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[3460] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[3461] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3462] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[3463] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3464] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3465] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[3466] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[3467] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[3468] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[3469] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3470] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[3471] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3472] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[3473] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[3474] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[3475] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[3476] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[3477] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[3478] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3479] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[3480] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[3481] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[3482] Train Loss:    tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[3483] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[3484] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[3485] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[3486] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[3487] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[3488] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3489] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[3490] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[3491] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[3492] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[3493] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[3494] Train Loss:    tensor(0.7876, device='cuda:0', grad_fn=<NllLossBackward>)
[3495] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3496] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[3497] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[3498] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[3499] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[3500] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[3501] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[3502] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[3503] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[3504] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[3505] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3506] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3507] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[3508] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[3509] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[3510] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[3511] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[3512] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3513] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[3514] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[3515] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[3516] Train Loss:    tensor(0.7918, device='cuda:0', grad_fn=<NllLossBackward>)
[3517] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3518] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[3519] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[3520] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[3521] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[3522] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[3523] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[3524] Train Loss:    tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[3525] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[3526] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3527] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[3528] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[3529] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[3530] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[3531] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[3532] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[3533] Train Loss:    tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
[3534] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[3535] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[3536] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[3537] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3538] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[3539] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[3540] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3541] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[3542] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[3543] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[3544] Train Loss:    tensor(0.7873, device='cuda:0', grad_fn=<NllLossBackward>)
[3545] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[3546] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[3547] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[3548] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[3549] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[3550] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[3551] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[3552] Train Loss:    tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[3553] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[3554] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[3555] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[3556] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[3557] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[3558] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[3559] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[3560] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3561] Train Loss:   tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[3562] Train Loss:   tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[3563] Train Loss:   tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[3564] Train Loss:   tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3565] Train Loss:   tensor(0.7908, device='cuda:0', grad_fn=<NllLossBackward>)
[3566] Train Loss:   tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[3567] Train Loss:   tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3568] Train Loss:   tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[3569] Train Loss:   tensor(0.7806, device='cuda:0', grad_fn=<NllLossBackward>)
[3570] Train Loss:   tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[3571] Train Loss:   tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[3572] Train Loss:   tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[3573] Train Loss:   tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[3574] Train Loss:   tensor(0.8004, device='cuda:0', grad_fn=<NllLossBackward>)
[3575] Train Loss:   tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[3576] Train Loss:   tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[3577] Train Loss:   tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[3578] Train Loss:   tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[3579] Train Loss:   tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[3580] Train Loss:   tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[3581] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[3582] Train Loss:   tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[3583] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[3584] Train Loss:   tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[3585] Train Loss:   tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[3586] Train Loss:   tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[3587] Train Loss:   tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[3588] Train Loss:   tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[3589] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[3590] Train Loss:   tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3591] Train Loss:   tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[3592] Train Loss:   tensor(0.8016, device='cuda:0', grad_fn=<NllLossBackward>)
[3593] Train Loss:   tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[3594] Train Loss:   tensor(0.8151, device='cuda:0', grad_fn=<NllLossBackward>)
[3595] Train Loss:   tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[3596] Train Loss:   tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[3597] Train Loss:   tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[3598] Train Loss:   tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[3599] Train Loss:   tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[3600] Train Loss:   tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[3601] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[3602] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[3603] Train Loss:   tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[3604] Train Loss:   tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[3605] Train Loss:   tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[3606] Train Loss:   tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[3607] Train Loss:   tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3608] Train Loss:   tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3609] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[3610] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[3611] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[3612] Train Loss:    tensor(0.7950, device='cuda:0', grad_fn=<NllLossBackward>)
[3613] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[3614] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[3615] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[3616] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[3617] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[3618] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[3619] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[3620] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[3621] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[3622] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[3623] Train Loss:    tensor(0.7582, device='cuda:0', grad_fn=<NllLossBackward>)
[3624] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[3625] Train Loss:    tensor(0.7884, device='cuda:0', grad_fn=<NllLossBackward>)
[3626] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[3627] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[3628] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[3629] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[3630] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[3631] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[3632] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[3633] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[3634] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[3635] Train Loss:    tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[3636] Train Loss:    tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[3637] Train Loss:    tensor(0.7994, device='cuda:0', grad_fn=<NllLossBackward>)
[3638] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[3639] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3640] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3641] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[3642] Train Loss:    tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[3643] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3644] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[3645] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[3646] Train Loss:    tensor(0.7844, device='cuda:0', grad_fn=<NllLossBackward>)
[3647] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[3648] Train Loss:    tensor(0.8066, device='cuda:0', grad_fn=<NllLossBackward>)
[3649] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[3650] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[3651] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[3652] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[3653] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[3654] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[3655] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[3656] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3657] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[3658] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[3659] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[3660] Train Loss:    tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[3661] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[3662] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[3663] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[3664] Train Loss:    tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[3665] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[3666] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[3667] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[3668] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[3669] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[3670] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[3671] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[3672] Train Loss:    tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[3673] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[3674] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[3675] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[3676] Train Loss:    tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[3677] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[3678] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[3679] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[3680] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[3681] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[3682] Train Loss:    tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[3683] Train Loss:    tensor(0.7916, device='cuda:0', grad_fn=<NllLossBackward>)
[3684] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[3685] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3686] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[3687] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[3688] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[3689] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[3690] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[3691] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[3692] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[3693] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[3694] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[3695] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[3696] Train Loss:    tensor(0.7904, device='cuda:0', grad_fn=<NllLossBackward>)
[3697] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[3698] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[3699] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[3700] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[3701] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[3702] Train Loss:    tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[3703] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[3704] Train Loss:    tensor(0.7890, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3705] Train Loss:   tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[3706] Train Loss:   tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[3707] Train Loss:   tensor(0.8093, device='cuda:0', grad_fn=<NllLossBackward>)
[3708] Train Loss:   tensor(0.7988, device='cuda:0', grad_fn=<NllLossBackward>)
[3709] Train Loss:   tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[3710] Train Loss:   tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[3711] Train Loss:   tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[3712] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[3713] Train Loss:   tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3714] Train Loss:   tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[3715] Train Loss:   tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[3716] Train Loss:   tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[3717] Train Loss:   tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[3718] Train Loss:   tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[3719] Train Loss:   tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[3720] Train Loss:   tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[3721] Train Loss:   tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[3722] Train Loss:   tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[3723] Train Loss:   tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[3724] Train Loss:   tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[3725] Train Loss:   tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[3726] Train Loss:   tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[3727] Train Loss:   tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[3728] Train Loss:   tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[3729] Train Loss:   tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3730] Train Loss:   tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[3731] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[3732] Train Loss:   tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[3733] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[3734] Train Loss:   tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[3735] Train Loss:   tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[3736] Train Loss:   tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[3737] Train Loss:   tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[3738] Train Loss:   tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[3739] Train Loss:   tensor(0.7952, device='cuda:0', grad_fn=<NllLossBackward>)
[3740] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[3741] Train Loss:   tensor(0.8035, device='cuda:0', grad_fn=<NllLossBackward>)
[3742] Train Loss:   tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[3743] Train Loss:   tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[3744] Train Loss:   tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[3745] Train Loss:   tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[3746] Train Loss:   tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[3747] Train Loss:   tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[3748] Train Loss:   tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[3749] Train Loss:   tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[3750] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[3751] Train Loss:   tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[3752] Train Loss:   tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3753] Train Loss:   tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[3754] Train Loss:   tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[3755] Train Loss:   tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[3756] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[3757] Train Loss:   tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[3758] Train Loss:   tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[3759] Train Loss:   tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[3760] Train Loss:   tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[3761] Train Loss:   tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[3762] Train Loss:   tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3763] Train Loss:   tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[3764] Train Loss:   tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[3765] Train Loss:   tensor(0.7863, device='cuda:0', grad_fn=<NllLossBackward>)
[3766] Train Loss:   tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[3767] Train Loss:   tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[3768] Train Loss:   tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[3769] Train Loss:   tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[3770] Train Loss:   tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[3771] Train Loss:   tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[3772] Train Loss:   tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[3773] Train Loss:   tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[3774] Train Loss:   tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[3775] Train Loss:   tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[3776] Train Loss:   tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[3777] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[3778] Train Loss:   tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[3779] Train Loss:   tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[3780] Train Loss:   tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[3781] Train Loss:   tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[3782] Train Loss:   tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[3783] Train Loss:   tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[3784] Train Loss:   tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[3785] Train Loss:   tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[3786] Train Loss:   tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[3787] Train Loss:   tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[3788] Train Loss:   tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[3789] Train Loss:   tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[3790] Train Loss:   tensor(0.8105, device='cuda:0', grad_fn=<NllLossBackward>)
[3791] Train Loss:   tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[3792] Train Loss:   tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[3793] Train Loss:   tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[3794] Train Loss:   tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[3795] Train Loss:   tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[3796] Train Loss:   tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[3797] Train Loss:   tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[3798] Train Loss:   tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[3799] Train Loss:   tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[3800] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3801] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[3802] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[3803] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[3804] Train Loss:    tensor(0.7917, device='cuda:0', grad_fn=<NllLossBackward>)
[3805] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[3806] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[3807] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[3808] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[3809] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[3810] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[3811] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[3812] Train Loss:    tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[3813] Train Loss:    tensor(0.7845, device='cuda:0', grad_fn=<NllLossBackward>)
[3814] Train Loss:    tensor(0.7688, device='cuda:0', grad_fn=<NllLossBackward>)
[3815] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[3816] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[3817] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[3818] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[3819] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[3820] Train Loss:    tensor(0.8094, device='cuda:0', grad_fn=<NllLossBackward>)
[3821] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3822] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[3823] Train Loss:    tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[3824] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[3825] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[3826] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[3827] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[3828] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[3829] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[3830] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[3831] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[3832] Train Loss:    tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[3833] Train Loss:    tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[3834] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[3835] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[3836] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[3837] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[3838] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[3839] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[3840] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[3841] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[3842] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[3843] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[3844] Train Loss:    tensor(0.7926, device='cuda:0', grad_fn=<NllLossBackward>)
[3845] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[3846] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[3847] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[3848] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3849] Train Loss:   tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[3850] Train Loss:   tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[3851] Train Loss:   tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[3852] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[3853] Train Loss:   tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[3854] Train Loss:   tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[3855] Train Loss:   tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[3856] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[3857] Train Loss:   tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[3858] Train Loss:   tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[3859] Train Loss:   tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[3860] Train Loss:   tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[3861] Train Loss:   tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[3862] Train Loss:   tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[3863] Train Loss:   tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[3864] Train Loss:   tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3865] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[3866] Train Loss:   tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[3867] Train Loss:   tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[3868] Train Loss:   tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[3869] Train Loss:   tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[3870] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[3871] Train Loss:   tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[3872] Train Loss:   tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[3873] Train Loss:   tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[3874] Train Loss:   tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[3875] Train Loss:   tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[3876] Train Loss:   tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[3877] Train Loss:   tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[3878] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[3879] Train Loss:   tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[3880] Train Loss:   tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[3881] Train Loss:   tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3882] Train Loss:   tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[3883] Train Loss:   tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[3884] Train Loss:   tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[3885] Train Loss:   tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[3886] Train Loss:   tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[3887] Train Loss:   tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[3888] Train Loss:   tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[3889] Train Loss:   tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[3890] Train Loss:   tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[3891] Train Loss:   tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[3892] Train Loss:   tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3893] Train Loss:   tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[3894] Train Loss:   tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[3895] Train Loss:   tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[3896] Train Loss:   tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3897] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[3898] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[3899] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[3900] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[3901] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[3902] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[3903] Train Loss:    tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[3904] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[3905] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[3906] Train Loss:    tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[3907] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[3908] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3909] Train Loss:    tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[3910] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[3911] Train Loss:    tensor(0.7949, device='cuda:0', grad_fn=<NllLossBackward>)
[3912] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[3913] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[3914] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[3915] Train Loss:    tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[3916] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[3917] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[3918] Train Loss:    tensor(0.7920, device='cuda:0', grad_fn=<NllLossBackward>)
[3919] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[3920] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[3921] Train Loss:    tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[3922] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[3923] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[3924] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[3925] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[3926] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[3927] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[3928] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[3929] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[3930] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[3931] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[3932] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[3933] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[3934] Train Loss:    tensor(0.7886, device='cuda:0', grad_fn=<NllLossBackward>)
[3935] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[3936] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[3937] Train Loss:    tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[3938] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[3939] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[3940] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[3941] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[3942] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[3943] Train Loss:    tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[3944] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3945] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[3946] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[3947] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[3948] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3949] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[3950] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[3951] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[3952] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[3953] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3954] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[3955] Train Loss:    tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[3956] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3957] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[3958] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[3959] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[3960] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[3961] Train Loss:    tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[3962] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[3963] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[3964] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[3965] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[3966] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[3967] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[3968] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[3969] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[3970] Train Loss:    tensor(0.7732, device='cuda:0', grad_fn=<NllLossBackward>)
[3971] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[3972] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[3973] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[3974] Train Loss:    tensor(0.7912, device='cuda:0', grad_fn=<NllLossBackward>)
[3975] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[3976] Train Loss:    tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[3977] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[3978] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[3979] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[3980] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[3981] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[3982] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[3983] Train Loss:    tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[3984] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[3985] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[3986] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[3987] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[3988] Train Loss:    tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[3989] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[3990] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3991] Train Loss:    tensor(0.7831, device='cuda:0', grad_fn=<NllLossBackward>)
[3992] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3993] Train Loss:    tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[3994] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[3995] Train Loss:    tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[3996] Train Loss:    tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[3997] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[3998] Train Loss:    tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[3999] Train Loss:    tensor(0.8067, device='cuda:0', grad_fn=<NllLossBackward>)
[4000] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[4001] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[4002] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[4003] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[4004] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[4005] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[4006] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[4007] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[4008] Train Loss:    tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[4009] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[4010] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[4011] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[4012] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[4013] Train Loss:    tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[4014] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[4015] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[4016] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[4017] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[4018] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[4019] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[4020] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[4021] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[4022] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[4023] Train Loss:    tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[4024] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[4025] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[4026] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[4027] Train Loss:    tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[4028] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[4029] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[4030] Train Loss:    tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[4031] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[4032] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[4033] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[4034] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[4035] Train Loss:    tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[4036] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[4037] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[4038] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[4039] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[4040] Train Loss:    tensor(0.7984, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4041] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[4042] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[4043] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[4044] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[4045] Train Loss:    tensor(0.7752, device='cuda:0', grad_fn=<NllLossBackward>)
[4046] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[4047] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[4048] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[4049] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[4050] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[4051] Train Loss:    tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[4052] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[4053] Train Loss:    tensor(0.7977, device='cuda:0', grad_fn=<NllLossBackward>)
[4054] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[4055] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[4056] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[4057] Train Loss:    tensor(0.7955, device='cuda:0', grad_fn=<NllLossBackward>)
[4058] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[4059] Train Loss:    tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
[4060] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[4061] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[4062] Train Loss:    tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[4063] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[4064] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[4065] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[4066] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[4067] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[4068] Train Loss:    tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[4069] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[4070] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[4071] Train Loss:    tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[4072] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[4073] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[4074] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[4075] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[4076] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[4077] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[4078] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[4079] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[4080] Train Loss:    tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[4081] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[4082] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[4083] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[4084] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[4085] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[4086] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[4087] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[4088] Train Loss:    tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4089] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[4090] Train Loss:    tensor(0.7828, device='cuda:0', grad_fn=<NllLossBackward>)
[4091] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[4092] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[4093] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[4094] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[4095] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[4096] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[4097] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[4098] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[4099] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[4100] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[4101] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[4102] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[4103] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[4104] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[4105] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[4106] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[4107] Train Loss:    tensor(0.7846, device='cuda:0', grad_fn=<NllLossBackward>)
[4108] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[4109] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[4110] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[4111] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[4112] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[4113] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[4114] Train Loss:    tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[4115] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[4116] Train Loss:    tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[4117] Train Loss:    tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[4118] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[4119] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[4120] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[4121] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[4122] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[4123] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[4124] Train Loss:    tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[4125] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[4126] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[4127] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[4128] Train Loss:    tensor(0.7903, device='cuda:0', grad_fn=<NllLossBackward>)
[4129] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[4130] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[4131] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[4132] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[4133] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[4134] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[4135] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[4136] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4137] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[4138] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[4139] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[4140] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[4141] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[4142] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[4143] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[4144] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[4145] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[4146] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[4147] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[4148] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[4149] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[4150] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[4151] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[4152] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[4153] Train Loss:    tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[4154] Train Loss:    tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[4155] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[4156] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[4157] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[4158] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[4159] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[4160] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[4161] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[4162] Train Loss:    tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[4163] Train Loss:    tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[4164] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[4165] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[4166] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[4167] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[4168] Train Loss:    tensor(0.7801, device='cuda:0', grad_fn=<NllLossBackward>)
[4169] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[4170] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[4171] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[4172] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[4173] Train Loss:    tensor(0.7983, device='cuda:0', grad_fn=<NllLossBackward>)
[4174] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[4175] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[4176] Train Loss:    tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[4177] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[4178] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[4179] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[4180] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[4181] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[4182] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[4183] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[4184] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4185] Train Loss:    tensor(0.7948, device='cuda:0', grad_fn=<NllLossBackward>)
[4186] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[4187] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[4188] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[4189] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[4190] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[4191] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[4192] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[4193] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[4194] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[4195] Train Loss:    tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[4196] Train Loss:    tensor(0.7816, device='cuda:0', grad_fn=<NllLossBackward>)
[4197] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[4198] Train Loss:    tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[4199] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[4200] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[4201] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[4202] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[4203] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[4204] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[4205] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[4206] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[4207] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[4208] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[4209] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[4210] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[4211] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[4212] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[4213] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[4214] Train Loss:    tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[4215] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[4216] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[4217] Train Loss:    tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[4218] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[4219] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[4220] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[4221] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[4222] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[4223] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[4224] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[4225] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[4226] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[4227] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[4228] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[4229] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[4230] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[4231] Train Loss:    tensor(0.7781, device='cuda:0', grad_fn=<NllLossBackward>)
[4232] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4233] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[4234] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[4235] Train Loss:    tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[4236] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[4237] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[4238] Train Loss:    tensor(0.7872, device='cuda:0', grad_fn=<NllLossBackward>)
[4239] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[4240] Train Loss:    tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[4241] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[4242] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[4243] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[4244] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[4245] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[4246] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[4247] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[4248] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[4249] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[4250] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[4251] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[4252] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[4253] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[4254] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[4255] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[4256] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[4257] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[4258] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[4259] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[4260] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[4261] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[4262] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[4263] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[4264] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[4265] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[4266] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[4267] Train Loss:    tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[4268] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[4269] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[4270] Train Loss:    tensor(0.7844, device='cuda:0', grad_fn=<NllLossBackward>)
[4271] Train Loss:    tensor(0.7973, device='cuda:0', grad_fn=<NllLossBackward>)
[4272] Train Loss:    tensor(0.7803, device='cuda:0', grad_fn=<NllLossBackward>)
[4273] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[4274] Train Loss:    tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[4275] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[4276] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[4277] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[4278] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[4279] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[4280] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4281] Train Loss:   tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[4282] Train Loss:   tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[4283] Train Loss:   tensor(0.7582, device='cuda:0', grad_fn=<NllLossBackward>)
[4284] Train Loss:   tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[4285] Train Loss:   tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[4286] Train Loss:   tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[4287] Train Loss:   tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[4288] Train Loss:   tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
[4289] Train Loss:   tensor(0.7848, device='cuda:0', grad_fn=<NllLossBackward>)
[4290] Train Loss:   tensor(0.8047, device='cuda:0', grad_fn=<NllLossBackward>)
[4291] Train Loss:   tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[4292] Train Loss:   tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[4293] Train Loss:   tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[4294] Train Loss:   tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[4295] Train Loss:   tensor(0.7582, device='cuda:0', grad_fn=<NllLossBackward>)
[4296] Train Loss:   tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[4297] Train Loss:   tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[4298] Train Loss:   tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[4299] Train Loss:   tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[4300] Train Loss:   tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[4301] Train Loss:   tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[4302] Train Loss:   tensor(0.7826, device='cuda:0', grad_fn=<NllLossBackward>)
[4303] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[4304] Train Loss:   tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[4305] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[4306] Train Loss:   tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[4307] Train Loss:   tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[4308] Train Loss:   tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[4309] Train Loss:   tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[4310] Train Loss:   tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[4311] Train Loss:   tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[4312] Train Loss:   tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[4313] Train Loss:   tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[4314] Train Loss:   tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[4315] Train Loss:   tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[4316] Train Loss:   tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[4317] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[4318] Train Loss:   tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[4319] Train Loss:   tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[4320] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[4321] Train Loss:   tensor(0.7960, device='cuda:0', grad_fn=<NllLossBackward>)
[4322] Train Loss:   tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[4323] Train Loss:   tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[4324] Train Loss:   tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[4325] Train Loss:   tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[4326] Train Loss:   tensor(0.7997, device='cuda:0', grad_fn=<NllLossBackward>)
[4327] Train Loss:   tensor(0.7829, device='cuda:0', grad_fn=<NllLossBackward>)
[4328] Train Loss:   tensor(0.7888, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4329] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[4330] Train Loss:    tensor(0.7880, device='cuda:0', grad_fn=<NllLossBackward>)
[4331] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[4332] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[4333] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[4334] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[4335] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[4336] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[4337] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[4338] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[4339] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[4340] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[4341] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[4342] Train Loss:    tensor(0.7738, device='cuda:0', grad_fn=<NllLossBackward>)
[4343] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[4344] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[4345] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[4346] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[4347] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[4348] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[4349] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[4350] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[4351] Train Loss:    tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[4352] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[4353] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[4354] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[4355] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[4356] Train Loss:    tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[4357] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[4358] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[4359] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[4360] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[4361] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[4362] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[4363] Train Loss:    tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[4364] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[4365] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[4366] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[4367] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[4368] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[4369] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[4370] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[4371] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[4372] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[4373] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[4374] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[4375] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[4376] Train Loss:    tensor(0.8020, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4377] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[4378] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[4379] Train Loss:    tensor(0.7820, device='cuda:0', grad_fn=<NllLossBackward>)
[4380] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[4381] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[4382] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[4383] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[4384] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[4385] Train Loss:    tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[4386] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[4387] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[4388] Train Loss:    tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[4389] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[4390] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[4391] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[4392] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[4393] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[4394] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[4395] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[4396] Train Loss:    tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[4397] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[4398] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[4399] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[4400] Train Loss:    tensor(0.8096, device='cuda:0', grad_fn=<NllLossBackward>)
[4401] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[4402] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[4403] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[4404] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[4405] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[4406] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[4407] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[4408] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[4409] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[4410] Train Loss:    tensor(0.7874, device='cuda:0', grad_fn=<NllLossBackward>)
[4411] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[4412] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[4413] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[4414] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[4415] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[4416] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[4417] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[4418] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[4419] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[4420] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[4421] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[4422] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[4423] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[4424] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4425] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[4426] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[4427] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[4428] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[4429] Train Loss:    tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[4430] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[4431] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[4432] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[4433] Train Loss:    tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[4434] Train Loss:    tensor(0.7897, device='cuda:0', grad_fn=<NllLossBackward>)
[4435] Train Loss:    tensor(0.7841, device='cuda:0', grad_fn=<NllLossBackward>)
[4436] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[4437] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[4438] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[4439] Train Loss:    tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[4440] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[4441] Train Loss:    tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[4442] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[4443] Train Loss:    tensor(0.7938, device='cuda:0', grad_fn=<NllLossBackward>)
[4444] Train Loss:    tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[4445] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[4446] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[4447] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[4448] Train Loss:    tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[4449] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[4450] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[4451] Train Loss:    tensor(0.7930, device='cuda:0', grad_fn=<NllLossBackward>)
[4452] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[4453] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[4454] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[4455] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[4456] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[4457] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[4458] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[4459] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[4460] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[4461] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[4462] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[4463] Train Loss:    tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[4464] Train Loss:    tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
[4465] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[4466] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[4467] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[4468] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[4469] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[4470] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[4471] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[4472] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4473] Train Loss:   tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[4474] Train Loss:   tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[4475] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[4476] Train Loss:   tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[4477] Train Loss:   tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[4478] Train Loss:   tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[4479] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[4480] Train Loss:   tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[4481] Train Loss:   tensor(0.8187, device='cuda:0', grad_fn=<NllLossBackward>)
[4482] Train Loss:   tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[4483] Train Loss:   tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[4484] Train Loss:   tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[4485] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[4486] Train Loss:   tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[4487] Train Loss:   tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[4488] Train Loss:   tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[4489] Train Loss:   tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[4490] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[4491] Train Loss:   tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[4492] Train Loss:   tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[4493] Train Loss:   tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[4494] Train Loss:   tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[4495] Train Loss:   tensor(0.7804, device='cuda:0', grad_fn=<NllLossBackward>)
[4496] Train Loss:   tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[4497] Train Loss:   tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[4498] Train Loss:   tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[4499] Train Loss:   tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[4500] Train Loss:   tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[4501] Train Loss:   tensor(0.7963, device='cuda:0', grad_fn=<NllLossBackward>)
[4502] Train Loss:   tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[4503] Train Loss:   tensor(0.7900, device='cuda:0', grad_fn=<NllLossBackward>)
[4504] Train Loss:   tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[4505] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[4506] Train Loss:   tensor(0.8000, device='cuda:0', grad_fn=<NllLossBackward>)
[4507] Train Loss:   tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[4508] Train Loss:   tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[4509] Train Loss:   tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[4510] Train Loss:   tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[4511] Train Loss:   tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
Training Loss:  0.7603523143375597 Time:  2493.4964122772217 s
Training Accuaracy:  76.68655611944537 %
[0] Valid Loss:  tensor(0.7201, device='cuda:0')
[1] Valid Loss:  tensor(0.8828, device='cuda:0')
[2] Valid Loss:  tensor(0.7082, device='cuda:0')
[3] Valid Loss:  tensor(0.6290, device='cuda:0')
[4] Valid Loss:  tensor(0.7511, device='cuda:0')
[5] Valid Loss:  tensor(0.7353, device='cuda:0')
[6] Valid Loss:  tensor(0.7437, device='cuda:0')
```

```
[7]  Valid Loss:   tensor(0.8782, device='cuda:0')
[8]  Valid Loss:   tensor(0.8083, device='cuda:0')
[9]  Valid Loss:   tensor(0.8385, device='cuda:0')
[10] Valid Loss:   tensor(0.7014, device='cuda:0')
[11] Valid Loss:   tensor(0.6873, device='cuda:0')
[12] Valid Loss:   tensor(1.1374, device='cuda:0')
[13] Valid Loss:   tensor(1.0091, device='cuda:0')
[14] Valid Loss:   tensor(0.6497, device='cuda:0')
[15] Valid Loss:   tensor(0.7789, device='cuda:0')
[16] Valid Loss:   tensor(0.6454, device='cuda:0')
[17] Valid Loss:   tensor(0.7934, device='cuda:0')
[18] Valid Loss:   tensor(0.7027, device='cuda:0')
[19] Valid Loss:   tensor(0.6548, device='cuda:0')
[20] Valid Loss:   tensor(0.7075, device='cuda:0')
[21] Valid Loss:   tensor(0.6242, device='cuda:0')
[22] Valid Loss:   tensor(0.5962, device='cuda:0')
[23] Valid Loss:   tensor(0.8443, device='cuda:0')
[24] Valid Loss:   tensor(0.7425, device='cuda:0')
[25] Valid Loss:   tensor(0.9658, device='cuda:0')
[26] Valid Loss:   tensor(0.5521, device='cuda:0')
[27] Valid Loss:   tensor(0.9105, device='cuda:0')
[28] Valid Loss:   tensor(0.7132, device='cuda:0')
[29] Valid Loss:   tensor(0.7300, device='cuda:0')
[30] Valid Loss:   tensor(0.6405, device='cuda:0')
[31] Valid Loss:   tensor(0.7675, device='cuda:0')
[32] Valid Loss:   tensor(0.8494, device='cuda:0')
[33] Valid Loss:   tensor(0.8410, device='cuda:0')
[34] Valid Loss:   tensor(0.8007, device='cuda:0')
[35] Valid Loss:   tensor(0.8899, device='cuda:0')
[36] Valid Loss:   tensor(0.7012, device='cuda:0')
[37] Valid Loss:   tensor(1.0222, device='cuda:0')
[38] Valid Loss:   tensor(0.8763, device='cuda:0')
[39] Valid Loss:   tensor(0.6841, device='cuda:0')
[40] Valid Loss:   tensor(0.7875, device='cuda:0')
[41] Valid Loss:   tensor(0.8622, device='cuda:0')
[42] Valid Loss:   tensor(0.7937, device='cuda:0')
[43] Valid Loss:   tensor(0.5833, device='cuda:0')
[44] Valid Loss:   tensor(0.7065, device='cuda:0')
[45] Valid Loss:   tensor(0.6089, device='cuda:0')
[46] Valid Loss:   tensor(0.9137, device='cuda:0')
[47] Valid Loss:   tensor(0.8283, device='cuda:0')
[48] Valid Loss:   tensor(0.6985, device='cuda:0')
[49] Valid Loss:   tensor(0.7874, device='cuda:0')
[50] Valid Loss:   tensor(0.7193, device='cuda:0')
[51] Valid Loss:   tensor(0.6779, device='cuda:0')
[52] Valid Loss:   tensor(0.6352, device='cuda:0')
[53] Valid Loss:   tensor(0.9174, device='cuda:0')
[54] Valid Loss:   tensor(0.6220, device='cuda:0')
```

```
[55] Valid Loss:  tensor(0.5916, device='cuda:0')
[56] Valid Loss:  tensor(0.8975, device='cuda:0')
[57] Valid Loss:  tensor(0.8454, device='cuda:0')
[58] Valid Loss:  tensor(0.7903, device='cuda:0')
[59] Valid Loss:  tensor(0.6861, device='cuda:0')
[60] Valid Loss:  tensor(1.0760, device='cuda:0')
[61] Valid Loss:  tensor(0.7164, device='cuda:0')
[62] Valid Loss:  tensor(0.8858, device='cuda:0')
[63] Valid Loss:  tensor(0.6905, device='cuda:0')
[64] Valid Loss:  tensor(0.7574, device='cuda:0')
[65] Valid Loss:  tensor(0.9010, device='cuda:0')
[66] Valid Loss:  tensor(0.7595, device='cuda:0')
[67] Valid Loss:  tensor(0.7702, device='cuda:0')
[68] Valid Loss:  tensor(0.7378, device='cuda:0')
[69] Valid Loss:  tensor(0.5997, device='cuda:0')
[70] Valid Loss:  tensor(0.6098, device='cuda:0')
[71] Valid Loss:  tensor(0.8506, device='cuda:0')
[72] Valid Loss:  tensor(0.7623, device='cuda:0')
[73] Valid Loss:  tensor(0.7168, device='cuda:0')
[74] Valid Loss:  tensor(0.7824, device='cuda:0')
[75] Valid Loss:  tensor(0.8658, device='cuda:0')
[76] Valid Loss:  tensor(0.8100, device='cuda:0')
[77] Valid Loss:  tensor(0.7298, device='cuda:0')
[78] Valid Loss:  tensor(0.8224, device='cuda:0')
[79] Valid Loss:  tensor(0.6463, device='cuda:0')
[80] Valid Loss:  tensor(0.7504, device='cuda:0')
[81] Valid Loss:  tensor(0.6408, device='cuda:0')
[82] Valid Loss:  tensor(0.8904, device='cuda:0')
[83] Valid Loss:  tensor(0.8309, device='cuda:0')
[84] Valid Loss:  tensor(0.6366, device='cuda:0')
[85] Valid Loss:  tensor(0.7572, device='cuda:0')
[86] Valid Loss:  tensor(0.7797, device='cuda:0')
[87] Valid Loss:  tensor(0.8583, device='cuda:0')
[88] Valid Loss:  tensor(0.6286, device='cuda:0')
[89] Valid Loss:  tensor(0.8078, device='cuda:0')
[90] Valid Loss:  tensor(0.9174, device='cuda:0')
[91] Valid Loss:  tensor(0.7704, device='cuda:0')
[92] Valid Loss:  tensor(0.8099, device='cuda:0')
[93] Valid Loss:  tensor(0.8146, device='cuda:0')
[94] Valid Loss:  tensor(0.8345, device='cuda:0')
[95] Valid Loss:  tensor(0.7120, device='cuda:0')
[96] Valid Loss:  tensor(0.8968, device='cuda:0')
[97] Valid Loss:  tensor(0.8279, device='cuda:0')
[98] Valid Loss:  tensor(0.8484, device='cuda:0')
[99] Valid Loss:  tensor(0.6411, device='cuda:0')
[100] Valid Loss:  tensor(0.7908, device='cuda:0')
[101] Valid Loss:  tensor(0.7482, device='cuda:0')
[102] Valid Loss:  tensor(0.9397, device='cuda:0')
```

```
[103] Valid Loss:  tensor(0.8945, device='cuda:0')
[104] Valid Loss:  tensor(0.7072, device='cuda:0')
[105] Valid Loss:  tensor(0.7024, device='cuda:0')
[106] Valid Loss:  tensor(0.6083, device='cuda:0')
[107] Valid Loss:  tensor(0.7893, device='cuda:0')
[108] Valid Loss:  tensor(0.8636, device='cuda:0')
[109] Valid Loss:  tensor(0.7800, device='cuda:0')
[110] Valid Loss:  tensor(0.8281, device='cuda:0')
[111] Valid Loss:  tensor(0.7968, device='cuda:0')
[112] Valid Loss:  tensor(0.8405, device='cuda:0')
[113] Valid Loss:  tensor(0.7882, device='cuda:0')
[114] Valid Loss:  tensor(0.7503, device='cuda:0')
[115] Valid Loss:  tensor(0.8196, device='cuda:0')
[116] Valid Loss:  tensor(0.8556, device='cuda:0')
[117] Valid Loss:  tensor(0.8687, device='cuda:0')
[118] Valid Loss:  tensor(0.7092, device='cuda:0')
[119] Valid Loss:  tensor(0.6444, device='cuda:0')
[120] Valid Loss:  tensor(0.9777, device='cuda:0')
[121] Valid Loss:  tensor(0.7057, device='cuda:0')
[122] Valid Loss:  tensor(0.7526, device='cuda:0')
[123] Valid Loss:  tensor(0.6816, device='cuda:0')
[124] Valid Loss:  tensor(0.9271, device='cuda:0')
[125] Valid Loss:  tensor(0.9635, device='cuda:0')
[126] Valid Loss:  tensor(0.6014, device='cuda:0')
[127] Valid Loss:  tensor(0.6263, device='cuda:0')
[128] Valid Loss:  tensor(0.8980, device='cuda:0')
[129] Valid Loss:  tensor(0.7332, device='cuda:0')
[130] Valid Loss:  tensor(0.8286, device='cuda:0')
[131] Valid Loss:  tensor(0.7052, device='cuda:0')
[132] Valid Loss:  tensor(0.7142, device='cuda:0')
[133] Valid Loss:  tensor(0.7870, device='cuda:0')
[134] Valid Loss:  tensor(0.6943, device='cuda:0')
[135] Valid Loss:  tensor(0.6095, device='cuda:0')
[136] Valid Loss:  tensor(0.6451, device='cuda:0')
[137] Valid Loss:  tensor(0.9734, device='cuda:0')
[138] Valid Loss:  tensor(0.8722, device='cuda:0')
[139] Valid Loss:  tensor(1.0691, device='cuda:0')
[140] Valid Loss:  tensor(0.8267, device='cuda:0')
[141] Valid Loss:  tensor(0.6779, device='cuda:0')
[142] Valid Loss:  tensor(0.9408, device='cuda:0')
[143] Valid Loss:  tensor(0.6521, device='cuda:0')
[144] Valid Loss:  tensor(0.7846, device='cuda:0')
[145] Valid Loss:  tensor(0.9715, device='cuda:0')
[146] Valid Loss:  tensor(0.6969, device='cuda:0')
[147] Valid Loss:  tensor(0.8260, device='cuda:0')
[148] Valid Loss:  tensor(0.7142, device='cuda:0')
[149] Valid Loss:  tensor(0.7701, device='cuda:0')
[150] Valid Loss:  tensor(0.7829, device='cuda:0')
```

```
[151] Valid Loss:  tensor(0.7608, device='cuda:0')
[152] Valid Loss:  tensor(0.6791, device='cuda:0')
[153] Valid Loss:  tensor(0.7079, device='cuda:0')
[154] Valid Loss:  tensor(0.7277, device='cuda:0')
[155] Valid Loss:  tensor(0.7977, device='cuda:0')
[156] Valid Loss:  tensor(0.8544, device='cuda:0')
[157] Valid Loss:  tensor(0.6836, device='cuda:0')
[158] Valid Loss:  tensor(0.7471, device='cuda:0')
[159] Valid Loss:  tensor(0.7445, device='cuda:0')
[160] Valid Loss:  tensor(0.6133, device='cuda:0')
[161] Valid Loss:  tensor(0.8699, device='cuda:0')
[162] Valid Loss:  tensor(0.7753, device='cuda:0')
[163] Valid Loss:  tensor(0.6676, device='cuda:0')
[164] Valid Loss:  tensor(0.9775, device='cuda:0')
[165] Valid Loss:  tensor(0.7498, device='cuda:0')
[166] Valid Loss:  tensor(0.7419, device='cuda:0')
[167] Valid Loss:  tensor(0.6568, device='cuda:0')
[168] Valid Loss:  tensor(0.7264, device='cuda:0')
[169] Valid Loss:  tensor(0.7120, device='cuda:0')
[170] Valid Loss:  tensor(0.7260, device='cuda:0')
[171] Valid Loss:  tensor(0.6021, device='cuda:0')
[172] Valid Loss:  tensor(0.8125, device='cuda:0')
[173] Valid Loss:  tensor(0.6547, device='cuda:0')
[174] Valid Loss:  tensor(0.7807, device='cuda:0')
[175] Valid Loss:  tensor(0.6715, device='cuda:0')
[176] Valid Loss:  tensor(0.5944, device='cuda:0')
[177] Valid Loss:  tensor(0.7959, device='cuda:0')
[178] Valid Loss:  tensor(0.8162, device='cuda:0')
[179] Valid Loss:  tensor(0.8022, device='cuda:0')
[180] Valid Loss:  tensor(0.8455, device='cuda:0')
[181] Valid Loss:  tensor(0.7950, device='cuda:0')
[182] Valid Loss:  tensor(0.7874, device='cuda:0')
[183] Valid Loss:  tensor(0.7721, device='cuda:0')
[184] Valid Loss:  tensor(0.9187, device='cuda:0')
[185] Valid Loss:  tensor(0.8390, device='cuda:0')
[186] Valid Loss:  tensor(0.7346, device='cuda:0')
[187] Valid Loss:  tensor(0.8320, device='cuda:0')
[188] Valid Loss:  tensor(0.9341, device='cuda:0')
[189] Valid Loss:  tensor(0.7213, device='cuda:0')
[190] Valid Loss:  tensor(0.7754, device='cuda:0')
[191] Valid Loss:  tensor(0.6319, device='cuda:0')
[192] Valid Loss:  tensor(1.0190, device='cuda:0')
[193] Valid Loss:  tensor(0.7754, device='cuda:0')
[194] Valid Loss:  tensor(0.8623, device='cuda:0')
[195] Valid Loss:  tensor(0.8732, device='cuda:0')
[196] Valid Loss:  tensor(0.6259, device='cuda:0')
[197] Valid Loss:  tensor(0.7985, device='cuda:0')
[198] Valid Loss:  tensor(0.7622, device='cuda:0')
```

```
[199] Valid Loss:   tensor(0.8241, device='cuda:0')
[200] Valid Loss:   tensor(0.7719, device='cuda:0')
[201] Valid Loss:   tensor(0.7242, device='cuda:0')
[202] Valid Loss:   tensor(0.6754, device='cuda:0')
[203] Valid Loss:   tensor(0.9121, device='cuda:0')
[204] Valid Loss:   tensor(0.8665, device='cuda:0')
[205] Valid Loss:   tensor(0.8664, device='cuda:0')
[206] Valid Loss:   tensor(0.9503, device='cuda:0')
[207] Valid Loss:   tensor(0.7960, device='cuda:0')
[208] Valid Loss:   tensor(0.6632, device='cuda:0')
[209] Valid Loss:   tensor(0.7456, device='cuda:0')
[210] Valid Loss:   tensor(1.1415, device='cuda:0')
[211] Valid Loss:   tensor(0.7855, device='cuda:0')
[212] Valid Loss:   tensor(0.6442, device='cuda:0')
[213] Valid Loss:   tensor(0.7288, device='cuda:0')
[214] Valid Loss:   tensor(0.7182, device='cuda:0')
[215] Valid Loss:   tensor(0.8238, device='cuda:0')
[216] Valid Loss:   tensor(0.7695, device='cuda:0')
[217] Valid Loss:   tensor(0.7788, device='cuda:0')
[218] Valid Loss:   tensor(0.6143, device='cuda:0')
[219] Valid Loss:   tensor(0.7279, device='cuda:0')
[220] Valid Loss:   tensor(0.6938, device='cuda:0')
[221] Valid Loss:   tensor(0.7813, device='cuda:0')
[222] Valid Loss:   tensor(0.6057, device='cuda:0')
[223] Valid Loss:   tensor(0.7516, device='cuda:0')
[224] Valid Loss:   tensor(0.8263, device='cuda:0')
[225] Valid Loss:   tensor(0.6166, device='cuda:0')
[226] Valid Loss:   tensor(0.7668, device='cuda:0')
[227] Valid Loss:   tensor(0.6284, device='cuda:0')
[228] Valid Loss:   tensor(0.8352, device='cuda:0')
[229] Valid Loss:   tensor(0.7146, device='cuda:0')
[230] Valid Loss:   tensor(0.7983, device='cuda:0')
[231] Valid Loss:   tensor(0.7247, device='cuda:0')
[232] Valid Loss:   tensor(0.6192, device='cuda:0')
[233] Valid Loss:   tensor(0.7069, device='cuda:0')
[234] Valid Loss:   tensor(0.6326, device='cuda:0')
[235] Valid Loss:   tensor(0.6239, device='cuda:0')
[236] Valid Loss:   tensor(0.7880, device='cuda:0')
[237] Valid Loss:   tensor(0.8174, device='cuda:0')
[238] Valid Loss:   tensor(0.5573, device='cuda:0')
[239] Valid Loss:   tensor(0.8340, device='cuda:0')
[240] Valid Loss:   tensor(0.7343, device='cuda:0')
[241] Valid Loss:   tensor(0.6607, device='cuda:0')
[242] Valid Loss:   tensor(0.5628, device='cuda:0')
[243] Valid Loss:   tensor(1.1722, device='cuda:0')
[244] Valid Loss:   tensor(0.8159, device='cuda:0')
[245] Valid Loss:   tensor(0.7209, device='cuda:0')
[246] Valid Loss:   tensor(0.6717, device='cuda:0')
```

```
[247] Valid Loss:  tensor(0.7192, device='cuda:0')
[248] Valid Loss:  tensor(0.9559, device='cuda:0')
[249] Valid Loss:  tensor(0.7331, device='cuda:0')
[250] Valid Loss:  tensor(0.7889, device='cuda:0')
[251] Valid Loss:  tensor(0.8241, device='cuda:0')
[252] Valid Loss:  tensor(0.6245, device='cuda:0')
[253] Valid Loss:  tensor(0.6568, device='cuda:0')
[254] Valid Loss:  tensor(0.6530, device='cuda:0')
[255] Valid Loss:  tensor(0.8010, device='cuda:0')
[256] Valid Loss:  tensor(0.8329, device='cuda:0')
[257] Valid Loss:  tensor(0.6991, device='cuda:0')
[258] Valid Loss:  tensor(0.8822, device='cuda:0')
[259] Valid Loss:  tensor(0.7479, device='cuda:0')
[260] Valid Loss:  tensor(0.9449, device='cuda:0')
[261] Valid Loss:  tensor(0.7838, device='cuda:0')
[262] Valid Loss:  tensor(0.9341, device='cuda:0')
[263] Valid Loss:  tensor(0.9032, device='cuda:0')
[264] Valid Loss:  tensor(0.8424, device='cuda:0')
[265] Valid Loss:  tensor(0.8494, device='cuda:0')
[266] Valid Loss:  tensor(0.7402, device='cuda:0')
[267] Valid Loss:  tensor(0.7464, device='cuda:0')
[268] Valid Loss:  tensor(0.7215, device='cuda:0')
[269] Valid Loss:  tensor(0.7024, device='cuda:0')
[270] Valid Loss:  tensor(0.6383, device='cuda:0')
[271] Valid Loss:  tensor(0.9878, device='cuda:0')
[272] Valid Loss:  tensor(0.7511, device='cuda:0')
[273] Valid Loss:  tensor(0.8912, device='cuda:0')
[274] Valid Loss:  tensor(0.7351, device='cuda:0')
[275] Valid Loss:  tensor(0.7610, device='cuda:0')
[276] Valid Loss:  tensor(0.6755, device='cuda:0')
[277] Valid Loss:  tensor(0.6648, device='cuda:0')
[278] Valid Loss:  tensor(0.9378, device='cuda:0')
[279] Valid Loss:  tensor(0.8309, device='cuda:0')
[280] Valid Loss:  tensor(0.9258, device='cuda:0')
[281] Valid Loss:  tensor(0.6893, device='cuda:0')
[282] Valid Loss:  tensor(0.7957, device='cuda:0')
[283] Valid Loss:  tensor(0.9125, device='cuda:0')
[284] Valid Loss:  tensor(0.6916, device='cuda:0')
[285] Valid Loss:  tensor(0.6101, device='cuda:0')
[286] Valid Loss:  tensor(0.8305, device='cuda:0')
[287] Valid Loss:  tensor(0.9570, device='cuda:0')
[288] Valid Loss:  tensor(0.5727, device='cuda:0')
[289] Valid Loss:  tensor(0.7987, device='cuda:0')
[290] Valid Loss:  tensor(0.6267, device='cuda:0')
[291] Valid Loss:  tensor(0.7328, device='cuda:0')
[292] Valid Loss:  tensor(0.8627, device='cuda:0')
[293] Valid Loss:  tensor(0.6830, device='cuda:0')
[294] Valid Loss:  tensor(0.8874, device='cuda:0')
```

```
[295] Valid Loss:  tensor(0.6943, device='cuda:0')
[296] Valid Loss:  tensor(0.7790, device='cuda:0')
[297] Valid Loss:  tensor(0.8235, device='cuda:0')
[298] Valid Loss:  tensor(0.6964, device='cuda:0')
[299] Valid Loss:  tensor(0.8059, device='cuda:0')
[300] Valid Loss:  tensor(0.9391, device='cuda:0')
[301] Valid Loss:  tensor(0.6156, device='cuda:0')
[302] Valid Loss:  tensor(0.5895, device='cuda:0')
[303] Valid Loss:  tensor(0.7057, device='cuda:0')
[304] Valid Loss:  tensor(0.7051, device='cuda:0')
[305] Valid Loss:  tensor(0.8043, device='cuda:0')
[306] Valid Loss:  tensor(0.7927, device='cuda:0')
[307] Valid Loss:  tensor(0.6120, device='cuda:0')
[308] Valid Loss:  tensor(0.6808, device='cuda:0')
[309] Valid Loss:  tensor(0.7156, device='cuda:0')
[310] Valid Loss:  tensor(0.7352, device='cuda:0')
[311] Valid Loss:  tensor(0.6449, device='cuda:0')
[312] Valid Loss:  tensor(0.6497, device='cuda:0')
[313] Valid Loss:  tensor(0.7458, device='cuda:0')
[314] Valid Loss:  tensor(0.7307, device='cuda:0')
[315] Valid Loss:  tensor(1.0275, device='cuda:0')
[316] Valid Loss:  tensor(0.8549, device='cuda:0')
[317] Valid Loss:  tensor(0.7431, device='cuda:0')
[318] Valid Loss:  tensor(0.7117, device='cuda:0')
[319] Valid Loss:  tensor(0.6954, device='cuda:0')
[320] Valid Loss:  tensor(0.6961, device='cuda:0')
[321] Valid Loss:  tensor(0.7161, device='cuda:0')
[322] Valid Loss:  tensor(0.6512, device='cuda:0')
[323] Valid Loss:  tensor(0.8363, device='cuda:0')
[324] Valid Loss:  tensor(0.7847, device='cuda:0')
[325] Valid Loss:  tensor(0.6529, device='cuda:0')
[326] Valid Loss:  tensor(0.7337, device='cuda:0')
[327] Valid Loss:  tensor(0.4898, device='cuda:0')
[328] Valid Loss:  tensor(0.6979, device='cuda:0')
[329] Valid Loss:  tensor(0.8222, device='cuda:0')
[330] Valid Loss:  tensor(0.6287, device='cuda:0')
[331] Valid Loss:  tensor(0.9318, device='cuda:0')
[332] Valid Loss:  tensor(0.6590, device='cuda:0')
[333] Valid Loss:  tensor(0.6012, device='cuda:0')
[334] Valid Loss:  tensor(0.6467, device='cuda:0')
[335] Valid Loss:  tensor(0.7350, device='cuda:0')
[336] Valid Loss:  tensor(0.8214, device='cuda:0')
[337] Valid Loss:  tensor(0.9181, device='cuda:0')
[338] Valid Loss:  tensor(0.7851, device='cuda:0')
[339] Valid Loss:  tensor(0.7408, device='cuda:0')
[340] Valid Loss:  tensor(0.7720, device='cuda:0')
[341] Valid Loss:  tensor(0.8616, device='cuda:0')
[342] Valid Loss:  tensor(0.9305, device='cuda:0')
```

```
[343] Valid Loss:  tensor(0.7693, device='cuda:0')
[344] Valid Loss:  tensor(0.6737, device='cuda:0')
[345] Valid Loss:  tensor(0.5836, device='cuda:0')
[346] Valid Loss:  tensor(0.7527, device='cuda:0')
[347] Valid Loss:  tensor(0.8507, device='cuda:0')
[348] Valid Loss:  tensor(0.7081, device='cuda:0')
[349] Valid Loss:  tensor(0.6986, device='cuda:0')
[350] Valid Loss:  tensor(0.7354, device='cuda:0')
[351] Valid Loss:  tensor(0.8401, device='cuda:0')
[352] Valid Loss:  tensor(0.6579, device='cuda:0')
[353] Valid Loss:  tensor(0.7476, device='cuda:0')
[354] Valid Loss:  tensor(0.9532, device='cuda:0')
[355] Valid Loss:  tensor(0.5627, device='cuda:0')
[356] Valid Loss:  tensor(0.6835, device='cuda:0')
[357] Valid Loss:  tensor(0.8502, device='cuda:0')
[358] Valid Loss:  tensor(0.7147, device='cuda:0')
[359] Valid Loss:  tensor(0.7975, device='cuda:0')
[360] Valid Loss:  tensor(0.7351, device='cuda:0')
[361] Valid Loss:  tensor(0.7720, device='cuda:0')
[362] Valid Loss:  tensor(0.7323, device='cuda:0')
[363] Valid Loss:  tensor(0.8095, device='cuda:0')
[364] Valid Loss:  tensor(0.9244, device='cuda:0')
[365] Valid Loss:  tensor(0.7067, device='cuda:0')
[366] Valid Loss:  tensor(0.7053, device='cuda:0')
[367] Valid Loss:  tensor(0.8131, device='cuda:0')
[368] Valid Loss:  tensor(0.8520, device='cuda:0')
[369] Valid Loss:  tensor(0.7444, device='cuda:0')
[370] Valid Loss:  tensor(0.6423, device='cuda:0')
[371] Valid Loss:  tensor(0.6466, device='cuda:0')
[372] Valid Loss:  tensor(0.8415, device='cuda:0')
[373] Valid Loss:  tensor(0.5776, device='cuda:0')
[374] Valid Loss:  tensor(0.7553, device='cuda:0')
[375] Valid Loss:  tensor(0.8235, device='cuda:0')
[376] Valid Loss:  tensor(0.6811, device='cuda:0')
[377] Valid Loss:  tensor(0.9614, device='cuda:0')
[378] Valid Loss:  tensor(0.6614, device='cuda:0')
[379] Valid Loss:  tensor(0.8379, device='cuda:0')
[380] Valid Loss:  tensor(1.1041, device='cuda:0')
Testing Loss:  0.7667813588940878
Testing Accuaracy:  76.32896039080033 %
Epoch Number:  5
[0] Train Loss:  tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[1] Train Loss:  tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[2] Train Loss:  tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[3] Train Loss:  tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[4] Train Loss:  tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[5] Train Loss:  tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[6] Train Loss:  tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[7] Train Loss:  tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[8] Train Loss:  tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[9] Train Loss:  tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[10] Train Loss:  tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[11] Train Loss:  tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[12] Train Loss:  tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[13] Train Loss:  tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[14] Train Loss:  tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[15] Train Loss:  tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[16] Train Loss:  tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[17] Train Loss:  tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[18] Train Loss:  tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[19] Train Loss:  tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[20] Train Loss:  tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[21] Train Loss:  tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[22] Train Loss:  tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[23] Train Loss:  tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[24] Train Loss:  tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[25] Train Loss:  tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[26] Train Loss:  tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[27] Train Loss:  tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[28] Train Loss:  tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[29] Train Loss:  tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[30] Train Loss:  tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[31] Train Loss:  tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[32] Train Loss:  tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[33] Train Loss:  tensor(0.7852, device='cuda:0', grad_fn=<NllLossBackward>)
[34] Train Loss:  tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[35] Train Loss:  tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[36] Train Loss:  tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[37] Train Loss:  tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[38] Train Loss:  tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[39] Train Loss:  tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[40] Train Loss:  tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[41] Train Loss:  tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[42] Train Loss:  tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[43] Train Loss:  tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[44] Train Loss:  tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[45] Train Loss:  tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[46] Train Loss:  tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[47] Train Loss:  tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[48] Train Loss:  tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[49] Train Loss:  tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[50] Train Loss:  tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[51] Train Loss:  tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[52] Train Loss:  tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[53] Train Loss:  tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[54] Train Loss:  tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[55] Train Loss:   tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[56] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[57] Train Loss:   tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[58] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[59] Train Loss:   tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[60] Train Loss:   tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[61] Train Loss:   tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[62] Train Loss:   tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[63] Train Loss:   tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[64] Train Loss:   tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[65] Train Loss:   tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[66] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[67] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[68] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[69] Train Loss:   tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[70] Train Loss:   tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[71] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[72] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[73] Train Loss:   tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[74] Train Loss:   tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[75] Train Loss:   tensor(0.7941, device='cuda:0', grad_fn=<NllLossBackward>)
[76] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[77] Train Loss:   tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[78] Train Loss:   tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[79] Train Loss:   tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[80] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[81] Train Loss:   tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[82] Train Loss:   tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[83] Train Loss:   tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[84] Train Loss:   tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[85] Train Loss:   tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[86] Train Loss:   tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[87] Train Loss:   tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[88] Train Loss:   tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[89] Train Loss:   tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[90] Train Loss:   tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[91] Train Loss:   tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[92] Train Loss:   tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[93] Train Loss:   tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[94] Train Loss:   tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[95] Train Loss:   tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[96] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[97] Train Loss:   tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[98] Train Loss:   tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[99] Train Loss:   tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[100] Train Loss:  tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[101] Train Loss:  tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[102] Train Loss:  tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[103] Train Loss:  tensor(0.7795, device='cuda:0', grad_fn=<NllLossBackward>)
[104] Train Loss:  tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[105] Train Loss:  tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[106] Train Loss:  tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[107] Train Loss:  tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[108] Train Loss:  tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[109] Train Loss:  tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[110] Train Loss:  tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[111] Train Loss:  tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[112] Train Loss:  tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[113] Train Loss:  tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[114] Train Loss:  tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[115] Train Loss:  tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[116] Train Loss:  tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[117] Train Loss:  tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[118] Train Loss:  tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[119] Train Loss:  tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[120] Train Loss:  tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[121] Train Loss:  tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[122] Train Loss:  tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[123] Train Loss:  tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[124] Train Loss:  tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[125] Train Loss:  tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[126] Train Loss:  tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[127] Train Loss:  tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[128] Train Loss:  tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[129] Train Loss:  tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[130] Train Loss:  tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[131] Train Loss:  tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[132] Train Loss:  tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[133] Train Loss:  tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[134] Train Loss:  tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[135] Train Loss:  tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[136] Train Loss:  tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[137] Train Loss:  tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[138] Train Loss:  tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[139] Train Loss:  tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[140] Train Loss:  tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[141] Train Loss:  tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[142] Train Loss:  tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[143] Train Loss:  tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[144] Train Loss:  tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[145] Train Loss:  tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[146] Train Loss:  tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[147] Train Loss:  tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[148] Train Loss:  tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[149] Train Loss:  tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[150] Train Loss:  tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[151] Train Loss:   tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[152] Train Loss:   tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[153] Train Loss:   tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[154] Train Loss:   tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[155] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[156] Train Loss:   tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[157] Train Loss:   tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[158] Train Loss:   tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[159] Train Loss:   tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[160] Train Loss:   tensor(0.7582, device='cuda:0', grad_fn=<NllLossBackward>)
[161] Train Loss:   tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[162] Train Loss:   tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[163] Train Loss:   tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[164] Train Loss:   tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[165] Train Loss:   tensor(0.7921, device='cuda:0', grad_fn=<NllLossBackward>)
[166] Train Loss:   tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[167] Train Loss:   tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[168] Train Loss:   tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[169] Train Loss:   tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[170] Train Loss:   tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[171] Train Loss:   tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[172] Train Loss:   tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[173] Train Loss:   tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[174] Train Loss:   tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[175] Train Loss:   tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[176] Train Loss:   tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[177] Train Loss:   tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[178] Train Loss:   tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[179] Train Loss:   tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[180] Train Loss:   tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[181] Train Loss:   tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[182] Train Loss:   tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[183] Train Loss:   tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[184] Train Loss:   tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[185] Train Loss:   tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[186] Train Loss:   tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[187] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[188] Train Loss:   tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[189] Train Loss:   tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[190] Train Loss:   tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[191] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[192] Train Loss:   tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[193] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[194] Train Loss:   tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[195] Train Loss:   tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[196] Train Loss:   tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[197] Train Loss:   tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[198] Train Loss:   tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[199] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[200] Train Loss:   tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[201] Train Loss:   tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[202] Train Loss:   tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[203] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[204] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[205] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[206] Train Loss:   tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[207] Train Loss:   tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[208] Train Loss:   tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[209] Train Loss:   tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[210] Train Loss:   tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[211] Train Loss:   tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[212] Train Loss:   tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[213] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[214] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[215] Train Loss:   tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[216] Train Loss:   tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[217] Train Loss:   tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[218] Train Loss:   tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[219] Train Loss:   tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[220] Train Loss:   tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[221] Train Loss:   tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[222] Train Loss:   tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[223] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[224] Train Loss:   tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[225] Train Loss:   tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[226] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[227] Train Loss:   tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[228] Train Loss:   tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[229] Train Loss:   tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[230] Train Loss:   tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[231] Train Loss:   tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[232] Train Loss:   tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[233] Train Loss:   tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[234] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[235] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[236] Train Loss:   tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[237] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[238] Train Loss:   tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[239] Train Loss:   tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[240] Train Loss:   tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[241] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[242] Train Loss:   tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[243] Train Loss:   tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[244] Train Loss:   tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[245] Train Loss:   tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[246] Train Loss:   tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[247] Train Loss:   tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[248] Train Loss:   tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[249] Train Loss:   tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[250] Train Loss:   tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[251] Train Loss:   tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[252] Train Loss:   tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[253] Train Loss:   tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[254] Train Loss:   tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[255] Train Loss:   tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[256] Train Loss:   tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[257] Train Loss:   tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[258] Train Loss:   tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[259] Train Loss:   tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[260] Train Loss:   tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[261] Train Loss:   tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[262] Train Loss:   tensor(0.7932, device='cuda:0', grad_fn=<NllLossBackward>)
[263] Train Loss:   tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[264] Train Loss:   tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[265] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[266] Train Loss:   tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[267] Train Loss:   tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[268] Train Loss:   tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[269] Train Loss:   tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[270] Train Loss:   tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[271] Train Loss:   tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[272] Train Loss:   tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[273] Train Loss:   tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[274] Train Loss:   tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[275] Train Loss:   tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[276] Train Loss:   tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[277] Train Loss:   tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[278] Train Loss:   tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[279] Train Loss:   tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[280] Train Loss:   tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[281] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[282] Train Loss:   tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[283] Train Loss:   tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[284] Train Loss:   tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[285] Train Loss:   tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[286] Train Loss:   tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[287] Train Loss:   tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[288] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[289] Train Loss:   tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[290] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[291] Train Loss:   tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[292] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[293] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[294] Train Loss:   tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[295] Train Loss:   tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[296] Train Loss:   tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[297] Train Loss:   tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[298] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[299] Train Loss:   tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[300] Train Loss:   tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[301] Train Loss:   tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[302] Train Loss:   tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[303] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[304] Train Loss:   tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[305] Train Loss:   tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[306] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[307] Train Loss:   tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[308] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[309] Train Loss:   tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[310] Train Loss:   tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[311] Train Loss:   tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[312] Train Loss:   tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[313] Train Loss:   tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[314] Train Loss:   tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[315] Train Loss:   tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[316] Train Loss:   tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[317] Train Loss:   tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[318] Train Loss:   tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[319] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[320] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[321] Train Loss:   tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[322] Train Loss:   tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[323] Train Loss:   tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[324] Train Loss:   tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[325] Train Loss:   tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[326] Train Loss:   tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[327] Train Loss:   tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[328] Train Loss:   tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[329] Train Loss:   tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[330] Train Loss:   tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[331] Train Loss:   tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[332] Train Loss:   tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[333] Train Loss:   tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[334] Train Loss:   tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[335] Train Loss:   tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[336] Train Loss:   tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
[337] Train Loss:   tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[338] Train Loss:   tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[339] Train Loss:   tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[340] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[341] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[342] Train Loss:   tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[343] Train Loss:   tensor(0.7731, device='cuda:0', grad_fn=<NllLossBackward>)
[344] Train Loss:   tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[345] Train Loss:   tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[346] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[347] Train Loss:   tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[348] Train Loss:   tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[349] Train Loss:   tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[350] Train Loss:   tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[351] Train Loss:   tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[352] Train Loss:   tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[353] Train Loss:   tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[354] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[355] Train Loss:   tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[356] Train Loss:   tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[357] Train Loss:   tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[358] Train Loss:   tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[359] Train Loss:   tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[360] Train Loss:   tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[361] Train Loss:   tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[362] Train Loss:   tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[363] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[364] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[365] Train Loss:   tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[366] Train Loss:   tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[367] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[368] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[369] Train Loss:   tensor(0.7754, device='cuda:0', grad_fn=<NllLossBackward>)
[370] Train Loss:   tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[371] Train Loss:   tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[372] Train Loss:   tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[373] Train Loss:   tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[374] Train Loss:   tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[375] Train Loss:   tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[376] Train Loss:   tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[377] Train Loss:   tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[378] Train Loss:   tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[379] Train Loss:   tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[380] Train Loss:   tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[381] Train Loss:   tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[382] Train Loss:   tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[383] Train Loss:   tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[384] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[385] Train Loss:   tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[386] Train Loss:   tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[387] Train Loss:   tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[388] Train Loss:   tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[389] Train Loss:   tensor(0.7782, device='cuda:0', grad_fn=<NllLossBackward>)
[390] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[391] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[392] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[393] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[394] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[395] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[396] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[397] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[398] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[399] Train Loss:    tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[400] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[401] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[402] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[403] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[404] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[405] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[406] Train Loss:    tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[407] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[408] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[409] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[410] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[411] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[412] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[413] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[414] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[415] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[416] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[417] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[418] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[419] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[420] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[421] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[422] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[423] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[424] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[425] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[426] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[427] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[428] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[429] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[430] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[431] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[432] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[433] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[434] Train Loss:    tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[435] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[436] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[437] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[438] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[439] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[440] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[441] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[442] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[443] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[444] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[445] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[446] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[447] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[448] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[449] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[450] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[451] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[452] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[453] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[454] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[455] Train Loss:    tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[456] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[457] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[458] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[459] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[460] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[461] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[462] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[463] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[464] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[465] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[466] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[467] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[468] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[469] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[470] Train Loss:    tensor(0.7582, device='cuda:0', grad_fn=<NllLossBackward>)
[471] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[472] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[473] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[474] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[475] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[476] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[477] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[478] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[479] Train Loss:    tensor(0.7859, device='cuda:0', grad_fn=<NllLossBackward>)
[480] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[481] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[482] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[483] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[484] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[485] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[486] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[487] Train Loss:   tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[488] Train Loss:   tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[489] Train Loss:   tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[490] Train Loss:   tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[491] Train Loss:   tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[492] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[493] Train Loss:   tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[494] Train Loss:   tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[495] Train Loss:   tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[496] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[497] Train Loss:   tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[498] Train Loss:   tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[499] Train Loss:   tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[500] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[501] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[502] Train Loss:   tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[503] Train Loss:   tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[504] Train Loss:   tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[505] Train Loss:   tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[506] Train Loss:   tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[507] Train Loss:   tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[508] Train Loss:   tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[509] Train Loss:   tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[510] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[511] Train Loss:   tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[512] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[513] Train Loss:   tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[514] Train Loss:   tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[515] Train Loss:   tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[516] Train Loss:   tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[517] Train Loss:   tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[518] Train Loss:   tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[519] Train Loss:   tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[520] Train Loss:   tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[521] Train Loss:   tensor(0.7840, device='cuda:0', grad_fn=<NllLossBackward>)
[522] Train Loss:   tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[523] Train Loss:   tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[524] Train Loss:   tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[525] Train Loss:   tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[526] Train Loss:   tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[527] Train Loss:   tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[528] Train Loss:   tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[529] Train Loss:   tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[530] Train Loss:   tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[531] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[532] Train Loss:   tensor(0.7974, device='cuda:0', grad_fn=<NllLossBackward>)
[533] Train Loss:   tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[534] Train Loss:   tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[535] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[536] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[537] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[538] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[539] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[540] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[541] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[542] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[543] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[544] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[545] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[546] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[547] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[548] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[549] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[550] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[551] Train Loss:    tensor(0.7862, device='cuda:0', grad_fn=<NllLossBackward>)
[552] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[553] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[554] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[555] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[556] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[557] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[558] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[559] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[560] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[561] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[562] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[563] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[564] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[565] Train Loss:    tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[566] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[567] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[568] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[569] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[570] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[571] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[572] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[573] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[574] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[575] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[576] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[577] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[578] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[579] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[580] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[581] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[582] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[583] Train Loss:  tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[584] Train Loss:  tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[585] Train Loss:  tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[586] Train Loss:  tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[587] Train Loss:  tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[588] Train Loss:  tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[589] Train Loss:  tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[590] Train Loss:  tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[591] Train Loss:  tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[592] Train Loss:  tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[593] Train Loss:  tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[594] Train Loss:  tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[595] Train Loss:  tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[596] Train Loss:  tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[597] Train Loss:  tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[598] Train Loss:  tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[599] Train Loss:  tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[600] Train Loss:  tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[601] Train Loss:  tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[602] Train Loss:  tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[603] Train Loss:  tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[604] Train Loss:  tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[605] Train Loss:  tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[606] Train Loss:  tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[607] Train Loss:  tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[608] Train Loss:  tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[609] Train Loss:  tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[610] Train Loss:  tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[611] Train Loss:  tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[612] Train Loss:  tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[613] Train Loss:  tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[614] Train Loss:  tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[615] Train Loss:  tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[616] Train Loss:  tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[617] Train Loss:  tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[618] Train Loss:  tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[619] Train Loss:  tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[620] Train Loss:  tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[621] Train Loss:  tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[622] Train Loss:  tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[623] Train Loss:  tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[624] Train Loss:  tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[625] Train Loss:  tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[626] Train Loss:  tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[627] Train Loss:  tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[628] Train Loss:  tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[629] Train Loss:  tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[630] Train Loss:  tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[631] Train Loss:  tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[632] Train Loss:  tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[633] Train Loss:  tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[634] Train Loss:  tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[635] Train Loss:  tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[636] Train Loss:  tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[637] Train Loss:  tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[638] Train Loss:  tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[639] Train Loss:  tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[640] Train Loss:  tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[641] Train Loss:  tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[642] Train Loss:  tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[643] Train Loss:  tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[644] Train Loss:  tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[645] Train Loss:  tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[646] Train Loss:  tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[647] Train Loss:  tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[648] Train Loss:  tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[649] Train Loss:  tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[650] Train Loss:  tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[651] Train Loss:  tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[652] Train Loss:  tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[653] Train Loss:  tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[654] Train Loss:  tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[655] Train Loss:  tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[656] Train Loss:  tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[657] Train Loss:  tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[658] Train Loss:  tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[659] Train Loss:  tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[660] Train Loss:  tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[661] Train Loss:  tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[662] Train Loss:  tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[663] Train Loss:  tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[664] Train Loss:  tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[665] Train Loss:  tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[666] Train Loss:  tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[667] Train Loss:  tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[668] Train Loss:  tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[669] Train Loss:  tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[670] Train Loss:  tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[671] Train Loss:  tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[672] Train Loss:  tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[673] Train Loss:  tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[674] Train Loss:  tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[675] Train Loss:  tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[676] Train Loss:  tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[677] Train Loss:  tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[678] Train Loss:  tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[679] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[680] Train Loss:   tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[681] Train Loss:   tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[682] Train Loss:   tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[683] Train Loss:   tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[684] Train Loss:   tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[685] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[686] Train Loss:   tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[687] Train Loss:   tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[688] Train Loss:   tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[689] Train Loss:   tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[690] Train Loss:   tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[691] Train Loss:   tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[692] Train Loss:   tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[693] Train Loss:   tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[694] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[695] Train Loss:   tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[696] Train Loss:   tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[697] Train Loss:   tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[698] Train Loss:   tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[699] Train Loss:   tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[700] Train Loss:   tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[701] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[702] Train Loss:   tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[703] Train Loss:   tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[704] Train Loss:   tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[705] Train Loss:   tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[706] Train Loss:   tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[707] Train Loss:   tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[708] Train Loss:   tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[709] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[710] Train Loss:   tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[711] Train Loss:   tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[712] Train Loss:   tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[713] Train Loss:   tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[714] Train Loss:   tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[715] Train Loss:   tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[716] Train Loss:   tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[717] Train Loss:   tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[718] Train Loss:   tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[719] Train Loss:   tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[720] Train Loss:   tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[721] Train Loss:   tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[722] Train Loss:   tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[723] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[724] Train Loss:   tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[725] Train Loss:   tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[726] Train Loss:   tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[727] Train Loss:   tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[728] Train Loss:   tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[729] Train Loss:   tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[730] Train Loss:   tensor(0.7981, device='cuda:0', grad_fn=<NllLossBackward>)
[731] Train Loss:   tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[732] Train Loss:   tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[733] Train Loss:   tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[734] Train Loss:   tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[735] Train Loss:   tensor(0.7801, device='cuda:0', grad_fn=<NllLossBackward>)
[736] Train Loss:   tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[737] Train Loss:   tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[738] Train Loss:   tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[739] Train Loss:   tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[740] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[741] Train Loss:   tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[742] Train Loss:   tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[743] Train Loss:   tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[744] Train Loss:   tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[745] Train Loss:   tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[746] Train Loss:   tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[747] Train Loss:   tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[748] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[749] Train Loss:   tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[750] Train Loss:   tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[751] Train Loss:   tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[752] Train Loss:   tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[753] Train Loss:   tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[754] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[755] Train Loss:   tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[756] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[757] Train Loss:   tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[758] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[759] Train Loss:   tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[760] Train Loss:   tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[761] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[762] Train Loss:   tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[763] Train Loss:   tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[764] Train Loss:   tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[765] Train Loss:   tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[766] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[767] Train Loss:   tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[768] Train Loss:   tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[769] Train Loss:   tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[770] Train Loss:   tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[771] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[772] Train Loss:   tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[773] Train Loss:   tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[774] Train Loss:   tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[775] Train Loss:   tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[776] Train Loss:   tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[777] Train Loss:   tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[778] Train Loss:   tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[779] Train Loss:   tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[780] Train Loss:   tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[781] Train Loss:   tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[782] Train Loss:   tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[783] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[784] Train Loss:   tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[785] Train Loss:   tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[786] Train Loss:   tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[787] Train Loss:   tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[788] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[789] Train Loss:   tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[790] Train Loss:   tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[791] Train Loss:   tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[792] Train Loss:   tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[793] Train Loss:   tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[794] Train Loss:   tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[795] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[796] Train Loss:   tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[797] Train Loss:   tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[798] Train Loss:   tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[799] Train Loss:   tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[800] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[801] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[802] Train Loss:   tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[803] Train Loss:   tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[804] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[805] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[806] Train Loss:   tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[807] Train Loss:   tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[808] Train Loss:   tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[809] Train Loss:   tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[810] Train Loss:   tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[811] Train Loss:   tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[812] Train Loss:   tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[813] Train Loss:   tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[814] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[815] Train Loss:   tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[816] Train Loss:   tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[817] Train Loss:   tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[818] Train Loss:   tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[819] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[820] Train Loss:   tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[821] Train Loss:   tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[822] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[823] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[824] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[825] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[826] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[827] Train Loss:    tensor(0.7833, device='cuda:0', grad_fn=<NllLossBackward>)
[828] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[829] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[830] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[831] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[832] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[833] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[834] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[835] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[836] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[837] Train Loss:    tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[838] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[839] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[840] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[841] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[842] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[843] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[844] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[845] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[846] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[847] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[848] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[849] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[850] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[851] Train Loss:    tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[852] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[853] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[854] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[855] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[856] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[857] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[858] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[859] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[860] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[861] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[862] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[863] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[864] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[865] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[866] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[867] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[868] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[869] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[870] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[871] Train Loss:   tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[872] Train Loss:   tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[873] Train Loss:   tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[874] Train Loss:   tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[875] Train Loss:   tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[876] Train Loss:   tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[877] Train Loss:   tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[878] Train Loss:   tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[879] Train Loss:   tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[880] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[881] Train Loss:   tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[882] Train Loss:   tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[883] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[884] Train Loss:   tensor(0.7878, device='cuda:0', grad_fn=<NllLossBackward>)
[885] Train Loss:   tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[886] Train Loss:   tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[887] Train Loss:   tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[888] Train Loss:   tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[889] Train Loss:   tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[890] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[891] Train Loss:   tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[892] Train Loss:   tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[893] Train Loss:   tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[894] Train Loss:   tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[895] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[896] Train Loss:   tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[897] Train Loss:   tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[898] Train Loss:   tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[899] Train Loss:   tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[900] Train Loss:   tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[901] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[902] Train Loss:   tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[903] Train Loss:   tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[904] Train Loss:   tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[905] Train Loss:   tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[906] Train Loss:   tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[907] Train Loss:   tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[908] Train Loss:   tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[909] Train Loss:   tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[910] Train Loss:   tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[911] Train Loss:   tensor(0.7785, device='cuda:0', grad_fn=<NllLossBackward>)
[912] Train Loss:   tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[913] Train Loss:   tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[914] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[915] Train Loss:   tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[916] Train Loss:   tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[917] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[918] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[919] Train Loss:   tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[920] Train Loss:   tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[921] Train Loss:   tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[922] Train Loss:   tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[923] Train Loss:   tensor(0.7715, device='cuda:0', grad_fn=<NllLossBackward>)
[924] Train Loss:   tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[925] Train Loss:   tensor(0.7786, device='cuda:0', grad_fn=<NllLossBackward>)
[926] Train Loss:   tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[927] Train Loss:   tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[928] Train Loss:   tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[929] Train Loss:   tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[930] Train Loss:   tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[931] Train Loss:   tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[932] Train Loss:   tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[933] Train Loss:   tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[934] Train Loss:   tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[935] Train Loss:   tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[936] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[937] Train Loss:   tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[938] Train Loss:   tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[939] Train Loss:   tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[940] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[941] Train Loss:   tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[942] Train Loss:   tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[943] Train Loss:   tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[944] Train Loss:   tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[945] Train Loss:   tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[946] Train Loss:   tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[947] Train Loss:   tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[948] Train Loss:   tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[949] Train Loss:   tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[950] Train Loss:   tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[951] Train Loss:   tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[952] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[953] Train Loss:   tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[954] Train Loss:   tensor(0.7933, device='cuda:0', grad_fn=<NllLossBackward>)
[955] Train Loss:   tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[956] Train Loss:   tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[957] Train Loss:   tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[958] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[959] Train Loss:   tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[960] Train Loss:   tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[961] Train Loss:   tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[962] Train Loss:   tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[963] Train Loss:   tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[964] Train Loss:   tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[965] Train Loss:   tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[966] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[967] Train Loss:  tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[968] Train Loss:  tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[969] Train Loss:  tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[970] Train Loss:  tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[971] Train Loss:  tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[972] Train Loss:  tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[973] Train Loss:  tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[974] Train Loss:  tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[975] Train Loss:  tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[976] Train Loss:  tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[977] Train Loss:  tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[978] Train Loss:  tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[979] Train Loss:  tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[980] Train Loss:  tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[981] Train Loss:  tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[982] Train Loss:  tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[983] Train Loss:  tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[984] Train Loss:  tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[985] Train Loss:  tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[986] Train Loss:  tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[987] Train Loss:  tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[988] Train Loss:  tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[989] Train Loss:  tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[990] Train Loss:  tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[991] Train Loss:  tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[992] Train Loss:  tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[993] Train Loss:  tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[994] Train Loss:  tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[995] Train Loss:  tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[996] Train Loss:  tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[997] Train Loss:  tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[998] Train Loss:  tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[999] Train Loss:  tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[1000] Train Loss:  tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[1001] Train Loss:  tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[1002] Train Loss:  tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[1003] Train Loss:  tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[1004] Train Loss:  tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[1005] Train Loss:  tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[1006] Train Loss:  tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[1007] Train Loss:  tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[1008] Train Loss:  tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[1009] Train Loss:  tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[1010] Train Loss:  tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[1011] Train Loss:  tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[1012] Train Loss:  tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[1013] Train Loss:  tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[1014] Train Loss:  tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1015] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[1016] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[1017] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[1018] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[1019] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[1020] Train Loss:    tensor(0.7929, device='cuda:0', grad_fn=<NllLossBackward>)
[1021] Train Loss:    tensor(0.7809, device='cuda:0', grad_fn=<NllLossBackward>)
[1022] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1023] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[1024] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[1025] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[1026] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[1027] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[1028] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[1029] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[1030] Train Loss:    tensor(0.7871, device='cuda:0', grad_fn=<NllLossBackward>)
[1031] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[1032] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[1033] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[1034] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[1035] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[1036] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[1037] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[1038] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[1039] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[1040] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1041] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[1042] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[1043] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[1044] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[1045] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[1046] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[1047] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[1048] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[1049] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[1050] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[1051] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[1052] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[1053] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[1054] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[1055] Train Loss:    tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[1056] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1057] Train Loss:    tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[1058] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1059] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[1060] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[1061] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1062] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1063] Train Loss:   tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[1064] Train Loss:   tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1065] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[1066] Train Loss:   tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[1067] Train Loss:   tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[1068] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[1069] Train Loss:   tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[1070] Train Loss:   tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[1071] Train Loss:   tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1072] Train Loss:   tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[1073] Train Loss:   tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[1074] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[1075] Train Loss:   tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[1076] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[1077] Train Loss:   tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[1078] Train Loss:   tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[1079] Train Loss:   tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[1080] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[1081] Train Loss:   tensor(0.7667, device='cuda:0', grad_fn=<NllLossBackward>)
[1082] Train Loss:   tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[1083] Train Loss:   tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[1084] Train Loss:   tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[1085] Train Loss:   tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[1086] Train Loss:   tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[1087] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[1088] Train Loss:   tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[1089] Train Loss:   tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[1090] Train Loss:   tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[1091] Train Loss:   tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[1092] Train Loss:   tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[1093] Train Loss:   tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[1094] Train Loss:   tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[1095] Train Loss:   tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[1096] Train Loss:   tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[1097] Train Loss:   tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[1098] Train Loss:   tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[1099] Train Loss:   tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[1100] Train Loss:   tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[1101] Train Loss:   tensor(0.7582, device='cuda:0', grad_fn=<NllLossBackward>)
[1102] Train Loss:   tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[1103] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1104] Train Loss:   tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[1105] Train Loss:   tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[1106] Train Loss:   tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[1107] Train Loss:   tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1108] Train Loss:   tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[1109] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[1110] Train Loss:   tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1111] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[1112] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[1113] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[1114] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[1115] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[1116] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[1117] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[1118] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[1119] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[1120] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[1121] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1122] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[1123] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[1124] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[1125] Train Loss:    tensor(0.7680, device='cuda:0', grad_fn=<NllLossBackward>)
[1126] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[1127] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[1128] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[1129] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[1130] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[1131] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[1132] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[1133] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[1134] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1135] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[1136] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[1137] Train Loss:    tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[1138] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[1139] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[1140] Train Loss:    tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[1141] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1142] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[1143] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[1144] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[1145] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[1146] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[1147] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[1148] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[1149] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[1150] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[1151] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[1152] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[1153] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[1154] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[1155] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[1156] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[1157] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1158] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1159] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[1160] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[1161] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[1162] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[1163] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[1164] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[1165] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[1166] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[1167] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[1168] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[1169] Train Loss:    tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[1170] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[1171] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1172] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[1173] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[1174] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[1175] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[1176] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[1177] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[1178] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[1179] Train Loss:    tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[1180] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[1181] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[1182] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[1183] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[1184] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1185] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[1186] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[1187] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[1188] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[1189] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[1190] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[1191] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[1192] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[1193] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[1194] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[1195] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1196] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[1197] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1198] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[1199] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[1200] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[1201] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[1202] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[1203] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[1204] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[1205] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[1206] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1207] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[1208] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1209] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[1210] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[1211] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[1212] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[1213] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[1214] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[1215] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[1216] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[1217] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[1218] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[1219] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[1220] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[1221] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[1222] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[1223] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[1224] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[1225] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[1226] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[1227] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[1228] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[1229] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[1230] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[1231] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[1232] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[1233] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[1234] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[1235] Train Loss:    tensor(0.7822, device='cuda:0', grad_fn=<NllLossBackward>)
[1236] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[1237] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[1238] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[1239] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[1240] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[1241] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[1242] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[1243] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[1244] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[1245] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[1246] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[1247] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[1248] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[1249] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[1250] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[1251] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[1252] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[1253] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[1254] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1255] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[1256] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[1257] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[1258] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[1259] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[1260] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[1261] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[1262] Train Loss:    tensor(0.7889, device='cuda:0', grad_fn=<NllLossBackward>)
[1263] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[1264] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[1265] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1266] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[1267] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[1268] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[1269] Train Loss:    tensor(0.7860, device='cuda:0', grad_fn=<NllLossBackward>)
[1270] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[1271] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[1272] Train Loss:    tensor(0.7865, device='cuda:0', grad_fn=<NllLossBackward>)
[1273] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[1274] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1275] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[1276] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[1277] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[1278] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[1279] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[1280] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1281] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[1282] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[1283] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[1284] Train Loss:    tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[1285] Train Loss:    tensor(0.7753, device='cuda:0', grad_fn=<NllLossBackward>)
[1286] Train Loss:    tensor(0.7936, device='cuda:0', grad_fn=<NllLossBackward>)
[1287] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[1288] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[1289] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[1290] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1291] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[1292] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[1293] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[1294] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[1295] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[1296] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[1297] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[1298] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[1299] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[1300] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[1301] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[1302] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1303] Train Loss:    tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[1304] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[1305] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[1306] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[1307] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[1308] Train Loss:    tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[1309] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[1310] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[1311] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1312] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[1313] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[1314] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[1315] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[1316] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[1317] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[1318] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[1319] Train Loss:    tensor(0.7802, device='cuda:0', grad_fn=<NllLossBackward>)
[1320] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[1321] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[1322] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[1323] Train Loss:    tensor(0.7760, device='cuda:0', grad_fn=<NllLossBackward>)
[1324] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[1325] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[1326] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[1327] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[1328] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[1329] Train Loss:    tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[1330] Train Loss:    tensor(0.7777, device='cuda:0', grad_fn=<NllLossBackward>)
[1331] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[1332] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[1333] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[1334] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[1335] Train Loss:    tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[1336] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[1337] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[1338] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[1339] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[1340] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[1341] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[1342] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[1343] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[1344] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1345] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[1346] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[1347] Train Loss:    tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[1348] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[1349] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[1350] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1351] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[1352] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[1353] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[1354] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[1355] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[1356] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[1357] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1358] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[1359] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[1360] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[1361] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[1362] Train Loss:    tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[1363] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[1364] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[1365] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[1366] Train Loss:    tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[1367] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[1368] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[1369] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[1370] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[1371] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[1372] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[1373] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[1374] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[1375] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[1376] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[1377] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1378] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1379] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[1380] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[1381] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1382] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1383] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[1384] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[1385] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[1386] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1387] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[1388] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[1389] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[1390] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1391] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[1392] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[1393] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[1394] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[1395] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[1396] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[1397] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[1398] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1399] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[1400] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[1401] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[1402] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[1403] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[1404] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[1405] Train Loss:    tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[1406] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[1407] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[1408] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[1409] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[1410] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[1411] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[1412] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[1413] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[1414] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[1415] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[1416] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[1417] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[1418] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[1419] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[1420] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[1421] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[1422] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[1423] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[1424] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[1425] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[1426] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[1427] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[1428] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1429] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[1430] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[1431] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[1432] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[1433] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[1434] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[1435] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[1436] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[1437] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[1438] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[1439] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[1440] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[1441] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[1442] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[1443] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[1444] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[1445] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[1446] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1447] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[1448] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[1449] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[1450] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[1451] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[1452] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[1453] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[1454] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[1455] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[1456] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[1457] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[1458] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[1459] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[1460] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[1461] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[1462] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[1463] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[1464] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[1465] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[1466] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[1467] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[1468] Train Loss:    tensor(0.7870, device='cuda:0', grad_fn=<NllLossBackward>)
[1469] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[1470] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[1471] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[1472] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[1473] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[1474] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[1475] Train Loss:    tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[1476] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[1477] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[1478] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[1479] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[1480] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[1481] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[1482] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[1483] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[1484] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[1485] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[1486] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[1487] Train Loss:    tensor(0.7805, device='cuda:0', grad_fn=<NllLossBackward>)
[1488] Train Loss:    tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[1489] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[1490] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[1491] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[1492] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[1493] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[1494] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1495] Train Loss:   tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1496] Train Loss:   tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[1497] Train Loss:   tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[1498] Train Loss:   tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[1499] Train Loss:   tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[1500] Train Loss:   tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[1501] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[1502] Train Loss:   tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[1503] Train Loss:   tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[1504] Train Loss:   tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[1505] Train Loss:   tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[1506] Train Loss:   tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[1507] Train Loss:   tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[1508] Train Loss:   tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[1509] Train Loss:   tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[1510] Train Loss:   tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[1511] Train Loss:   tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[1512] Train Loss:   tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[1513] Train Loss:   tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[1514] Train Loss:   tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[1515] Train Loss:   tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[1516] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[1517] Train Loss:   tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[1518] Train Loss:   tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[1519] Train Loss:   tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[1520] Train Loss:   tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[1521] Train Loss:   tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[1522] Train Loss:   tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[1523] Train Loss:   tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[1524] Train Loss:   tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[1525] Train Loss:   tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[1526] Train Loss:   tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[1527] Train Loss:   tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[1528] Train Loss:   tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[1529] Train Loss:   tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[1530] Train Loss:   tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[1531] Train Loss:   tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[1532] Train Loss:   tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[1533] Train Loss:   tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[1534] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1535] Train Loss:   tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[1536] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1537] Train Loss:   tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[1538] Train Loss:   tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[1539] Train Loss:   tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[1540] Train Loss:   tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[1541] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[1542] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1543] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[1544] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[1545] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[1546] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[1547] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[1548] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1549] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[1550] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[1551] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[1552] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[1553] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[1554] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[1555] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[1556] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1557] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1558] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[1559] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[1560] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[1561] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[1562] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[1563] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[1564] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[1565] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[1566] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[1567] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[1568] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[1569] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[1570] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[1571] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1572] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[1573] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1574] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[1575] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[1576] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1577] Train Loss:    tensor(0.7850, device='cuda:0', grad_fn=<NllLossBackward>)
[1578] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[1579] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[1580] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[1581] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[1582] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[1583] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[1584] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[1585] Train Loss:    tensor(0.7709, device='cuda:0', grad_fn=<NllLossBackward>)
[1586] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[1587] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1588] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[1589] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1590] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1591] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[1592] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[1593] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[1594] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[1595] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[1596] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[1597] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[1598] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[1599] Train Loss:    tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[1600] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[1601] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[1602] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[1603] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[1604] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[1605] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[1606] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[1607] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[1608] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[1609] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[1610] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[1611] Train Loss:    tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[1612] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[1613] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[1614] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[1615] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[1616] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[1617] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[1618] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[1619] Train Loss:    tensor(0.7767, device='cuda:0', grad_fn=<NllLossBackward>)
[1620] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[1621] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[1622] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[1623] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1624] Train Loss:    tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[1625] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[1626] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[1627] Train Loss:    tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[1628] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[1629] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[1630] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[1631] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[1632] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[1633] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[1634] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[1635] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[1636] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1637] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[1638] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1639] Train Loss:    tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[1640] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[1641] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[1642] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[1643] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[1644] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[1645] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[1646] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[1647] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[1648] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[1649] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[1650] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[1651] Train Loss:    tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[1652] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[1653] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[1654] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[1655] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[1656] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[1657] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[1658] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[1659] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[1660] Train Loss:    tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[1661] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[1662] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[1663] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[1664] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[1665] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[1666] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[1667] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[1668] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[1669] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[1670] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[1671] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[1672] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[1673] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[1674] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[1675] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[1676] Train Loss:    tensor(0.7706, device='cuda:0', grad_fn=<NllLossBackward>)
[1677] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[1678] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[1679] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[1680] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[1681] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[1682] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[1683] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[1684] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[1685] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[1686] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1687] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[1688] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[1689] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[1690] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[1691] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[1692] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[1693] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[1694] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[1695] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[1696] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1697] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[1698] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[1699] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[1700] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[1701] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[1702] Train Loss:    tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[1703] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[1704] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[1705] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[1706] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[1707] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[1708] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[1709] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[1710] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[1711] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[1712] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[1713] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1714] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[1715] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[1716] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[1717] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[1718] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[1719] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[1720] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[1721] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[1722] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[1723] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[1724] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1725] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[1726] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[1727] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[1728] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1729] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[1730] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[1731] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[1732] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[1733] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[1734] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1735] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[1736] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[1737] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[1738] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[1739] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[1740] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[1741] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[1742] Train Loss:    tensor(0.7922, device='cuda:0', grad_fn=<NllLossBackward>)
[1743] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[1744] Train Loss:    tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[1745] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[1746] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[1747] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[1748] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[1749] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[1750] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[1751] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[1752] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[1753] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[1754] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1755] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[1756] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[1757] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[1758] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[1759] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[1760] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[1761] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[1762] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1763] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[1764] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[1765] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[1766] Train Loss:    tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[1767] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[1768] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[1769] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1770] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[1771] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[1772] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[1773] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[1774] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[1775] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[1776] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[1777] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[1778] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[1779] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[1780] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1781] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[1782] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1783] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[1784] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1785] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[1786] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[1787] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[1788] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[1789] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[1790] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[1791] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[1792] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[1793] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[1794] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[1795] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[1796] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[1797] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[1798] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[1799] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[1800] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[1801] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[1802] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[1803] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[1804] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[1805] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[1806] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[1807] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[1808] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[1809] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[1810] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[1811] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[1812] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[1813] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[1814] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[1815] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[1816] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[1817] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[1818] Train Loss:    tensor(0.7746, device='cuda:0', grad_fn=<NllLossBackward>)
[1819] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[1820] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[1821] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[1822] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[1823] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[1824] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[1825] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[1826] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[1827] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[1828] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[1829] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[1830] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1831] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[1832] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[1833] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[1834] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[1835] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[1836] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[1837] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[1838] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[1839] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[1840] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[1841] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[1842] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[1843] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[1844] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[1845] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[1846] Train Loss:    tensor(0.7865, device='cuda:0', grad_fn=<NllLossBackward>)
[1847] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[1848] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[1849] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[1850] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[1851] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1852] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
[1853] Train Loss:    tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[1854] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[1855] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[1856] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[1857] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[1858] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[1859] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[1860] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[1861] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[1862] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[1863] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[1864] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[1865] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[1866] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[1867] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[1868] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[1869] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[1870] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[1871] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[1872] Train Loss:    tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[1873] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[1874] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[1875] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[1876] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1877] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[1878] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1879] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[1880] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[1881] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[1882] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[1883] Train Loss:    tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[1884] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[1885] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[1886] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[1887] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[1888] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[1889] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[1890] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[1891] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[1892] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[1893] Train Loss:    tensor(0.7915, device='cuda:0', grad_fn=<NllLossBackward>)
[1894] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[1895] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1896] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1897] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[1898] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[1899] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[1900] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[1901] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[1902] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[1903] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[1904] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[1905] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[1906] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1907] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[1908] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[1909] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[1910] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[1911] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[1912] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[1913] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[1914] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[1915] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[1916] Train Loss:    tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[1917] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[1918] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1919] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[1920] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[1921] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[1922] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[1923] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[1924] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[1925] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[1926] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1927] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[1928] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[1929] Train Loss:    tensor(0.7832, device='cuda:0', grad_fn=<NllLossBackward>)
[1930] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[1931] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[1932] Train Loss:    tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[1933] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[1934] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[1935] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[1936] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[1937] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[1938] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[1939] Train Loss:    tensor(0.7861, device='cuda:0', grad_fn=<NllLossBackward>)
[1940] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[1941] Train Loss:    tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[1942] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[1943] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[1944] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1945] Train Loss:    tensor(0.7763, device='cuda:0', grad_fn=<NllLossBackward>)
[1946] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[1947] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[1948] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[1949] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[1950] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[1951] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[1952] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[1953] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[1954] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[1955] Train Loss:    tensor(0.7791, device='cuda:0', grad_fn=<NllLossBackward>)
[1956] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[1957] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[1958] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[1959] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[1960] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[1961] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[1962] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[1963] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[1964] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[1965] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[1966] Train Loss:    tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[1967] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[1968] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[1969] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[1970] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[1971] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[1972] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[1973] Train Loss:    tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[1974] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1975] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[1976] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[1977] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[1978] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1979] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[1980] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[1981] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[1982] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[1983] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[1984] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[1985] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[1986] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[1987] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[1988] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[1989] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[1990] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[1991] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1992] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[1993] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[1994] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[1995] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[1996] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[1997] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[1998] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[1999] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[2000] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[2001] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[2002] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[2003] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[2004] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[2005] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[2006] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[2007] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[2008] Train Loss:    tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[2009] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[2010] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[2011] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[2012] Train Loss:    tensor(0.7849, device='cuda:0', grad_fn=<NllLossBackward>)
[2013] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[2014] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[2015] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[2016] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[2017] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[2018] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[2019] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[2020] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[2021] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[2022] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2023] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[2024] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[2025] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[2026] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[2027] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[2028] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[2029] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[2030] Train Loss:    tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[2031] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[2032] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[2033] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[2034] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[2035] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2036] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[2037] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[2038] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[2039] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[2040] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[2041] Train Loss:    tensor(0.7727, device='cuda:0', grad_fn=<NllLossBackward>)
[2042] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[2043] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2044] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[2045] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[2046] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[2047] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[2048] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[2049] Train Loss:    tensor(0.7747, device='cuda:0', grad_fn=<NllLossBackward>)
[2050] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[2051] Train Loss:    tensor(0.7769, device='cuda:0', grad_fn=<NllLossBackward>)
[2052] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[2053] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[2054] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2055] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[2056] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[2057] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[2058] Train Loss:    tensor(0.7901, device='cuda:0', grad_fn=<NllLossBackward>)
[2059] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[2060] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[2061] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[2062] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[2063] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[2064] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[2065] Train Loss:    tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[2066] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[2067] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[2068] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[2069] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[2070] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2071] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[2072] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[2073] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[2074] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[2075] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[2076] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[2077] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[2078] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[2079] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[2080] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[2081] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[2082] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[2083] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[2084] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[2085] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[2086] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[2087] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[2088] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[2089] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[2090] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[2091] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[2092] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[2093] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2094] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[2095] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2096] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[2097] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[2098] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[2099] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[2100] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[2101] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[2102] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[2103] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[2104] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[2105] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[2106] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[2107] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[2108] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[2109] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[2110] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[2111] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[2112] Train Loss:    tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[2113] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[2114] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2115] Train Loss:    tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[2116] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[2117] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[2118] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2119] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[2120] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[2121] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[2122] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[2123] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[2124] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[2125] Train Loss:    tensor(0.7717, device='cuda:0', grad_fn=<NllLossBackward>)
[2126] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[2127] Train Loss:    tensor(0.7722, device='cuda:0', grad_fn=<NllLossBackward>)
[2128] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[2129] Train Loss:    tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[2130] Train Loss:    tensor(0.7784, device='cuda:0', grad_fn=<NllLossBackward>)
[2131] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[2132] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2133] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[2134] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[2135] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[2136] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[2137] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[2138] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2139] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[2140] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[2141] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[2142] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[2143] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[2144] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[2145] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[2146] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[2147] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[2148] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[2149] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[2150] Train Loss:    tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[2151] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[2152] Train Loss:    tensor(0.7847, device='cuda:0', grad_fn=<NllLossBackward>)
[2153] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[2154] Train Loss:    tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[2155] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[2156] Train Loss:    tensor(0.7731, device='cuda:0', grad_fn=<NllLossBackward>)
[2157] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[2158] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[2159] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[2160] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2161] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[2162] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2163] Train Loss:    tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[2164] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2165] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[2166] Train Loss:    tensor(0.7839, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2167] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[2168] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[2169] Train Loss:    tensor(0.7815, device='cuda:0', grad_fn=<NllLossBackward>)
[2170] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[2171] Train Loss:    tensor(0.7674, device='cuda:0', grad_fn=<NllLossBackward>)
[2172] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[2173] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[2174] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[2175] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[2176] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[2177] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[2178] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[2179] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[2180] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[2181] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[2182] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[2183] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[2184] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[2185] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[2186] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[2187] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[2188] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[2189] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[2190] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[2191] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2192] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[2193] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[2194] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[2195] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[2196] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[2197] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[2198] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[2199] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[2200] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[2201] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[2202] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[2203] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[2204] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[2205] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[2206] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2207] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2208] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[2209] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[2210] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[2211] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[2212] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[2213] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[2214] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2215] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[2216] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[2217] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[2218] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[2219] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[2220] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[2221] Train Loss:    tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[2222] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[2223] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[2224] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[2225] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[2226] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[2227] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[2228] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[2229] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[2230] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[2231] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[2232] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[2233] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[2234] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[2235] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[2236] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[2237] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[2238] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[2239] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[2240] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[2241] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[2242] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[2243] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[2244] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[2245] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[2246] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[2247] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[2248] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[2249] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[2250] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[2251] Train Loss:    tensor(0.7788, device='cuda:0', grad_fn=<NllLossBackward>)
[2252] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[2253] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[2254] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[2255] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[2256] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2257] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[2258] Train Loss:    tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[2259] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[2260] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[2261] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[2262] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2263] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[2264] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[2265] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[2266] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[2267] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[2268] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[2269] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[2270] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[2271] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[2272] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[2273] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[2274] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[2275] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[2276] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[2277] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[2278] Train Loss:    tensor(0.7743, device='cuda:0', grad_fn=<NllLossBackward>)
[2279] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[2280] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2281] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[2282] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[2283] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[2284] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[2285] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[2286] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[2287] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[2288] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[2289] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[2290] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[2291] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[2292] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[2293] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2294] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2295] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[2296] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[2297] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[2298] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[2299] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[2300] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[2301] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[2302] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[2303] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[2304] Train Loss:    tensor(0.7758, device='cuda:0', grad_fn=<NllLossBackward>)
[2305] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[2306] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[2307] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[2308] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[2309] Train Loss:    tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[2310] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2311] Train Loss:   tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[2312] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[2313] Train Loss:   tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2314] Train Loss:   tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[2315] Train Loss:   tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[2316] Train Loss:   tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[2317] Train Loss:   tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[2318] Train Loss:   tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[2319] Train Loss:   tensor(0.7707, device='cuda:0', grad_fn=<NllLossBackward>)
[2320] Train Loss:   tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[2321] Train Loss:   tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[2322] Train Loss:   tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[2323] Train Loss:   tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2324] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[2325] Train Loss:   tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[2326] Train Loss:   tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[2327] Train Loss:   tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[2328] Train Loss:   tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[2329] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[2330] Train Loss:   tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[2331] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[2332] Train Loss:   tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[2333] Train Loss:   tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2334] Train Loss:   tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2335] Train Loss:   tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[2336] Train Loss:   tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[2337] Train Loss:   tensor(0.7893, device='cuda:0', grad_fn=<NllLossBackward>)
[2338] Train Loss:   tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[2339] Train Loss:   tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[2340] Train Loss:   tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2341] Train Loss:   tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[2342] Train Loss:   tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[2343] Train Loss:   tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[2344] Train Loss:   tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[2345] Train Loss:   tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2346] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[2347] Train Loss:   tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[2348] Train Loss:   tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[2349] Train Loss:   tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[2350] Train Loss:   tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[2351] Train Loss:   tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[2352] Train Loss:   tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[2353] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[2354] Train Loss:   tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2355] Train Loss:   tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[2356] Train Loss:   tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[2357] Train Loss:   tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[2358] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2359] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[2360] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2361] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[2362] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[2363] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[2364] Train Loss:    tensor(0.8068, device='cuda:0', grad_fn=<NllLossBackward>)
[2365] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[2366] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2367] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[2368] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2369] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[2370] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[2371] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[2372] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[2373] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[2374] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[2375] Train Loss:    tensor(0.7672, device='cuda:0', grad_fn=<NllLossBackward>)
[2376] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[2377] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[2378] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[2379] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[2380] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[2381] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[2382] Train Loss:    tensor(0.7990, device='cuda:0', grad_fn=<NllLossBackward>)
[2383] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[2384] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[2385] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[2386] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[2387] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[2388] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[2389] Train Loss:    tensor(0.7648, device='cuda:0', grad_fn=<NllLossBackward>)
[2390] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[2391] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[2392] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[2393] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[2394] Train Loss:    tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[2395] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2396] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[2397] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[2398] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[2399] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[2400] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[2401] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[2402] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[2403] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[2404] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[2405] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[2406] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2407] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[2408] Train Loss:   tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[2409] Train Loss:   tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2410] Train Loss:   tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[2411] Train Loss:   tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[2412] Train Loss:   tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[2413] Train Loss:   tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[2414] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2415] Train Loss:   tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[2416] Train Loss:   tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[2417] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[2418] Train Loss:   tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[2419] Train Loss:   tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2420] Train Loss:   tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[2421] Train Loss:   tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[2422] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[2423] Train Loss:   tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[2424] Train Loss:   tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[2425] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[2426] Train Loss:   tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[2427] Train Loss:   tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[2428] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2429] Train Loss:   tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[2430] Train Loss:   tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[2431] Train Loss:   tensor(0.7686, device='cuda:0', grad_fn=<NllLossBackward>)
[2432] Train Loss:   tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[2433] Train Loss:   tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[2434] Train Loss:   tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[2435] Train Loss:   tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[2436] Train Loss:   tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[2437] Train Loss:   tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[2438] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[2439] Train Loss:   tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[2440] Train Loss:   tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[2441] Train Loss:   tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[2442] Train Loss:   tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[2443] Train Loss:   tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[2444] Train Loss:   tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[2445] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[2446] Train Loss:   tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[2447] Train Loss:   tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[2448] Train Loss:   tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[2449] Train Loss:   tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[2450] Train Loss:   tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[2451] Train Loss:   tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[2452] Train Loss:   tensor(0.7857, device='cuda:0', grad_fn=<NllLossBackward>)
[2453] Train Loss:   tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[2454] Train Loss:   tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2455] Train Loss:   tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2456] Train Loss:   tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[2457] Train Loss:   tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[2458] Train Loss:   tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2459] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[2460] Train Loss:   tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[2461] Train Loss:   tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[2462] Train Loss:   tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[2463] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[2464] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[2465] Train Loss:   tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[2466] Train Loss:   tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[2467] Train Loss:   tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[2468] Train Loss:   tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[2469] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[2470] Train Loss:   tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[2471] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[2472] Train Loss:   tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[2473] Train Loss:   tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[2474] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[2475] Train Loss:   tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[2476] Train Loss:   tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[2477] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[2478] Train Loss:   tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[2479] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[2480] Train Loss:   tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[2481] Train Loss:   tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[2482] Train Loss:   tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[2483] Train Loss:   tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[2484] Train Loss:   tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[2485] Train Loss:   tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[2486] Train Loss:   tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[2487] Train Loss:   tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[2488] Train Loss:   tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[2489] Train Loss:   tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[2490] Train Loss:   tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[2491] Train Loss:   tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[2492] Train Loss:   tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[2493] Train Loss:   tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[2494] Train Loss:   tensor(0.7657, device='cuda:0', grad_fn=<NllLossBackward>)
[2495] Train Loss:   tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2496] Train Loss:   tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[2497] Train Loss:   tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[2498] Train Loss:   tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[2499] Train Loss:   tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[2500] Train Loss:   tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[2501] Train Loss:   tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[2502] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2503] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[2504] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[2505] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[2506] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[2507] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[2508] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[2509] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[2510] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[2511] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[2512] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[2513] Train Loss:    tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[2514] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2515] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[2516] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[2517] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[2518] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[2519] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[2520] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[2521] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[2522] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[2523] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[2524] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[2525] Train Loss:    tensor(0.7774, device='cuda:0', grad_fn=<NllLossBackward>)
[2526] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[2527] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2528] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2529] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[2530] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[2531] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[2532] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2533] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[2534] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[2535] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[2536] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[2537] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[2538] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[2539] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2540] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[2541] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[2542] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[2543] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[2544] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[2545] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[2546] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[2547] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[2548] Train Loss:    tensor(0.7762, device='cuda:0', grad_fn=<NllLossBackward>)
[2549] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[2550] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2551] Train Loss:    tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[2552] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[2553] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[2554] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[2555] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[2556] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[2557] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2558] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[2559] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[2560] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[2561] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[2562] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[2563] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[2564] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[2565] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[2566] Train Loss:    tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[2567] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[2568] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[2569] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[2570] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2571] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[2572] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[2573] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[2574] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[2575] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[2576] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[2577] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[2578] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[2579] Train Loss:    tensor(0.7761, device='cuda:0', grad_fn=<NllLossBackward>)
[2580] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[2581] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[2582] Train Loss:    tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[2583] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[2584] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[2585] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[2586] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[2587] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[2588] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[2589] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[2590] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[2591] Train Loss:    tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[2592] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[2593] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[2594] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[2595] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[2596] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[2597] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[2598] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2599] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[2600] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[2601] Train Loss:    tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[2602] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[2603] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[2604] Train Loss:    tensor(0.7776, device='cuda:0', grad_fn=<NllLossBackward>)
[2605] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[2606] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[2607] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[2608] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[2609] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[2610] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[2611] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[2612] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[2613] Train Loss:    tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[2614] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2615] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[2616] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[2617] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[2618] Train Loss:    tensor(0.7867, device='cuda:0', grad_fn=<NllLossBackward>)
[2619] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[2620] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[2621] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[2622] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[2623] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[2624] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2625] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[2626] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[2627] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2628] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2629] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[2630] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[2631] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[2632] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[2633] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[2634] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[2635] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[2636] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[2637] Train Loss:    tensor(0.7693, device='cuda:0', grad_fn=<NllLossBackward>)
[2638] Train Loss:    tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[2639] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[2640] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[2641] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[2642] Train Loss:    tensor(0.7855, device='cuda:0', grad_fn=<NllLossBackward>)
[2643] Train Loss:    tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[2644] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2645] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[2646] Train Loss:    tensor(0.7794, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2647] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[2648] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2649] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[2650] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[2651] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2652] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[2653] Train Loss:    tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[2654] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2655] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[2656] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[2657] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[2658] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[2659] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[2660] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[2661] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2662] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[2663] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[2664] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[2665] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[2666] Train Loss:    tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[2667] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[2668] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2669] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[2670] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2671] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[2672] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2673] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[2674] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[2675] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[2676] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[2677] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[2678] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[2679] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[2680] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[2681] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[2682] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[2683] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[2684] Train Loss:    tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[2685] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[2686] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[2687] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[2688] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[2689] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[2690] Train Loss:    tensor(0.7765, device='cuda:0', grad_fn=<NllLossBackward>)
[2691] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[2692] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[2693] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[2694] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2695] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2696] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[2697] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[2698] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[2699] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[2700] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[2701] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[2702] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[2703] Train Loss:    tensor(0.7924, device='cuda:0', grad_fn=<NllLossBackward>)
[2704] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[2705] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[2706] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[2707] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[2708] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[2709] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[2710] Train Loss:    tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[2711] Train Loss:    tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[2712] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[2713] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[2714] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[2715] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[2716] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[2717] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[2718] Train Loss:    tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[2719] Train Loss:    tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[2720] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[2721] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[2722] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[2723] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[2724] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[2725] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[2726] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[2727] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[2728] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[2729] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[2730] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[2731] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[2732] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[2733] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[2734] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[2735] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[2736] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[2737] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[2738] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[2739] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[2740] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[2741] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[2742] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2743] Train Loss:   tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[2744] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[2745] Train Loss:   tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[2746] Train Loss:   tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[2747] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[2748] Train Loss:   tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[2749] Train Loss:   tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[2750] Train Loss:   tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[2751] Train Loss:   tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[2752] Train Loss:   tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[2753] Train Loss:   tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[2754] Train Loss:   tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[2755] Train Loss:   tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[2756] Train Loss:   tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[2757] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[2758] Train Loss:   tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[2759] Train Loss:   tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[2760] Train Loss:   tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[2761] Train Loss:   tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[2762] Train Loss:   tensor(0.7817, device='cuda:0', grad_fn=<NllLossBackward>)
[2763] Train Loss:   tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[2764] Train Loss:   tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[2765] Train Loss:   tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[2766] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[2767] Train Loss:   tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[2768] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[2769] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[2770] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[2771] Train Loss:   tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[2772] Train Loss:   tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[2773] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[2774] Train Loss:   tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2775] Train Loss:   tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[2776] Train Loss:   tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[2777] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[2778] Train Loss:   tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[2779] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[2780] Train Loss:   tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[2781] Train Loss:   tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[2782] Train Loss:   tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[2783] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[2784] Train Loss:   tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[2785] Train Loss:   tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[2786] Train Loss:   tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[2787] Train Loss:   tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[2788] Train Loss:   tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[2789] Train Loss:   tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[2790] Train Loss:   tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2791] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[2792] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[2793] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[2794] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[2795] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[2796] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[2797] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[2798] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[2799] Train Loss:    tensor(0.7799, device='cuda:0', grad_fn=<NllLossBackward>)
[2800] Train Loss:    tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[2801] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[2802] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[2803] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[2804] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[2805] Train Loss:    tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[2806] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[2807] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[2808] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[2809] Train Loss:    tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[2810] Train Loss:    tensor(0.7800, device='cuda:0', grad_fn=<NllLossBackward>)
[2811] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[2812] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[2813] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[2814] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[2815] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[2816] Train Loss:    tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[2817] Train Loss:    tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[2818] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[2819] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[2820] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[2821] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[2822] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[2823] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[2824] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[2825] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[2826] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[2827] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[2828] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[2829] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[2830] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[2831] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[2832] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[2833] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[2834] Train Loss:    tensor(0.7797, device='cuda:0', grad_fn=<NllLossBackward>)
[2835] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[2836] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[2837] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[2838] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2839] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[2840] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[2841] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2842] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[2843] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[2844] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2845] Train Loss:    tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[2846] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2847] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[2848] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[2849] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[2850] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[2851] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[2852] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[2853] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[2854] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[2855] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[2856] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[2857] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[2858] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[2859] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[2860] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[2861] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[2862] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[2863] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2864] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[2865] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[2866] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[2867] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[2868] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[2869] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[2870] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[2871] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[2872] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[2873] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[2874] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[2875] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[2876] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[2877] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[2878] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[2879] Train Loss:    tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[2880] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[2881] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[2882] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[2883] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[2884] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[2885] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[2886] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2887] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[2888] Train Loss:    tensor(0.7756, device='cuda:0', grad_fn=<NllLossBackward>)
[2889] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[2890] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[2891] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[2892] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[2893] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[2894] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[2895] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2896] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[2897] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[2898] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[2899] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[2900] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[2901] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2902] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[2903] Train Loss:    tensor(0.7687, device='cuda:0', grad_fn=<NllLossBackward>)
[2904] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[2905] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[2906] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[2907] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[2908] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[2909] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[2910] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[2911] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[2912] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[2913] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[2914] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[2915] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[2916] Train Loss:    tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[2917] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[2918] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[2919] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[2920] Train Loss:    tensor(0.7735, device='cuda:0', grad_fn=<NllLossBackward>)
[2921] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[2922] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[2923] Train Loss:    tensor(0.7729, device='cuda:0', grad_fn=<NllLossBackward>)
[2924] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[2925] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[2926] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2927] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[2928] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[2929] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[2930] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[2931] Train Loss:    tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[2932] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[2933] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[2934] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2935] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2936] Train Loss:    tensor(0.7901, device='cuda:0', grad_fn=<NllLossBackward>)
[2937] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[2938] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[2939] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[2940] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[2941] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[2942] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[2943] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[2944] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[2945] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[2946] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[2947] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[2948] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[2949] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[2950] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2951] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[2952] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[2953] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[2954] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[2955] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[2956] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[2957] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[2958] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[2959] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[2960] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[2961] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[2962] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[2963] Train Loss:    tensor(0.7701, device='cuda:0', grad_fn=<NllLossBackward>)
[2964] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[2965] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[2966] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[2967] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[2968] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[2969] Train Loss:    tensor(0.7771, device='cuda:0', grad_fn=<NllLossBackward>)
[2970] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[2971] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2972] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[2973] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[2974] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[2975] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[2976] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[2977] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[2978] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[2979] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[2980] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[2981] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[2982] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2983] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[2984] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2985] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[2986] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[2987] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[2988] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[2989] Train Loss:    tensor(0.7812, device='cuda:0', grad_fn=<NllLossBackward>)
[2990] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[2991] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[2992] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2993] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[2994] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[2995] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[2996] Train Loss:    tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[2997] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[2998] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[2999] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3000] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[3001] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[3002] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[3003] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[3004] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[3005] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[3006] Train Loss:    tensor(0.7766, device='cuda:0', grad_fn=<NllLossBackward>)
[3007] Train Loss:    tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[3008] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[3009] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[3010] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[3011] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[3012] Train Loss:    tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[3013] Train Loss:    tensor(0.7748, device='cuda:0', grad_fn=<NllLossBackward>)
[3014] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[3015] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3016] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[3017] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[3018] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[3019] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[3020] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[3021] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[3022] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[3023] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[3024] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3025] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[3026] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[3027] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[3028] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[3029] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[3030] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3031] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[3032] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[3033] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3034] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[3035] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3036] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3037] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[3038] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[3039] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[3040] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[3041] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[3042] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[3043] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[3044] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[3045] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[3046] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[3047] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[3048] Train Loss:    tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[3049] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[3050] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[3051] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[3052] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[3053] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[3054] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[3055] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[3056] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3057] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3058] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[3059] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[3060] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[3061] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[3062] Train Loss:    tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[3063] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[3064] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3065] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[3066] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[3067] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[3068] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[3069] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[3070] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[3071] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[3072] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[3073] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3074] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[3075] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[3076] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[3077] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[3078] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3079] Train Loss:   tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[3080] Train Loss:   tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[3081] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[3082] Train Loss:   tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[3083] Train Loss:   tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[3084] Train Loss:   tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[3085] Train Loss:   tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[3086] Train Loss:   tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[3087] Train Loss:   tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[3088] Train Loss:   tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[3089] Train Loss:   tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[3090] Train Loss:   tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3091] Train Loss:   tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[3092] Train Loss:   tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3093] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3094] Train Loss:   tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[3095] Train Loss:   tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[3096] Train Loss:   tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[3097] Train Loss:   tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[3098] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[3099] Train Loss:   tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[3100] Train Loss:   tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[3101] Train Loss:   tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[3102] Train Loss:   tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[3103] Train Loss:   tensor(0.7734, device='cuda:0', grad_fn=<NllLossBackward>)
[3104] Train Loss:   tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[3105] Train Loss:   tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[3106] Train Loss:   tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[3107] Train Loss:   tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[3108] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[3109] Train Loss:   tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[3110] Train Loss:   tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[3111] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3112] Train Loss:   tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[3113] Train Loss:   tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[3114] Train Loss:   tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[3115] Train Loss:   tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[3116] Train Loss:   tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[3117] Train Loss:   tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[3118] Train Loss:   tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[3119] Train Loss:   tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[3120] Train Loss:   tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[3121] Train Loss:   tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[3122] Train Loss:   tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[3123] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[3124] Train Loss:   tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[3125] Train Loss:   tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[3126] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3127] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[3128] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[3129] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[3130] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[3131] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[3132] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[3133] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[3134] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[3135] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[3136] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[3137] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[3138] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[3139] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[3140] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[3141] Train Loss:    tensor(0.7928, device='cuda:0', grad_fn=<NllLossBackward>)
[3142] Train Loss:    tensor(0.7627, device='cuda:0', grad_fn=<NllLossBackward>)
[3143] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[3144] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[3145] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[3146] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3147] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[3148] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[3149] Train Loss:    tensor(0.7696, device='cuda:0', grad_fn=<NllLossBackward>)
[3150] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[3151] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[3152] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[3153] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[3154] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[3155] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[3156] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[3157] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[3158] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[3159] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[3160] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[3161] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[3162] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[3163] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3164] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[3165] Train Loss:    tensor(0.7868, device='cuda:0', grad_fn=<NllLossBackward>)
[3166] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[3167] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[3168] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[3169] Train Loss:    tensor(0.7654, device='cuda:0', grad_fn=<NllLossBackward>)
[3170] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[3171] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[3172] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[3173] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3174] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3175] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[3176] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3177] Train Loss:   tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[3178] Train Loss:   tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[3179] Train Loss:   tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[3180] Train Loss:   tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[3181] Train Loss:   tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[3182] Train Loss:   tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[3183] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3184] Train Loss:   tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[3185] Train Loss:   tensor(0.7764, device='cuda:0', grad_fn=<NllLossBackward>)
[3186] Train Loss:   tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[3187] Train Loss:   tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[3188] Train Loss:   tensor(0.7663, device='cuda:0', grad_fn=<NllLossBackward>)
[3189] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[3190] Train Loss:   tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[3191] Train Loss:   tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3192] Train Loss:   tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[3193] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[3194] Train Loss:   tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[3195] Train Loss:   tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[3196] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[3197] Train Loss:   tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[3198] Train Loss:   tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[3199] Train Loss:   tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[3200] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[3201] Train Loss:   tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[3202] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[3203] Train Loss:   tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[3204] Train Loss:   tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[3205] Train Loss:   tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[3206] Train Loss:   tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[3207] Train Loss:   tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[3208] Train Loss:   tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[3209] Train Loss:   tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[3210] Train Loss:   tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[3211] Train Loss:   tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[3212] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[3213] Train Loss:   tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3214] Train Loss:   tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[3215] Train Loss:   tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[3216] Train Loss:   tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[3217] Train Loss:   tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[3218] Train Loss:   tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[3219] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3220] Train Loss:   tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[3221] Train Loss:   tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[3222] Train Loss:   tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3223] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[3224] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3225] Train Loss:    tensor(0.7712, device='cuda:0', grad_fn=<NllLossBackward>)
[3226] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[3227] Train Loss:    tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[3228] Train Loss:    tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[3229] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[3230] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[3231] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3232] Train Loss:    tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[3233] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[3234] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[3235] Train Loss:    tensor(0.7724, device='cuda:0', grad_fn=<NllLossBackward>)
[3236] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[3237] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3238] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[3239] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[3240] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[3241] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3242] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[3243] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[3244] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[3245] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3246] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3247] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[3248] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3249] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[3250] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[3251] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3252] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[3253] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3254] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[3255] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[3256] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[3257] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[3258] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[3259] Train Loss:    tensor(0.7705, device='cuda:0', grad_fn=<NllLossBackward>)
[3260] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[3261] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[3262] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[3263] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3264] Train Loss:    tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[3265] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[3266] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[3267] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[3268] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[3269] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[3270] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3271] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[3272] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[3273] Train Loss:   tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3274] Train Loss:   tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3275] Train Loss:   tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3276] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3277] Train Loss:   tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[3278] Train Loss:   tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[3279] Train Loss:   tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[3280] Train Loss:   tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[3281] Train Loss:   tensor(0.7719, device='cuda:0', grad_fn=<NllLossBackward>)
[3282] Train Loss:   tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[3283] Train Loss:   tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[3284] Train Loss:   tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[3285] Train Loss:   tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[3286] Train Loss:   tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[3287] Train Loss:   tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[3288] Train Loss:   tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3289] Train Loss:   tensor(0.7638, device='cuda:0', grad_fn=<NllLossBackward>)
[3290] Train Loss:   tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[3291] Train Loss:   tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[3292] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[3293] Train Loss:   tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[3294] Train Loss:   tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[3295] Train Loss:   tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[3296] Train Loss:   tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[3297] Train Loss:   tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[3298] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[3299] Train Loss:   tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[3300] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[3301] Train Loss:   tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3302] Train Loss:   tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[3303] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[3304] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[3305] Train Loss:   tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[3306] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[3307] Train Loss:   tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[3308] Train Loss:   tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[3309] Train Loss:   tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[3310] Train Loss:   tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[3311] Train Loss:   tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[3312] Train Loss:   tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3313] Train Loss:   tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3314] Train Loss:   tensor(0.7993, device='cuda:0', grad_fn=<NllLossBackward>)
[3315] Train Loss:   tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[3316] Train Loss:   tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[3317] Train Loss:   tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[3318] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3319] Train Loss:   tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3320] Train Loss:   tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[3321] Train Loss:   tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[3322] Train Loss:   tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[3323] Train Loss:   tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[3324] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3325] Train Loss:   tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[3326] Train Loss:   tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[3327] Train Loss:   tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[3328] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[3329] Train Loss:   tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[3330] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[3331] Train Loss:   tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3332] Train Loss:   tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[3333] Train Loss:   tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[3334] Train Loss:   tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[3335] Train Loss:   tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[3336] Train Loss:   tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[3337] Train Loss:   tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[3338] Train Loss:   tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[3339] Train Loss:   tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[3340] Train Loss:   tensor(0.7891, device='cuda:0', grad_fn=<NllLossBackward>)
[3341] Train Loss:   tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[3342] Train Loss:   tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[3343] Train Loss:   tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[3344] Train Loss:   tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[3345] Train Loss:   tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[3346] Train Loss:   tensor(0.7824, device='cuda:0', grad_fn=<NllLossBackward>)
[3347] Train Loss:   tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[3348] Train Loss:   tensor(0.7661, device='cuda:0', grad_fn=<NllLossBackward>)
[3349] Train Loss:   tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3350] Train Loss:   tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[3351] Train Loss:   tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[3352] Train Loss:   tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[3353] Train Loss:   tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[3354] Train Loss:   tensor(0.7792, device='cuda:0', grad_fn=<NllLossBackward>)
[3355] Train Loss:   tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[3356] Train Loss:   tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[3357] Train Loss:   tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[3358] Train Loss:   tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[3359] Train Loss:   tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[3360] Train Loss:   tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[3361] Train Loss:   tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[3362] Train Loss:   tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[3363] Train Loss:   tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[3364] Train Loss:   tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[3365] Train Loss:   tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[3366] Train Loss:   tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3367] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[3368] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[3369] Train Loss:    tensor(0.7778, device='cuda:0', grad_fn=<NllLossBackward>)
[3370] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[3371] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[3372] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[3373] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[3374] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[3375] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[3376] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[3377] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[3378] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[3379] Train Loss:    tensor(0.7894, device='cuda:0', grad_fn=<NllLossBackward>)
[3380] Train Loss:    tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[3381] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[3382] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[3383] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[3384] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[3385] Train Loss:    tensor(0.7737, device='cuda:0', grad_fn=<NllLossBackward>)
[3386] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[3387] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[3388] Train Loss:    tensor(0.7827, device='cuda:0', grad_fn=<NllLossBackward>)
[3389] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[3390] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[3391] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[3392] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3393] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[3394] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[3395] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[3396] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[3397] Train Loss:    tensor(0.7842, device='cuda:0', grad_fn=<NllLossBackward>)
[3398] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[3399] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[3400] Train Loss:    tensor(0.7711, device='cuda:0', grad_fn=<NllLossBackward>)
[3401] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[3402] Train Loss:    tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[3403] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[3404] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[3405] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[3406] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3407] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[3408] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[3409] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[3410] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[3411] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[3412] Train Loss:    tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[3413] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[3414] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3415] Train Loss:    tensor(0.7700, device='cuda:0', grad_fn=<NllLossBackward>)
[3416] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[3417] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[3418] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3419] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[3420] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[3421] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[3422] Train Loss:    tensor(0.7714, device='cuda:0', grad_fn=<NllLossBackward>)
[3423] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3424] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3425] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[3426] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[3427] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[3428] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[3429] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3430] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[3431] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[3432] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[3433] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[3434] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[3435] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[3436] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[3437] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3438] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[3439] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[3440] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[3441] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[3442] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[3443] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[3444] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[3445] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[3446] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[3447] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[3448] Train Loss:    tensor(0.7704, device='cuda:0', grad_fn=<NllLossBackward>)
[3449] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[3450] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[3451] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[3452] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[3453] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[3454] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[3455] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[3456] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[3457] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[3458] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3459] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[3460] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[3461] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[3462] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3463] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3464] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[3465] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[3466] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[3467] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[3468] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[3469] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3470] Train Loss:    tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[3471] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[3472] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[3473] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[3474] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[3475] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[3476] Train Loss:    tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[3477] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[3478] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[3479] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[3480] Train Loss:    tensor(0.7851, device='cuda:0', grad_fn=<NllLossBackward>)
[3481] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[3482] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[3483] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[3484] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[3485] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[3486] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[3487] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[3488] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[3489] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[3490] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[3491] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[3492] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3493] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[3494] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[3495] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[3496] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[3497] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[3498] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[3499] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[3500] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3501] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3502] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[3503] Train Loss:    tensor(0.7740, device='cuda:0', grad_fn=<NllLossBackward>)
[3504] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[3505] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[3506] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3507] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[3508] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[3509] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[3510] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3511] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[3512] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[3513] Train Loss:    tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[3514] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[3515] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[3516] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[3517] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[3518] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[3519] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[3520] Train Loss:    tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[3521] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3522] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[3523] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[3524] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3525] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[3526] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[3527] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[3528] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3529] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[3530] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[3531] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[3532] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[3533] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[3534] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[3535] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[3536] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[3537] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[3538] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[3539] Train Loss:    tensor(0.7798, device='cuda:0', grad_fn=<NllLossBackward>)
[3540] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[3541] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[3542] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[3543] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3544] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[3545] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[3546] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[3547] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[3548] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[3549] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[3550] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3551] Train Loss:    tensor(0.7716, device='cuda:0', grad_fn=<NllLossBackward>)
[3552] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[3553] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[3554] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[3555] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3556] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[3557] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[3558] Train Loss:    tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3559] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[3560] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[3561] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3562] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3563] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[3564] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[3565] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[3566] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[3567] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[3568] Train Loss:    tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[3569] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[3570] Train Loss:    tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[3571] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[3572] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[3573] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[3574] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[3575] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[3576] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[3577] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[3578] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3579] Train Loss:    tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[3580] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3581] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[3582] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3583] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[3584] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[3585] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[3586] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[3587] Train Loss:    tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[3588] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[3589] Train Loss:    tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[3590] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[3591] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[3592] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[3593] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[3594] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[3595] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[3596] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[3597] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[3598] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[3599] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[3600] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[3601] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[3602] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[3603] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[3604] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[3605] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[3606] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3607] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[3608] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[3609] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3610] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[3611] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[3612] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[3613] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[3614] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[3615] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[3616] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[3617] Train Loss:    tensor(0.7641, device='cuda:0', grad_fn=<NllLossBackward>)
[3618] Train Loss:    tensor(0.7821, device='cuda:0', grad_fn=<NllLossBackward>)
[3619] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[3620] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3621] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3622] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[3623] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[3624] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3625] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[3626] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[3627] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[3628] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[3629] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[3630] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[3631] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[3632] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[3633] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[3634] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[3635] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[3636] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[3637] Train Loss:    tensor(0.7883, device='cuda:0', grad_fn=<NllLossBackward>)
[3638] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[3639] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[3640] Train Loss:    tensor(0.7813, device='cuda:0', grad_fn=<NllLossBackward>)
[3641] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3642] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[3643] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[3644] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[3645] Train Loss:    tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[3646] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[3647] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3648] Train Loss:    tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[3649] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[3650] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[3651] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[3652] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3653] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[3654] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3655] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[3656] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[3657] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[3658] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[3659] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3660] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[3661] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[3662] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[3663] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[3664] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[3665] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3666] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[3667] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[3668] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[3669] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[3670] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[3671] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[3672] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[3673] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3674] Train Loss:    tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[3675] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[3676] Train Loss:    tensor(0.7667, device='cuda:0', grad_fn=<NllLossBackward>)
[3677] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3678] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[3679] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[3680] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[3681] Train Loss:    tensor(0.7739, device='cuda:0', grad_fn=<NllLossBackward>)
[3682] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[3683] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[3684] Train Loss:    tensor(0.7906, device='cuda:0', grad_fn=<NllLossBackward>)
[3685] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[3686] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[3687] Train Loss:    tensor(0.7678, device='cuda:0', grad_fn=<NllLossBackward>)
[3688] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[3689] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[3690] Train Loss:    tensor(0.7807, device='cuda:0', grad_fn=<NllLossBackward>)
[3691] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3692] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[3693] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[3694] Train Loss:    tensor(0.8007, device='cuda:0', grad_fn=<NllLossBackward>)
[3695] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[3696] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[3697] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[3698] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[3699] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3700] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[3701] Train Loss:    tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[3702] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3703] Train Loss:   tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[3704] Train Loss:   tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[3705] Train Loss:   tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[3706] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[3707] Train Loss:   tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[3708] Train Loss:   tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[3709] Train Loss:   tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[3710] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3711] Train Loss:   tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[3712] Train Loss:   tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3713] Train Loss:   tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[3714] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3715] Train Loss:   tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3716] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[3717] Train Loss:   tensor(0.7887, device='cuda:0', grad_fn=<NllLossBackward>)
[3718] Train Loss:   tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[3719] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[3720] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[3721] Train Loss:   tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[3722] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3723] Train Loss:   tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
[3724] Train Loss:   tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[3725] Train Loss:   tensor(0.7888, device='cuda:0', grad_fn=<NllLossBackward>)
[3726] Train Loss:   tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[3727] Train Loss:   tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[3728] Train Loss:   tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[3729] Train Loss:   tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[3730] Train Loss:   tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[3731] Train Loss:   tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[3732] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[3733] Train Loss:   tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[3734] Train Loss:   tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[3735] Train Loss:   tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[3736] Train Loss:   tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[3737] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[3738] Train Loss:   tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[3739] Train Loss:   tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[3740] Train Loss:   tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[3741] Train Loss:   tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[3742] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[3743] Train Loss:   tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[3744] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[3745] Train Loss:   tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[3746] Train Loss:   tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[3747] Train Loss:   tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[3748] Train Loss:   tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[3749] Train Loss:   tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[3750] Train Loss:   tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3751] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[3752] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[3753] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[3754] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[3755] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[3756] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[3757] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[3758] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[3759] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[3760] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[3761] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[3762] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[3763] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[3764] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[3765] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[3766] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[3767] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[3768] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[3769] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[3770] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[3771] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[3772] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[3773] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3774] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[3775] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[3776] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[3777] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[3778] Train Loss:    tensor(0.7702, device='cuda:0', grad_fn=<NllLossBackward>)
[3779] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[3780] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3781] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[3782] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[3783] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[3784] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[3785] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[3786] Train Loss:    tensor(0.7911, device='cuda:0', grad_fn=<NllLossBackward>)
[3787] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3788] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[3789] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3790] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[3791] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[3792] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[3793] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[3794] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[3795] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[3796] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[3797] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[3798] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3799] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[3800] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[3801] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[3802] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[3803] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[3804] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[3805] Train Loss:    tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[3806] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[3807] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[3808] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[3809] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[3810] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[3811] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[3812] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[3813] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[3814] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[3815] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[3816] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[3817] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[3818] Train Loss:    tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[3819] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3820] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[3821] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[3822] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3823] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[3824] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[3825] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[3826] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[3827] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[3828] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[3829] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[3830] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[3831] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[3832] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[3833] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[3834] Train Loss:    tensor(0.7755, device='cuda:0', grad_fn=<NllLossBackward>)
[3835] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[3836] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[3837] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[3838] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3839] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[3840] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[3841] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[3842] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[3843] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[3844] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[3845] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[3846] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3847] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[3848] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[3849] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[3850] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[3851] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[3852] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[3853] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3854] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[3855] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[3856] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[3857] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[3858] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[3859] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[3860] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[3861] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[3862] Train Loss:    tensor(0.7671, device='cuda:0', grad_fn=<NllLossBackward>)
[3863] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[3864] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[3865] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[3866] Train Loss:    tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[3867] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[3868] Train Loss:    tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[3869] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[3870] Train Loss:    tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[3871] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[3872] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[3873] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[3874] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[3875] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[3876] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[3877] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[3878] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[3879] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[3880] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[3881] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3882] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[3883] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[3884] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[3885] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[3886] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[3887] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3888] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[3889] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[3890] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[3891] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[3892] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[3893] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[3894] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3895] Train Loss:   tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3896] Train Loss:   tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[3897] Train Loss:   tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[3898] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3899] Train Loss:   tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[3900] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3901] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[3902] Train Loss:   tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[3903] Train Loss:   tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3904] Train Loss:   tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[3905] Train Loss:   tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[3906] Train Loss:   tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[3907] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[3908] Train Loss:   tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[3909] Train Loss:   tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[3910] Train Loss:   tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[3911] Train Loss:   tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[3912] Train Loss:   tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[3913] Train Loss:   tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[3914] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[3915] Train Loss:   tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[3916] Train Loss:   tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[3917] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[3918] Train Loss:   tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[3919] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[3920] Train Loss:   tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[3921] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3922] Train Loss:   tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[3923] Train Loss:   tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[3924] Train Loss:   tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[3925] Train Loss:   tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[3926] Train Loss:   tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[3927] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[3928] Train Loss:   tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[3929] Train Loss:   tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[3930] Train Loss:   tensor(0.7772, device='cuda:0', grad_fn=<NllLossBackward>)
[3931] Train Loss:   tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[3932] Train Loss:   tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[3933] Train Loss:   tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[3934] Train Loss:   tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[3935] Train Loss:   tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3936] Train Loss:   tensor(0.7744, device='cuda:0', grad_fn=<NllLossBackward>)
[3937] Train Loss:   tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[3938] Train Loss:   tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[3939] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[3940] Train Loss:   tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[3941] Train Loss:   tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[3942] Train Loss:   tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3943] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[3944] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[3945] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[3946] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[3947] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[3948] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[3949] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3950] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[3951] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[3952] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[3953] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3954] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[3955] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[3956] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[3957] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[3958] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[3959] Train Loss:    tensor(0.7703, device='cuda:0', grad_fn=<NllLossBackward>)
[3960] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[3961] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3962] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3963] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[3964] Train Loss:    tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[3965] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[3966] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[3967] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3968] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[3969] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[3970] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[3971] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[3972] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[3973] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[3974] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[3975] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[3976] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[3977] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[3978] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[3979] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[3980] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[3981] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[3982] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[3983] Train Loss:    tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[3984] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[3985] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[3986] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[3987] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[3988] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[3989] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[3990] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3991] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[3992] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[3993] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[3994] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[3995] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[3996] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[3997] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[3998] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[3999] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[4000] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[4001] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[4002] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[4003] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[4004] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[4005] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[4006] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[4007] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[4008] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[4009] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[4010] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[4011] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[4012] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[4013] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[4014] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[4015] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[4016] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[4017] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[4018] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[4019] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[4020] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[4021] Train Loss:    tensor(0.7681, device='cuda:0', grad_fn=<NllLossBackward>)
[4022] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[4023] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[4024] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[4025] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[4026] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[4027] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[4028] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[4029] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[4030] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[4031] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[4032] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[4033] Train Loss:    tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[4034] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[4035] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[4036] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[4037] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[4038] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4039] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[4040] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[4041] Train Loss:    tensor(0.7808, device='cuda:0', grad_fn=<NllLossBackward>)
[4042] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[4043] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[4044] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[4045] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[4046] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[4047] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[4048] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[4049] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[4050] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[4051] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[4052] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[4053] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[4054] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[4055] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[4056] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[4057] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[4058] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[4059] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[4060] Train Loss:    tensor(0.7613, device='cuda:0', grad_fn=<NllLossBackward>)
[4061] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[4062] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[4063] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[4064] Train Loss:    tensor(0.7775, device='cuda:0', grad_fn=<NllLossBackward>)
[4065] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[4066] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[4067] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[4068] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[4069] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[4070] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[4071] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[4072] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[4073] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[4074] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[4075] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[4076] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[4077] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[4078] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[4079] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[4080] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[4081] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[4082] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[4083] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[4084] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[4085] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[4086] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4087] Train Loss:   tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[4088] Train Loss:   tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[4089] Train Loss:   tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[4090] Train Loss:   tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[4091] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[4092] Train Loss:   tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[4093] Train Loss:   tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[4094] Train Loss:   tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[4095] Train Loss:   tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[4096] Train Loss:   tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[4097] Train Loss:   tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[4098] Train Loss:   tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[4099] Train Loss:   tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[4100] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[4101] Train Loss:   tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[4102] Train Loss:   tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[4103] Train Loss:   tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[4104] Train Loss:   tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[4105] Train Loss:   tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[4106] Train Loss:   tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[4107] Train Loss:   tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[4108] Train Loss:   tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[4109] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[4110] Train Loss:   tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[4111] Train Loss:   tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[4112] Train Loss:   tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[4113] Train Loss:   tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[4114] Train Loss:   tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[4115] Train Loss:   tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[4116] Train Loss:   tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[4117] Train Loss:   tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[4118] Train Loss:   tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[4119] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[4120] Train Loss:   tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[4121] Train Loss:   tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[4122] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[4123] Train Loss:   tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[4124] Train Loss:   tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[4125] Train Loss:   tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[4126] Train Loss:   tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[4127] Train Loss:   tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[4128] Train Loss:   tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[4129] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[4130] Train Loss:   tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[4131] Train Loss:   tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[4132] Train Loss:   tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[4133] Train Loss:   tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[4134] Train Loss:   tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4135] Train Loss:   tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[4136] Train Loss:   tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[4137] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[4138] Train Loss:   tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[4139] Train Loss:   tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[4140] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[4141] Train Loss:   tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[4142] Train Loss:   tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[4143] Train Loss:   tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[4144] Train Loss:   tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[4145] Train Loss:   tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[4146] Train Loss:   tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[4147] Train Loss:   tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[4148] Train Loss:   tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[4149] Train Loss:   tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[4150] Train Loss:   tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[4151] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[4152] Train Loss:   tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[4153] Train Loss:   tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[4154] Train Loss:   tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[4155] Train Loss:   tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[4156] Train Loss:   tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[4157] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[4158] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[4159] Train Loss:   tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[4160] Train Loss:   tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[4161] Train Loss:   tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[4162] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[4163] Train Loss:   tensor(0.7885, device='cuda:0', grad_fn=<NllLossBackward>)
[4164] Train Loss:   tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[4165] Train Loss:   tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[4166] Train Loss:   tensor(0.7836, device='cuda:0', grad_fn=<NllLossBackward>)
[4167] Train Loss:   tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[4168] Train Loss:   tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[4169] Train Loss:   tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[4170] Train Loss:   tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[4171] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[4172] Train Loss:   tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[4173] Train Loss:   tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[4174] Train Loss:   tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[4175] Train Loss:   tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[4176] Train Loss:   tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[4177] Train Loss:   tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[4178] Train Loss:   tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[4179] Train Loss:   tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[4180] Train Loss:   tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[4181] Train Loss:   tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[4182] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4183] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[4184] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[4185] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[4186] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[4187] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[4188] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[4189] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[4190] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[4191] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[4192] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[4193] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[4194] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[4195] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[4196] Train Loss:    tensor(0.7698, device='cuda:0', grad_fn=<NllLossBackward>)
[4197] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[4198] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[4199] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[4200] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[4201] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[4202] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[4203] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[4204] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[4205] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[4206] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[4207] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[4208] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[4209] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[4210] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[4211] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[4212] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[4213] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[4214] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[4215] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[4216] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[4217] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[4218] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[4219] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[4220] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[4221] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[4222] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[4223] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[4224] Train Loss:    tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[4225] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[4226] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[4227] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[4228] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[4229] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[4230] Train Loss:    tensor(0.7668, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4231] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[4232] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[4233] Train Loss:   tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[4234] Train Loss:   tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[4235] Train Loss:   tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[4236] Train Loss:   tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[4237] Train Loss:   tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[4238] Train Loss:   tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[4239] Train Loss:   tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[4240] Train Loss:   tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[4241] Train Loss:   tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[4242] Train Loss:   tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[4243] Train Loss:   tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[4244] Train Loss:   tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[4245] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[4246] Train Loss:   tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[4247] Train Loss:   tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[4248] Train Loss:   tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[4249] Train Loss:   tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[4250] Train Loss:   tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[4251] Train Loss:   tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[4252] Train Loss:   tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[4253] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[4254] Train Loss:   tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[4255] Train Loss:   tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[4256] Train Loss:   tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[4257] Train Loss:   tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[4258] Train Loss:   tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[4259] Train Loss:   tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[4260] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[4261] Train Loss:   tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[4262] Train Loss:   tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[4263] Train Loss:   tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[4264] Train Loss:   tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[4265] Train Loss:   tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[4266] Train Loss:   tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[4267] Train Loss:   tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[4268] Train Loss:   tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[4269] Train Loss:   tensor(0.7790, device='cuda:0', grad_fn=<NllLossBackward>)
[4270] Train Loss:   tensor(0.7625, device='cuda:0', grad_fn=<NllLossBackward>)
[4271] Train Loss:   tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[4272] Train Loss:   tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[4273] Train Loss:   tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[4274] Train Loss:   tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[4275] Train Loss:   tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[4276] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[4277] Train Loss:   tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[4278] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4279] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[4280] Train Loss:   tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[4281] Train Loss:   tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[4282] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[4283] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[4284] Train Loss:   tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[4285] Train Loss:   tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[4286] Train Loss:   tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[4287] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[4288] Train Loss:   tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[4289] Train Loss:   tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[4290] Train Loss:   tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[4291] Train Loss:   tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[4292] Train Loss:   tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[4293] Train Loss:   tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[4294] Train Loss:   tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[4295] Train Loss:   tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[4296] Train Loss:   tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[4297] Train Loss:   tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[4298] Train Loss:   tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[4299] Train Loss:   tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[4300] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[4301] Train Loss:   tensor(0.7676, device='cuda:0', grad_fn=<NllLossBackward>)
[4302] Train Loss:   tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[4303] Train Loss:   tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[4304] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[4305] Train Loss:   tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[4306] Train Loss:   tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[4307] Train Loss:   tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[4308] Train Loss:   tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[4309] Train Loss:   tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[4310] Train Loss:   tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[4311] Train Loss:   tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[4312] Train Loss:   tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[4313] Train Loss:   tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[4314] Train Loss:   tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[4315] Train Loss:   tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[4316] Train Loss:   tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[4317] Train Loss:   tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[4318] Train Loss:   tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[4319] Train Loss:   tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[4320] Train Loss:   tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[4321] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[4322] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[4323] Train Loss:   tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[4324] Train Loss:   tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[4325] Train Loss:   tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[4326] Train Loss:   tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4327] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[4328] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[4329] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[4330] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[4331] Train Loss:    tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[4332] Train Loss:    tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[4333] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[4334] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[4335] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[4336] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[4337] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[4338] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[4339] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[4340] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[4341] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[4342] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[4343] Train Loss:    tensor(0.7811, device='cuda:0', grad_fn=<NllLossBackward>)
[4344] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[4345] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[4346] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[4347] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[4348] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[4349] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[4350] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[4351] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[4352] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[4353] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[4354] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[4355] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[4356] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[4357] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[4358] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[4359] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[4360] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[4361] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[4362] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[4363] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[4364] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[4365] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[4366] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[4367] Train Loss:    tensor(0.7682, device='cuda:0', grad_fn=<NllLossBackward>)
[4368] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[4369] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[4370] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[4371] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[4372] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[4373] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[4374] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4375] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[4376] Train Loss:    tensor(0.7819, device='cuda:0', grad_fn=<NllLossBackward>)
[4377] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[4378] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[4379] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[4380] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[4381] Train Loss:    tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[4382] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[4383] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[4384] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[4385] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[4386] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[4387] Train Loss:    tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[4388] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[4389] Train Loss:    tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[4390] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[4391] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[4392] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[4393] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[4394] Train Loss:    tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[4395] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[4396] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[4397] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[4398] Train Loss:    tensor(0.7548, device='cuda:0', grad_fn=<NllLossBackward>)
[4399] Train Loss:    tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[4400] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[4401] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[4402] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[4403] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[4404] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[4405] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[4406] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[4407] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[4408] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[4409] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[4410] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[4411] Train Loss:    tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[4412] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[4413] Train Loss:    tensor(0.7759, device='cuda:0', grad_fn=<NllLossBackward>)
[4414] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[4415] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[4416] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[4417] Train Loss:    tensor(0.7913, device='cuda:0', grad_fn=<NllLossBackward>)
[4418] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[4419] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[4420] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[4421] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[4422] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4423] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[4424] Train Loss:    tensor(0.7757, device='cuda:0', grad_fn=<NllLossBackward>)
[4425] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[4426] Train Loss:    tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[4427] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[4428] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[4429] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[4430] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[4431] Train Loss:    tensor(0.7710, device='cuda:0', grad_fn=<NllLossBackward>)
[4432] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[4433] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[4434] Train Loss:    tensor(0.7720, device='cuda:0', grad_fn=<NllLossBackward>)
[4435] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[4436] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[4437] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[4438] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[4439] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[4440] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[4441] Train Loss:    tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[4442] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[4443] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[4444] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[4445] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[4446] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[4447] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[4448] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[4449] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[4450] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[4451] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[4452] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[4453] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[4454] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[4455] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[4456] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[4457] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[4458] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[4459] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[4460] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[4461] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[4462] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[4463] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[4464] Train Loss:    tensor(0.7723, device='cuda:0', grad_fn=<NllLossBackward>)
[4465] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[4466] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[4467] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[4468] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[4469] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[4470] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4471] Train Loss:  tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[4472] Train Loss:  tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[4473] Train Loss:  tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[4474] Train Loss:  tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[4475] Train Loss:  tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[4476] Train Loss:  tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[4477] Train Loss:  tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[4478] Train Loss:  tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[4479] Train Loss:  tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[4480] Train Loss:  tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[4481] Train Loss:  tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[4482] Train Loss:  tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[4483] Train Loss:  tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[4484] Train Loss:  tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[4485] Train Loss:  tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[4486] Train Loss:  tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[4487] Train Loss:  tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[4488] Train Loss:  tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[4489] Train Loss:  tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[4490] Train Loss:  tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[4491] Train Loss:  tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[4492] Train Loss:  tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[4493] Train Loss:  tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[4494] Train Loss:  tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[4495] Train Loss:  tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[4496] Train Loss:  tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[4497] Train Loss:  tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[4498] Train Loss:  tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[4499] Train Loss:  tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[4500] Train Loss:  tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[4501] Train Loss:  tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[4502] Train Loss:  tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[4503] Train Loss:  tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[4504] Train Loss:  tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[4505] Train Loss:  tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[4506] Train Loss:  tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[4507] Train Loss:  tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[4508] Train Loss:  tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[4509] Train Loss:  tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[4510] Train Loss:  tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[4511] Train Loss:  tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
Training Loss:  0.739794924577817 Time:  2444.287858724594 s
Training Accuaracy:  77.27410066212322 %
[0] Valid Loss:  tensor(0.7129, device='cuda:0')
[1] Valid Loss:  tensor(0.8671, device='cuda:0')
[2] Valid Loss:  tensor(0.7162, device='cuda:0')
[3] Valid Loss:  tensor(0.6360, device='cuda:0')
[4] Valid Loss:  tensor(0.7088, device='cuda:0')
```

```
[5] Valid Loss:  tensor(0.7213, device='cuda:0')
[6] Valid Loss:  tensor(0.7400, device='cuda:0')
[7] Valid Loss:  tensor(0.8393, device='cuda:0')
[8] Valid Loss:  tensor(0.8005, device='cuda:0')
[9] Valid Loss:  tensor(0.8303, device='cuda:0')
[10] Valid Loss:  tensor(0.6940, device='cuda:0')
[11] Valid Loss:  tensor(0.6670, device='cuda:0')
[12] Valid Loss:  tensor(1.1169, device='cuda:0')
[13] Valid Loss:  tensor(0.9715, device='cuda:0')
[14] Valid Loss:  tensor(0.6325, device='cuda:0')
[15] Valid Loss:  tensor(0.7555, device='cuda:0')
[16] Valid Loss:  tensor(0.6528, device='cuda:0')
[17] Valid Loss:  tensor(0.8018, device='cuda:0')
[18] Valid Loss:  tensor(0.6871, device='cuda:0')
[19] Valid Loss:  tensor(0.6255, device='cuda:0')
[20] Valid Loss:  tensor(0.6953, device='cuda:0')
[21] Valid Loss:  tensor(0.6055, device='cuda:0')
[22] Valid Loss:  tensor(0.5874, device='cuda:0')
[23] Valid Loss:  tensor(0.8447, device='cuda:0')
[24] Valid Loss:  tensor(0.7506, device='cuda:0')
[25] Valid Loss:  tensor(0.9666, device='cuda:0')
[26] Valid Loss:  tensor(0.5451, device='cuda:0')
[27] Valid Loss:  tensor(0.9019, device='cuda:0')
[28] Valid Loss:  tensor(0.6962, device='cuda:0')
[29] Valid Loss:  tensor(0.7277, device='cuda:0')
[30] Valid Loss:  tensor(0.6360, device='cuda:0')
[31] Valid Loss:  tensor(0.7416, device='cuda:0')
[32] Valid Loss:  tensor(0.8395, device='cuda:0')
[33] Valid Loss:  tensor(0.7989, device='cuda:0')
[34] Valid Loss:  tensor(0.8063, device='cuda:0')
[35] Valid Loss:  tensor(0.8835, device='cuda:0')
[36] Valid Loss:  tensor(0.6911, device='cuda:0')
[37] Valid Loss:  tensor(1.0372, device='cuda:0')
[38] Valid Loss:  tensor(0.8470, device='cuda:0')
[39] Valid Loss:  tensor(0.6872, device='cuda:0')
[40] Valid Loss:  tensor(0.7810, device='cuda:0')
[41] Valid Loss:  tensor(0.8511, device='cuda:0')
[42] Valid Loss:  tensor(0.7833, device='cuda:0')
[43] Valid Loss:  tensor(0.5581, device='cuda:0')
[44] Valid Loss:  tensor(0.6957, device='cuda:0')
[45] Valid Loss:  tensor(0.5998, device='cuda:0')
[46] Valid Loss:  tensor(0.8794, device='cuda:0')
[47] Valid Loss:  tensor(0.8010, device='cuda:0')
[48] Valid Loss:  tensor(0.6804, device='cuda:0')
[49] Valid Loss:  tensor(0.7679, device='cuda:0')
[50] Valid Loss:  tensor(0.7049, device='cuda:0')
[51] Valid Loss:  tensor(0.6746, device='cuda:0')
[52] Valid Loss:  tensor(0.5993, device='cuda:0')
```

```
[53] Valid Loss:  tensor(0.9024, device='cuda:0')
[54] Valid Loss:  tensor(0.6261, device='cuda:0')
[55] Valid Loss:  tensor(0.5767, device='cuda:0')
[56] Valid Loss:  tensor(0.8692, device='cuda:0')
[57] Valid Loss:  tensor(0.8408, device='cuda:0')
[58] Valid Loss:  tensor(0.7741, device='cuda:0')
[59] Valid Loss:  tensor(0.6777, device='cuda:0')
[60] Valid Loss:  tensor(1.0393, device='cuda:0')
[61] Valid Loss:  tensor(0.6938, device='cuda:0')
[62] Valid Loss:  tensor(0.8771, device='cuda:0')
[63] Valid Loss:  tensor(0.6765, device='cuda:0')
[64] Valid Loss:  tensor(0.7529, device='cuda:0')
[65] Valid Loss:  tensor(0.8842, device='cuda:0')
[66] Valid Loss:  tensor(0.7370, device='cuda:0')
[67] Valid Loss:  tensor(0.7633, device='cuda:0')
[68] Valid Loss:  tensor(0.7243, device='cuda:0')
[69] Valid Loss:  tensor(0.5883, device='cuda:0')
[70] Valid Loss:  tensor(0.5972, device='cuda:0')
[71] Valid Loss:  tensor(0.8395, device='cuda:0')
[72] Valid Loss:  tensor(0.7705, device='cuda:0')
[73] Valid Loss:  tensor(0.6986, device='cuda:0')
[74] Valid Loss:  tensor(0.7694, device='cuda:0')
[75] Valid Loss:  tensor(0.8644, device='cuda:0')
[76] Valid Loss:  tensor(0.8038, device='cuda:0')
[77] Valid Loss:  tensor(0.6990, device='cuda:0')
[78] Valid Loss:  tensor(0.7934, device='cuda:0')
[79] Valid Loss:  tensor(0.6361, device='cuda:0')
[80] Valid Loss:  tensor(0.7440, device='cuda:0')
[81] Valid Loss:  tensor(0.6288, device='cuda:0')
[82] Valid Loss:  tensor(0.8908, device='cuda:0')
[83] Valid Loss:  tensor(0.8355, device='cuda:0')
[84] Valid Loss:  tensor(0.6366, device='cuda:0')
[85] Valid Loss:  tensor(0.7402, device='cuda:0')
[86] Valid Loss:  tensor(0.7550, device='cuda:0')
[87] Valid Loss:  tensor(0.8366, device='cuda:0')
[88] Valid Loss:  tensor(0.5994, device='cuda:0')
[89] Valid Loss:  tensor(0.8064, device='cuda:0')
[90] Valid Loss:  tensor(0.9539, device='cuda:0')
[91] Valid Loss:  tensor(0.7340, device='cuda:0')
[92] Valid Loss:  tensor(0.8058, device='cuda:0')
[93] Valid Loss:  tensor(0.7911, device='cuda:0')
[94] Valid Loss:  tensor(0.8541, device='cuda:0')
[95] Valid Loss:  tensor(0.7031, device='cuda:0')
[96] Valid Loss:  tensor(0.8722, device='cuda:0')
[97] Valid Loss:  tensor(0.8082, device='cuda:0')
[98] Valid Loss:  tensor(0.8128, device='cuda:0')
[99] Valid Loss:  tensor(0.6373, device='cuda:0')
[100] Valid Loss:  tensor(0.7539, device='cuda:0')
```

```
[101] Valid Loss:  tensor(0.7204, device='cuda:0')
[102] Valid Loss:  tensor(0.9183, device='cuda:0')
[103] Valid Loss:  tensor(0.8988, device='cuda:0')
[104] Valid Loss:  tensor(0.6795, device='cuda:0')
[105] Valid Loss:  tensor(0.6864, device='cuda:0')
[106] Valid Loss:  tensor(0.6121, device='cuda:0')
[107] Valid Loss:  tensor(0.7837, device='cuda:0')
[108] Valid Loss:  tensor(0.8503, device='cuda:0')
[109] Valid Loss:  tensor(0.7692, device='cuda:0')
[110] Valid Loss:  tensor(0.8151, device='cuda:0')
[111] Valid Loss:  tensor(0.7713, device='cuda:0')
[112] Valid Loss:  tensor(0.8201, device='cuda:0')
[113] Valid Loss:  tensor(0.7880, device='cuda:0')
[114] Valid Loss:  tensor(0.7474, device='cuda:0')
[115] Valid Loss:  tensor(0.8294, device='cuda:0')
[116] Valid Loss:  tensor(0.8289, device='cuda:0')
[117] Valid Loss:  tensor(0.8558, device='cuda:0')
[118] Valid Loss:  tensor(0.6921, device='cuda:0')
[119] Valid Loss:  tensor(0.6233, device='cuda:0')
[120] Valid Loss:  tensor(0.9840, device='cuda:0')
[121] Valid Loss:  tensor(0.6897, device='cuda:0')
[122] Valid Loss:  tensor(0.7382, device='cuda:0')
[123] Valid Loss:  tensor(0.6818, device='cuda:0')
[124] Valid Loss:  tensor(0.9084, device='cuda:0')
[125] Valid Loss:  tensor(0.9539, device='cuda:0')
[126] Valid Loss:  tensor(0.5962, device='cuda:0')
[127] Valid Loss:  tensor(0.6333, device='cuda:0')
[128] Valid Loss:  tensor(0.8640, device='cuda:0')
[129] Valid Loss:  tensor(0.6974, device='cuda:0')
[130] Valid Loss:  tensor(0.8348, device='cuda:0')
[131] Valid Loss:  tensor(0.6950, device='cuda:0')
[132] Valid Loss:  tensor(0.7387, device='cuda:0')
[133] Valid Loss:  tensor(0.7876, device='cuda:0')
[134] Valid Loss:  tensor(0.6767, device='cuda:0')
[135] Valid Loss:  tensor(0.5935, device='cuda:0')
[136] Valid Loss:  tensor(0.6330, device='cuda:0')
[137] Valid Loss:  tensor(0.9501, device='cuda:0')
[138] Valid Loss:  tensor(0.8530, device='cuda:0')
[139] Valid Loss:  tensor(1.0611, device='cuda:0')
[140] Valid Loss:  tensor(0.8112, device='cuda:0')
[141] Valid Loss:  tensor(0.6706, device='cuda:0')
[142] Valid Loss:  tensor(0.9370, device='cuda:0')
[143] Valid Loss:  tensor(0.6627, device='cuda:0')
[144] Valid Loss:  tensor(0.7686, device='cuda:0')
[145] Valid Loss:  tensor(0.9596, device='cuda:0')
[146] Valid Loss:  tensor(0.6731, device='cuda:0')
[147] Valid Loss:  tensor(0.7956, device='cuda:0')
[148] Valid Loss:  tensor(0.7051, device='cuda:0')
```

```
[149] Valid Loss:   tensor(0.7807, device='cuda:0')
[150] Valid Loss:   tensor(0.7745, device='cuda:0')
[151] Valid Loss:   tensor(0.7704, device='cuda:0')
[152] Valid Loss:   tensor(0.6743, device='cuda:0')
[153] Valid Loss:   tensor(0.6877, device='cuda:0')
[154] Valid Loss:   tensor(0.7124, device='cuda:0')
[155] Valid Loss:   tensor(0.7701, device='cuda:0')
[156] Valid Loss:   tensor(0.8235, device='cuda:0')
[157] Valid Loss:   tensor(0.6662, device='cuda:0')
[158] Valid Loss:   tensor(0.7288, device='cuda:0')
[159] Valid Loss:   tensor(0.7209, device='cuda:0')
[160] Valid Loss:   tensor(0.6006, device='cuda:0')
[161] Valid Loss:   tensor(0.8166, device='cuda:0')
[162] Valid Loss:   tensor(0.7615, device='cuda:0')
[163] Valid Loss:   tensor(0.6631, device='cuda:0')
[164] Valid Loss:   tensor(0.9600, device='cuda:0')
[165] Valid Loss:   tensor(0.7204, device='cuda:0')
[166] Valid Loss:   tensor(0.7328, device='cuda:0')
[167] Valid Loss:   tensor(0.6288, device='cuda:0')
[168] Valid Loss:   tensor(0.7168, device='cuda:0')
[169] Valid Loss:   tensor(0.6837, device='cuda:0')
[170] Valid Loss:   tensor(0.6964, device='cuda:0')
[171] Valid Loss:   tensor(0.6030, device='cuda:0')
[172] Valid Loss:   tensor(0.8059, device='cuda:0')
[173] Valid Loss:   tensor(0.6376, device='cuda:0')
[174] Valid Loss:   tensor(0.7622, device='cuda:0')
[175] Valid Loss:   tensor(0.6783, device='cuda:0')
[176] Valid Loss:   tensor(0.5958, device='cuda:0')
[177] Valid Loss:   tensor(0.7865, device='cuda:0')
[178] Valid Loss:   tensor(0.7905, device='cuda:0')
[179] Valid Loss:   tensor(0.7954, device='cuda:0')
[180] Valid Loss:   tensor(0.8278, device='cuda:0')
[181] Valid Loss:   tensor(0.7761, device='cuda:0')
[182] Valid Loss:   tensor(0.7493, device='cuda:0')
[183] Valid Loss:   tensor(0.7595, device='cuda:0')
[184] Valid Loss:   tensor(0.9010, device='cuda:0')
[185] Valid Loss:   tensor(0.8384, device='cuda:0')
[186] Valid Loss:   tensor(0.7211, device='cuda:0')
[187] Valid Loss:   tensor(0.8468, device='cuda:0')
[188] Valid Loss:   tensor(0.9109, device='cuda:0')
[189] Valid Loss:   tensor(0.7357, device='cuda:0')
[190] Valid Loss:   tensor(0.7652, device='cuda:0')
[191] Valid Loss:   tensor(0.6264, device='cuda:0')
[192] Valid Loss:   tensor(1.0037, device='cuda:0')
[193] Valid Loss:   tensor(0.7605, device='cuda:0')
[194] Valid Loss:   tensor(0.8372, device='cuda:0')
[195] Valid Loss:   tensor(0.8654, device='cuda:0')
[196] Valid Loss:   tensor(0.6264, device='cuda:0')
```

```
[197] Valid Loss:   tensor(0.7811, device='cuda:0')
[198] Valid Loss:   tensor(0.7430, device='cuda:0')
[199] Valid Loss:   tensor(0.8121, device='cuda:0')
[200] Valid Loss:   tensor(0.7428, device='cuda:0')
[201] Valid Loss:   tensor(0.7064, device='cuda:0')
[202] Valid Loss:   tensor(0.6797, device='cuda:0')
[203] Valid Loss:   tensor(0.8986, device='cuda:0')
[204] Valid Loss:   tensor(0.8504, device='cuda:0')
[205] Valid Loss:   tensor(0.8439, device='cuda:0')
[206] Valid Loss:   tensor(0.9385, device='cuda:0')
[207] Valid Loss:   tensor(0.7757, device='cuda:0')
[208] Valid Loss:   tensor(0.6309, device='cuda:0')
[209] Valid Loss:   tensor(0.7491, device='cuda:0')
[210] Valid Loss:   tensor(1.1263, device='cuda:0')
[211] Valid Loss:   tensor(0.7762, device='cuda:0')
[212] Valid Loss:   tensor(0.6399, device='cuda:0')
[213] Valid Loss:   tensor(0.7241, device='cuda:0')
[214] Valid Loss:   tensor(0.7021, device='cuda:0')
[215] Valid Loss:   tensor(0.8184, device='cuda:0')
[216] Valid Loss:   tensor(0.7525, device='cuda:0')
[217] Valid Loss:   tensor(0.7478, device='cuda:0')
[218] Valid Loss:   tensor(0.6059, device='cuda:0')
[219] Valid Loss:   tensor(0.7053, device='cuda:0')
[220] Valid Loss:   tensor(0.6800, device='cuda:0')
[221] Valid Loss:   tensor(0.7785, device='cuda:0')
[222] Valid Loss:   tensor(0.6041, device='cuda:0')
[223] Valid Loss:   tensor(0.7047, device='cuda:0')
[224] Valid Loss:   tensor(0.8131, device='cuda:0')
[225] Valid Loss:   tensor(0.6314, device='cuda:0')
[226] Valid Loss:   tensor(0.7365, device='cuda:0')
[227] Valid Loss:   tensor(0.6257, device='cuda:0')
[228] Valid Loss:   tensor(0.8042, device='cuda:0')
[229] Valid Loss:   tensor(0.6866, device='cuda:0')
[230] Valid Loss:   tensor(0.8008, device='cuda:0')
[231] Valid Loss:   tensor(0.6891, device='cuda:0')
[232] Valid Loss:   tensor(0.6007, device='cuda:0')
[233] Valid Loss:   tensor(0.6846, device='cuda:0')
[234] Valid Loss:   tensor(0.6208, device='cuda:0')
[235] Valid Loss:   tensor(0.6163, device='cuda:0')
[236] Valid Loss:   tensor(0.7746, device='cuda:0')
[237] Valid Loss:   tensor(0.8085, device='cuda:0')
[238] Valid Loss:   tensor(0.5534, device='cuda:0')
[239] Valid Loss:   tensor(0.8153, device='cuda:0')
[240] Valid Loss:   tensor(0.7111, device='cuda:0')
[241] Valid Loss:   tensor(0.6614, device='cuda:0')
[242] Valid Loss:   tensor(0.5373, device='cuda:0')
[243] Valid Loss:   tensor(1.1817, device='cuda:0')
[244] Valid Loss:   tensor(0.8160, device='cuda:0')
```

```
[245] Valid Loss:  tensor(0.7026, device='cuda:0')
[246] Valid Loss:  tensor(0.6651, device='cuda:0')
[247] Valid Loss:  tensor(0.6955, device='cuda:0')
[248] Valid Loss:  tensor(0.9614, device='cuda:0')
[249] Valid Loss:  tensor(0.7256, device='cuda:0')
[250] Valid Loss:  tensor(0.7756, device='cuda:0')
[251] Valid Loss:  tensor(0.8314, device='cuda:0')
[252] Valid Loss:  tensor(0.6274, device='cuda:0')
[253] Valid Loss:  tensor(0.6599, device='cuda:0')
[254] Valid Loss:  tensor(0.6525, device='cuda:0')
[255] Valid Loss:  tensor(0.7953, device='cuda:0')
[256] Valid Loss:  tensor(0.8387, device='cuda:0')
[257] Valid Loss:  tensor(0.7001, device='cuda:0')
[258] Valid Loss:  tensor(0.8612, device='cuda:0')
[259] Valid Loss:  tensor(0.7382, device='cuda:0')
[260] Valid Loss:  tensor(0.9254, device='cuda:0')
[261] Valid Loss:  tensor(0.7642, device='cuda:0')
[262] Valid Loss:  tensor(0.8970, device='cuda:0')
[263] Valid Loss:  tensor(0.8647, device='cuda:0')
[264] Valid Loss:  tensor(0.8351, device='cuda:0')
[265] Valid Loss:  tensor(0.8019, device='cuda:0')
[266] Valid Loss:  tensor(0.7161, device='cuda:0')
[267] Valid Loss:  tensor(0.7272, device='cuda:0')
[268] Valid Loss:  tensor(0.7226, device='cuda:0')
[269] Valid Loss:  tensor(0.6880, device='cuda:0')
[270] Valid Loss:  tensor(0.6316, device='cuda:0')
[271] Valid Loss:  tensor(0.9798, device='cuda:0')
[272] Valid Loss:  tensor(0.7423, device='cuda:0')
[273] Valid Loss:  tensor(0.8809, device='cuda:0')
[274] Valid Loss:  tensor(0.7152, device='cuda:0')
[275] Valid Loss:  tensor(0.7520, device='cuda:0')
[276] Valid Loss:  tensor(0.6733, device='cuda:0')
[277] Valid Loss:  tensor(0.6588, device='cuda:0')
[278] Valid Loss:  tensor(0.9213, device='cuda:0')
[279] Valid Loss:  tensor(0.8140, device='cuda:0')
[280] Valid Loss:  tensor(0.8992, device='cuda:0')
[281] Valid Loss:  tensor(0.6834, device='cuda:0')
[282] Valid Loss:  tensor(0.7919, device='cuda:0')
[283] Valid Loss:  tensor(0.8813, device='cuda:0')
[284] Valid Loss:  tensor(0.6857, device='cuda:0')
[285] Valid Loss:  tensor(0.6031, device='cuda:0')
[286] Valid Loss:  tensor(0.8269, device='cuda:0')
[287] Valid Loss:  tensor(0.9359, device='cuda:0')
[288] Valid Loss:  tensor(0.5683, device='cuda:0')
[289] Valid Loss:  tensor(0.7742, device='cuda:0')
[290] Valid Loss:  tensor(0.6217, device='cuda:0')
[291] Valid Loss:  tensor(0.7270, device='cuda:0')
[292] Valid Loss:  tensor(0.8647, device='cuda:0')
```

```
[293] Valid Loss:  tensor(0.6720, device='cuda:0')
[294] Valid Loss:  tensor(0.8747, device='cuda:0')
[295] Valid Loss:  tensor(0.6722, device='cuda:0')
[296] Valid Loss:  tensor(0.7751, device='cuda:0')
[297] Valid Loss:  tensor(0.8035, device='cuda:0')
[298] Valid Loss:  tensor(0.6678, device='cuda:0')
[299] Valid Loss:  tensor(0.7781, device='cuda:0')
[300] Valid Loss:  tensor(0.9372, device='cuda:0')
[301] Valid Loss:  tensor(0.6175, device='cuda:0')
[302] Valid Loss:  tensor(0.5891, device='cuda:0')
[303] Valid Loss:  tensor(0.7011, device='cuda:0')
[304] Valid Loss:  tensor(0.6636, device='cuda:0')
[305] Valid Loss:  tensor(0.7673, device='cuda:0')
[306] Valid Loss:  tensor(0.7891, device='cuda:0')
[307] Valid Loss:  tensor(0.5971, device='cuda:0')
[308] Valid Loss:  tensor(0.6616, device='cuda:0')
[309] Valid Loss:  tensor(0.7009, device='cuda:0')
[310] Valid Loss:  tensor(0.7225, device='cuda:0')
[311] Valid Loss:  tensor(0.6413, device='cuda:0')
[312] Valid Loss:  tensor(0.6205, device='cuda:0')
[313] Valid Loss:  tensor(0.7228, device='cuda:0')
[314] Valid Loss:  tensor(0.7297, device='cuda:0')
[315] Valid Loss:  tensor(1.0074, device='cuda:0')
[316] Valid Loss:  tensor(0.8469, device='cuda:0')
[317] Valid Loss:  tensor(0.7209, device='cuda:0')
[318] Valid Loss:  tensor(0.6894, device='cuda:0')
[319] Valid Loss:  tensor(0.6852, device='cuda:0')
[320] Valid Loss:  tensor(0.6831, device='cuda:0')
[321] Valid Loss:  tensor(0.6871, device='cuda:0')
[322] Valid Loss:  tensor(0.6564, device='cuda:0')
[323] Valid Loss:  tensor(0.8395, device='cuda:0')
[324] Valid Loss:  tensor(0.7666, device='cuda:0')
[325] Valid Loss:  tensor(0.6143, device='cuda:0')
[326] Valid Loss:  tensor(0.7301, device='cuda:0')
[327] Valid Loss:  tensor(0.4777, device='cuda:0')
[328] Valid Loss:  tensor(0.6855, device='cuda:0')
[329] Valid Loss:  tensor(0.8043, device='cuda:0')
[330] Valid Loss:  tensor(0.6117, device='cuda:0')
[331] Valid Loss:  tensor(0.9333, device='cuda:0')
[332] Valid Loss:  tensor(0.6352, device='cuda:0')
[333] Valid Loss:  tensor(0.6065, device='cuda:0')
[334] Valid Loss:  tensor(0.6139, device='cuda:0')
[335] Valid Loss:  tensor(0.7143, device='cuda:0')
[336] Valid Loss:  tensor(0.8009, device='cuda:0')
[337] Valid Loss:  tensor(0.8960, device='cuda:0')
[338] Valid Loss:  tensor(0.7584, device='cuda:0')
[339] Valid Loss:  tensor(0.7248, device='cuda:0')
[340] Valid Loss:  tensor(0.7539, device='cuda:0')
```

```
[341] Valid Loss:  tensor(0.8577, device='cuda:0')
[342] Valid Loss:  tensor(0.9103, device='cuda:0')
[343] Valid Loss:  tensor(0.7665, device='cuda:0')
[344] Valid Loss:  tensor(0.6517, device='cuda:0')
[345] Valid Loss:  tensor(0.5764, device='cuda:0')
[346] Valid Loss:  tensor(0.7350, device='cuda:0')
[347] Valid Loss:  tensor(0.8374, device='cuda:0')
[348] Valid Loss:  tensor(0.7168, device='cuda:0')
[349] Valid Loss:  tensor(0.7049, device='cuda:0')
[350] Valid Loss:  tensor(0.7357, device='cuda:0')
[351] Valid Loss:  tensor(0.8310, device='cuda:0')
[352] Valid Loss:  tensor(0.6516, device='cuda:0')
[353] Valid Loss:  tensor(0.7551, device='cuda:0')
[354] Valid Loss:  tensor(0.9384, device='cuda:0')
[355] Valid Loss:  tensor(0.5408, device='cuda:0')
[356] Valid Loss:  tensor(0.6942, device='cuda:0')
[357] Valid Loss:  tensor(0.8248, device='cuda:0')
[358] Valid Loss:  tensor(0.6909, device='cuda:0')
[359] Valid Loss:  tensor(0.7949, device='cuda:0')
[360] Valid Loss:  tensor(0.7027, device='cuda:0')
[361] Valid Loss:  tensor(0.7654, device='cuda:0')
[362] Valid Loss:  tensor(0.7430, device='cuda:0')
[363] Valid Loss:  tensor(0.7981, device='cuda:0')
[364] Valid Loss:  tensor(0.9127, device='cuda:0')
[365] Valid Loss:  tensor(0.6861, device='cuda:0')
[366] Valid Loss:  tensor(0.6959, device='cuda:0')
[367] Valid Loss:  tensor(0.8012, device='cuda:0')
[368] Valid Loss:  tensor(0.8419, device='cuda:0')
[369] Valid Loss:  tensor(0.7439, device='cuda:0')
[370] Valid Loss:  tensor(0.6021, device='cuda:0')
[371] Valid Loss:  tensor(0.6314, device='cuda:0')
[372] Valid Loss:  tensor(0.8216, device='cuda:0')
[373] Valid Loss:  tensor(0.5571, device='cuda:0')
[374] Valid Loss:  tensor(0.7110, device='cuda:0')
[375] Valid Loss:  tensor(0.7945, device='cuda:0')
[376] Valid Loss:  tensor(0.6469, device='cuda:0')
[377] Valid Loss:  tensor(0.9406, device='cuda:0')
[378] Valid Loss:  tensor(0.6534, device='cuda:0')
[379] Valid Loss:  tensor(0.7988, device='cuda:0')
[380] Valid Loss:  tensor(1.0951, device='cuda:0')
Testing Loss:  0.7536907086535076
Testing Accuaracy:  76.73978565248095 %
Epoch Number:  6
[0] Train Loss:  tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1] Train Loss:  tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[2] Train Loss:  tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[3] Train Loss:  tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[4] Train Loss:  tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[5] Train Loss:  tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[6] Train Loss:  tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[7] Train Loss:  tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[8] Train Loss:  tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[9] Train Loss:  tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[10] Train Loss:  tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[11] Train Loss:  tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[12] Train Loss:  tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[13] Train Loss:  tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[14] Train Loss:  tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[15] Train Loss:  tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[16] Train Loss:  tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[17] Train Loss:  tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[18] Train Loss:  tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[19] Train Loss:  tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[20] Train Loss:  tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[21] Train Loss:  tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[22] Train Loss:  tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[23] Train Loss:  tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[24] Train Loss:  tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[25] Train Loss:  tensor(0.7591, device='cuda:0', grad_fn=<NllLossBackward>)
[26] Train Loss:  tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[27] Train Loss:  tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[28] Train Loss:  tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[29] Train Loss:  tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[30] Train Loss:  tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[31] Train Loss:  tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[32] Train Loss:  tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[33] Train Loss:  tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[34] Train Loss:  tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[35] Train Loss:  tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[36] Train Loss:  tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[37] Train Loss:  tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[38] Train Loss:  tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[39] Train Loss:  tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[40] Train Loss:  tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[41] Train Loss:  tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[42] Train Loss:  tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[43] Train Loss:  tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[44] Train Loss:  tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[45] Train Loss:  tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[46] Train Loss:  tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[47] Train Loss:  tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[48] Train Loss:  tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[49] Train Loss:  tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[50] Train Loss:  tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[51] Train Loss:  tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[52] Train Loss:  tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[53] Train Loss:  tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[54] Train Loss:  tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[55] Train Loss:  tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[56] Train Loss:  tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[57] Train Loss:  tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[58] Train Loss:  tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[59] Train Loss:  tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[60] Train Loss:  tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[61] Train Loss:  tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[62] Train Loss:  tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[63] Train Loss:  tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[64] Train Loss:  tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[65] Train Loss:  tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[66] Train Loss:  tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[67] Train Loss:  tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[68] Train Loss:  tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[69] Train Loss:  tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[70] Train Loss:  tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[71] Train Loss:  tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[72] Train Loss:  tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[73] Train Loss:  tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[74] Train Loss:  tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[75] Train Loss:  tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[76] Train Loss:  tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[77] Train Loss:  tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[78] Train Loss:  tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[79] Train Loss:  tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[80] Train Loss:  tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[81] Train Loss:  tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[82] Train Loss:  tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[83] Train Loss:  tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[84] Train Loss:  tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[85] Train Loss:  tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[86] Train Loss:  tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[87] Train Loss:  tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[88] Train Loss:  tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[89] Train Loss:  tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[90] Train Loss:  tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[91] Train Loss:  tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[92] Train Loss:  tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[93] Train Loss:  tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[94] Train Loss:  tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[95] Train Loss:  tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[96] Train Loss:  tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[97] Train Loss:  tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[98] Train Loss:  tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[99] Train Loss:  tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[100] Train Loss:  tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[101] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[102] Train Loss:   tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[103] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[104] Train Loss:   tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[105] Train Loss:   tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[106] Train Loss:   tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[107] Train Loss:   tensor(0.7569, device='cuda:0', grad_fn=<NllLossBackward>)
[108] Train Loss:   tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[109] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[110] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[111] Train Loss:   tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[112] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[113] Train Loss:   tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[114] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[115] Train Loss:   tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[116] Train Loss:   tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[117] Train Loss:   tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[118] Train Loss:   tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[119] Train Loss:   tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[120] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[121] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[122] Train Loss:   tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[123] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[124] Train Loss:   tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[125] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[126] Train Loss:   tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[127] Train Loss:   tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[128] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[129] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[130] Train Loss:   tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[131] Train Loss:   tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[132] Train Loss:   tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[133] Train Loss:   tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[134] Train Loss:   tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[135] Train Loss:   tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[136] Train Loss:   tensor(0.7554, device='cuda:0', grad_fn=<NllLossBackward>)
[137] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[138] Train Loss:   tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[139] Train Loss:   tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[140] Train Loss:   tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[141] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[142] Train Loss:   tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[143] Train Loss:   tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[144] Train Loss:   tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[145] Train Loss:   tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[146] Train Loss:   tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[147] Train Loss:   tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[148] Train Loss:   tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[149] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[150] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[151] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[152] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[153] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[154] Train Loss:    tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[155] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[156] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[157] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[158] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[159] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[160] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[161] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[162] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[163] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[164] Train Loss:    tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[165] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[166] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[167] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[168] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[169] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[170] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[171] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[172] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[173] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[174] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[175] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[176] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[177] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[178] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[179] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[180] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[181] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[182] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[183] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[184] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[185] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[186] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[187] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[188] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[189] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[190] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[191] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[192] Train Loss:    tensor(0.7590, device='cuda:0', grad_fn=<NllLossBackward>)
[193] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[194] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[195] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[196] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[197] Train Loss:   tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[198] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[199] Train Loss:   tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[200] Train Loss:   tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[201] Train Loss:   tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[202] Train Loss:   tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[203] Train Loss:   tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[204] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[205] Train Loss:   tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[206] Train Loss:   tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[207] Train Loss:   tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[208] Train Loss:   tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[209] Train Loss:   tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[210] Train Loss:   tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[211] Train Loss:   tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[212] Train Loss:   tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[213] Train Loss:   tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[214] Train Loss:   tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[215] Train Loss:   tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[216] Train Loss:   tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[217] Train Loss:   tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[218] Train Loss:   tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[219] Train Loss:   tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[220] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[221] Train Loss:   tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[222] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[223] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[224] Train Loss:   tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[225] Train Loss:   tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[226] Train Loss:   tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[227] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[228] Train Loss:   tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[229] Train Loss:   tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[230] Train Loss:   tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[231] Train Loss:   tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[232] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[233] Train Loss:   tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[234] Train Loss:   tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[235] Train Loss:   tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[236] Train Loss:   tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[237] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[238] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[239] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[240] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[241] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[242] Train Loss:   tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[243] Train Loss:   tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[244] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[245] Train Loss:   tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[246] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[247] Train Loss:   tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[248] Train Loss:   tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[249] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[250] Train Loss:   tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[251] Train Loss:   tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[252] Train Loss:   tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[253] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[254] Train Loss:   tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[255] Train Loss:   tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[256] Train Loss:   tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[257] Train Loss:   tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[258] Train Loss:   tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[259] Train Loss:   tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[260] Train Loss:   tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[261] Train Loss:   tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[262] Train Loss:   tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[263] Train Loss:   tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[264] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[265] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[266] Train Loss:   tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[267] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[268] Train Loss:   tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[269] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[270] Train Loss:   tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[271] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[272] Train Loss:   tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[273] Train Loss:   tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[274] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[275] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[276] Train Loss:   tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[277] Train Loss:   tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[278] Train Loss:   tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[279] Train Loss:   tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[280] Train Loss:   tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[281] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[282] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[283] Train Loss:   tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[284] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[285] Train Loss:   tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[286] Train Loss:   tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[287] Train Loss:   tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[288] Train Loss:   tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[289] Train Loss:   tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[290] Train Loss:   tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[291] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[292] Train Loss:   tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[293] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[294] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[295] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[296] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[297] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[298] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[299] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[300] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[301] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[302] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[303] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[304] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[305] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[306] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[307] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[308] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[309] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[310] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[311] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[312] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[313] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[314] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[315] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[316] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[317] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[318] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[319] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[320] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[321] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[322] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[323] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[324] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[325] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[326] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[327] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[328] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[329] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[330] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[331] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[332] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[333] Train Loss:    tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[334] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[335] Train Loss:    tensor(0.6715, device='cuda:0', grad_fn=<NllLossBackward>)
[336] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[337] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[338] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[339] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[340] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[341] Train Loss:   tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[342] Train Loss:   tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[343] Train Loss:   tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[344] Train Loss:   tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[345] Train Loss:   tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[346] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[347] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[348] Train Loss:   tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[349] Train Loss:   tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[350] Train Loss:   tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[351] Train Loss:   tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[352] Train Loss:   tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[353] Train Loss:   tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[354] Train Loss:   tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[355] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[356] Train Loss:   tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[357] Train Loss:   tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[358] Train Loss:   tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[359] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[360] Train Loss:   tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[361] Train Loss:   tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[362] Train Loss:   tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[363] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[364] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[365] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[366] Train Loss:   tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[367] Train Loss:   tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[368] Train Loss:   tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[369] Train Loss:   tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[370] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[371] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[372] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[373] Train Loss:   tensor(0.6704, device='cuda:0', grad_fn=<NllLossBackward>)
[374] Train Loss:   tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[375] Train Loss:   tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[376] Train Loss:   tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[377] Train Loss:   tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[378] Train Loss:   tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[379] Train Loss:   tensor(0.7617, device='cuda:0', grad_fn=<NllLossBackward>)
[380] Train Loss:   tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[381] Train Loss:   tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[382] Train Loss:   tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[383] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[384] Train Loss:   tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[385] Train Loss:   tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[386] Train Loss:   tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[387] Train Loss:   tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[388] Train Loss:   tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[389] Train Loss:   tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[390] Train Loss:   tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[391] Train Loss:   tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[392] Train Loss:   tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[393] Train Loss:   tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[394] Train Loss:   tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[395] Train Loss:   tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[396] Train Loss:   tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[397] Train Loss:   tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[398] Train Loss:   tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[399] Train Loss:   tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[400] Train Loss:   tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[401] Train Loss:   tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[402] Train Loss:   tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[403] Train Loss:   tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[404] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[405] Train Loss:   tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[406] Train Loss:   tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[407] Train Loss:   tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[408] Train Loss:   tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[409] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[410] Train Loss:   tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[411] Train Loss:   tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[412] Train Loss:   tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[413] Train Loss:   tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[414] Train Loss:   tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[415] Train Loss:   tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[416] Train Loss:   tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[417] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[418] Train Loss:   tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[419] Train Loss:   tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[420] Train Loss:   tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[421] Train Loss:   tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[422] Train Loss:   tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[423] Train Loss:   tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[424] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[425] Train Loss:   tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[426] Train Loss:   tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[427] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[428] Train Loss:   tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[429] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[430] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[431] Train Loss:   tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[432] Train Loss:   tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[433] Train Loss:   tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[434] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[435] Train Loss:   tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[436] Train Loss:   tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[437] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[438] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[439] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[440] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[441] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[442] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[443] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[444] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[445] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[446] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[447] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[448] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[449] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[450] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[451] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[452] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[453] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[454] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[455] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[456] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[457] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[458] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[459] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[460] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[461] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[462] Train Loss:    tensor(0.7653, device='cuda:0', grad_fn=<NllLossBackward>)
[463] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[464] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[465] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[466] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[467] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[468] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[469] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[470] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[471] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[472] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[473] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[474] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[475] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[476] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[477] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[478] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[479] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[480] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[481] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[482] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[483] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[484] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[485] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[486] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[487] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[488] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[489] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[490] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[491] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[492] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[493] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[494] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[495] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[496] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[497] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[498] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[499] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[500] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[501] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[502] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[503] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[504] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[505] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[506] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[507] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[508] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[509] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[510] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[511] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[512] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[513] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[514] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[515] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[516] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[517] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[518] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[519] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[520] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[521] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[522] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[523] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[524] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[525] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[526] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[527] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[528] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[529] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[530] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[531] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[532] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[533] Train Loss:   tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[534] Train Loss:   tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[535] Train Loss:   tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[536] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[537] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[538] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[539] Train Loss:   tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[540] Train Loss:   tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[541] Train Loss:   tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[542] Train Loss:   tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[543] Train Loss:   tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[544] Train Loss:   tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[545] Train Loss:   tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[546] Train Loss:   tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[547] Train Loss:   tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[548] Train Loss:   tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[549] Train Loss:   tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[550] Train Loss:   tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[551] Train Loss:   tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[552] Train Loss:   tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[553] Train Loss:   tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[554] Train Loss:   tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[555] Train Loss:   tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[556] Train Loss:   tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[557] Train Loss:   tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[558] Train Loss:   tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[559] Train Loss:   tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[560] Train Loss:   tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[561] Train Loss:   tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[562] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[563] Train Loss:   tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[564] Train Loss:   tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[565] Train Loss:   tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[566] Train Loss:   tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[567] Train Loss:   tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[568] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[569] Train Loss:   tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[570] Train Loss:   tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[571] Train Loss:   tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[572] Train Loss:   tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[573] Train Loss:   tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[574] Train Loss:   tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[575] Train Loss:   tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[576] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[577] Train Loss:   tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[578] Train Loss:   tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[579] Train Loss:   tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[580] Train Loss:   tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[581] Train Loss:   tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[582] Train Loss:   tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[583] Train Loss:   tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[584] Train Loss:   tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[585] Train Loss:   tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[586] Train Loss:   tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[587] Train Loss:   tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[588] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[589] Train Loss:   tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[590] Train Loss:   tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[591] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[592] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[593] Train Loss:   tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[594] Train Loss:   tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[595] Train Loss:   tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[596] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[597] Train Loss:   tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[598] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[599] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[600] Train Loss:   tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[601] Train Loss:   tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[602] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[603] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[604] Train Loss:   tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[605] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[606] Train Loss:   tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[607] Train Loss:   tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[608] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[609] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[610] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[611] Train Loss:   tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[612] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[613] Train Loss:   tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[614] Train Loss:   tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[615] Train Loss:   tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[616] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[617] Train Loss:   tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[618] Train Loss:   tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[619] Train Loss:   tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[620] Train Loss:   tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[621] Train Loss:   tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[622] Train Loss:   tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[623] Train Loss:   tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[624] Train Loss:   tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[625] Train Loss:   tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[626] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[627] Train Loss:   tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[628] Train Loss:   tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[629] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[630] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[631] Train Loss:   tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[632] Train Loss:   tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[633] Train Loss:   tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[634] Train Loss:   tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[635] Train Loss:   tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[636] Train Loss:   tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[637] Train Loss:   tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[638] Train Loss:   tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[639] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[640] Train Loss:   tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[641] Train Loss:   tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[642] Train Loss:   tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[643] Train Loss:   tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[644] Train Loss:   tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[645] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[646] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[647] Train Loss:   tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[648] Train Loss:   tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[649] Train Loss:   tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[650] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[651] Train Loss:   tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[652] Train Loss:   tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[653] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[654] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[655] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[656] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[657] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[658] Train Loss:   tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[659] Train Loss:   tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[660] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[661] Train Loss:   tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[662] Train Loss:   tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[663] Train Loss:   tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[664] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[665] Train Loss:   tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[666] Train Loss:   tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[667] Train Loss:   tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[668] Train Loss:   tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[669] Train Loss:   tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[670] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[671] Train Loss:   tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[672] Train Loss:   tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[673] Train Loss:   tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[674] Train Loss:   tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[675] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[676] Train Loss:   tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[677] Train Loss:  tensor(0.7621, device='cuda:0', grad_fn=<NllLossBackward>)
[678] Train Loss:  tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[679] Train Loss:  tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[680] Train Loss:  tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[681] Train Loss:  tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[682] Train Loss:  tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[683] Train Loss:  tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[684] Train Loss:  tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[685] Train Loss:  tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[686] Train Loss:  tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[687] Train Loss:  tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[688] Train Loss:  tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[689] Train Loss:  tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[690] Train Loss:  tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[691] Train Loss:  tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[692] Train Loss:  tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[693] Train Loss:  tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[694] Train Loss:  tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[695] Train Loss:  tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[696] Train Loss:  tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[697] Train Loss:  tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[698] Train Loss:  tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[699] Train Loss:  tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[700] Train Loss:  tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[701] Train Loss:  tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[702] Train Loss:  tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[703] Train Loss:  tensor(0.7659, device='cuda:0', grad_fn=<NllLossBackward>)
[704] Train Loss:  tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[705] Train Loss:  tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[706] Train Loss:  tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[707] Train Loss:  tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[708] Train Loss:  tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[709] Train Loss:  tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[710] Train Loss:  tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[711] Train Loss:  tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[712] Train Loss:  tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[713] Train Loss:  tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[714] Train Loss:  tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[715] Train Loss:  tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[716] Train Loss:  tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[717] Train Loss:  tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[718] Train Loss:  tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[719] Train Loss:  tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[720] Train Loss:  tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[721] Train Loss:  tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[722] Train Loss:  tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[723] Train Loss:  tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[724] Train Loss:  tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[725] Train Loss:   tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[726] Train Loss:   tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[727] Train Loss:   tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[728] Train Loss:   tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[729] Train Loss:   tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[730] Train Loss:   tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[731] Train Loss:   tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[732] Train Loss:   tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[733] Train Loss:   tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[734] Train Loss:   tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[735] Train Loss:   tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[736] Train Loss:   tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[737] Train Loss:   tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[738] Train Loss:   tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[739] Train Loss:   tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[740] Train Loss:   tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[741] Train Loss:   tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[742] Train Loss:   tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[743] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[744] Train Loss:   tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[745] Train Loss:   tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[746] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[747] Train Loss:   tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[748] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[749] Train Loss:   tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[750] Train Loss:   tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[751] Train Loss:   tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[752] Train Loss:   tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[753] Train Loss:   tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[754] Train Loss:   tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[755] Train Loss:   tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[756] Train Loss:   tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[757] Train Loss:   tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[758] Train Loss:   tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[759] Train Loss:   tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[760] Train Loss:   tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[761] Train Loss:   tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[762] Train Loss:   tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[763] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[764] Train Loss:   tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[765] Train Loss:   tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[766] Train Loss:   tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[767] Train Loss:   tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[768] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[769] Train Loss:   tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[770] Train Loss:   tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[771] Train Loss:   tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[772] Train Loss:   tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[773] Train Loss:   tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[774] Train Loss:   tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[775] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[776] Train Loss:   tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[777] Train Loss:   tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[778] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[779] Train Loss:   tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[780] Train Loss:   tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[781] Train Loss:   tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[782] Train Loss:   tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[783] Train Loss:   tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[784] Train Loss:   tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[785] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[786] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[787] Train Loss:   tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[788] Train Loss:   tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[789] Train Loss:   tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[790] Train Loss:   tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[791] Train Loss:   tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[792] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[793] Train Loss:   tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[794] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[795] Train Loss:   tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[796] Train Loss:   tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[797] Train Loss:   tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[798] Train Loss:   tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[799] Train Loss:   tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[800] Train Loss:   tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[801] Train Loss:   tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[802] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[803] Train Loss:   tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[804] Train Loss:   tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[805] Train Loss:   tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[806] Train Loss:   tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[807] Train Loss:   tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[808] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[809] Train Loss:   tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[810] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[811] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[812] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[813] Train Loss:   tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[814] Train Loss:   tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[815] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[816] Train Loss:   tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[817] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[818] Train Loss:   tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[819] Train Loss:   tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[820] Train Loss:   tensor(0.7750, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[821] Train Loss:   tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[822] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[823] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[824] Train Loss:   tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[825] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[826] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[827] Train Loss:   tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[828] Train Loss:   tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[829] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[830] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[831] Train Loss:   tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[832] Train Loss:   tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[833] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[834] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[835] Train Loss:   tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[836] Train Loss:   tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[837] Train Loss:   tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[838] Train Loss:   tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[839] Train Loss:   tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[840] Train Loss:   tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[841] Train Loss:   tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[842] Train Loss:   tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[843] Train Loss:   tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[844] Train Loss:   tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[845] Train Loss:   tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[846] Train Loss:   tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[847] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[848] Train Loss:   tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[849] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[850] Train Loss:   tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[851] Train Loss:   tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[852] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[853] Train Loss:   tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[854] Train Loss:   tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[855] Train Loss:   tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[856] Train Loss:   tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[857] Train Loss:   tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[858] Train Loss:   tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[859] Train Loss:   tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[860] Train Loss:   tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[861] Train Loss:   tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[862] Train Loss:   tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[863] Train Loss:   tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[864] Train Loss:   tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[865] Train Loss:   tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[866] Train Loss:   tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[867] Train Loss:   tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[868] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[869] Train Loss:   tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[870] Train Loss:   tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[871] Train Loss:   tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[872] Train Loss:   tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[873] Train Loss:   tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[874] Train Loss:   tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[875] Train Loss:   tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[876] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[877] Train Loss:   tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[878] Train Loss:   tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[879] Train Loss:   tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[880] Train Loss:   tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[881] Train Loss:   tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[882] Train Loss:   tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[883] Train Loss:   tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[884] Train Loss:   tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[885] Train Loss:   tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[886] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[887] Train Loss:   tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[888] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[889] Train Loss:   tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[890] Train Loss:   tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[891] Train Loss:   tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[892] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[893] Train Loss:   tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[894] Train Loss:   tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[895] Train Loss:   tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[896] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[897] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[898] Train Loss:   tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[899] Train Loss:   tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[900] Train Loss:   tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[901] Train Loss:   tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[902] Train Loss:   tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[903] Train Loss:   tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[904] Train Loss:   tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[905] Train Loss:   tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[906] Train Loss:   tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[907] Train Loss:   tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[908] Train Loss:   tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[909] Train Loss:   tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[910] Train Loss:   tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[911] Train Loss:   tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[912] Train Loss:   tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[913] Train Loss:   tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[914] Train Loss:   tensor(0.6643, device='cuda:0', grad_fn=<NllLossBackward>)
[915] Train Loss:   tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[916] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[917] Train Loss:   tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[918] Train Loss:   tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[919] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[920] Train Loss:   tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[921] Train Loss:   tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[922] Train Loss:   tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[923] Train Loss:   tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[924] Train Loss:   tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[925] Train Loss:   tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[926] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[927] Train Loss:   tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[928] Train Loss:   tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[929] Train Loss:   tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[930] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[931] Train Loss:   tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[932] Train Loss:   tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[933] Train Loss:   tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[934] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[935] Train Loss:   tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[936] Train Loss:   tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[937] Train Loss:   tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[938] Train Loss:   tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[939] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[940] Train Loss:   tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[941] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[942] Train Loss:   tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[943] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[944] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[945] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[946] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[947] Train Loss:   tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[948] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[949] Train Loss:   tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[950] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[951] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[952] Train Loss:   tensor(0.7692, device='cuda:0', grad_fn=<NllLossBackward>)
[953] Train Loss:   tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[954] Train Loss:   tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[955] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[956] Train Loss:   tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[957] Train Loss:   tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[958] Train Loss:   tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[959] Train Loss:   tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[960] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[961] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[962] Train Loss:   tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[963] Train Loss:   tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[964] Train Loss:   tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[965] Train Loss:  tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[966] Train Loss:  tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[967] Train Loss:  tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[968] Train Loss:  tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[969] Train Loss:  tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[970] Train Loss:  tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[971] Train Loss:  tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[972] Train Loss:  tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[973] Train Loss:  tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[974] Train Loss:  tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[975] Train Loss:  tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[976] Train Loss:  tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[977] Train Loss:  tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[978] Train Loss:  tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[979] Train Loss:  tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[980] Train Loss:  tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[981] Train Loss:  tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[982] Train Loss:  tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[983] Train Loss:  tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[984] Train Loss:  tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[985] Train Loss:  tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[986] Train Loss:  tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[987] Train Loss:  tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[988] Train Loss:  tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[989] Train Loss:  tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[990] Train Loss:  tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[991] Train Loss:  tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[992] Train Loss:  tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[993] Train Loss:  tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[994] Train Loss:  tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[995] Train Loss:  tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[996] Train Loss:  tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[997] Train Loss:  tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[998] Train Loss:  tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[999] Train Loss:  tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[1000] Train Loss:  tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[1001] Train Loss:  tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[1002] Train Loss:  tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[1003] Train Loss:  tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[1004] Train Loss:  tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1005] Train Loss:  tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[1006] Train Loss:  tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[1007] Train Loss:  tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[1008] Train Loss:  tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[1009] Train Loss:  tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[1010] Train Loss:  tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[1011] Train Loss:  tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[1012] Train Loss:  tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1013] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[1014] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[1015] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[1016] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[1017] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[1018] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[1019] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[1020] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[1021] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[1022] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[1023] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1024] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[1025] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[1026] Train Loss:    tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[1027] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1028] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1029] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[1030] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[1031] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[1032] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[1033] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[1034] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[1035] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[1036] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[1037] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[1038] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[1039] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[1040] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[1041] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[1042] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[1043] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[1044] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[1045] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1046] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1047] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[1048] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[1049] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[1050] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[1051] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1052] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[1053] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[1054] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[1055] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[1056] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[1057] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[1058] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[1059] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[1060] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1061] Train Loss:   tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[1062] Train Loss:   tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[1063] Train Loss:   tensor(0.7578, device='cuda:0', grad_fn=<NllLossBackward>)
[1064] Train Loss:   tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[1065] Train Loss:   tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[1066] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1067] Train Loss:   tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[1068] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[1069] Train Loss:   tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[1070] Train Loss:   tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[1071] Train Loss:   tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[1072] Train Loss:   tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[1073] Train Loss:   tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[1074] Train Loss:   tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[1075] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[1076] Train Loss:   tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[1077] Train Loss:   tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[1078] Train Loss:   tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[1079] Train Loss:   tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[1080] Train Loss:   tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[1081] Train Loss:   tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[1082] Train Loss:   tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[1083] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[1084] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[1085] Train Loss:   tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[1086] Train Loss:   tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[1087] Train Loss:   tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[1088] Train Loss:   tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[1089] Train Loss:   tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[1090] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[1091] Train Loss:   tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[1092] Train Loss:   tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[1093] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[1094] Train Loss:   tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[1095] Train Loss:   tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[1096] Train Loss:   tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[1097] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[1098] Train Loss:   tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[1099] Train Loss:   tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[1100] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[1101] Train Loss:   tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[1102] Train Loss:   tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[1103] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[1104] Train Loss:   tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[1105] Train Loss:   tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[1106] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[1107] Train Loss:   tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[1108] Train Loss:   tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1109] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[1110] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[1111] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[1112] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[1113] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[1114] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[1115] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[1116] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[1117] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1118] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[1119] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[1120] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[1121] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1122] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[1123] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[1124] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[1125] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[1126] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[1127] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[1128] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[1129] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[1130] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1131] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[1132] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[1133] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[1134] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[1135] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[1136] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[1137] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[1138] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1139] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[1140] Train Loss:    tensor(0.7780, device='cuda:0', grad_fn=<NllLossBackward>)
[1141] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[1142] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1143] Train Loss:    tensor(0.7749, device='cuda:0', grad_fn=<NllLossBackward>)
[1144] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[1145] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[1146] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[1147] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[1148] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[1149] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[1150] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[1151] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1152] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[1153] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[1154] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[1155] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[1156] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1157] Train Loss:   tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[1158] Train Loss:   tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[1159] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[1160] Train Loss:   tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[1161] Train Loss:   tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[1162] Train Loss:   tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[1163] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[1164] Train Loss:   tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[1165] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1166] Train Loss:   tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[1167] Train Loss:   tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1168] Train Loss:   tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[1169] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[1170] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[1171] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[1172] Train Loss:   tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1173] Train Loss:   tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[1174] Train Loss:   tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[1175] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1176] Train Loss:   tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[1177] Train Loss:   tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[1178] Train Loss:   tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[1179] Train Loss:   tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[1180] Train Loss:   tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[1181] Train Loss:   tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[1182] Train Loss:   tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[1183] Train Loss:   tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[1184] Train Loss:   tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[1185] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[1186] Train Loss:   tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1187] Train Loss:   tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[1188] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[1189] Train Loss:   tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[1190] Train Loss:   tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[1191] Train Loss:   tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1192] Train Loss:   tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[1193] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[1194] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[1195] Train Loss:   tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[1196] Train Loss:   tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[1197] Train Loss:   tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[1198] Train Loss:   tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[1199] Train Loss:   tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[1200] Train Loss:   tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[1201] Train Loss:   tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[1202] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[1203] Train Loss:   tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[1204] Train Loss:   tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1205] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[1206] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[1207] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[1208] Train Loss:    tensor(0.7576, device='cuda:0', grad_fn=<NllLossBackward>)
[1209] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[1210] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[1211] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[1212] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[1213] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[1214] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[1215] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[1216] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[1217] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1218] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[1219] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[1220] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[1221] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[1222] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[1223] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[1224] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[1225] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[1226] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[1227] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[1228] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[1229] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[1230] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[1231] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[1232] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[1233] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[1234] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[1235] Train Loss:    tensor(0.7679, device='cuda:0', grad_fn=<NllLossBackward>)
[1236] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[1237] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[1238] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[1239] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[1240] Train Loss:    tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[1241] Train Loss:    tensor(0.6722, device='cuda:0', grad_fn=<NllLossBackward>)
[1242] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[1243] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[1244] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[1245] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1246] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1247] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[1248] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[1249] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[1250] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1251] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1252] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1253] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[1254] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[1255] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[1256] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[1257] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[1258] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1259] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[1260] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[1261] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[1262] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[1263] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[1264] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[1265] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[1266] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1267] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[1268] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[1269] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[1270] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[1271] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[1272] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[1273] Train Loss:    tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[1274] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[1275] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[1276] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1277] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[1278] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[1279] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[1280] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[1281] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[1282] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[1283] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[1284] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[1285] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[1286] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[1287] Train Loss:    tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[1288] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[1289] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[1290] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[1291] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[1292] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[1293] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[1294] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[1295] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1296] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[1297] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[1298] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1299] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[1300] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1301] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1302] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[1303] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[1304] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[1305] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[1306] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[1307] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[1308] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1309] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[1310] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[1311] Train Loss:    tensor(0.7602, device='cuda:0', grad_fn=<NllLossBackward>)
[1312] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[1313] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1314] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[1315] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[1316] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[1317] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[1318] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[1319] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[1320] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[1321] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1322] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[1323] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[1324] Train Loss:    tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[1325] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[1326] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[1327] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[1328] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1329] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[1330] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[1331] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[1332] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[1333] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[1334] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[1335] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[1336] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[1337] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[1338] Train Loss:    tensor(0.6544, device='cuda:0', grad_fn=<NllLossBackward>)
[1339] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[1340] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[1341] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[1342] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[1343] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1344] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[1345] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[1346] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[1347] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[1348] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1349] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[1350] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[1351] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[1352] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[1353] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[1354] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1355] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[1356] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[1357] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[1358] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[1359] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[1360] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[1361] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[1362] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[1363] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[1364] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[1365] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[1366] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[1367] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[1368] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[1369] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[1370] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[1371] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[1372] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[1373] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[1374] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[1375] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1376] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[1377] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[1378] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[1379] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[1380] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[1381] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[1382] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[1383] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[1384] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[1385] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[1386] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[1387] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[1388] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[1389] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[1390] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[1391] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[1392] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[1393] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[1394] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[1395] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[1396] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1397] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[1398] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[1399] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[1400] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[1401] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[1402] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1403] Train Loss:    tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[1404] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[1405] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[1406] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[1407] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[1408] Train Loss:    tensor(0.7616, device='cuda:0', grad_fn=<NllLossBackward>)
[1409] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[1410] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[1411] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[1412] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1413] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[1414] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[1415] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[1416] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[1417] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[1418] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[1419] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[1420] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[1421] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[1422] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[1423] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1424] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[1425] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[1426] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[1427] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[1428] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[1429] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[1430] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[1431] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[1432] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[1433] Train Loss:    tensor(0.7544, device='cuda:0', grad_fn=<NllLossBackward>)
[1434] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[1435] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[1436] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[1437] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[1438] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[1439] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[1440] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[1441] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[1442] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[1443] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[1444] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1445] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[1446] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[1447] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[1448] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[1449] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[1450] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[1451] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[1452] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[1453] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[1454] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[1455] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[1456] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[1457] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[1458] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[1459] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[1460] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[1461] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[1462] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1463] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[1464] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[1465] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[1466] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[1467] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[1468] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[1469] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[1470] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[1471] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[1472] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[1473] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[1474] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[1475] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[1476] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1477] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[1478] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[1479] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[1480] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[1481] Train Loss:    tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[1482] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[1483] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[1484] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[1485] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[1486] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[1487] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[1488] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[1489] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[1490] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[1491] Train Loss:    tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[1492] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1493] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[1494] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[1495] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[1496] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[1497] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[1498] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[1499] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[1500] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[1501] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[1502] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[1503] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[1504] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[1505] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[1506] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[1507] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[1508] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[1509] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1510] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[1511] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1512] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[1513] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[1514] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[1515] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[1516] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[1517] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[1518] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[1519] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[1520] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1521] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[1522] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[1523] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[1524] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[1525] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1526] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[1527] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[1528] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1529] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[1530] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[1531] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[1532] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[1533] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[1534] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1535] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[1536] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[1537] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[1538] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[1539] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[1540] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1541] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[1542] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[1543] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[1544] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[1545] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[1546] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[1547] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[1548] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[1549] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[1550] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[1551] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[1552] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[1553] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[1554] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[1555] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[1556] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1557] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[1558] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[1559] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[1560] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[1561] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[1562] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[1563] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[1564] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[1565] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[1566] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[1567] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[1568] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[1569] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[1570] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[1571] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[1572] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1573] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[1574] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[1575] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[1576] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[1577] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[1578] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[1579] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1580] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[1581] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[1582] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[1583] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[1584] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[1585] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1586] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[1587] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[1588] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1589] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[1590] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[1591] Train Loss:    tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[1592] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[1593] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[1594] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1595] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[1596] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1597] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1598] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[1599] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[1600] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[1601] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[1602] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[1603] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[1604] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[1605] Train Loss:    tensor(0.7618, device='cuda:0', grad_fn=<NllLossBackward>)
[1606] Train Loss:    tensor(0.7634, device='cuda:0', grad_fn=<NllLossBackward>)
[1607] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[1608] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[1609] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[1610] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[1611] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[1612] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[1613] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[1614] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1615] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[1616] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[1617] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1618] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[1619] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1620] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[1621] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[1622] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[1623] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[1624] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[1625] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[1626] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[1627] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[1628] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[1629] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[1630] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1631] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[1632] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[1633] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[1634] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1635] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[1636] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
```

653

```
[1637] Train Loss:   tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[1638] Train Loss:   tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1639] Train Loss:   tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[1640] Train Loss:   tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[1641] Train Loss:   tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[1642] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[1643] Train Loss:   tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[1644] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[1645] Train Loss:   tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[1646] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[1647] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[1648] Train Loss:   tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[1649] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[1650] Train Loss:   tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[1651] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[1652] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1653] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[1654] Train Loss:   tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[1655] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1656] Train Loss:   tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1657] Train Loss:   tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[1658] Train Loss:   tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1659] Train Loss:   tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1660] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[1661] Train Loss:   tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[1662] Train Loss:   tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[1663] Train Loss:   tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[1664] Train Loss:   tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[1665] Train Loss:   tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[1666] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[1667] Train Loss:   tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1668] Train Loss:   tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1669] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[1670] Train Loss:   tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[1671] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[1672] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[1673] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1674] Train Loss:   tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[1675] Train Loss:   tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[1676] Train Loss:   tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[1677] Train Loss:   tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[1678] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[1679] Train Loss:   tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[1680] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[1681] Train Loss:   tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1682] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[1683] Train Loss:   tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[1684] Train Loss:   tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1685] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[1686] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[1687] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[1688] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[1689] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[1690] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[1691] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[1692] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[1693] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1694] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[1695] Train Loss:    tensor(0.7742, device='cuda:0', grad_fn=<NllLossBackward>)
[1696] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[1697] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[1698] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1699] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[1700] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[1701] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[1702] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[1703] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[1704] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[1705] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[1706] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[1707] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[1708] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[1709] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1710] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[1711] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[1712] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[1713] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[1714] Train Loss:    tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
[1715] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[1716] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[1717] Train Loss:    tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[1718] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1719] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[1720] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[1721] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[1722] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[1723] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[1724] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[1725] Train Loss:    tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[1726] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[1727] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1728] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[1729] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[1730] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[1731] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[1732] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1733] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[1734] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[1735] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[1736] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[1737] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[1738] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[1739] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[1740] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[1741] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[1742] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[1743] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1744] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[1745] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[1746] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[1747] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[1748] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[1749] Train Loss:    tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[1750] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[1751] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[1752] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[1753] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[1754] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[1755] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[1756] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[1757] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[1758] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[1759] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[1760] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[1761] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[1762] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[1763] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[1764] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[1765] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[1766] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[1767] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[1768] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[1769] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[1770] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[1771] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[1772] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1773] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[1774] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[1775] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[1776] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[1777] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[1778] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[1779] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[1780] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1781] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[1782] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[1783] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[1784] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1785] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[1786] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[1787] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[1788] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1789] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[1790] Train Loss:    tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[1791] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[1792] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[1793] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[1794] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[1795] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1796] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[1797] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[1798] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[1799] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[1800] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[1801] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1802] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1803] Train Loss:    tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[1804] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[1805] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[1806] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[1807] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[1808] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[1809] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[1810] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[1811] Train Loss:    tensor(0.6633, device='cuda:0', grad_fn=<NllLossBackward>)
[1812] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[1813] Train Loss:    tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[1814] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[1815] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[1816] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[1817] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[1818] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[1819] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1820] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1821] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[1822] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[1823] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1824] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[1825] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[1826] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[1827] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[1828] Train Loss:    tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1829] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[1830] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[1831] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[1832] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[1833] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[1834] Train Loss:    tensor(0.7594, device='cuda:0', grad_fn=<NllLossBackward>)
[1835] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[1836] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[1837] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[1838] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[1839] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[1840] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[1841] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1842] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[1843] Train Loss:    tensor(0.7665, device='cuda:0', grad_fn=<NllLossBackward>)
[1844] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[1845] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[1846] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[1847] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1848] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[1849] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[1850] Train Loss:    tensor(0.7669, device='cuda:0', grad_fn=<NllLossBackward>)
[1851] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[1852] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[1853] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[1854] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[1855] Train Loss:    tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[1856] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[1857] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[1858] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[1859] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[1860] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[1861] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[1862] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1863] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[1864] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[1865] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[1866] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[1867] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[1868] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[1869] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1870] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[1871] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[1872] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[1873] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[1874] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[1875] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[1876] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1877] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[1878] Train Loss:   tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1879] Train Loss:   tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[1880] Train Loss:   tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[1881] Train Loss:   tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[1882] Train Loss:   tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[1883] Train Loss:   tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[1884] Train Loss:   tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[1885] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[1886] Train Loss:   tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[1887] Train Loss:   tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[1888] Train Loss:   tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[1889] Train Loss:   tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[1890] Train Loss:   tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[1891] Train Loss:   tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[1892] Train Loss:   tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[1893] Train Loss:   tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[1894] Train Loss:   tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1895] Train Loss:   tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[1896] Train Loss:   tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[1897] Train Loss:   tensor(0.7633, device='cuda:0', grad_fn=<NllLossBackward>)
[1898] Train Loss:   tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[1899] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[1900] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[1901] Train Loss:   tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[1902] Train Loss:   tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1903] Train Loss:   tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[1904] Train Loss:   tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[1905] Train Loss:   tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[1906] Train Loss:   tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[1907] Train Loss:   tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[1908] Train Loss:   tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[1909] Train Loss:   tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[1910] Train Loss:   tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[1911] Train Loss:   tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[1912] Train Loss:   tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[1913] Train Loss:   tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[1914] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1915] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[1916] Train Loss:   tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1917] Train Loss:   tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1918] Train Loss:   tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[1919] Train Loss:   tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[1920] Train Loss:   tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[1921] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[1922] Train Loss:   tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[1923] Train Loss:   tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[1924] Train Loss:   tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1925] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[1926] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[1927] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1928] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[1929] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1930] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[1931] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[1932] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[1933] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1934] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[1935] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[1936] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[1937] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1938] Train Loss:    tensor(0.7751, device='cuda:0', grad_fn=<NllLossBackward>)
[1939] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[1940] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[1941] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[1942] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[1943] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[1944] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[1945] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[1946] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1947] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[1948] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[1949] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[1950] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[1951] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[1952] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[1953] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[1954] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[1955] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[1956] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1957] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[1958] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[1959] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[1960] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[1961] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[1962] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[1963] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[1964] Train Loss:    tensor(0.7667, device='cuda:0', grad_fn=<NllLossBackward>)
[1965] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[1966] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[1967] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[1968] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[1969] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[1970] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[1971] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[1972] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1973] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[1974] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[1975] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[1976] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1977] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[1978] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[1979] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[1980] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[1981] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1982] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[1983] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1984] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[1985] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[1986] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[1987] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[1988] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[1989] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[1990] Train Loss:    tensor(0.7610, device='cuda:0', grad_fn=<NllLossBackward>)
[1991] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[1992] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[1993] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[1994] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[1995] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[1996] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[1997] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[1998] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[1999] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2000] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[2001] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[2002] Train Loss:    tensor(0.7882, device='cuda:0', grad_fn=<NllLossBackward>)
[2003] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[2004] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[2005] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[2006] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2007] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[2008] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[2009] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[2010] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[2011] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[2012] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[2013] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2014] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[2015] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[2016] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[2017] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2018] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[2019] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[2020] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2021] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[2022] Train Loss:    tensor(0.7830, device='cuda:0', grad_fn=<NllLossBackward>)
[2023] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[2024] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[2025] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[2026] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[2027] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[2028] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2029] Train Loss:    tensor(0.6479, device='cuda:0', grad_fn=<NllLossBackward>)
[2030] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[2031] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2032] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2033] Train Loss:    tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[2034] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[2035] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[2036] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[2037] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[2038] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[2039] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[2040] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2041] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[2042] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[2043] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[2044] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[2045] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[2046] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[2047] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[2048] Train Loss:    tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[2049] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[2050] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[2051] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[2052] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[2053] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2054] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2055] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[2056] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[2057] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[2058] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[2059] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[2060] Train Loss:    tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[2061] Train Loss:    tensor(0.7535, device='cuda:0', grad_fn=<NllLossBackward>)
[2062] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[2063] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[2064] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[2065] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[2066] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2067] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[2068] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2069] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[2070] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[2071] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2072] Train Loss:    tensor(0.7892, device='cuda:0', grad_fn=<NllLossBackward>)
[2073] Train Loss:    tensor(0.7779, device='cuda:0', grad_fn=<NllLossBackward>)
[2074] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2075] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[2076] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[2077] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2078] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[2079] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[2080] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[2081] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[2082] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[2083] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[2084] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[2085] Train Loss:    tensor(0.7564, device='cuda:0', grad_fn=<NllLossBackward>)
[2086] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2087] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[2088] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[2089] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[2090] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2091] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[2092] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2093] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[2094] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[2095] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[2096] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[2097] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[2098] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[2099] Train Loss:    tensor(0.7519, device='cuda:0', grad_fn=<NllLossBackward>)
[2100] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[2101] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[2102] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[2103] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[2104] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[2105] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[2106] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[2107] Train Loss:    tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[2108] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2109] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[2110] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[2111] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[2112] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[2113] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[2114] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[2115] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[2116] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2117] Train Loss:   tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[2118] Train Loss:   tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[2119] Train Loss:   tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[2120] Train Loss:   tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[2121] Train Loss:   tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[2122] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[2123] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[2124] Train Loss:   tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[2125] Train Loss:   tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[2126] Train Loss:   tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2127] Train Loss:   tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[2128] Train Loss:   tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[2129] Train Loss:   tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[2130] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2131] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2132] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[2133] Train Loss:   tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[2134] Train Loss:   tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[2135] Train Loss:   tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[2136] Train Loss:   tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[2137] Train Loss:   tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[2138] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[2139] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2140] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[2141] Train Loss:   tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[2142] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[2143] Train Loss:   tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[2144] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[2145] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[2146] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[2147] Train Loss:   tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[2148] Train Loss:   tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[2149] Train Loss:   tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[2150] Train Loss:   tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[2151] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2152] Train Loss:   tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[2153] Train Loss:   tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[2154] Train Loss:   tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[2155] Train Loss:   tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[2156] Train Loss:   tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[2157] Train Loss:   tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[2158] Train Loss:   tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[2159] Train Loss:   tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[2160] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[2161] Train Loss:   tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[2162] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2163] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[2164] Train Loss:   tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2165] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2166] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[2167] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[2168] Train Loss:    tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[2169] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[2170] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[2171] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[2172] Train Loss:    tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[2173] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2174] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[2175] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[2176] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[2177] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[2178] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[2179] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[2180] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[2181] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[2182] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[2183] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[2184] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[2185] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[2186] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[2187] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[2188] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[2189] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[2190] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[2191] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[2192] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[2193] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2194] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[2195] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[2196] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[2197] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[2198] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[2199] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[2200] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[2201] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[2202] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[2203] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[2204] Train Loss:    tensor(0.7567, device='cuda:0', grad_fn=<NllLossBackward>)
[2205] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[2206] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[2207] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[2208] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[2209] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[2210] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2211] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[2212] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2213] Train Loss:   tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[2214] Train Loss:   tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[2215] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[2216] Train Loss:   tensor(0.7478, device='cuda:0', grad_fn=<NllLossBackward>)
[2217] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[2218] Train Loss:   tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[2219] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2220] Train Loss:   tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[2221] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[2222] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[2223] Train Loss:   tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[2224] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[2225] Train Loss:   tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[2226] Train Loss:   tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[2227] Train Loss:   tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[2228] Train Loss:   tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[2229] Train Loss:   tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[2230] Train Loss:   tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[2231] Train Loss:   tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2232] Train Loss:   tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[2233] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[2234] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2235] Train Loss:   tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[2236] Train Loss:   tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[2237] Train Loss:   tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[2238] Train Loss:   tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[2239] Train Loss:   tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[2240] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[2241] Train Loss:   tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[2242] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[2243] Train Loss:   tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[2244] Train Loss:   tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[2245] Train Loss:   tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[2246] Train Loss:   tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[2247] Train Loss:   tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[2248] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2249] Train Loss:   tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[2250] Train Loss:   tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[2251] Train Loss:   tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[2252] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[2253] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2254] Train Loss:   tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[2255] Train Loss:   tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[2256] Train Loss:   tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[2257] Train Loss:   tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[2258] Train Loss:   tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[2259] Train Loss:   tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[2260] Train Loss:   tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2261] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[2262] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2263] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[2264] Train Loss:    tensor(0.7536, device='cuda:0', grad_fn=<NllLossBackward>)
[2265] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[2266] Train Loss:    tensor(0.7579, device='cuda:0', grad_fn=<NllLossBackward>)
[2267] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[2268] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2269] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[2270] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2271] Train Loss:    tensor(0.7689, device='cuda:0', grad_fn=<NllLossBackward>)
[2272] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[2273] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[2274] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[2275] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2276] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2277] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[2278] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[2279] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[2280] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[2281] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2282] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[2283] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[2284] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[2285] Train Loss:    tensor(0.7527, device='cuda:0', grad_fn=<NllLossBackward>)
[2286] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[2287] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[2288] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[2289] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[2290] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[2291] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[2292] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[2293] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2294] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[2295] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[2296] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[2297] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[2298] Train Loss:    tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[2299] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[2300] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[2301] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[2302] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[2303] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[2304] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[2305] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[2306] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[2307] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[2308] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2309] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[2310] Train Loss:   tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[2311] Train Loss:   tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[2312] Train Loss:   tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[2313] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[2314] Train Loss:   tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[2315] Train Loss:   tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[2316] Train Loss:   tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[2317] Train Loss:   tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[2318] Train Loss:   tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[2319] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[2320] Train Loss:   tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[2321] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[2322] Train Loss:   tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[2323] Train Loss:   tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[2324] Train Loss:   tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[2325] Train Loss:   tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[2326] Train Loss:   tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[2327] Train Loss:   tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[2328] Train Loss:   tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[2329] Train Loss:   tensor(0.7573, device='cuda:0', grad_fn=<NllLossBackward>)
[2330] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[2331] Train Loss:   tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[2332] Train Loss:   tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[2333] Train Loss:   tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[2334] Train Loss:   tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[2335] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2336] Train Loss:   tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[2337] Train Loss:   tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[2338] Train Loss:   tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[2339] Train Loss:   tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2340] Train Loss:   tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[2341] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[2342] Train Loss:   tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[2343] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[2344] Train Loss:   tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[2345] Train Loss:   tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[2346] Train Loss:   tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[2347] Train Loss:   tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2348] Train Loss:   tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[2349] Train Loss:   tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[2350] Train Loss:   tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[2351] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[2352] Train Loss:   tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2353] Train Loss:   tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[2354] Train Loss:   tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[2355] Train Loss:   tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[2356] Train Loss:   tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2357] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[2358] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[2359] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2360] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[2361] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[2362] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[2363] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[2364] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[2365] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[2366] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[2367] Train Loss:    tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[2368] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[2369] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[2370] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[2371] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[2372] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[2373] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2374] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2375] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[2376] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2377] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[2378] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[2379] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
[2380] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[2381] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[2382] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[2383] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2384] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[2385] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[2386] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2387] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[2388] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[2389] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[2390] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[2391] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[2392] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[2393] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2394] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[2395] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[2396] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[2397] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[2398] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[2399] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[2400] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[2401] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[2402] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[2403] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[2404] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
```

669

```
[2405] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[2406] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[2407] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[2408] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[2409] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[2410] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[2411] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[2412] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[2413] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[2414] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[2415] Train Loss:    tensor(0.7691, device='cuda:0', grad_fn=<NllLossBackward>)
[2416] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2417] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[2418] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[2419] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2420] Train Loss:    tensor(0.7582, device='cuda:0', grad_fn=<NllLossBackward>)
[2421] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2422] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[2423] Train Loss:    tensor(0.6690, device='cuda:0', grad_fn=<NllLossBackward>)
[2424] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[2425] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[2426] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[2427] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2428] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[2429] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2430] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[2431] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2432] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[2433] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[2434] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[2435] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2436] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[2437] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[2438] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[2439] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2440] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[2441] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[2442] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[2443] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2444] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[2445] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[2446] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[2447] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[2448] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[2449] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2450] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[2451] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[2452] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2453] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[2454] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[2455] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[2456] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[2457] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[2458] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[2459] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[2460] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[2461] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[2462] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2463] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[2464] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[2465] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[2466] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[2467] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[2468] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[2469] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[2470] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[2471] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[2472] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[2473] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2474] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[2475] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[2476] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[2477] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[2478] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[2479] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[2480] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[2481] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[2482] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[2483] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[2484] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[2485] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[2486] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[2487] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[2488] Train Loss:    tensor(0.7619, device='cuda:0', grad_fn=<NllLossBackward>)
[2489] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[2490] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[2491] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[2492] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[2493] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[2494] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[2495] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[2496] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[2497] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[2498] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2499] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[2500] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2501] Train Loss:   tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[2502] Train Loss:   tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[2503] Train Loss:   tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[2504] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[2505] Train Loss:   tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[2506] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[2507] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[2508] Train Loss:   tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[2509] Train Loss:   tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[2510] Train Loss:   tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[2511] Train Loss:   tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[2512] Train Loss:   tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[2513] Train Loss:   tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[2514] Train Loss:   tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[2515] Train Loss:   tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[2516] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[2517] Train Loss:   tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[2518] Train Loss:   tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2519] Train Loss:   tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[2520] Train Loss:   tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[2521] Train Loss:   tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[2522] Train Loss:   tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[2523] Train Loss:   tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[2524] Train Loss:   tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[2525] Train Loss:   tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[2526] Train Loss:   tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[2527] Train Loss:   tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[2528] Train Loss:   tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[2529] Train Loss:   tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[2530] Train Loss:   tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[2531] Train Loss:   tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[2532] Train Loss:   tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[2533] Train Loss:   tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[2534] Train Loss:   tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[2535] Train Loss:   tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[2536] Train Loss:   tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2537] Train Loss:   tensor(0.7525, device='cuda:0', grad_fn=<NllLossBackward>)
[2538] Train Loss:   tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[2539] Train Loss:   tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[2540] Train Loss:   tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[2541] Train Loss:   tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[2542] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[2543] Train Loss:   tensor(0.7677, device='cuda:0', grad_fn=<NllLossBackward>)
[2544] Train Loss:   tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[2545] Train Loss:   tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2546] Train Loss:   tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[2547] Train Loss:   tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[2548] Train Loss:   tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2549] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[2550] Train Loss:   tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[2551] Train Loss:   tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[2552] Train Loss:   tensor(0.7580, device='cuda:0', grad_fn=<NllLossBackward>)
[2553] Train Loss:   tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[2554] Train Loss:   tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[2555] Train Loss:   tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2556] Train Loss:   tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[2557] Train Loss:   tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[2558] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[2559] Train Loss:   tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[2560] Train Loss:   tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[2561] Train Loss:   tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[2562] Train Loss:   tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[2563] Train Loss:   tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[2564] Train Loss:   tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[2565] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[2566] Train Loss:   tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2567] Train Loss:   tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[2568] Train Loss:   tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2569] Train Loss:   tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[2570] Train Loss:   tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[2571] Train Loss:   tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[2572] Train Loss:   tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[2573] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2574] Train Loss:   tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[2575] Train Loss:   tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[2576] Train Loss:   tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[2577] Train Loss:   tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[2578] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2579] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2580] Train Loss:   tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[2581] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[2582] Train Loss:   tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[2583] Train Loss:   tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[2584] Train Loss:   tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[2585] Train Loss:   tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[2586] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[2587] Train Loss:   tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[2588] Train Loss:   tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[2589] Train Loss:   tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[2590] Train Loss:   tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[2591] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2592] Train Loss:   tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[2593] Train Loss:   tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[2594] Train Loss:   tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[2595] Train Loss:   tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[2596] Train Loss:   tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2597] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[2598] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[2599] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[2600] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[2601] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[2602] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[2603] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[2604] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[2605] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[2606] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[2607] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[2608] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[2609] Train Loss:    tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[2610] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[2611] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[2612] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[2613] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[2614] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[2615] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[2616] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[2617] Train Loss:    tensor(0.7533, device='cuda:0', grad_fn=<NllLossBackward>)
[2618] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[2619] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[2620] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2621] Train Loss:    tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[2622] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[2623] Train Loss:    tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[2624] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[2625] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[2626] Train Loss:    tensor(0.7697, device='cuda:0', grad_fn=<NllLossBackward>)
[2627] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[2628] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[2629] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[2630] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2631] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[2632] Train Loss:    tensor(0.7637, device='cuda:0', grad_fn=<NllLossBackward>)
[2633] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[2634] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[2635] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[2636] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[2637] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[2638] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[2639] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[2640] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[2641] Train Loss:    tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[2642] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[2643] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[2644] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2645] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[2646] Train Loss:    tensor(0.7467, device='cuda:0', grad_fn=<NllLossBackward>)
[2647] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2648] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[2649] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[2650] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[2651] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[2652] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[2653] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[2654] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2655] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[2656] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[2657] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2658] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[2659] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[2660] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[2661] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[2662] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2663] Train Loss:    tensor(0.7643, device='cuda:0', grad_fn=<NllLossBackward>)
[2664] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[2665] Train Loss:    tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[2666] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[2667] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2668] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[2669] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[2670] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[2671] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[2672] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[2673] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[2674] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[2675] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2676] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[2677] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[2678] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[2679] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[2680] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[2681] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[2682] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[2683] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[2684] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2685] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[2686] Train Loss:    tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[2687] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[2688] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[2689] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[2690] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[2691] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[2692] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2693] Train Loss:   tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[2694] Train Loss:   tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[2695] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[2696] Train Loss:   tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[2697] Train Loss:   tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[2698] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[2699] Train Loss:   tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[2700] Train Loss:   tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[2701] Train Loss:   tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[2702] Train Loss:   tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[2703] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2704] Train Loss:   tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[2705] Train Loss:   tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[2706] Train Loss:   tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[2707] Train Loss:   tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[2708] Train Loss:   tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[2709] Train Loss:   tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[2710] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[2711] Train Loss:   tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[2712] Train Loss:   tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[2713] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[2714] Train Loss:   tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[2715] Train Loss:   tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[2716] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[2717] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[2718] Train Loss:   tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[2719] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[2720] Train Loss:   tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[2721] Train Loss:   tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[2722] Train Loss:   tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[2723] Train Loss:   tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[2724] Train Loss:   tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[2725] Train Loss:   tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[2726] Train Loss:   tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[2727] Train Loss:   tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[2728] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[2729] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2730] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[2731] Train Loss:   tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[2732] Train Loss:   tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[2733] Train Loss:   tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[2734] Train Loss:   tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[2735] Train Loss:   tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[2736] Train Loss:   tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[2737] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2738] Train Loss:   tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[2739] Train Loss:   tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[2740] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2741] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[2742] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[2743] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[2744] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[2745] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[2746] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[2747] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[2748] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2749] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2750] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[2751] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[2752] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[2753] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[2754] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[2755] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2756] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2757] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[2758] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[2759] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[2760] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[2761] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[2762] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[2763] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[2764] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[2765] Train Loss:    tensor(0.7635, device='cuda:0', grad_fn=<NllLossBackward>)
[2766] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[2767] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[2768] Train Loss:    tensor(0.7695, device='cuda:0', grad_fn=<NllLossBackward>)
[2769] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[2770] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[2771] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[2772] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[2773] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[2774] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[2775] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[2776] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[2777] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[2778] Train Loss:    tensor(0.7477, device='cuda:0', grad_fn=<NllLossBackward>)
[2779] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[2780] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[2781] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[2782] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[2783] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[2784] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
[2785] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[2786] Train Loss:    tensor(0.7596, device='cuda:0', grad_fn=<NllLossBackward>)
[2787] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[2788] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2789] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[2790] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[2791] Train Loss:    tensor(0.7470, device='cuda:0', grad_fn=<NllLossBackward>)
[2792] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[2793] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[2794] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[2795] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[2796] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[2797] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[2798] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[2799] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[2800] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[2801] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[2802] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[2803] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[2804] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[2805] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[2806] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[2807] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[2808] Train Loss:    tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[2809] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[2810] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[2811] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[2812] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[2813] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[2814] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[2815] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[2816] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[2817] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2818] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[2819] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[2820] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[2821] Train Loss:    tensor(0.7482, device='cuda:0', grad_fn=<NllLossBackward>)
[2822] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[2823] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[2824] Train Loss:    tensor(0.7652, device='cuda:0', grad_fn=<NllLossBackward>)
[2825] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[2826] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[2827] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[2828] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[2829] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[2830] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[2831] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[2832] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[2833] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2834] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2835] Train Loss:    tensor(0.7510, device='cuda:0', grad_fn=<NllLossBackward>)
[2836] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2837] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[2838] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[2839] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[2840] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[2841] Train Loss:    tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[2842] Train Loss:    tensor(0.7615, device='cuda:0', grad_fn=<NllLossBackward>)
[2843] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[2844] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[2845] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[2846] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[2847] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[2848] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[2849] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[2850] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[2851] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[2852] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[2853] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[2854] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[2855] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2856] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[2857] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[2858] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[2859] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[2860] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[2861] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[2862] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[2863] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[2864] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[2865] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[2866] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[2867] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2868] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[2869] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[2870] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[2871] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[2872] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[2873] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[2874] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[2875] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[2876] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[2877] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[2878] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[2879] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[2880] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[2881] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[2882] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[2883] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[2884] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2885] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[2886] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[2887] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[2888] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[2889] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[2890] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[2891] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[2892] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[2893] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[2894] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2895] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[2896] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[2897] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[2898] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[2899] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[2900] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[2901] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[2902] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[2903] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[2904] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2905] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[2906] Train Loss:    tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[2907] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[2908] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[2909] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[2910] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[2911] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[2912] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[2913] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[2914] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[2915] Train Loss:    tensor(0.7587, device='cuda:0', grad_fn=<NllLossBackward>)
[2916] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[2917] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2918] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[2919] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[2920] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[2921] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[2922] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[2923] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[2924] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[2925] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2926] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[2927] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[2928] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[2929] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[2930] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[2931] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[2932] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2933] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2934] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[2935] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[2936] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[2937] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[2938] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[2939] Train Loss:    tensor(0.7662, device='cuda:0', grad_fn=<NllLossBackward>)
[2940] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[2941] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[2942] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[2943] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[2944] Train Loss:    tensor(0.7708, device='cuda:0', grad_fn=<NllLossBackward>)
[2945] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2946] Train Loss:    tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[2947] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[2948] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2949] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[2950] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[2951] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[2952] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[2953] Train Loss:    tensor(0.6698, device='cuda:0', grad_fn=<NllLossBackward>)
[2954] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[2955] Train Loss:    tensor(0.7770, device='cuda:0', grad_fn=<NllLossBackward>)
[2956] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[2957] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[2958] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[2959] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[2960] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[2961] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[2962] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[2963] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[2964] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[2965] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[2966] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[2967] Train Loss:    tensor(0.7685, device='cuda:0', grad_fn=<NllLossBackward>)
[2968] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[2969] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[2970] Train Loss:    tensor(0.7612, device='cuda:0', grad_fn=<NllLossBackward>)
[2971] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[2972] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[2973] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[2974] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[2975] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[2976] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[2977] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[2978] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[2979] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[2980] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2981] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[2982] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[2983] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[2984] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[2985] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[2986] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[2987] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[2988] Train Loss:    tensor(0.7783, device='cuda:0', grad_fn=<NllLossBackward>)
[2989] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[2990] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[2991] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[2992] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2993] Train Loss:    tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[2994] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[2995] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[2996] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[2997] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[2998] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[2999] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[3000] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[3001] Train Loss:    tensor(0.6705, device='cuda:0', grad_fn=<NllLossBackward>)
[3002] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3003] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[3004] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[3005] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3006] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[3007] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[3008] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[3009] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[3010] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[3011] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[3012] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[3013] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[3014] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[3015] Train Loss:    tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[3016] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[3017] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[3018] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[3019] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3020] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[3021] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[3022] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[3023] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[3024] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[3025] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[3026] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[3027] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[3028] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3029] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[3030] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[3031] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[3032] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[3033] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[3034] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[3035] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3036] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[3037] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[3038] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[3039] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[3040] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[3041] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3042] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[3043] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[3044] Train Loss:    tensor(0.7606, device='cuda:0', grad_fn=<NllLossBackward>)
[3045] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3046] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[3047] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[3048] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[3049] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[3050] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[3051] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[3052] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[3053] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3054] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[3055] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[3056] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3057] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[3058] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[3059] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[3060] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[3061] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[3062] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[3063] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3064] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[3065] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[3066] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[3067] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[3068] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[3069] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[3070] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[3071] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[3072] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[3073] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[3074] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[3075] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[3076] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3077] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3078] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3079] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[3080] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[3081] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[3082] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[3083] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[3084] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[3085] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3086] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[3087] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[3088] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3089] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[3090] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[3091] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[3092] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[3093] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[3094] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[3095] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3096] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[3097] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[3098] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[3099] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[3100] Train Loss:    tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[3101] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[3102] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[3103] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[3104] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[3105] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[3106] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[3107] Train Loss:    tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[3108] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[3109] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[3110] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[3111] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[3112] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[3113] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[3114] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[3115] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[3116] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[3117] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[3118] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[3119] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[3120] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[3121] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[3122] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[3123] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[3124] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3125] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[3126] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3127] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3128] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[3129] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[3130] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[3131] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[3132] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[3133] Train Loss:    tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[3134] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[3135] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[3136] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[3137] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3138] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[3139] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[3140] Train Loss:    tensor(0.7656, device='cuda:0', grad_fn=<NllLossBackward>)
[3141] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[3142] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[3143] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[3144] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[3145] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[3146] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[3147] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[3148] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[3149] Train Loss:    tensor(0.7814, device='cuda:0', grad_fn=<NllLossBackward>)
[3150] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3151] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3152] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[3153] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[3154] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[3155] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3156] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[3157] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[3158] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[3159] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[3160] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[3161] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[3162] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[3163] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[3164] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3165] Train Loss:    tensor(0.7705, device='cuda:0', grad_fn=<NllLossBackward>)
[3166] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[3167] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3168] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[3169] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[3170] Train Loss:    tensor(0.7741, device='cuda:0', grad_fn=<NllLossBackward>)
[3171] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[3172] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3173] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[3174] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[3175] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[3176] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[3177] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[3178] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[3179] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[3180] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[3181] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[3182] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[3183] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[3184] Train Loss:    tensor(0.7628, device='cuda:0', grad_fn=<NllLossBackward>)
[3185] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3186] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[3187] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[3188] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[3189] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[3190] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[3191] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[3192] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[3193] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[3194] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[3195] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3196] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[3197] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3198] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[3199] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[3200] Train Loss:    tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[3201] Train Loss:    tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[3202] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3203] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[3204] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[3205] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[3206] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[3207] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[3208] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[3209] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[3210] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[3211] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[3212] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3213] Train Loss:    tensor(0.7834, device='cuda:0', grad_fn=<NllLossBackward>)
[3214] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[3215] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3216] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[3217] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3218] Train Loss:    tensor(0.7550, device='cuda:0', grad_fn=<NllLossBackward>)
[3219] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[3220] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3221] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3222] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[3223] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[3224] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[3225] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[3226] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[3227] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3228] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[3229] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[3230] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[3231] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[3232] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[3233] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[3234] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[3235] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[3236] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[3237] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[3238] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[3239] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[3240] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[3241] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[3242] Train Loss:    tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[3243] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[3244] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[3245] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[3246] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[3247] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[3248] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[3249] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3250] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[3251] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[3252] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3253] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[3254] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[3255] Train Loss:    tensor(0.7605, device='cuda:0', grad_fn=<NllLossBackward>)
[3256] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3257] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3258] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[3259] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[3260] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[3261] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[3262] Train Loss:    tensor(0.7592, device='cuda:0', grad_fn=<NllLossBackward>)
[3263] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[3264] Train Loss:    tensor(0.7561, device='cuda:0', grad_fn=<NllLossBackward>)
[3265] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3266] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[3267] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[3268] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3269] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[3270] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[3271] Train Loss:    tensor(0.7632, device='cuda:0', grad_fn=<NllLossBackward>)
[3272] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[3273] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3274] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[3275] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[3276] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[3277] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[3278] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[3279] Train Loss:    tensor(0.7553, device='cuda:0', grad_fn=<NllLossBackward>)
[3280] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[3281] Train Loss:    tensor(0.7773, device='cuda:0', grad_fn=<NllLossBackward>)
[3282] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[3283] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[3284] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[3285] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[3286] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[3287] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[3288] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3289] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3290] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[3291] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[3292] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[3293] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[3294] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[3295] Train Loss:    tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[3296] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[3297] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[3298] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[3299] Train Loss:    tensor(0.7624, device='cuda:0', grad_fn=<NllLossBackward>)
[3300] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[3301] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[3302] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[3303] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[3304] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[3305] Train Loss:    tensor(0.7629, device='cuda:0', grad_fn=<NllLossBackward>)
[3306] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[3307] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[3308] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[3309] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[3310] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[3311] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[3312] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3313] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3314] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[3315] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[3316] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3317] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[3318] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3319] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[3320] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[3321] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[3322] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[3323] Train Loss:    tensor(0.7816, device='cuda:0', grad_fn=<NllLossBackward>)
[3324] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[3325] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[3326] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[3327] Train Loss:    tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[3328] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[3329] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[3330] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[3331] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[3332] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3333] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[3334] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[3335] Train Loss:    tensor(0.6643, device='cuda:0', grad_fn=<NllLossBackward>)
[3336] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[3337] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[3338] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[3339] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[3340] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[3341] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[3342] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[3343] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[3344] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[3345] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[3346] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[3347] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3348] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[3349] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[3350] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[3351] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[3352] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[3353] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[3354] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3355] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[3356] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[3357] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[3358] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[3359] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[3360] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[3361] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[3362] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[3363] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[3364] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3365] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[3366] Train Loss:    tensor(0.7683, device='cuda:0', grad_fn=<NllLossBackward>)
[3367] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[3368] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3369] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3370] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[3371] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[3372] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3373] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[3374] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[3375] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[3376] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[3377] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[3378] Train Loss:    tensor(0.7513, device='cuda:0', grad_fn=<NllLossBackward>)
[3379] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[3380] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[3381] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[3382] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[3383] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3384] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[3385] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[3386] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[3387] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[3388] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[3389] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[3390] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[3391] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[3392] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[3393] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[3394] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3395] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[3396] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[3397] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[3398] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[3399] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[3400] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[3401] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[3402] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[3403] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[3404] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[3405] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[3406] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[3407] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[3408] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[3409] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[3410] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[3411] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3412] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3413] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[3414] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[3415] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[3416] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3417] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[3418] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[3419] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3420] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[3421] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[3422] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[3423] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[3424] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[3425] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[3426] Train Loss:    tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[3427] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[3428] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[3429] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[3430] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[3431] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[3432] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[3433] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[3434] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[3435] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[3436] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[3437] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[3438] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[3439] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[3440] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[3441] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[3442] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[3443] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[3444] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3445] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[3446] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[3447] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3448] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[3449] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[3450] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[3451] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[3452] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[3453] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[3454] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[3455] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3456] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[3457] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[3458] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[3459] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[3460] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3461] Train Loss:   tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[3462] Train Loss:   tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[3463] Train Loss:   tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[3464] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[3465] Train Loss:   tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[3466] Train Loss:   tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[3467] Train Loss:   tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[3468] Train Loss:   tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[3469] Train Loss:   tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[3470] Train Loss:   tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[3471] Train Loss:   tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[3472] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[3473] Train Loss:   tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[3474] Train Loss:   tensor(0.7673, device='cuda:0', grad_fn=<NllLossBackward>)
[3475] Train Loss:   tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[3476] Train Loss:   tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[3477] Train Loss:   tensor(0.7557, device='cuda:0', grad_fn=<NllLossBackward>)
[3478] Train Loss:   tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
[3479] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[3480] Train Loss:   tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[3481] Train Loss:   tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3482] Train Loss:   tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[3483] Train Loss:   tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[3484] Train Loss:   tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[3485] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3486] Train Loss:   tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[3487] Train Loss:   tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[3488] Train Loss:   tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[3489] Train Loss:   tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[3490] Train Loss:   tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[3491] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[3492] Train Loss:   tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3493] Train Loss:   tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[3494] Train Loss:   tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[3495] Train Loss:   tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[3496] Train Loss:   tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[3497] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[3498] Train Loss:   tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[3499] Train Loss:   tensor(0.7768, device='cuda:0', grad_fn=<NllLossBackward>)
[3500] Train Loss:   tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[3501] Train Loss:   tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[3502] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[3503] Train Loss:   tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[3504] Train Loss:   tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[3505] Train Loss:   tensor(0.7650, device='cuda:0', grad_fn=<NllLossBackward>)
[3506] Train Loss:   tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[3507] Train Loss:   tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[3508] Train Loss:   tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3509] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[3510] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[3511] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[3512] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[3513] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[3514] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[3515] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[3516] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[3517] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[3518] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[3519] Train Loss:    tensor(0.6641, device='cuda:0', grad_fn=<NllLossBackward>)
[3520] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[3521] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[3522] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[3523] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[3524] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[3525] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3526] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3527] Train Loss:    tensor(0.7572, device='cuda:0', grad_fn=<NllLossBackward>)
[3528] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[3529] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[3530] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[3531] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[3532] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[3533] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[3534] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3535] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[3536] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[3537] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[3538] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[3539] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[3540] Train Loss:    tensor(0.7485, device='cuda:0', grad_fn=<NllLossBackward>)
[3541] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[3542] Train Loss:    tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[3543] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[3544] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[3545] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3546] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[3547] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[3548] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[3549] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[3550] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3551] Train Loss:    tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[3552] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[3553] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[3554] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[3555] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3556] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3557] Train Loss:   tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[3558] Train Loss:   tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[3559] Train Loss:   tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[3560] Train Loss:   tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[3561] Train Loss:   tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[3562] Train Loss:   tensor(0.7728, device='cuda:0', grad_fn=<NllLossBackward>)
[3563] Train Loss:   tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3564] Train Loss:   tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[3565] Train Loss:   tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[3566] Train Loss:   tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[3567] Train Loss:   tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[3568] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3569] Train Loss:   tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3570] Train Loss:   tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[3571] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3572] Train Loss:   tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3573] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[3574] Train Loss:   tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3575] Train Loss:   tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[3576] Train Loss:   tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[3577] Train Loss:   tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[3578] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[3579] Train Loss:   tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[3580] Train Loss:   tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[3581] Train Loss:   tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3582] Train Loss:   tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[3583] Train Loss:   tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[3584] Train Loss:   tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[3585] Train Loss:   tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[3586] Train Loss:   tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[3587] Train Loss:   tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[3588] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3589] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[3590] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3591] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[3592] Train Loss:   tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[3593] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[3594] Train Loss:   tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[3595] Train Loss:   tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[3596] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[3597] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[3598] Train Loss:   tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[3599] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[3600] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[3601] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[3602] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3603] Train Loss:   tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[3604] Train Loss:   tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3605] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[3606] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[3607] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3608] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[3609] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3610] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[3611] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[3612] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[3613] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[3614] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3615] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[3616] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[3617] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[3618] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[3619] Train Loss:    tensor(0.7473, device='cuda:0', grad_fn=<NllLossBackward>)
[3620] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[3621] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[3622] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[3623] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[3624] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3625] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[3626] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3627] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3628] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[3629] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[3630] Train Loss:    tensor(0.7644, device='cuda:0', grad_fn=<NllLossBackward>)
[3631] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[3632] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[3633] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[3634] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[3635] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[3636] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3637] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[3638] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[3639] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[3640] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[3641] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[3642] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3643] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[3644] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[3645] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[3646] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3647] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3648] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[3649] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[3650] Train Loss:    tensor(0.7733, device='cuda:0', grad_fn=<NllLossBackward>)
[3651] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3652] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3653] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[3654] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[3655] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[3656] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[3657] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3658] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[3659] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[3660] Train Loss:    tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[3661] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[3662] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[3663] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3664] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[3665] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[3666] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[3667] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3668] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[3669] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[3670] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3671] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[3672] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[3673] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[3674] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[3675] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[3676] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[3677] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[3678] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[3679] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[3680] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[3681] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[3682] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[3683] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[3684] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[3685] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[3686] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[3687] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[3688] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[3689] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[3690] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[3691] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[3692] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3693] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[3694] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[3695] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[3696] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[3697] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[3698] Train Loss:    tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[3699] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3700] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3701] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[3702] Train Loss:    tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[3703] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3704] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[3705] Train Loss:    tensor(0.7452, device='cuda:0', grad_fn=<NllLossBackward>)
[3706] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[3707] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3708] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[3709] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[3710] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[3711] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[3712] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[3713] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[3714] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[3715] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3716] Train Loss:    tensor(0.7631, device='cuda:0', grad_fn=<NllLossBackward>)
[3717] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[3718] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[3719] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[3720] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[3721] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[3722] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[3723] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[3724] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[3725] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[3726] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3727] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[3728] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3729] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[3730] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[3731] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[3732] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[3733] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[3734] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[3735] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[3736] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[3737] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[3738] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[3739] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[3740] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[3741] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3742] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3743] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3744] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3745] Train Loss:    tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[3746] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[3747] Train Loss:    tensor(0.7534, device='cuda:0', grad_fn=<NllLossBackward>)
[3748] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3749] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[3750] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[3751] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[3752] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[3753] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[3754] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[3755] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[3756] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[3757] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3758] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[3759] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[3760] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[3761] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[3762] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[3763] Train Loss:    tensor(0.7649, device='cuda:0', grad_fn=<NllLossBackward>)
[3764] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[3765] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[3766] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[3767] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[3768] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[3769] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[3770] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[3771] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[3772] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[3773] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[3774] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[3775] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[3776] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3777] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[3778] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[3779] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[3780] Train Loss:    tensor(0.7423, device='cuda:0', grad_fn=<NllLossBackward>)
[3781] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[3782] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[3783] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[3784] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[3785] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3786] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[3787] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3788] Train Loss:    tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[3789] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[3790] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[3791] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[3792] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[3793] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[3794] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[3795] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[3796] Train Loss:    tensor(0.6631, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3797] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[3798] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[3799] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[3800] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[3801] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[3802] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[3803] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[3804] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[3805] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[3806] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[3807] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[3808] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[3809] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3810] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3811] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[3812] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[3813] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[3814] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[3815] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[3816] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[3817] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3818] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[3819] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[3820] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[3821] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[3822] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[3823] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[3824] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[3825] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[3826] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[3827] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[3828] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[3829] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[3830] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[3831] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[3832] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[3833] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[3834] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[3835] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[3836] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3837] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3838] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[3839] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[3840] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[3841] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[3842] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3843] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3844] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3845] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[3846] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[3847] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[3848] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3849] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[3850] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[3851] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[3852] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[3853] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[3854] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[3855] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[3856] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[3857] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[3858] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[3859] Train Loss:    tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[3860] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3861] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3862] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3863] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[3864] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3865] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3866] Train Loss:    tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[3867] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[3868] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[3869] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3870] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[3871] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[3872] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[3873] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[3874] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3875] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[3876] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[3877] Train Loss:    tensor(0.7438, device='cuda:0', grad_fn=<NllLossBackward>)
[3878] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[3879] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[3880] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[3881] Train Loss:    tensor(0.7684, device='cuda:0', grad_fn=<NllLossBackward>)
[3882] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[3883] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[3884] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3885] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[3886] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3887] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[3888] Train Loss:    tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[3889] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[3890] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[3891] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[3892] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3893] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[3894] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[3895] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[3896] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[3897] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[3898] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[3899] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[3900] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3901] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3902] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[3903] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[3904] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[3905] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3906] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[3907] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[3908] Train Loss:    tensor(0.6627, device='cuda:0', grad_fn=<NllLossBackward>)
[3909] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[3910] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[3911] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[3912] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[3913] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3914] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[3915] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3916] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[3917] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[3918] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3919] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[3920] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3921] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[3922] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[3923] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[3924] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[3925] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[3926] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[3927] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3928] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[3929] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[3930] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[3931] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[3932] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3933] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[3934] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[3935] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[3936] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[3937] Train Loss:    tensor(0.7725, device='cuda:0', grad_fn=<NllLossBackward>)
[3938] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[3939] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[3940] Train Loss:    tensor(0.6698, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3941] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[3942] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[3943] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[3944] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[3945] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[3946] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[3947] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3948] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3949] Train Loss:    tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
[3950] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[3951] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[3952] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[3953] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[3954] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[3955] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[3956] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[3957] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3958] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3959] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[3960] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[3961] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[3962] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[3963] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[3964] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[3965] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[3966] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[3967] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[3968] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[3969] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[3970] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[3971] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3972] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[3973] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[3974] Train Loss:    tensor(0.7511, device='cuda:0', grad_fn=<NllLossBackward>)
[3975] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[3976] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[3977] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[3978] Train Loss:    tensor(0.7825, device='cuda:0', grad_fn=<NllLossBackward>)
[3979] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3980] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3981] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[3982] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[3983] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3984] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[3985] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[3986] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[3987] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[3988] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3989] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3990] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[3991] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[3992] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[3993] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[3994] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[3995] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[3996] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[3997] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[3998] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[3999] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[4000] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[4001] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[4002] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[4003] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[4004] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[4005] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[4006] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[4007] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[4008] Train Loss:    tensor(0.7523, device='cuda:0', grad_fn=<NllLossBackward>)
[4009] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[4010] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[4011] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[4012] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[4013] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[4014] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[4015] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[4016] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[4017] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[4018] Train Loss:    tensor(0.7647, device='cuda:0', grad_fn=<NllLossBackward>)
[4019] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[4020] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[4021] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[4022] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[4023] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[4024] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[4025] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[4026] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[4027] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[4028] Train Loss:    tensor(0.7713, device='cuda:0', grad_fn=<NllLossBackward>)
[4029] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[4030] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[4031] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[4032] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[4033] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[4034] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[4035] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[4036] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4037] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[4038] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[4039] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[4040] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[4041] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[4042] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[4043] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[4044] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[4045] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[4046] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[4047] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[4048] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[4049] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[4050] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[4051] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[4052] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[4053] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[4054] Train Loss:    tensor(0.6591, device='cuda:0', grad_fn=<NllLossBackward>)
[4055] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[4056] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[4057] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[4058] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[4059] Train Loss:    tensor(0.7361, device='cuda:0', grad_fn=<NllLossBackward>)
[4060] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[4061] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[4062] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[4063] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[4064] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[4065] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[4066] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[4067] Train Loss:    tensor(0.7464, device='cuda:0', grad_fn=<NllLossBackward>)
[4068] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[4069] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[4070] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[4071] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[4072] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[4073] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[4074] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[4075] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[4076] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[4077] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[4078] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[4079] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[4080] Train Loss:    tensor(0.7655, device='cuda:0', grad_fn=<NllLossBackward>)
[4081] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[4082] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[4083] Train Loss:    tensor(0.7796, device='cuda:0', grad_fn=<NllLossBackward>)
[4084] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4085] Train Loss:    tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[4086] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[4087] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[4088] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[4089] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[4090] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[4091] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[4092] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[4093] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[4094] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[4095] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[4096] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[4097] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[4098] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[4099] Train Loss:    tensor(0.7577, device='cuda:0', grad_fn=<NllLossBackward>)
[4100] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[4101] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[4102] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[4103] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[4104] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[4105] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[4106] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[4107] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[4108] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[4109] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[4110] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[4111] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[4112] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[4113] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[4114] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[4115] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[4116] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[4117] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[4118] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[4119] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[4120] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[4121] Train Loss:    tensor(0.7608, device='cuda:0', grad_fn=<NllLossBackward>)
[4122] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[4123] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[4124] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[4125] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[4126] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[4127] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[4128] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[4129] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[4130] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[4131] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[4132] Train Loss:    tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4133] Train Loss:   tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[4134] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[4135] Train Loss:   tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[4136] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[4137] Train Loss:   tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[4138] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[4139] Train Loss:   tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[4140] Train Loss:   tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[4141] Train Loss:   tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[4142] Train Loss:   tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[4143] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[4144] Train Loss:   tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[4145] Train Loss:   tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[4146] Train Loss:   tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[4147] Train Loss:   tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[4148] Train Loss:   tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[4149] Train Loss:   tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[4150] Train Loss:   tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[4151] Train Loss:   tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[4152] Train Loss:   tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[4153] Train Loss:   tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[4154] Train Loss:   tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[4155] Train Loss:   tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[4156] Train Loss:   tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[4157] Train Loss:   tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[4158] Train Loss:   tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[4159] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[4160] Train Loss:   tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[4161] Train Loss:   tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[4162] Train Loss:   tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[4163] Train Loss:   tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[4164] Train Loss:   tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[4165] Train Loss:   tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[4166] Train Loss:   tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[4167] Train Loss:   tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[4168] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[4169] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[4170] Train Loss:   tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[4171] Train Loss:   tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[4172] Train Loss:   tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[4173] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[4174] Train Loss:   tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[4175] Train Loss:   tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[4176] Train Loss:   tensor(0.7593, device='cuda:0', grad_fn=<NllLossBackward>)
[4177] Train Loss:   tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[4178] Train Loss:   tensor(0.7899, device='cuda:0', grad_fn=<NllLossBackward>)
[4179] Train Loss:   tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[4180] Train Loss:   tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4181] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[4182] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[4183] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[4184] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[4185] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[4186] Train Loss:    tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[4187] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[4188] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[4189] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[4190] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[4191] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[4192] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[4193] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[4194] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[4195] Train Loss:    tensor(0.7660, device='cuda:0', grad_fn=<NllLossBackward>)
[4196] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[4197] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[4198] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[4199] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[4200] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[4201] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[4202] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[4203] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[4204] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[4205] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[4206] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[4207] Train Loss:    tensor(0.7670, device='cuda:0', grad_fn=<NllLossBackward>)
[4208] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[4209] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[4210] Train Loss:    tensor(0.7529, device='cuda:0', grad_fn=<NllLossBackward>)
[4211] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[4212] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[4213] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[4214] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[4215] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[4216] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[4217] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[4218] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[4219] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[4220] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[4221] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[4222] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[4223] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[4224] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[4225] Train Loss:    tensor(0.7607, device='cuda:0', grad_fn=<NllLossBackward>)
[4226] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
[4227] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[4228] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4229] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[4230] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[4231] Train Loss:    tensor(0.7512, device='cuda:0', grad_fn=<NllLossBackward>)
[4232] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[4233] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[4234] Train Loss:    tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[4235] Train Loss:    tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[4236] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[4237] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[4238] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[4239] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[4240] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[4241] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[4242] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[4243] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[4244] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[4245] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[4246] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[4247] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[4248] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[4249] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[4250] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[4251] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[4252] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[4253] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[4254] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[4255] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[4256] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[4257] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[4258] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[4259] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[4260] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[4261] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[4262] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[4263] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[4264] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[4265] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[4266] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[4267] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[4268] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[4269] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[4270] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[4271] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[4272] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[4273] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[4274] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[4275] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[4276] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4277] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[4278] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[4279] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[4280] Train Loss:    tensor(0.7620, device='cuda:0', grad_fn=<NllLossBackward>)
[4281] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[4282] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[4283] Train Loss:    tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[4284] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[4285] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[4286] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[4287] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[4288] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[4289] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[4290] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[4291] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[4292] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[4293] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[4294] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[4295] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[4296] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[4297] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[4298] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[4299] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[4300] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[4301] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[4302] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[4303] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[4304] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[4305] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[4306] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[4307] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[4308] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[4309] Train Loss:    tensor(0.7471, device='cuda:0', grad_fn=<NllLossBackward>)
[4310] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[4311] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[4312] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[4313] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[4314] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[4315] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[4316] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[4317] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[4318] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[4319] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[4320] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[4321] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[4322] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[4323] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[4324] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4325] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[4326] Train Loss:   tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[4327] Train Loss:   tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[4328] Train Loss:   tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[4329] Train Loss:   tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[4330] Train Loss:   tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[4331] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[4332] Train Loss:   tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[4333] Train Loss:   tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[4334] Train Loss:   tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[4335] Train Loss:   tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[4336] Train Loss:   tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[4337] Train Loss:   tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[4338] Train Loss:   tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[4339] Train Loss:   tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[4340] Train Loss:   tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[4341] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[4342] Train Loss:   tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[4343] Train Loss:   tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[4344] Train Loss:   tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[4345] Train Loss:   tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[4346] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[4347] Train Loss:   tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[4348] Train Loss:   tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[4349] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[4350] Train Loss:   tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[4351] Train Loss:   tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[4352] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[4353] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[4354] Train Loss:   tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[4355] Train Loss:   tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[4356] Train Loss:   tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[4357] Train Loss:   tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[4358] Train Loss:   tensor(0.7556, device='cuda:0', grad_fn=<NllLossBackward>)
[4359] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[4360] Train Loss:   tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[4361] Train Loss:   tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[4362] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[4363] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[4364] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[4365] Train Loss:   tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[4366] Train Loss:   tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[4367] Train Loss:   tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[4368] Train Loss:   tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[4369] Train Loss:   tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[4370] Train Loss:   tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[4371] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[4372] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4373] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[4374] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[4375] Train Loss:    tensor(0.7542, device='cuda:0', grad_fn=<NllLossBackward>)
[4376] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[4377] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[4378] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[4379] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[4380] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[4381] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[4382] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[4383] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[4384] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[4385] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[4386] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[4387] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[4388] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[4389] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[4390] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[4391] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[4392] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[4393] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[4394] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[4395] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[4396] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[4397] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[4398] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[4399] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[4400] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[4401] Train Loss:    tensor(0.7694, device='cuda:0', grad_fn=<NllLossBackward>)
[4402] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[4403] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[4404] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[4405] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[4406] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[4407] Train Loss:    tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[4408] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[4409] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[4410] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[4411] Train Loss:    tensor(0.7571, device='cuda:0', grad_fn=<NllLossBackward>)
[4412] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[4413] Train Loss:    tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[4414] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[4415] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[4416] Train Loss:    tensor(0.7539, device='cuda:0', grad_fn=<NllLossBackward>)
[4417] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[4418] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[4419] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[4420] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4421] Train Loss:   tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[4422] Train Loss:   tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[4423] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[4424] Train Loss:   tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[4425] Train Loss:   tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[4426] Train Loss:   tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[4427] Train Loss:   tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[4428] Train Loss:   tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[4429] Train Loss:   tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[4430] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[4431] Train Loss:   tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[4432] Train Loss:   tensor(0.7583, device='cuda:0', grad_fn=<NllLossBackward>)
[4433] Train Loss:   tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[4434] Train Loss:   tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[4435] Train Loss:   tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[4436] Train Loss:   tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[4437] Train Loss:   tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[4438] Train Loss:   tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[4439] Train Loss:   tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[4440] Train Loss:   tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[4441] Train Loss:   tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[4442] Train Loss:   tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[4443] Train Loss:   tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[4444] Train Loss:   tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[4445] Train Loss:   tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[4446] Train Loss:   tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[4447] Train Loss:   tensor(0.7623, device='cuda:0', grad_fn=<NllLossBackward>)
[4448] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[4449] Train Loss:   tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[4450] Train Loss:   tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[4451] Train Loss:   tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[4452] Train Loss:   tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[4453] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[4454] Train Loss:   tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[4455] Train Loss:   tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[4456] Train Loss:   tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[4457] Train Loss:   tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[4458] Train Loss:   tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[4459] Train Loss:   tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[4460] Train Loss:   tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[4461] Train Loss:   tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[4462] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[4463] Train Loss:   tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[4464] Train Loss:   tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[4465] Train Loss:   tensor(0.7603, device='cuda:0', grad_fn=<NllLossBackward>)
[4466] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[4467] Train Loss:   tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[4468] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4469] Train Loss:  tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[4470] Train Loss:  tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[4471] Train Loss:  tensor(0.7718, device='cuda:0', grad_fn=<NllLossBackward>)
[4472] Train Loss:  tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[4473] Train Loss:  tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[4474] Train Loss:  tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[4475] Train Loss:  tensor(0.7651, device='cuda:0', grad_fn=<NllLossBackward>)
[4476] Train Loss:  tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[4477] Train Loss:  tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[4478] Train Loss:  tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[4479] Train Loss:  tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[4480] Train Loss:  tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[4481] Train Loss:  tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[4482] Train Loss:  tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[4483] Train Loss:  tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[4484] Train Loss:  tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[4485] Train Loss:  tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[4486] Train Loss:  tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[4487] Train Loss:  tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[4488] Train Loss:  tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[4489] Train Loss:  tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[4490] Train Loss:  tensor(0.7626, device='cuda:0', grad_fn=<NllLossBackward>)
[4491] Train Loss:  tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[4492] Train Loss:  tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[4493] Train Loss:  tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[4494] Train Loss:  tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[4495] Train Loss:  tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[4496] Train Loss:  tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[4497] Train Loss:  tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[4498] Train Loss:  tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[4499] Train Loss:  tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[4500] Train Loss:  tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[4501] Train Loss:  tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[4502] Train Loss:  tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[4503] Train Loss:  tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[4504] Train Loss:  tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[4505] Train Loss:  tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[4506] Train Loss:  tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[4507] Train Loss:  tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[4508] Train Loss:  tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[4509] Train Loss:  tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[4510] Train Loss:  tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[4511] Train Loss:  tensor(0.7658, device='cuda:0', grad_fn=<NllLossBackward>)
Training Loss:  0.72342161624197 Time:  2435.638948202133 s
Training Accuaracy:  77.738914760292 %
[0] Valid Loss:  tensor(0.7181, device='cuda:0')
[1] Valid Loss:  tensor(0.8816, device='cuda:0')
[2] Valid Loss:  tensor(0.6870, device='cuda:0')
```

```
[3] Valid Loss:  tensor(0.6074, device='cuda:0')
[4] Valid Loss:  tensor(0.7339, device='cuda:0')
[5] Valid Loss:  tensor(0.7210, device='cuda:0')
[6] Valid Loss:  tensor(0.7107, device='cuda:0')
[7] Valid Loss:  tensor(0.8448, device='cuda:0')
[8] Valid Loss:  tensor(0.8114, device='cuda:0')
[9] Valid Loss:  tensor(0.8186, device='cuda:0')
[10] Valid Loss:  tensor(0.6907, device='cuda:0')
[11] Valid Loss:  tensor(0.6764, device='cuda:0')
[12] Valid Loss:  tensor(1.1317, device='cuda:0')
[13] Valid Loss:  tensor(0.9858, device='cuda:0')
[14] Valid Loss:  tensor(0.6367, device='cuda:0')
[15] Valid Loss:  tensor(0.7441, device='cuda:0')
[16] Valid Loss:  tensor(0.6338, device='cuda:0')
[17] Valid Loss:  tensor(0.7887, device='cuda:0')
[18] Valid Loss:  tensor(0.6712, device='cuda:0')
[19] Valid Loss:  tensor(0.6278, device='cuda:0')
[20] Valid Loss:  tensor(0.7045, device='cuda:0')
[21] Valid Loss:  tensor(0.5921, device='cuda:0')
[22] Valid Loss:  tensor(0.5877, device='cuda:0')
[23] Valid Loss:  tensor(0.8453, device='cuda:0')
[24] Valid Loss:  tensor(0.7214, device='cuda:0')
[25] Valid Loss:  tensor(0.9929, device='cuda:0')
[26] Valid Loss:  tensor(0.5656, device='cuda:0')
[27] Valid Loss:  tensor(0.9097, device='cuda:0')
[28] Valid Loss:  tensor(0.6851, device='cuda:0')
[29] Valid Loss:  tensor(0.7116, device='cuda:0')
[30] Valid Loss:  tensor(0.6182, device='cuda:0')
[31] Valid Loss:  tensor(0.7451, device='cuda:0')
[32] Valid Loss:  tensor(0.8534, device='cuda:0')
[33] Valid Loss:  tensor(0.8041, device='cuda:0')
[34] Valid Loss:  tensor(0.8017, device='cuda:0')
[35] Valid Loss:  tensor(0.8818, device='cuda:0')
[36] Valid Loss:  tensor(0.6784, device='cuda:0')
[37] Valid Loss:  tensor(1.0222, device='cuda:0')
[38] Valid Loss:  tensor(0.8519, device='cuda:0')
[39] Valid Loss:  tensor(0.6781, device='cuda:0')
[40] Valid Loss:  tensor(0.7841, device='cuda:0')
[41] Valid Loss:  tensor(0.8491, device='cuda:0')
[42] Valid Loss:  tensor(0.7813, device='cuda:0')
[43] Valid Loss:  tensor(0.5711, device='cuda:0')
[44] Valid Loss:  tensor(0.6876, device='cuda:0')
[45] Valid Loss:  tensor(0.6001, device='cuda:0')
[46] Valid Loss:  tensor(0.8954, device='cuda:0')
[47] Valid Loss:  tensor(0.8066, device='cuda:0')
[48] Valid Loss:  tensor(0.6884, device='cuda:0')
[49] Valid Loss:  tensor(0.7736, device='cuda:0')
[50] Valid Loss:  tensor(0.7019, device='cuda:0')
```

```
[51] Valid Loss:   tensor(0.6628, device='cuda:0')
[52] Valid Loss:   tensor(0.6010, device='cuda:0')
[53] Valid Loss:   tensor(0.9107, device='cuda:0')
[54] Valid Loss:   tensor(0.6188, device='cuda:0')
[55] Valid Loss:   tensor(0.5554, device='cuda:0')
[56] Valid Loss:   tensor(0.8808, device='cuda:0')
[57] Valid Loss:   tensor(0.8207, device='cuda:0')
[58] Valid Loss:   tensor(0.7631, device='cuda:0')
[59] Valid Loss:   tensor(0.6722, device='cuda:0')
[60] Valid Loss:   tensor(1.0673, device='cuda:0')
[61] Valid Loss:   tensor(0.7115, device='cuda:0')
[62] Valid Loss:   tensor(0.8907, device='cuda:0')
[63] Valid Loss:   tensor(0.6644, device='cuda:0')
[64] Valid Loss:   tensor(0.7377, device='cuda:0')
[65] Valid Loss:   tensor(0.8892, device='cuda:0')
[66] Valid Loss:   tensor(0.7307, device='cuda:0')
[67] Valid Loss:   tensor(0.7549, device='cuda:0')
[68] Valid Loss:   tensor(0.7275, device='cuda:0')
[69] Valid Loss:   tensor(0.5885, device='cuda:0')
[70] Valid Loss:   tensor(0.5996, device='cuda:0')
[71] Valid Loss:   tensor(0.8320, device='cuda:0')
[72] Valid Loss:   tensor(0.7572, device='cuda:0')
[73] Valid Loss:   tensor(0.6641, device='cuda:0')
[74] Valid Loss:   tensor(0.7863, device='cuda:0')
[75] Valid Loss:   tensor(0.8725, device='cuda:0')
[76] Valid Loss:   tensor(0.8018, device='cuda:0')
[77] Valid Loss:   tensor(0.7077, device='cuda:0')
[78] Valid Loss:   tensor(0.8089, device='cuda:0')
[79] Valid Loss:   tensor(0.6362, device='cuda:0')
[80] Valid Loss:   tensor(0.7466, device='cuda:0')
[81] Valid Loss:   tensor(0.6399, device='cuda:0')
[82] Valid Loss:   tensor(0.8635, device='cuda:0')
[83] Valid Loss:   tensor(0.8291, device='cuda:0')
[84] Valid Loss:   tensor(0.6203, device='cuda:0')
[85] Valid Loss:   tensor(0.7297, device='cuda:0')
[86] Valid Loss:   tensor(0.7515, device='cuda:0')
[87] Valid Loss:   tensor(0.8102, device='cuda:0')
[88] Valid Loss:   tensor(0.5825, device='cuda:0')
[89] Valid Loss:   tensor(0.8169, device='cuda:0')
[90] Valid Loss:   tensor(0.9081, device='cuda:0')
[91] Valid Loss:   tensor(0.7394, device='cuda:0')
[92] Valid Loss:   tensor(0.7886, device='cuda:0')
[93] Valid Loss:   tensor(0.7959, device='cuda:0')
[94] Valid Loss:   tensor(0.8456, device='cuda:0')
[95] Valid Loss:   tensor(0.7068, device='cuda:0')
[96] Valid Loss:   tensor(0.8887, device='cuda:0')
[97] Valid Loss:   tensor(0.8413, device='cuda:0')
[98] Valid Loss:   tensor(0.8407, device='cuda:0')
```

```
[99] Valid Loss:  tensor(0.6409, device='cuda:0')
[100] Valid Loss:  tensor(0.7617, device='cuda:0')
[101] Valid Loss:  tensor(0.7236, device='cuda:0')
[102] Valid Loss:  tensor(0.9101, device='cuda:0')
[103] Valid Loss:  tensor(0.8729, device='cuda:0')
[104] Valid Loss:  tensor(0.6806, device='cuda:0')
[105] Valid Loss:  tensor(0.6728, device='cuda:0')
[106] Valid Loss:  tensor(0.5892, device='cuda:0')
[107] Valid Loss:  tensor(0.7712, device='cuda:0')
[108] Valid Loss:  tensor(0.8436, device='cuda:0')
[109] Valid Loss:  tensor(0.7694, device='cuda:0')
[110] Valid Loss:  tensor(0.8067, device='cuda:0')
[111] Valid Loss:  tensor(0.7803, device='cuda:0')
[112] Valid Loss:  tensor(0.8245, device='cuda:0')
[113] Valid Loss:  tensor(0.7744, device='cuda:0')
[114] Valid Loss:  tensor(0.7181, device='cuda:0')
[115] Valid Loss:  tensor(0.7989, device='cuda:0')
[116] Valid Loss:  tensor(0.8257, device='cuda:0')
[117] Valid Loss:  tensor(0.8681, device='cuda:0')
[118] Valid Loss:  tensor(0.6782, device='cuda:0')
[119] Valid Loss:  tensor(0.6116, device='cuda:0')
[120] Valid Loss:  tensor(0.9934, device='cuda:0')
[121] Valid Loss:  tensor(0.6881, device='cuda:0')
[122] Valid Loss:  tensor(0.7462, device='cuda:0')
[123] Valid Loss:  tensor(0.6758, device='cuda:0')
[124] Valid Loss:  tensor(0.9106, device='cuda:0')
[125] Valid Loss:  tensor(0.9556, device='cuda:0')
[126] Valid Loss:  tensor(0.5700, device='cuda:0')
[127] Valid Loss:  tensor(0.6219, device='cuda:0')
[128] Valid Loss:  tensor(0.8646, device='cuda:0')
[129] Valid Loss:  tensor(0.7103, device='cuda:0')
[130] Valid Loss:  tensor(0.8277, device='cuda:0')
[131] Valid Loss:  tensor(0.6790, device='cuda:0')
[132] Valid Loss:  tensor(0.7183, device='cuda:0')
[133] Valid Loss:  tensor(0.7807, device='cuda:0')
[134] Valid Loss:  tensor(0.6626, device='cuda:0')
[135] Valid Loss:  tensor(0.5970, device='cuda:0')
[136] Valid Loss:  tensor(0.6244, device='cuda:0')
[137] Valid Loss:  tensor(0.9552, device='cuda:0')
[138] Valid Loss:  tensor(0.8511, device='cuda:0')
[139] Valid Loss:  tensor(1.0589, device='cuda:0')
[140] Valid Loss:  tensor(0.8151, device='cuda:0')
[141] Valid Loss:  tensor(0.6602, device='cuda:0')
[142] Valid Loss:  tensor(0.9150, device='cuda:0')
[143] Valid Loss:  tensor(0.6458, device='cuda:0')
[144] Valid Loss:  tensor(0.7694, device='cuda:0')
[145] Valid Loss:  tensor(0.9499, device='cuda:0')
[146] Valid Loss:  tensor(0.6719, device='cuda:0')
```

```
[147] Valid Loss:   tensor(0.8021, device='cuda:0')
[148] Valid Loss:   tensor(0.6881, device='cuda:0')
[149] Valid Loss:   tensor(0.7874, device='cuda:0')
[150] Valid Loss:   tensor(0.7570, device='cuda:0')
[151] Valid Loss:   tensor(0.7496, device='cuda:0')
[152] Valid Loss:   tensor(0.6783, device='cuda:0')
[153] Valid Loss:   tensor(0.6896, device='cuda:0')
[154] Valid Loss:   tensor(0.7010, device='cuda:0')
[155] Valid Loss:   tensor(0.7839, device='cuda:0')
[156] Valid Loss:   tensor(0.8343, device='cuda:0')
[157] Valid Loss:   tensor(0.6538, device='cuda:0')
[158] Valid Loss:   tensor(0.7359, device='cuda:0')
[159] Valid Loss:   tensor(0.7303, device='cuda:0')
[160] Valid Loss:   tensor(0.6037, device='cuda:0')
[161] Valid Loss:   tensor(0.8195, device='cuda:0')
[162] Valid Loss:   tensor(0.7740, device='cuda:0')
[163] Valid Loss:   tensor(0.6492, device='cuda:0')
[164] Valid Loss:   tensor(0.9588, device='cuda:0')
[165] Valid Loss:   tensor(0.7233, device='cuda:0')
[166] Valid Loss:   tensor(0.7234, device='cuda:0')
[167] Valid Loss:   tensor(0.6369, device='cuda:0')
[168] Valid Loss:   tensor(0.6980, device='cuda:0')
[169] Valid Loss:   tensor(0.6821, device='cuda:0')
[170] Valid Loss:   tensor(0.6986, device='cuda:0')
[171] Valid Loss:   tensor(0.5997, device='cuda:0')
[172] Valid Loss:   tensor(0.7879, device='cuda:0')
[173] Valid Loss:   tensor(0.6374, device='cuda:0')
[174] Valid Loss:   tensor(0.7628, device='cuda:0')
[175] Valid Loss:   tensor(0.6634, device='cuda:0')
[176] Valid Loss:   tensor(0.5926, device='cuda:0')
[177] Valid Loss:   tensor(0.7774, device='cuda:0')
[178] Valid Loss:   tensor(0.7890, device='cuda:0')
[179] Valid Loss:   tensor(0.7997, device='cuda:0')
[180] Valid Loss:   tensor(0.8295, device='cuda:0')
[181] Valid Loss:   tensor(0.7752, device='cuda:0')
[182] Valid Loss:   tensor(0.7366, device='cuda:0')
[183] Valid Loss:   tensor(0.7396, device='cuda:0')
[184] Valid Loss:   tensor(0.8900, device='cuda:0')
[185] Valid Loss:   tensor(0.8253, device='cuda:0')
[186] Valid Loss:   tensor(0.7169, device='cuda:0')
[187] Valid Loss:   tensor(0.8452, device='cuda:0')
[188] Valid Loss:   tensor(0.9094, device='cuda:0')
[189] Valid Loss:   tensor(0.7184, device='cuda:0')
[190] Valid Loss:   tensor(0.7449, device='cuda:0')
[191] Valid Loss:   tensor(0.6299, device='cuda:0')
[192] Valid Loss:   tensor(0.9912, device='cuda:0')
[193] Valid Loss:   tensor(0.7698, device='cuda:0')
[194] Valid Loss:   tensor(0.8449, device='cuda:0')
```

```
[195] Valid Loss:    tensor(0.8479, device='cuda:0')
[196] Valid Loss:    tensor(0.5997, device='cuda:0')
[197] Valid Loss:    tensor(0.7721, device='cuda:0')
[198] Valid Loss:    tensor(0.7584, device='cuda:0')
[199] Valid Loss:    tensor(0.7870, device='cuda:0')
[200] Valid Loss:    tensor(0.7704, device='cuda:0')
[201] Valid Loss:    tensor(0.7128, device='cuda:0')
[202] Valid Loss:    tensor(0.6748, device='cuda:0')
[203] Valid Loss:    tensor(0.8928, device='cuda:0')
[204] Valid Loss:    tensor(0.8696, device='cuda:0')
[205] Valid Loss:    tensor(0.8464, device='cuda:0')
[206] Valid Loss:    tensor(0.9438, device='cuda:0')
[207] Valid Loss:    tensor(0.7938, device='cuda:0')
[208] Valid Loss:    tensor(0.6066, device='cuda:0')
[209] Valid Loss:    tensor(0.7310, device='cuda:0')
[210] Valid Loss:    tensor(1.1162, device='cuda:0')
[211] Valid Loss:    tensor(0.7757, device='cuda:0')
[212] Valid Loss:    tensor(0.6382, device='cuda:0')
[213] Valid Loss:    tensor(0.7110, device='cuda:0')
[214] Valid Loss:    tensor(0.6894, device='cuda:0')
[215] Valid Loss:    tensor(0.8269, device='cuda:0')
[216] Valid Loss:    tensor(0.7459, device='cuda:0')
[217] Valid Loss:    tensor(0.7370, device='cuda:0')
[218] Valid Loss:    tensor(0.5956, device='cuda:0')
[219] Valid Loss:    tensor(0.7126, device='cuda:0')
[220] Valid Loss:    tensor(0.6834, device='cuda:0')
[221] Valid Loss:    tensor(0.7574, device='cuda:0')
[222] Valid Loss:    tensor(0.6014, device='cuda:0')
[223] Valid Loss:    tensor(0.7119, device='cuda:0')
[224] Valid Loss:    tensor(0.8092, device='cuda:0')
[225] Valid Loss:    tensor(0.6176, device='cuda:0')
[226] Valid Loss:    tensor(0.7763, device='cuda:0')
[227] Valid Loss:    tensor(0.6209, device='cuda:0')
[228] Valid Loss:    tensor(0.8066, device='cuda:0')
[229] Valid Loss:    tensor(0.6940, device='cuda:0')
[230] Valid Loss:    tensor(0.7908, device='cuda:0')
[231] Valid Loss:    tensor(0.6961, device='cuda:0')
[232] Valid Loss:    tensor(0.5981, device='cuda:0')
[233] Valid Loss:    tensor(0.6875, device='cuda:0')
[234] Valid Loss:    tensor(0.6054, device='cuda:0')
[235] Valid Loss:    tensor(0.6217, device='cuda:0')
[236] Valid Loss:    tensor(0.7677, device='cuda:0')
[237] Valid Loss:    tensor(0.8209, device='cuda:0')
[238] Valid Loss:    tensor(0.5449, device='cuda:0')
[239] Valid Loss:    tensor(0.8206, device='cuda:0')
[240] Valid Loss:    tensor(0.7021, device='cuda:0')
[241] Valid Loss:    tensor(0.6384, device='cuda:0')
[242] Valid Loss:    tensor(0.5359, device='cuda:0')
```

```
[243] Valid Loss:  tensor(1.1486, device='cuda:0')
[244] Valid Loss:  tensor(0.7986, device='cuda:0')
[245] Valid Loss:  tensor(0.6951, device='cuda:0')
[246] Valid Loss:  tensor(0.6446, device='cuda:0')
[247] Valid Loss:  tensor(0.7202, device='cuda:0')
[248] Valid Loss:  tensor(0.9407, device='cuda:0')
[249] Valid Loss:  tensor(0.7173, device='cuda:0')
[250] Valid Loss:  tensor(0.7794, device='cuda:0')
[251] Valid Loss:  tensor(0.8229, device='cuda:0')
[252] Valid Loss:  tensor(0.6203, device='cuda:0')
[253] Valid Loss:  tensor(0.6391, device='cuda:0')
[254] Valid Loss:  tensor(0.6386, device='cuda:0')
[255] Valid Loss:  tensor(0.7741, device='cuda:0')
[256] Valid Loss:  tensor(0.8162, device='cuda:0')
[257] Valid Loss:  tensor(0.6769, device='cuda:0')
[258] Valid Loss:  tensor(0.8777, device='cuda:0')
[259] Valid Loss:  tensor(0.7519, device='cuda:0')
[260] Valid Loss:  tensor(0.9066, device='cuda:0')
[261] Valid Loss:  tensor(0.7706, device='cuda:0')
[262] Valid Loss:  tensor(0.9037, device='cuda:0')
[263] Valid Loss:  tensor(0.8926, device='cuda:0')
[264] Valid Loss:  tensor(0.8357, device='cuda:0')
[265] Valid Loss:  tensor(0.8168, device='cuda:0')
[266] Valid Loss:  tensor(0.7083, device='cuda:0')
[267] Valid Loss:  tensor(0.7497, device='cuda:0')
[268] Valid Loss:  tensor(0.7230, device='cuda:0')
[269] Valid Loss:  tensor(0.6994, device='cuda:0')
[270] Valid Loss:  tensor(0.6413, device='cuda:0')
[271] Valid Loss:  tensor(0.9717, device='cuda:0')
[272] Valid Loss:  tensor(0.7476, device='cuda:0')
[273] Valid Loss:  tensor(0.8704, device='cuda:0')
[274] Valid Loss:  tensor(0.6940, device='cuda:0')
[275] Valid Loss:  tensor(0.7591, device='cuda:0')
[276] Valid Loss:  tensor(0.6820, device='cuda:0')
[277] Valid Loss:  tensor(0.6457, device='cuda:0')
[278] Valid Loss:  tensor(0.9211, device='cuda:0')
[279] Valid Loss:  tensor(0.8062, device='cuda:0')
[280] Valid Loss:  tensor(0.9249, device='cuda:0')
[281] Valid Loss:  tensor(0.6691, device='cuda:0')
[282] Valid Loss:  tensor(0.7750, device='cuda:0')
[283] Valid Loss:  tensor(0.8799, device='cuda:0')
[284] Valid Loss:  tensor(0.6941, device='cuda:0')
[285] Valid Loss:  tensor(0.5895, device='cuda:0')
[286] Valid Loss:  tensor(0.8173, device='cuda:0')
[287] Valid Loss:  tensor(0.9431, device='cuda:0')
[288] Valid Loss:  tensor(0.5608, device='cuda:0')
[289] Valid Loss:  tensor(0.7835, device='cuda:0')
[290] Valid Loss:  tensor(0.5935, device='cuda:0')
```

```
[291] Valid Loss:  tensor(0.7343, device='cuda:0')
[292] Valid Loss:  tensor(0.8539, device='cuda:0')
[293] Valid Loss:  tensor(0.6715, device='cuda:0')
[294] Valid Loss:  tensor(0.8742, device='cuda:0')
[295] Valid Loss:  tensor(0.6785, device='cuda:0')
[296] Valid Loss:  tensor(0.7678, device='cuda:0')
[297] Valid Loss:  tensor(0.8136, device='cuda:0')
[298] Valid Loss:  tensor(0.6718, device='cuda:0')
[299] Valid Loss:  tensor(0.7799, device='cuda:0')
[300] Valid Loss:  tensor(0.9438, device='cuda:0')
[301] Valid Loss:  tensor(0.6066, device='cuda:0')
[302] Valid Loss:  tensor(0.5728, device='cuda:0')
[303] Valid Loss:  tensor(0.6950, device='cuda:0')
[304] Valid Loss:  tensor(0.6648, device='cuda:0')
[305] Valid Loss:  tensor(0.7529, device='cuda:0')
[306] Valid Loss:  tensor(0.7790, device='cuda:0')
[307] Valid Loss:  tensor(0.5971, device='cuda:0')
[308] Valid Loss:  tensor(0.6546, device='cuda:0')
[309] Valid Loss:  tensor(0.6873, device='cuda:0')
[310] Valid Loss:  tensor(0.7056, device='cuda:0')
[311] Valid Loss:  tensor(0.6173, device='cuda:0')
[312] Valid Loss:  tensor(0.6213, device='cuda:0')
[313] Valid Loss:  tensor(0.7143, device='cuda:0')
[314] Valid Loss:  tensor(0.7364, device='cuda:0')
[315] Valid Loss:  tensor(1.0386, device='cuda:0')
[316] Valid Loss:  tensor(0.8578, device='cuda:0')
[317] Valid Loss:  tensor(0.7240, device='cuda:0')
[318] Valid Loss:  tensor(0.6734, device='cuda:0')
[319] Valid Loss:  tensor(0.6690, device='cuda:0')
[320] Valid Loss:  tensor(0.6601, device='cuda:0')
[321] Valid Loss:  tensor(0.6888, device='cuda:0')
[322] Valid Loss:  tensor(0.6577, device='cuda:0')
[323] Valid Loss:  tensor(0.8276, device='cuda:0')
[324] Valid Loss:  tensor(0.7637, device='cuda:0')
[325] Valid Loss:  tensor(0.6363, device='cuda:0')
[326] Valid Loss:  tensor(0.7093, device='cuda:0')
[327] Valid Loss:  tensor(0.4631, device='cuda:0')
[328] Valid Loss:  tensor(0.6887, device='cuda:0')
[329] Valid Loss:  tensor(0.8233, device='cuda:0')
[330] Valid Loss:  tensor(0.5986, device='cuda:0')
[331] Valid Loss:  tensor(0.9194, device='cuda:0')
[332] Valid Loss:  tensor(0.6145, device='cuda:0')
[333] Valid Loss:  tensor(0.6001, device='cuda:0')
[334] Valid Loss:  tensor(0.6269, device='cuda:0')
[335] Valid Loss:  tensor(0.7121, device='cuda:0')
[336] Valid Loss:  tensor(0.8054, device='cuda:0')
[337] Valid Loss:  tensor(0.9034, device='cuda:0')
[338] Valid Loss:  tensor(0.7563, device='cuda:0')
```

```
[339] Valid Loss:  tensor(0.7132, device='cuda:0')
[340] Valid Loss:  tensor(0.7533, device='cuda:0')
[341] Valid Loss:  tensor(0.8499, device='cuda:0')
[342] Valid Loss:  tensor(0.9280, device='cuda:0')
[343] Valid Loss:  tensor(0.7676, device='cuda:0')
[344] Valid Loss:  tensor(0.6500, device='cuda:0')
[345] Valid Loss:  tensor(0.5561, device='cuda:0')
[346] Valid Loss:  tensor(0.7270, device='cuda:0')
[347] Valid Loss:  tensor(0.8539, device='cuda:0')
[348] Valid Loss:  tensor(0.7100, device='cuda:0')
[349] Valid Loss:  tensor(0.6903, device='cuda:0')
[350] Valid Loss:  tensor(0.7093, device='cuda:0')
[351] Valid Loss:  tensor(0.8353, device='cuda:0')
[352] Valid Loss:  tensor(0.6376, device='cuda:0')
[353] Valid Loss:  tensor(0.7308, device='cuda:0')
[354] Valid Loss:  tensor(0.9343, device='cuda:0')
[355] Valid Loss:  tensor(0.5281, device='cuda:0')
[356] Valid Loss:  tensor(0.6929, device='cuda:0')
[357] Valid Loss:  tensor(0.8318, device='cuda:0')
[358] Valid Loss:  tensor(0.6987, device='cuda:0')
[359] Valid Loss:  tensor(0.7974, device='cuda:0')
[360] Valid Loss:  tensor(0.7151, device='cuda:0')
[361] Valid Loss:  tensor(0.7445, device='cuda:0')
[362] Valid Loss:  tensor(0.7405, device='cuda:0')
[363] Valid Loss:  tensor(0.7914, device='cuda:0')
[364] Valid Loss:  tensor(0.9123, device='cuda:0')
[365] Valid Loss:  tensor(0.6953, device='cuda:0')
[366] Valid Loss:  tensor(0.6923, device='cuda:0')
[367] Valid Loss:  tensor(0.7817, device='cuda:0')
[368] Valid Loss:  tensor(0.8550, device='cuda:0')
[369] Valid Loss:  tensor(0.7214, device='cuda:0')
[370] Valid Loss:  tensor(0.6293, device='cuda:0')
[371] Valid Loss:  tensor(0.6326, device='cuda:0')
[372] Valid Loss:  tensor(0.8259, device='cuda:0')
[373] Valid Loss:  tensor(0.5546, device='cuda:0')
[374] Valid Loss:  tensor(0.7287, device='cuda:0')
[375] Valid Loss:  tensor(0.8029, device='cuda:0')
[376] Valid Loss:  tensor(0.6388, device='cuda:0')
[377] Valid Loss:  tensor(0.9376, device='cuda:0')
[378] Valid Loss:  tensor(0.6612, device='cuda:0')
[379] Valid Loss:  tensor(0.8111, device='cuda:0')
[380] Valid Loss:  tensor(1.0980, device='cuda:0')
Testing Loss:  0.750731614630992
Testing Accuaracy:  76.8419628018732 %
Epoch Number:  7
[0] Train Loss:  tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[1] Train Loss:  tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[2] Train Loss:  tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3] Train Loss:  tensor(0.6573, device='cuda:0', grad_fn=<NllLossBackward>)
[4] Train Loss:  tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[5] Train Loss:  tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[6] Train Loss:  tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[7] Train Loss:  tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[8] Train Loss:  tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[9] Train Loss:  tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[10] Train Loss:  tensor(0.6729, device='cuda:0', grad_fn=<NllLossBackward>)
[11] Train Loss:  tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[12] Train Loss:  tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[13] Train Loss:  tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[14] Train Loss:  tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[15] Train Loss:  tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[16] Train Loss:  tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[17] Train Loss:  tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[18] Train Loss:  tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[19] Train Loss:  tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[20] Train Loss:  tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[21] Train Loss:  tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[22] Train Loss:  tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[23] Train Loss:  tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[24] Train Loss:  tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[25] Train Loss:  tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[26] Train Loss:  tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[27] Train Loss:  tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[28] Train Loss:  tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[29] Train Loss:  tensor(0.6629, device='cuda:0', grad_fn=<NllLossBackward>)
[30] Train Loss:  tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[31] Train Loss:  tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[32] Train Loss:  tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[33] Train Loss:  tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[34] Train Loss:  tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[35] Train Loss:  tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[36] Train Loss:  tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[37] Train Loss:  tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[38] Train Loss:  tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[39] Train Loss:  tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[40] Train Loss:  tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[41] Train Loss:  tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[42] Train Loss:  tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[43] Train Loss:  tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[44] Train Loss:  tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[45] Train Loss:  tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[46] Train Loss:  tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[47] Train Loss:  tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[48] Train Loss:  tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[49] Train Loss:  tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[50] Train Loss:  tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[51] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[52] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[53] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[54] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[55] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[56] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[57] Train Loss:    tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
[58] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[59] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[60] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[61] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[62] Train Loss:    tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[63] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[64] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[65] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[66] Train Loss:    tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[67] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[68] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[69] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[70] Train Loss:    tensor(0.7459, device='cuda:0', grad_fn=<NllLossBackward>)
[71] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[72] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[73] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[74] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[75] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[76] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[77] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[78] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[79] Train Loss:    tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[80] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[81] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[82] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[83] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[84] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[85] Train Loss:    tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[86] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[87] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[88] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[89] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[90] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[91] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[92] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[93] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[94] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[95] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[96] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[97] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[98] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[99] Train Loss:  tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[100] Train Loss:  tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[101] Train Loss:  tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[102] Train Loss:  tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[103] Train Loss:  tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[104] Train Loss:  tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[105] Train Loss:  tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[106] Train Loss:  tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[107] Train Loss:  tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[108] Train Loss:  tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[109] Train Loss:  tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[110] Train Loss:  tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[111] Train Loss:  tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[112] Train Loss:  tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[113] Train Loss:  tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[114] Train Loss:  tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[115] Train Loss:  tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[116] Train Loss:  tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[117] Train Loss:  tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[118] Train Loss:  tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[119] Train Loss:  tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[120] Train Loss:  tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[121] Train Loss:  tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[122] Train Loss:  tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[123] Train Loss:  tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[124] Train Loss:  tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[125] Train Loss:  tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[126] Train Loss:  tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[127] Train Loss:  tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[128] Train Loss:  tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[129] Train Loss:  tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[130] Train Loss:  tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[131] Train Loss:  tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[132] Train Loss:  tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[133] Train Loss:  tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[134] Train Loss:  tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[135] Train Loss:  tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[136] Train Loss:  tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[137] Train Loss:  tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[138] Train Loss:  tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[139] Train Loss:  tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[140] Train Loss:  tensor(0.7642, device='cuda:0', grad_fn=<NllLossBackward>)
[141] Train Loss:  tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[142] Train Loss:  tensor(0.6715, device='cuda:0', grad_fn=<NllLossBackward>)
[143] Train Loss:  tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[144] Train Loss:  tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[145] Train Loss:  tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[146] Train Loss:  tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[147] Train Loss:   tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[148] Train Loss:   tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[149] Train Loss:   tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[150] Train Loss:   tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[151] Train Loss:   tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[152] Train Loss:   tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[153] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[154] Train Loss:   tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[155] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[156] Train Loss:   tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[157] Train Loss:   tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[158] Train Loss:   tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[159] Train Loss:   tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[160] Train Loss:   tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[161] Train Loss:   tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[162] Train Loss:   tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[163] Train Loss:   tensor(0.7483, device='cuda:0', grad_fn=<NllLossBackward>)
[164] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[165] Train Loss:   tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[166] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[167] Train Loss:   tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[168] Train Loss:   tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[169] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[170] Train Loss:   tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[171] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[172] Train Loss:   tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[173] Train Loss:   tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[174] Train Loss:   tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[175] Train Loss:   tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[176] Train Loss:   tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[177] Train Loss:   tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[178] Train Loss:   tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[179] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[180] Train Loss:   tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[181] Train Loss:   tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[182] Train Loss:   tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[183] Train Loss:   tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[184] Train Loss:   tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[185] Train Loss:   tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[186] Train Loss:   tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[187] Train Loss:   tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[188] Train Loss:   tensor(0.7552, device='cuda:0', grad_fn=<NllLossBackward>)
[189] Train Loss:   tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[190] Train Loss:   tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[191] Train Loss:   tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[192] Train Loss:   tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[193] Train Loss:   tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[194] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[195] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[196] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[197] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[198] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[199] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[200] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[201] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[202] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[203] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[204] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[205] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[206] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[207] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[208] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[209] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[210] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[211] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[212] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[213] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[214] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[215] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[216] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[217] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[218] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[219] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[220] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[221] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[222] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[223] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[224] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[225] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[226] Train Loss:    tensor(0.6706, device='cuda:0', grad_fn=<NllLossBackward>)
[227] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[228] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[229] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[230] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[231] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[232] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[233] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[234] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[235] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[236] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[237] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[238] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[239] Train Loss:    tensor(0.7543, device='cuda:0', grad_fn=<NllLossBackward>)
[240] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[241] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[242] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[243] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[244] Train Loss:   tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[245] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[246] Train Loss:   tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[247] Train Loss:   tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[248] Train Loss:   tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[249] Train Loss:   tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[250] Train Loss:   tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[251] Train Loss:   tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[252] Train Loss:   tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[253] Train Loss:   tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[254] Train Loss:   tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[255] Train Loss:   tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[256] Train Loss:   tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[257] Train Loss:   tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[258] Train Loss:   tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[259] Train Loss:   tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[260] Train Loss:   tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[261] Train Loss:   tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[262] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[263] Train Loss:   tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[264] Train Loss:   tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[265] Train Loss:   tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[266] Train Loss:   tensor(0.6709, device='cuda:0', grad_fn=<NllLossBackward>)
[267] Train Loss:   tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[268] Train Loss:   tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[269] Train Loss:   tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[270] Train Loss:   tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[271] Train Loss:   tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[272] Train Loss:   tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[273] Train Loss:   tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[274] Train Loss:   tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[275] Train Loss:   tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[276] Train Loss:   tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[277] Train Loss:   tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[278] Train Loss:   tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[279] Train Loss:   tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[280] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[281] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[282] Train Loss:   tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[283] Train Loss:   tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[284] Train Loss:   tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[285] Train Loss:   tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[286] Train Loss:   tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[287] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[288] Train Loss:   tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[289] Train Loss:   tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[290] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[291] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[292] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[293] Train Loss:    tensor(0.6613, device='cuda:0', grad_fn=<NllLossBackward>)
[294] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[295] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[296] Train Loss:    tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[297] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[298] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[299] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[300] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[301] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[302] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[303] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[304] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[305] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[306] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[307] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[308] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[309] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[310] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[311] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[312] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[313] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[314] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[315] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[316] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[317] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[318] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[319] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[320] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[321] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[322] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[323] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[324] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[325] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[326] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[327] Train Loss:    tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[328] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[329] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[330] Train Loss:    tensor(0.7545, device='cuda:0', grad_fn=<NllLossBackward>)
[331] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[332] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[333] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[334] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[335] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[336] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[337] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[338] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[339] Train Loss:  tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[340] Train Loss:  tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[341] Train Loss:  tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[342] Train Loss:  tensor(0.7690, device='cuda:0', grad_fn=<NllLossBackward>)
[343] Train Loss:  tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[344] Train Loss:  tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[345] Train Loss:  tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[346] Train Loss:  tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[347] Train Loss:  tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[348] Train Loss:  tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[349] Train Loss:  tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[350] Train Loss:  tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[351] Train Loss:  tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[352] Train Loss:  tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[353] Train Loss:  tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[354] Train Loss:  tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[355] Train Loss:  tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[356] Train Loss:  tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[357] Train Loss:  tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[358] Train Loss:  tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[359] Train Loss:  tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[360] Train Loss:  tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[361] Train Loss:  tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[362] Train Loss:  tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[363] Train Loss:  tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[364] Train Loss:  tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[365] Train Loss:  tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[366] Train Loss:  tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[367] Train Loss:  tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[368] Train Loss:  tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[369] Train Loss:  tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[370] Train Loss:  tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[371] Train Loss:  tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[372] Train Loss:  tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[373] Train Loss:  tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[374] Train Loss:  tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[375] Train Loss:  tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[376] Train Loss:  tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[377] Train Loss:  tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[378] Train Loss:  tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[379] Train Loss:  tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[380] Train Loss:  tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[381] Train Loss:  tensor(0.7614, device='cuda:0', grad_fn=<NllLossBackward>)
[382] Train Loss:  tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[383] Train Loss:  tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[384] Train Loss:  tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[385] Train Loss:  tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[386] Train Loss:  tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[387] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[388] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[389] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[390] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[391] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[392] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[393] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[394] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[395] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[396] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[397] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[398] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[399] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[400] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[401] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[402] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[403] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[404] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[405] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[406] Train Loss:    tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[407] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[408] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[409] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[410] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[411] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[412] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[413] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[414] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[415] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[416] Train Loss:    tensor(0.6529, device='cuda:0', grad_fn=<NllLossBackward>)
[417] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[418] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[419] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[420] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[421] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[422] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[423] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[424] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[425] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[426] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[427] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[428] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[429] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[430] Train Loss:    tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[431] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[432] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[433] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[434] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[435] Train Loss:   tensor(0.7787, device='cuda:0', grad_fn=<NllLossBackward>)
[436] Train Loss:   tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[437] Train Loss:   tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[438] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[439] Train Loss:   tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[440] Train Loss:   tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[441] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[442] Train Loss:   tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[443] Train Loss:   tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[444] Train Loss:   tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[445] Train Loss:   tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[446] Train Loss:   tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[447] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[448] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[449] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[450] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[451] Train Loss:   tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[452] Train Loss:   tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[453] Train Loss:   tensor(0.6733, device='cuda:0', grad_fn=<NllLossBackward>)
[454] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[455] Train Loss:   tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[456] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[457] Train Loss:   tensor(0.6604, device='cuda:0', grad_fn=<NllLossBackward>)
[458] Train Loss:   tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[459] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[460] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[461] Train Loss:   tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[462] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[463] Train Loss:   tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[464] Train Loss:   tensor(0.7675, device='cuda:0', grad_fn=<NllLossBackward>)
[465] Train Loss:   tensor(0.7516, device='cuda:0', grad_fn=<NllLossBackward>)
[466] Train Loss:   tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[467] Train Loss:   tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[468] Train Loss:   tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[469] Train Loss:   tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[470] Train Loss:   tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[471] Train Loss:   tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[472] Train Loss:   tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[473] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[474] Train Loss:   tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[475] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[476] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[477] Train Loss:   tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[478] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[479] Train Loss:   tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[480] Train Loss:   tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[481] Train Loss:   tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[482] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[483] Train Loss:   tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[484] Train Loss:   tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[485] Train Loss:   tensor(0.7522, device='cuda:0', grad_fn=<NllLossBackward>)
[486] Train Loss:   tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[487] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[488] Train Loss:   tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[489] Train Loss:   tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[490] Train Loss:   tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[491] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[492] Train Loss:   tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[493] Train Loss:   tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[494] Train Loss:   tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[495] Train Loss:   tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[496] Train Loss:   tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[497] Train Loss:   tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[498] Train Loss:   tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[499] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[500] Train Loss:   tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[501] Train Loss:   tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[502] Train Loss:   tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[503] Train Loss:   tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[504] Train Loss:   tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[505] Train Loss:   tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[506] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[507] Train Loss:   tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[508] Train Loss:   tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[509] Train Loss:   tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[510] Train Loss:   tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[511] Train Loss:   tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[512] Train Loss:   tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[513] Train Loss:   tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[514] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[515] Train Loss:   tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[516] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[517] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[518] Train Loss:   tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[519] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[520] Train Loss:   tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[521] Train Loss:   tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[522] Train Loss:   tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[523] Train Loss:   tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[524] Train Loss:   tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[525] Train Loss:   tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[526] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[527] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[528] Train Loss:   tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[529] Train Loss:   tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[530] Train Loss:   tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[531] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[532] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[533] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[534] Train Loss:    tensor(0.7530, device='cuda:0', grad_fn=<NllLossBackward>)
[535] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[536] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[537] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[538] Train Loss:    tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[539] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[540] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[541] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[542] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[543] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[544] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[545] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[546] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[547] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[548] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[549] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[550] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[551] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[552] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[553] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[554] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[555] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[556] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[557] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[558] Train Loss:    tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[559] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[560] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[561] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[562] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[563] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[564] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[565] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[566] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[567] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[568] Train Loss:    tensor(0.7488, device='cuda:0', grad_fn=<NllLossBackward>)
[569] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[570] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[571] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[572] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[573] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[574] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[575] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[576] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[577] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[578] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[579] Train Loss:   tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[580] Train Loss:   tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[581] Train Loss:   tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[582] Train Loss:   tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[583] Train Loss:   tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[584] Train Loss:   tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[585] Train Loss:   tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[586] Train Loss:   tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[587] Train Loss:   tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[588] Train Loss:   tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[589] Train Loss:   tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[590] Train Loss:   tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[591] Train Loss:   tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[592] Train Loss:   tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[593] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[594] Train Loss:   tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[595] Train Loss:   tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[596] Train Loss:   tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[597] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[598] Train Loss:   tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[599] Train Loss:   tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[600] Train Loss:   tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[601] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[602] Train Loss:   tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[603] Train Loss:   tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[604] Train Loss:   tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[605] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[606] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[607] Train Loss:   tensor(0.6639, device='cuda:0', grad_fn=<NllLossBackward>)
[608] Train Loss:   tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[609] Train Loss:   tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[610] Train Loss:   tensor(0.6684, device='cuda:0', grad_fn=<NllLossBackward>)
[611] Train Loss:   tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[612] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[613] Train Loss:   tensor(0.6601, device='cuda:0', grad_fn=<NllLossBackward>)
[614] Train Loss:   tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[615] Train Loss:   tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[616] Train Loss:   tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[617] Train Loss:   tensor(0.6629, device='cuda:0', grad_fn=<NllLossBackward>)
[618] Train Loss:   tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[619] Train Loss:   tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[620] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[621] Train Loss:   tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[622] Train Loss:   tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[623] Train Loss:   tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[624] Train Loss:   tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[625] Train Loss:   tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[626] Train Loss:   tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[627] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[628] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[629] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[630] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[631] Train Loss:    tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[632] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[633] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[634] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[635] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[636] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[637] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[638] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[639] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[640] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[641] Train Loss:    tensor(0.6625, device='cuda:0', grad_fn=<NllLossBackward>)
[642] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[643] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[644] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[645] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[646] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[647] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[648] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[649] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[650] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[651] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[652] Train Loss:    tensor(0.6609, device='cuda:0', grad_fn=<NllLossBackward>)
[653] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[654] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[655] Train Loss:    tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[656] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[657] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[658] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[659] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[660] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[661] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[662] Train Loss:    tensor(0.7575, device='cuda:0', grad_fn=<NllLossBackward>)
[663] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[664] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[665] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[666] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[667] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[668] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[669] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[670] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[671] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[672] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[673] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[674] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[675] Train Loss:   tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[676] Train Loss:   tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[677] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[678] Train Loss:   tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[679] Train Loss:   tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[680] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[681] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[682] Train Loss:   tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[683] Train Loss:   tensor(0.6583, device='cuda:0', grad_fn=<NllLossBackward>)
[684] Train Loss:   tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[685] Train Loss:   tensor(0.7585, device='cuda:0', grad_fn=<NllLossBackward>)
[686] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[687] Train Loss:   tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[688] Train Loss:   tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[689] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[690] Train Loss:   tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[691] Train Loss:   tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[692] Train Loss:   tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[693] Train Loss:   tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[694] Train Loss:   tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[695] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[696] Train Loss:   tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[697] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[698] Train Loss:   tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[699] Train Loss:   tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[700] Train Loss:   tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[701] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[702] Train Loss:   tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[703] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[704] Train Loss:   tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[705] Train Loss:   tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[706] Train Loss:   tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[707] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[708] Train Loss:   tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[709] Train Loss:   tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[710] Train Loss:   tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[711] Train Loss:   tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[712] Train Loss:   tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[713] Train Loss:   tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[714] Train Loss:   tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[715] Train Loss:   tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[716] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[717] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[718] Train Loss:   tensor(0.6663, device='cuda:0', grad_fn=<NllLossBackward>)
[719] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[720] Train Loss:   tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[721] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[722] Train Loss:   tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[723] Train Loss:   tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[724] Train Loss:   tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[725] Train Loss:   tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[726] Train Loss:   tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[727] Train Loss:   tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[728] Train Loss:   tensor(0.6664, device='cuda:0', grad_fn=<NllLossBackward>)
[729] Train Loss:   tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[730] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[731] Train Loss:   tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[732] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[733] Train Loss:   tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[734] Train Loss:   tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[735] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[736] Train Loss:   tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[737] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[738] Train Loss:   tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[739] Train Loss:   tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[740] Train Loss:   tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[741] Train Loss:   tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[742] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[743] Train Loss:   tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[744] Train Loss:   tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[745] Train Loss:   tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[746] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[747] Train Loss:   tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[748] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[749] Train Loss:   tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[750] Train Loss:   tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[751] Train Loss:   tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[752] Train Loss:   tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[753] Train Loss:   tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[754] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[755] Train Loss:   tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[756] Train Loss:   tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[757] Train Loss:   tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[758] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[759] Train Loss:   tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[760] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[761] Train Loss:   tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[762] Train Loss:   tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[763] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[764] Train Loss:   tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[765] Train Loss:   tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[766] Train Loss:   tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[767] Train Loss:   tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[768] Train Loss:   tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[769] Train Loss:   tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[770] Train Loss:   tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[771] Train Loss:   tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[772] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[773] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[774] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[775] Train Loss:   tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[776] Train Loss:   tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[777] Train Loss:   tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[778] Train Loss:   tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[779] Train Loss:   tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[780] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[781] Train Loss:   tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[782] Train Loss:   tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[783] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[784] Train Loss:   tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[785] Train Loss:   tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[786] Train Loss:   tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[787] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[788] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[789] Train Loss:   tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[790] Train Loss:   tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[791] Train Loss:   tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[792] Train Loss:   tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[793] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[794] Train Loss:   tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[795] Train Loss:   tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[796] Train Loss:   tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[797] Train Loss:   tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[798] Train Loss:   tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[799] Train Loss:   tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[800] Train Loss:   tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[801] Train Loss:   tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[802] Train Loss:   tensor(0.6636, device='cuda:0', grad_fn=<NllLossBackward>)
[803] Train Loss:   tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[804] Train Loss:   tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[805] Train Loss:   tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[806] Train Loss:   tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[807] Train Loss:   tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[808] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[809] Train Loss:   tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[810] Train Loss:   tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[811] Train Loss:   tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[812] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[813] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[814] Train Loss:   tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[815] Train Loss:   tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[816] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[817] Train Loss:   tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[818] Train Loss:   tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[819] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[820] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[821] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[822] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[823] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[824] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[825] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[826] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[827] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[828] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[829] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[830] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[831] Train Loss:    tensor(0.7495, device='cuda:0', grad_fn=<NllLossBackward>)
[832] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[833] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[834] Train Loss:    tensor(0.7468, device='cuda:0', grad_fn=<NllLossBackward>)
[835] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[836] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[837] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[838] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[839] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[840] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[841] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[842] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[843] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[844] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[845] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[846] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[847] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[848] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[849] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[850] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[851] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[852] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[853] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[854] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[855] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[856] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[857] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[858] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[859] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[860] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[861] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[862] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[863] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[864] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[865] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[866] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[867] Train Loss:  tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[868] Train Loss:  tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[869] Train Loss:  tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[870] Train Loss:  tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[871] Train Loss:  tensor(0.6679, device='cuda:0', grad_fn=<NllLossBackward>)
[872] Train Loss:  tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[873] Train Loss:  tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[874] Train Loss:  tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[875] Train Loss:  tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[876] Train Loss:  tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[877] Train Loss:  tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[878] Train Loss:  tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[879] Train Loss:  tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[880] Train Loss:  tensor(0.6676, device='cuda:0', grad_fn=<NllLossBackward>)
[881] Train Loss:  tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[882] Train Loss:  tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[883] Train Loss:  tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[884] Train Loss:  tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[885] Train Loss:  tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[886] Train Loss:  tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[887] Train Loss:  tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[888] Train Loss:  tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[889] Train Loss:  tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[890] Train Loss:  tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[891] Train Loss:  tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[892] Train Loss:  tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[893] Train Loss:  tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[894] Train Loss:  tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[895] Train Loss:  tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[896] Train Loss:  tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[897] Train Loss:  tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[898] Train Loss:  tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[899] Train Loss:  tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[900] Train Loss:  tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[901] Train Loss:  tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[902] Train Loss:  tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[903] Train Loss:  tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[904] Train Loss:  tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[905] Train Loss:  tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[906] Train Loss:  tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[907] Train Loss:  tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[908] Train Loss:  tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[909] Train Loss:  tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[910] Train Loss:  tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[911] Train Loss:  tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[912] Train Loss:  tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[913] Train Loss:  tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[914] Train Loss:  tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[915] Train Loss:   tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[916] Train Loss:   tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[917] Train Loss:   tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[918] Train Loss:   tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[919] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[920] Train Loss:   tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[921] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[922] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[923] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[924] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[925] Train Loss:   tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[926] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[927] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[928] Train Loss:   tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[929] Train Loss:   tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[930] Train Loss:   tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[931] Train Loss:   tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[932] Train Loss:   tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[933] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[934] Train Loss:   tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[935] Train Loss:   tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[936] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[937] Train Loss:   tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[938] Train Loss:   tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[939] Train Loss:   tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[940] Train Loss:   tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[941] Train Loss:   tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[942] Train Loss:   tensor(0.7520, device='cuda:0', grad_fn=<NllLossBackward>)
[943] Train Loss:   tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[944] Train Loss:   tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[945] Train Loss:   tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[946] Train Loss:   tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[947] Train Loss:   tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[948] Train Loss:   tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[949] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[950] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[951] Train Loss:   tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[952] Train Loss:   tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[953] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[954] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[955] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[956] Train Loss:   tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[957] Train Loss:   tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[958] Train Loss:   tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[959] Train Loss:   tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[960] Train Loss:   tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[961] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[962] Train Loss:   tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[963] Train Loss:  tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[964] Train Loss:  tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[965] Train Loss:  tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[966] Train Loss:  tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[967] Train Loss:  tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[968] Train Loss:  tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[969] Train Loss:  tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[970] Train Loss:  tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[971] Train Loss:  tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[972] Train Loss:  tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[973] Train Loss:  tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[974] Train Loss:  tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[975] Train Loss:  tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[976] Train Loss:  tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[977] Train Loss:  tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[978] Train Loss:  tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[979] Train Loss:  tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[980] Train Loss:  tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[981] Train Loss:  tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[982] Train Loss:  tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[983] Train Loss:  tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[984] Train Loss:  tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[985] Train Loss:  tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[986] Train Loss:  tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[987] Train Loss:  tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[988] Train Loss:  tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[989] Train Loss:  tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[990] Train Loss:  tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[991] Train Loss:  tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[992] Train Loss:  tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[993] Train Loss:  tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[994] Train Loss:  tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[995] Train Loss:  tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[996] Train Loss:  tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[997] Train Loss:  tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[998] Train Loss:  tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[999] Train Loss:  tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[1000] Train Loss:  tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[1001] Train Loss:  tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[1002] Train Loss:  tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[1003] Train Loss:  tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[1004] Train Loss:  tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[1005] Train Loss:  tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1006] Train Loss:  tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[1007] Train Loss:  tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[1008] Train Loss:  tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[1009] Train Loss:  tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1010] Train Loss:  tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1011] Train Loss:   tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[1012] Train Loss:   tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[1013] Train Loss:   tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[1014] Train Loss:   tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[1015] Train Loss:   tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[1016] Train Loss:   tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[1017] Train Loss:   tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[1018] Train Loss:   tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[1019] Train Loss:   tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[1020] Train Loss:   tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[1021] Train Loss:   tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[1022] Train Loss:   tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[1023] Train Loss:   tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[1024] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[1025] Train Loss:   tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[1026] Train Loss:   tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[1027] Train Loss:   tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[1028] Train Loss:   tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[1029] Train Loss:   tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1030] Train Loss:   tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[1031] Train Loss:   tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[1032] Train Loss:   tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[1033] Train Loss:   tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[1034] Train Loss:   tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[1035] Train Loss:   tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[1036] Train Loss:   tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[1037] Train Loss:   tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[1038] Train Loss:   tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[1039] Train Loss:   tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[1040] Train Loss:   tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[1041] Train Loss:   tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[1042] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[1043] Train Loss:   tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[1044] Train Loss:   tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[1045] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[1046] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[1047] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[1048] Train Loss:   tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[1049] Train Loss:   tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[1050] Train Loss:   tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1051] Train Loss:   tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1052] Train Loss:   tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[1053] Train Loss:   tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[1054] Train Loss:   tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[1055] Train Loss:   tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1056] Train Loss:   tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[1057] Train Loss:   tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[1058] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1059] Train Loss:   tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[1060] Train Loss:   tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[1061] Train Loss:   tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[1062] Train Loss:   tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[1063] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[1064] Train Loss:   tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[1065] Train Loss:   tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1066] Train Loss:   tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[1067] Train Loss:   tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[1068] Train Loss:   tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[1069] Train Loss:   tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[1070] Train Loss:   tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[1071] Train Loss:   tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1072] Train Loss:   tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[1073] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[1074] Train Loss:   tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[1075] Train Loss:   tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[1076] Train Loss:   tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[1077] Train Loss:   tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[1078] Train Loss:   tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[1079] Train Loss:   tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[1080] Train Loss:   tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[1081] Train Loss:   tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[1082] Train Loss:   tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[1083] Train Loss:   tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[1084] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[1085] Train Loss:   tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[1086] Train Loss:   tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[1087] Train Loss:   tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[1088] Train Loss:   tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[1089] Train Loss:   tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[1090] Train Loss:   tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[1091] Train Loss:   tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1092] Train Loss:   tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[1093] Train Loss:   tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[1094] Train Loss:   tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1095] Train Loss:   tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[1096] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1097] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[1098] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1099] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[1100] Train Loss:   tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[1101] Train Loss:   tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[1102] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[1103] Train Loss:   tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[1104] Train Loss:   tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[1105] Train Loss:   tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[1106] Train Loss:   tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1107] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1108] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[1109] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[1110] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[1111] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[1112] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[1113] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[1114] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[1115] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[1116] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[1117] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[1118] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[1119] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[1120] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[1121] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[1122] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[1123] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[1124] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[1125] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[1126] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[1127] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1128] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[1129] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[1130] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[1131] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[1132] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[1133] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[1134] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[1135] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[1136] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[1137] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[1138] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1139] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[1140] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[1141] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[1142] Train Loss:    tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[1143] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[1144] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[1145] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[1146] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[1147] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[1148] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[1149] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[1150] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[1151] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[1152] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[1153] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[1154] Train Loss:    tensor(0.7504, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1155] Train Loss:    tensor(0.6605, device='cuda:0', grad_fn=<NllLossBackward>)
[1156] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[1157] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[1158] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[1159] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[1160] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[1161] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1162] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[1163] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[1164] Train Loss:    tensor(0.7492, device='cuda:0', grad_fn=<NllLossBackward>)
[1165] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1166] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[1167] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[1168] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[1169] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[1170] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[1171] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[1172] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[1173] Train Loss:    tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[1174] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[1175] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[1176] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[1177] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[1178] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[1179] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1180] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[1181] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[1182] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[1183] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[1184] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[1185] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[1186] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[1187] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1188] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[1189] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[1190] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[1191] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[1192] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[1193] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1194] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[1195] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[1196] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[1197] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[1198] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[1199] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[1200] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[1201] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[1202] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1203] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[1204] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[1205] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[1206] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[1207] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[1208] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1209] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[1210] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[1211] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[1212] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[1213] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[1214] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[1215] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[1216] Train Loss:    tensor(0.7664, device='cuda:0', grad_fn=<NllLossBackward>)
[1217] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[1218] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[1219] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1220] Train Loss:    tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[1221] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1222] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[1223] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[1224] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[1225] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[1226] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[1227] Train Loss:    tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[1228] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[1229] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
[1230] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[1231] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[1232] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[1233] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1234] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[1235] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[1236] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[1237] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[1238] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[1239] Train Loss:    tensor(0.7440, device='cuda:0', grad_fn=<NllLossBackward>)
[1240] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[1241] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[1242] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[1243] Train Loss:    tensor(0.7646, device='cuda:0', grad_fn=<NllLossBackward>)
[1244] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[1245] Train Loss:    tensor(0.6748, device='cuda:0', grad_fn=<NllLossBackward>)
[1246] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[1247] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[1248] Train Loss:    tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[1249] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[1250] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1251] Train Loss:   tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[1252] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[1253] Train Loss:   tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[1254] Train Loss:   tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[1255] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[1256] Train Loss:   tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[1257] Train Loss:   tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[1258] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[1259] Train Loss:   tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[1260] Train Loss:   tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1261] Train Loss:   tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[1262] Train Loss:   tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[1263] Train Loss:   tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1264] Train Loss:   tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[1265] Train Loss:   tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[1266] Train Loss:   tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[1267] Train Loss:   tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1268] Train Loss:   tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[1269] Train Loss:   tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1270] Train Loss:   tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[1271] Train Loss:   tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[1272] Train Loss:   tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[1273] Train Loss:   tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[1274] Train Loss:   tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[1275] Train Loss:   tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1276] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[1277] Train Loss:   tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[1278] Train Loss:   tensor(0.7905, device='cuda:0', grad_fn=<NllLossBackward>)
[1279] Train Loss:   tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[1280] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[1281] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[1282] Train Loss:   tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[1283] Train Loss:   tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[1284] Train Loss:   tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1285] Train Loss:   tensor(0.6705, device='cuda:0', grad_fn=<NllLossBackward>)
[1286] Train Loss:   tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[1287] Train Loss:   tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[1288] Train Loss:   tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[1289] Train Loss:   tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[1290] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[1291] Train Loss:   tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[1292] Train Loss:   tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[1293] Train Loss:   tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[1294] Train Loss:   tensor(0.7598, device='cuda:0', grad_fn=<NllLossBackward>)
[1295] Train Loss:   tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[1296] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[1297] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[1298] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1299] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[1300] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[1301] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1302] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[1303] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[1304] Train Loss:    tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[1305] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[1306] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[1307] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[1308] Train Loss:    tensor(0.6665, device='cuda:0', grad_fn=<NllLossBackward>)
[1309] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[1310] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[1311] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[1312] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[1313] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[1314] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[1315] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[1316] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[1317] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[1318] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[1319] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1320] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[1321] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[1322] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[1323] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[1324] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[1325] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[1326] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[1327] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[1328] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[1329] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[1330] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[1331] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[1332] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[1333] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[1334] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[1335] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[1336] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[1337] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[1338] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[1339] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[1340] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[1341] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[1342] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[1343] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[1344] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[1345] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[1346] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1347] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[1348] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[1349] Train Loss:    tensor(0.6688, device='cuda:0', grad_fn=<NllLossBackward>)
[1350] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[1351] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[1352] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[1353] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[1354] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[1355] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[1356] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[1357] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[1358] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[1359] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[1360] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[1361] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[1362] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1363] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[1364] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[1365] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[1366] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[1367] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[1368] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[1369] Train Loss:    tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[1370] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[1371] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1372] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[1373] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[1374] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[1375] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[1376] Train Loss:    tensor(0.6704, device='cuda:0', grad_fn=<NllLossBackward>)
[1377] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[1378] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[1379] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[1380] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[1381] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[1382] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[1383] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[1384] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[1385] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[1386] Train Loss:    tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[1387] Train Loss:    tensor(0.7366, device='cuda:0', grad_fn=<NllLossBackward>)
[1388] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[1389] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[1390] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[1391] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[1392] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[1393] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[1394] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1395] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[1396] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[1397] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[1398] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[1399] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[1400] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[1401] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[1402] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[1403] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1404] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[1405] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[1406] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1407] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[1408] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[1409] Train Loss:    tensor(0.7430, device='cuda:0', grad_fn=<NllLossBackward>)
[1410] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[1411] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[1412] Train Loss:    tensor(0.7493, device='cuda:0', grad_fn=<NllLossBackward>)
[1413] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1414] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1415] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[1416] Train Loss:    tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[1417] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[1418] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[1419] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[1420] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[1421] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[1422] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[1423] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1424] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[1425] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[1426] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[1427] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[1428] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[1429] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1430] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[1431] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[1432] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[1433] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[1434] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[1435] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[1436] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[1437] Train Loss:    tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[1438] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[1439] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[1440] Train Loss:    tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[1441] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[1442] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1443] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[1444] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[1445] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[1446] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[1447] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[1448] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[1449] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[1450] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[1451] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[1452] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[1453] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[1454] Train Loss:    tensor(0.6748, device='cuda:0', grad_fn=<NllLossBackward>)
[1455] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[1456] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[1457] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[1458] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[1459] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[1460] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[1461] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[1462] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[1463] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[1464] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[1465] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[1466] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1467] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[1468] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[1469] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[1470] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[1471] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[1472] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[1473] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[1474] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[1475] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1476] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[1477] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[1478] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[1479] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1480] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[1481] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[1482] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[1483] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[1484] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[1485] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[1486] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1487] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[1488] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[1489] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[1490] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1491] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[1492] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[1493] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[1494] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[1495] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[1496] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[1497] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[1498] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[1499] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[1500] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[1501] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[1502] Train Loss:    tensor(0.7454, device='cuda:0', grad_fn=<NllLossBackward>)
[1503] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[1504] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[1505] Train Loss:    tensor(0.7730, device='cuda:0', grad_fn=<NllLossBackward>)
[1506] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[1507] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[1508] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[1509] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[1510] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[1511] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[1512] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[1513] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[1514] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[1515] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[1516] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[1517] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[1518] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[1519] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[1520] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[1521] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1522] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[1523] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[1524] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1525] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[1526] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[1527] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[1528] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[1529] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[1530] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[1531] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[1532] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[1533] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[1534] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[1535] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[1536] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[1537] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[1538] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1539] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[1540] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[1541] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[1542] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[1543] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[1544] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[1545] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[1546] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[1547] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[1548] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[1549] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[1550] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[1551] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[1552] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[1553] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[1554] Train Loss:    tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[1555] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[1556] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[1557] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[1558] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[1559] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[1560] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[1561] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[1562] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[1563] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[1564] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1565] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[1566] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[1567] Train Loss:    tensor(0.6463, device='cuda:0', grad_fn=<NllLossBackward>)
[1568] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[1569] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[1570] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1571] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[1572] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[1573] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[1574] Train Loss:    tensor(0.6609, device='cuda:0', grad_fn=<NllLossBackward>)
[1575] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[1576] Train Loss:    tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[1577] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[1578] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[1579] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[1580] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[1581] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1582] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[1583] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[1584] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[1585] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1586] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1587] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[1588] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[1589] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[1590] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[1591] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1592] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[1593] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[1594] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[1595] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1596] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[1597] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[1598] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[1599] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[1600] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1601] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[1602] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[1603] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[1604] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[1605] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1606] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[1607] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[1608] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[1609] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[1610] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1611] Train Loss:    tensor(0.6679, device='cuda:0', grad_fn=<NllLossBackward>)
[1612] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[1613] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[1614] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1615] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[1616] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[1617] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[1618] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[1619] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[1620] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1621] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[1622] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[1623] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[1624] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[1625] Train Loss:    tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[1626] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[1627] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[1628] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[1629] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[1630] Train Loss:    tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[1631] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[1632] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[1633] Train Loss:    tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[1634] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1635] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[1636] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[1637] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[1638] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[1639] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[1640] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[1641] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1642] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[1643] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[1644] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[1645] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[1646] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[1647] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[1648] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[1649] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[1650] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[1651] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1652] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[1653] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[1654] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[1655] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[1656] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[1657] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[1658] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[1659] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[1660] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[1661] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[1662] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[1663] Train Loss:    tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[1664] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[1665] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[1666] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[1667] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[1668] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[1669] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[1670] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[1671] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[1672] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[1673] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[1674] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[1675] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[1676] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[1677] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[1678] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[1679] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[1680] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1681] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[1682] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1683] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[1684] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[1685] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[1686] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[1687] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[1688] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[1689] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[1690] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[1691] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[1692] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[1693] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[1694] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[1695] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[1696] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[1697] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[1698] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[1699] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[1700] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1701] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[1702] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1703] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1704] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[1705] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[1706] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[1707] Train Loss:    tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[1708] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[1709] Train Loss:    tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[1710] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[1711] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[1712] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[1713] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[1714] Train Loss:    tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[1715] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[1716] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[1717] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1718] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[1719] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[1720] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[1721] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[1722] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[1723] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[1724] Train Loss:    tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
[1725] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[1726] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[1727] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[1728] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[1729] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[1730] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1731] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[1732] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1733] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[1734] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[1735] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[1736] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[1737] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[1738] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[1739] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[1740] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[1741] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[1742] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[1743] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1744] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[1745] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[1746] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[1747] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[1748] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[1749] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[1750] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
[1751] Train Loss:    tensor(0.6719, device='cuda:0', grad_fn=<NllLossBackward>)
[1752] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[1753] Train Loss:    tensor(0.6672, device='cuda:0', grad_fn=<NllLossBackward>)
[1754] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[1755] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[1756] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[1757] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[1758] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[1759] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[1760] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[1761] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[1762] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[1763] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[1764] Train Loss:    tensor(0.7474, device='cuda:0', grad_fn=<NllLossBackward>)
[1765] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[1766] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[1767] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[1768] Train Loss:    tensor(0.7600, device='cuda:0', grad_fn=<NllLossBackward>)
[1769] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[1770] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[1771] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[1772] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[1773] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[1774] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[1775] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[1776] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[1777] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[1778] Train Loss:    tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1779] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[1780] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[1781] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[1782] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[1783] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[1784] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[1785] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1786] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[1787] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[1788] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[1789] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1790] Train Loss:    tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[1791] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[1792] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[1793] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[1794] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[1795] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[1796] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[1797] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[1798] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[1799] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1800] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[1801] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[1802] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[1803] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[1804] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[1805] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[1806] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[1807] Train Loss:    tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[1808] Train Loss:    tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[1809] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[1810] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[1811] Train Loss:    tensor(0.7463, device='cuda:0', grad_fn=<NllLossBackward>)
[1812] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[1813] Train Loss:    tensor(0.6643, device='cuda:0', grad_fn=<NllLossBackward>)
[1814] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[1815] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[1816] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[1817] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[1818] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[1819] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[1820] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[1821] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1822] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[1823] Train Loss:    tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[1824] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[1825] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[1826] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1827] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1828] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[1829] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[1830] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[1831] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[1832] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[1833] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[1834] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[1835] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[1836] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[1837] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[1838] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[1839] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[1840] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[1841] Train Loss:    tensor(0.7526, device='cuda:0', grad_fn=<NllLossBackward>)
[1842] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[1843] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[1844] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[1845] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[1846] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1847] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[1848] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[1849] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[1850] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[1851] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[1852] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[1853] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[1854] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[1855] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[1856] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[1857] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[1858] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[1859] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1860] Train Loss:    tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[1861] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[1862] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[1863] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[1864] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[1865] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[1866] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[1867] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[1868] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[1869] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[1870] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[1871] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[1872] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[1873] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[1874] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1875] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[1876] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[1877] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[1878] Train Loss:    tensor(0.7640, device='cuda:0', grad_fn=<NllLossBackward>)
[1879] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[1880] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[1881] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[1882] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1883] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[1884] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[1885] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[1886] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[1887] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[1888] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[1889] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[1890] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[1891] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[1892] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[1893] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[1894] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[1895] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[1896] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[1897] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[1898] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1899] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[1900] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[1901] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[1902] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[1903] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1904] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1905] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[1906] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[1907] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[1908] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[1909] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[1910] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[1911] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[1912] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[1913] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[1914] Train Loss:    tensor(0.7439, device='cuda:0', grad_fn=<NllLossBackward>)
[1915] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[1916] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[1917] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[1918] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[1919] Train Loss:    tensor(0.7645, device='cuda:0', grad_fn=<NllLossBackward>)
[1920] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[1921] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[1922] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1923] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[1924] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[1925] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[1926] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[1927] Train Loss:    tensor(0.6644, device='cuda:0', grad_fn=<NllLossBackward>)
[1928] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[1929] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[1930] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[1931] Train Loss:    tensor(0.7515, device='cuda:0', grad_fn=<NllLossBackward>)
[1932] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[1933] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[1934] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[1935] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[1936] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[1937] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[1938] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[1939] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[1940] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1941] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[1942] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[1943] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[1944] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[1945] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[1946] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[1947] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1948] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[1949] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[1950] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[1951] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[1952] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[1953] Train Loss:    tensor(0.6733, device='cuda:0', grad_fn=<NllLossBackward>)
[1954] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[1955] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[1956] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[1957] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[1958] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[1959] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[1960] Train Loss:    tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[1961] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1962] Train Loss:    tensor(0.6623, device='cuda:0', grad_fn=<NllLossBackward>)
[1963] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[1964] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[1965] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1966] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[1967] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[1968] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[1969] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[1970] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1971] Train Loss:   tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[1972] Train Loss:   tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[1973] Train Loss:   tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[1974] Train Loss:   tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[1975] Train Loss:   tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[1976] Train Loss:   tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[1977] Train Loss:   tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[1978] Train Loss:   tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[1979] Train Loss:   tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[1980] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[1981] Train Loss:   tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[1982] Train Loss:   tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1983] Train Loss:   tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[1984] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[1985] Train Loss:   tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[1986] Train Loss:   tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[1987] Train Loss:   tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[1988] Train Loss:   tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[1989] Train Loss:   tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[1990] Train Loss:   tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[1991] Train Loss:   tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[1992] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[1993] Train Loss:   tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1994] Train Loss:   tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[1995] Train Loss:   tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[1996] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[1997] Train Loss:   tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[1998] Train Loss:   tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[1999] Train Loss:   tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2000] Train Loss:   tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[2001] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[2002] Train Loss:   tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[2003] Train Loss:   tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[2004] Train Loss:   tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[2005] Train Loss:   tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[2006] Train Loss:   tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[2007] Train Loss:   tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[2008] Train Loss:   tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[2009] Train Loss:   tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[2010] Train Loss:   tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[2011] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[2012] Train Loss:   tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[2013] Train Loss:   tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[2014] Train Loss:   tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[2015] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[2016] Train Loss:   tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2017] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2018] Train Loss:   tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2019] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[2020] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[2021] Train Loss:   tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[2022] Train Loss:   tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2023] Train Loss:   tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[2024] Train Loss:   tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[2025] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[2026] Train Loss:   tensor(0.6691, device='cuda:0', grad_fn=<NllLossBackward>)
[2027] Train Loss:   tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[2028] Train Loss:   tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[2029] Train Loss:   tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[2030] Train Loss:   tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[2031] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2032] Train Loss:   tensor(0.7505, device='cuda:0', grad_fn=<NllLossBackward>)
[2033] Train Loss:   tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2034] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2035] Train Loss:   tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[2036] Train Loss:   tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[2037] Train Loss:   tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[2038] Train Loss:   tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[2039] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[2040] Train Loss:   tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[2041] Train Loss:   tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[2042] Train Loss:   tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[2043] Train Loss:   tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[2044] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[2045] Train Loss:   tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[2046] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[2047] Train Loss:   tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[2048] Train Loss:   tensor(0.7507, device='cuda:0', grad_fn=<NllLossBackward>)
[2049] Train Loss:   tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[2050] Train Loss:   tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[2051] Train Loss:   tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[2052] Train Loss:   tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[2053] Train Loss:   tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[2054] Train Loss:   tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[2055] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[2056] Train Loss:   tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[2057] Train Loss:   tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[2058] Train Loss:   tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[2059] Train Loss:   tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[2060] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[2061] Train Loss:   tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[2062] Train Loss:   tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[2063] Train Loss:   tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[2064] Train Loss:   tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[2065] Train Loss:   tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[2066] Train Loss:   tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
```

[2067] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2068] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[2069] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2070] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[2071] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[2072] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[2073] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[2074] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[2075] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[2076] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2077] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[2078] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[2079] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[2080] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2081] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[2082] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[2083] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[2084] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2085] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2086] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[2087] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2088] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[2089] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[2090] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[2091] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2092] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[2093] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[2094] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[2095] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[2096] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[2097] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[2098] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[2099] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[2100] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2101] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[2102] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[2103] Train Loss:    tensor(0.7379, device='cuda:0', grad_fn=<NllLossBackward>)
[2104] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[2105] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[2106] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[2107] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[2108] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[2109] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[2110] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[2111] Train Loss:    tensor(0.6599, device='cuda:0', grad_fn=<NllLossBackward>)
[2112] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[2113] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[2114] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)

```
[2115] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[2116] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[2117] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[2118] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2119] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[2120] Train Loss:    tensor(0.6668, device='cuda:0', grad_fn=<NllLossBackward>)
[2121] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[2122] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[2123] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[2124] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[2125] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2126] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[2127] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[2128] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2129] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[2130] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[2131] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[2132] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[2133] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[2134] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[2135] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[2136] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[2137] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[2138] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[2139] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[2140] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[2141] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[2142] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[2143] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[2144] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[2145] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[2146] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[2147] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[2148] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[2149] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2150] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[2151] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[2152] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[2153] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[2154] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[2155] Train Loss:    tensor(0.7501, device='cuda:0', grad_fn=<NllLossBackward>)
[2156] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[2157] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[2158] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[2159] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[2160] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2161] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[2162] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2163] Train Loss:    tensor(0.6612, device='cuda:0', grad_fn=<NllLossBackward>)
[2164] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[2165] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2166] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[2167] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[2168] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[2169] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[2170] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[2171] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[2172] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[2173] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[2174] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[2175] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[2176] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[2177] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[2178] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[2179] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[2180] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[2181] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[2182] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2183] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[2184] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[2185] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[2186] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[2187] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[2188] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[2189] Train Loss:    tensor(0.6686, device='cuda:0', grad_fn=<NllLossBackward>)
[2190] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[2191] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[2192] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[2193] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[2194] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[2195] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[2196] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[2197] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[2198] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[2199] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[2200] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[2201] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[2202] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[2203] Train Loss:    tensor(0.6608, device='cuda:0', grad_fn=<NllLossBackward>)
[2204] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[2205] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[2206] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[2207] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2208] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[2209] Train Loss:    tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[2210] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2211] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2212] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[2213] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[2214] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[2215] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[2216] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[2217] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[2218] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[2219] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[2220] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[2221] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[2222] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[2223] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2224] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[2225] Train Loss:    tensor(0.7503, device='cuda:0', grad_fn=<NllLossBackward>)
[2226] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[2227] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[2228] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[2229] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[2230] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[2231] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[2232] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[2233] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[2234] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2235] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[2236] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[2237] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[2238] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[2239] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[2240] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[2241] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[2242] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[2243] Train Loss:    tensor(0.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[2244] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[2245] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[2246] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[2247] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[2248] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2249] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[2250] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[2251] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[2252] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[2253] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[2254] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[2255] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[2256] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[2257] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[2258] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2259] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[2260] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[2261] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[2262] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[2263] Train Loss:    tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[2264] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[2265] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[2266] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[2267] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[2268] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[2269] Train Loss:    tensor(0.7558, device='cuda:0', grad_fn=<NllLossBackward>)
[2270] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[2271] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[2272] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[2273] Train Loss:    tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[2274] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[2275] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[2276] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2277] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[2278] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[2279] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[2280] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[2281] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[2282] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[2283] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[2284] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[2285] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[2286] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[2287] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[2288] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[2289] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2290] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[2291] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[2292] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[2293] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[2294] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2295] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2296] Train Loss:    tensor(0.7479, device='cuda:0', grad_fn=<NllLossBackward>)
[2297] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2298] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2299] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[2300] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[2301] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[2302] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[2303] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[2304] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[2305] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[2306] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2307] Train Loss:    tensor(0.6656, device='cuda:0', grad_fn=<NllLossBackward>)
[2308] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[2309] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[2310] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[2311] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[2312] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[2313] Train Loss:    tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[2314] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[2315] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[2316] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[2317] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[2318] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[2319] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2320] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[2321] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2322] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[2323] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[2324] Train Loss:    tensor(0.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[2325] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[2326] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[2327] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[2328] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[2329] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[2330] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[2331] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[2332] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[2333] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[2334] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[2335] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[2336] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2337] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[2338] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[2339] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[2340] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[2341] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[2342] Train Loss:    tensor(0.6688, device='cuda:0', grad_fn=<NllLossBackward>)
[2343] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[2344] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2345] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[2346] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[2347] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[2348] Train Loss:    tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
[2349] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[2350] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[2351] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[2352] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[2353] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[2354] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2355] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[2356] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[2357] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[2358] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[2359] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[2360] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[2361] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[2362] Train Loss:    tensor(0.6694, device='cuda:0', grad_fn=<NllLossBackward>)
[2363] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[2364] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[2365] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[2366] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[2367] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2368] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[2369] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[2370] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2371] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[2372] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[2373] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[2374] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[2375] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[2376] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[2377] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[2378] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[2379] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[2380] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[2381] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2382] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[2383] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[2384] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2385] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[2386] Train Loss:    tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[2387] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[2388] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[2389] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[2390] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2391] Train Loss:    tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[2392] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[2393] Train Loss:    tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[2394] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[2395] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[2396] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2397] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[2398] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[2399] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2400] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[2401] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[2402] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2403] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[2404] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[2405] Train Loss:    tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[2406] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[2407] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[2408] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[2409] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[2410] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[2411] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[2412] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[2413] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[2414] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[2415] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[2416] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[2417] Train Loss:    tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[2418] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[2419] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[2420] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[2421] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2422] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[2423] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[2424] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2425] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[2426] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[2427] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2428] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[2429] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[2430] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[2431] Train Loss:    tensor(0.7421, device='cuda:0', grad_fn=<NllLossBackward>)
[2432] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[2433] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[2434] Train Loss:    tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[2435] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[2436] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[2437] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2438] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[2439] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[2440] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[2441] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[2442] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2443] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2444] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[2445] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[2446] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[2447] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[2448] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[2449] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[2450] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2451] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[2452] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[2453] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[2454] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[2455] Train Loss:    tensor(0.7494, device='cuda:0', grad_fn=<NllLossBackward>)
[2456] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[2457] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[2458] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[2459] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2460] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[2461] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[2462] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2463] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[2464] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[2465] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[2466] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[2467] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[2468] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[2469] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2470] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[2471] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[2472] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[2473] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[2474] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[2475] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[2476] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[2477] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[2478] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[2479] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[2480] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[2481] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[2482] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[2483] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[2484] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[2485] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[2486] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[2487] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[2488] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[2489] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[2490] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[2491] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[2492] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[2493] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[2494] Train Loss:    tensor(0.7384, device='cuda:0', grad_fn=<NllLossBackward>)
[2495] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[2496] Train Loss:    tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[2497] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[2498] Train Loss:    tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2499] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[2500] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[2501] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[2502] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[2503] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[2504] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[2505] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[2506] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[2507] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[2508] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[2509] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[2510] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[2511] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[2512] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[2513] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[2514] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[2515] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[2516] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2517] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[2518] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[2519] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[2520] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[2521] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[2522] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[2523] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[2524] Train Loss:    tensor(0.7537, device='cuda:0', grad_fn=<NllLossBackward>)
[2525] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2526] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2527] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[2528] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[2529] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[2530] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2531] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[2532] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2533] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[2534] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[2535] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2536] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[2537] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[2538] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2539] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[2540] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[2541] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[2542] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2543] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[2544] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[2545] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[2546] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2547] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[2548] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[2549] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[2550] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[2551] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[2552] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[2553] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[2554] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[2555] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[2556] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[2557] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[2558] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[2559] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[2560] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[2561] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[2562] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2563] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2564] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[2565] Train Loss:    tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[2566] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[2567] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[2568] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[2569] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[2570] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[2571] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2572] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[2573] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[2574] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[2575] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[2576] Train Loss:    tensor(0.7407, device='cuda:0', grad_fn=<NllLossBackward>)
[2577] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[2578] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[2579] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[2580] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[2581] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[2582] Train Loss:    tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[2583] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[2584] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[2585] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[2586] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[2587] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2588] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[2589] Train Loss:    tensor(0.7559, device='cuda:0', grad_fn=<NllLossBackward>)
[2590] Train Loss:    tensor(0.6695, device='cuda:0', grad_fn=<NllLossBackward>)
[2591] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[2592] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[2593] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[2594] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2595] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[2596] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[2597] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[2598] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[2599] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[2600] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[2601] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[2602] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[2603] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[2604] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[2605] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2606] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[2607] Train Loss:    tensor(0.7395, device='cuda:0', grad_fn=<NllLossBackward>)
[2608] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[2609] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[2610] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2611] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[2612] Train Loss:    tensor(0.6610, device='cuda:0', grad_fn=<NllLossBackward>)
[2613] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[2614] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[2615] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[2616] Train Loss:    tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[2617] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[2618] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[2619] Train Loss:    tensor(0.7457, device='cuda:0', grad_fn=<NllLossBackward>)
[2620] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[2621] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[2622] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2623] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[2624] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[2625] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[2626] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[2627] Train Loss:    tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[2628] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[2629] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[2630] Train Loss:    tensor(0.6664, device='cuda:0', grad_fn=<NllLossBackward>)
[2631] Train Loss:    tensor(0.6686, device='cuda:0', grad_fn=<NllLossBackward>)
[2632] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[2633] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2634] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2635] Train Loss:    tensor(0.7736, device='cuda:0', grad_fn=<NllLossBackward>)
[2636] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2637] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[2638] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2639] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[2640] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[2641] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[2642] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2643] Train Loss:   tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[2644] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[2645] Train Loss:   tensor(0.7563, device='cuda:0', grad_fn=<NllLossBackward>)
[2646] Train Loss:   tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[2647] Train Loss:   tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[2648] Train Loss:   tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[2649] Train Loss:   tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[2650] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[2651] Train Loss:   tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[2652] Train Loss:   tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[2653] Train Loss:   tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[2654] Train Loss:   tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[2655] Train Loss:   tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[2656] Train Loss:   tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2657] Train Loss:   tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[2658] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[2659] Train Loss:   tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[2660] Train Loss:   tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[2661] Train Loss:   tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2662] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[2663] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2664] Train Loss:   tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[2665] Train Loss:   tensor(0.7622, device='cuda:0', grad_fn=<NllLossBackward>)
[2666] Train Loss:   tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[2667] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[2668] Train Loss:   tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[2669] Train Loss:   tensor(0.7546, device='cuda:0', grad_fn=<NllLossBackward>)
[2670] Train Loss:   tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[2671] Train Loss:   tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[2672] Train Loss:   tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[2673] Train Loss:   tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[2674] Train Loss:   tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[2675] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2676] Train Loss:   tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[2677] Train Loss:   tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[2678] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[2679] Train Loss:   tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[2680] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[2681] Train Loss:   tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[2682] Train Loss:   tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[2683] Train Loss:   tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[2684] Train Loss:   tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2685] Train Loss:   tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[2686] Train Loss:   tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[2687] Train Loss:   tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[2688] Train Loss:   tensor(0.7532, device='cuda:0', grad_fn=<NllLossBackward>)
[2689] Train Loss:   tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[2690] Train Loss:   tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2691] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[2692] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[2693] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[2694] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[2695] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[2696] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[2697] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2698] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2699] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[2700] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[2701] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[2702] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[2703] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[2704] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[2705] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2706] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[2707] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[2708] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[2709] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[2710] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2711] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2712] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[2713] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[2714] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[2715] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[2716] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[2717] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[2718] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[2719] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[2720] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[2721] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[2722] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2723] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[2724] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[2725] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[2726] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[2727] Train Loss:    tensor(0.7589, device='cuda:0', grad_fn=<NllLossBackward>)
[2728] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[2729] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[2730] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[2731] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[2732] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[2733] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[2734] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2735] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[2736] Train Loss:    tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[2737] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[2738] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2739] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[2740] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[2741] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[2742] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[2743] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[2744] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[2745] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[2746] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[2747] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[2748] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[2749] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[2750] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2751] Train Loss:    tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[2752] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[2753] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[2754] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[2755] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[2756] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[2757] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[2758] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[2759] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[2760] Train Loss:    tensor(0.7498, device='cuda:0', grad_fn=<NllLossBackward>)
[2761] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[2762] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[2763] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2764] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[2765] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[2766] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[2767] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[2768] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[2769] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2770] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[2771] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[2772] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[2773] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[2774] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[2775] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[2776] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[2777] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[2778] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[2779] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[2780] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[2781] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2782] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[2783] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[2784] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[2785] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2786] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2787] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[2788] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[2789] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[2790] Train Loss:    tensor(0.7599, device='cuda:0', grad_fn=<NllLossBackward>)
[2791] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[2792] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[2793] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[2794] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[2795] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[2796] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[2797] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2798] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[2799] Train Loss:    tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[2800] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[2801] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[2802] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[2803] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[2804] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[2805] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[2806] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[2807] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[2808] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[2809] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[2810] Train Loss:    tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[2811] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[2812] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[2813] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2814] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2815] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[2816] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[2817] Train Loss:    tensor(0.7597, device='cuda:0', grad_fn=<NllLossBackward>)
[2818] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[2819] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[2820] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[2821] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[2822] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[2823] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[2824] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[2825] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[2826] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[2827] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[2828] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2829] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[2830] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[2831] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[2832] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[2833] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[2834] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2835] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[2836] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[2837] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[2838] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[2839] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[2840] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[2841] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[2842] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[2843] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[2844] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[2845] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2846] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[2847] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[2848] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2849] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[2850] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[2851] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[2852] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[2853] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[2854] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[2855] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[2856] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[2857] Train Loss:    tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[2858] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[2859] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[2860] Train Loss:    tensor(0.7528, device='cuda:0', grad_fn=<NllLossBackward>)
[2861] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[2862] Train Loss:    tensor(0.7444, device='cuda:0', grad_fn=<NllLossBackward>)
[2863] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[2864] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[2865] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[2866] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[2867] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[2868] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[2869] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[2870] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[2871] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[2872] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[2873] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[2874] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[2875] Train Loss:    tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[2876] Train Loss:    tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[2877] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[2878] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[2879] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[2880] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[2881] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[2882] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2883] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[2884] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[2885] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[2886] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[2887] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[2888] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[2889] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[2890] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[2891] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[2892] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[2893] Train Loss:    tensor(0.7551, device='cuda:0', grad_fn=<NllLossBackward>)
[2894] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[2895] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[2896] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[2897] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[2898] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[2899] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[2900] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[2901] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[2902] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[2903] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2904] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2905] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2906] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[2907] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[2908] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[2909] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[2910] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[2911] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[2912] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2913] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2914] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[2915] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[2916] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2917] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[2918] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[2919] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2920] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[2921] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[2922] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[2923] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[2924] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[2925] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[2926] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[2927] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[2928] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[2929] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[2930] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2931] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[2932] Train Loss:    tensor(0.7568, device='cuda:0', grad_fn=<NllLossBackward>)
[2933] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[2934] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[2935] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2936] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[2937] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[2938] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[2939] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[2940] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[2941] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[2942] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[2943] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[2944] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[2945] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[2946] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[2947] Train Loss:    tensor(0.7496, device='cuda:0', grad_fn=<NllLossBackward>)
[2948] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[2949] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[2950] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[2951] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2952] Train Loss:    tensor(0.6606, device='cuda:0', grad_fn=<NllLossBackward>)
[2953] Train Loss:    tensor(0.6622, device='cuda:0', grad_fn=<NllLossBackward>)
[2954] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[2955] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2956] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[2957] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[2958] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[2959] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2960] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[2961] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[2962] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[2963] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[2964] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[2965] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[2966] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[2967] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[2968] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[2969] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[2970] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[2971] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[2972] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[2973] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[2974] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[2975] Train Loss:    tensor(0.6729, device='cuda:0', grad_fn=<NllLossBackward>)
[2976] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[2977] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[2978] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2979] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[2980] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[2981] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[2982] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[2983] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[2984] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[2985] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[2986] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[2987] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[2988] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[2989] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[2990] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[2991] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[2992] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[2993] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[2994] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[2995] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[2996] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[2997] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2998] Train Loss:    tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[2999] Train Loss:    tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[3000] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[3001] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[3002] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[3003] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[3004] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[3005] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[3006] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[3007] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[3008] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[3009] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[3010] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[3011] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[3012] Train Loss:    tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[3013] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[3014] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[3015] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[3016] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[3017] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[3018] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[3019] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[3020] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[3021] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[3022] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[3023] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[3024] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[3025] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[3026] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3027] Train Loss:  tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[3028] Train Loss:  tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3029] Train Loss:  tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[3030] Train Loss:  tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[3031] Train Loss:  tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[3032] Train Loss:  tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[3033] Train Loss:  tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3034] Train Loss:  tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[3035] Train Loss:  tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[3036] Train Loss:  tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[3037] Train Loss:  tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[3038] Train Loss:  tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[3039] Train Loss:  tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[3040] Train Loss:  tensor(0.6701, device='cuda:0', grad_fn=<NllLossBackward>)
[3041] Train Loss:  tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[3042] Train Loss:  tensor(0.6639, device='cuda:0', grad_fn=<NllLossBackward>)
[3043] Train Loss:  tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[3044] Train Loss:  tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[3045] Train Loss:  tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[3046] Train Loss:  tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[3047] Train Loss:  tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[3048] Train Loss:  tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[3049] Train Loss:  tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[3050] Train Loss:  tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[3051] Train Loss:  tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[3052] Train Loss:  tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3053] Train Loss:  tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[3054] Train Loss:  tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[3055] Train Loss:  tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[3056] Train Loss:  tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3057] Train Loss:  tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[3058] Train Loss:  tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[3059] Train Loss:  tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[3060] Train Loss:  tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3061] Train Loss:  tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[3062] Train Loss:  tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[3063] Train Loss:  tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3064] Train Loss:  tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[3065] Train Loss:  tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3066] Train Loss:  tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[3067] Train Loss:  tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[3068] Train Loss:  tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[3069] Train Loss:  tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[3070] Train Loss:  tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[3071] Train Loss:  tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[3072] Train Loss:  tensor(0.7609, device='cuda:0', grad_fn=<NllLossBackward>)
[3073] Train Loss:  tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[3074] Train Loss:  tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3075] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3076] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[3077] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[3078] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3079] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[3080] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[3081] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[3082] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[3083] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[3084] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[3085] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3086] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[3087] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[3088] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[3089] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[3090] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[3091] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[3092] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[3093] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[3094] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[3095] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[3096] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[3097] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[3098] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3099] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[3100] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[3101] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[3102] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[3103] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[3104] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3105] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[3106] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[3107] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[3108] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[3109] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[3110] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[3111] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[3112] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[3113] Train Loss:    tensor(0.7428, device='cuda:0', grad_fn=<NllLossBackward>)
[3114] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3115] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[3116] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[3117] Train Loss:    tensor(0.6642, device='cuda:0', grad_fn=<NllLossBackward>)
[3118] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[3119] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[3120] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[3121] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3122] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3123] Train Loss:   tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[3124] Train Loss:   tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[3125] Train Loss:   tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[3126] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[3127] Train Loss:   tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[3128] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3129] Train Loss:   tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[3130] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3131] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3132] Train Loss:   tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[3133] Train Loss:   tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[3134] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[3135] Train Loss:   tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3136] Train Loss:   tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[3137] Train Loss:   tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[3138] Train Loss:   tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[3139] Train Loss:   tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3140] Train Loss:   tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[3141] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3142] Train Loss:   tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[3143] Train Loss:   tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[3144] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[3145] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[3146] Train Loss:   tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[3147] Train Loss:   tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[3148] Train Loss:   tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3149] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[3150] Train Loss:   tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[3151] Train Loss:   tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[3152] Train Loss:   tensor(0.7435, device='cuda:0', grad_fn=<NllLossBackward>)
[3153] Train Loss:   tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[3154] Train Loss:   tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[3155] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[3156] Train Loss:   tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[3157] Train Loss:   tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[3158] Train Loss:   tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3159] Train Loss:   tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[3160] Train Loss:   tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[3161] Train Loss:   tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[3162] Train Loss:   tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[3163] Train Loss:   tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3164] Train Loss:   tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[3165] Train Loss:   tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[3166] Train Loss:   tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[3167] Train Loss:   tensor(0.6615, device='cuda:0', grad_fn=<NllLossBackward>)
[3168] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3169] Train Loss:   tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3170] Train Loss:   tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3171] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[3172] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[3173] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[3174] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[3175] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3176] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[3177] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3178] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[3179] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[3180] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[3181] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[3182] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[3183] Train Loss:    tensor(0.7514, device='cuda:0', grad_fn=<NllLossBackward>)
[3184] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[3185] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[3186] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[3187] Train Loss:    tensor(0.7398, device='cuda:0', grad_fn=<NllLossBackward>)
[3188] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[3189] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[3190] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3191] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[3192] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[3193] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[3194] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[3195] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[3196] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[3197] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[3198] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[3199] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[3200] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[3201] Train Loss:    tensor(0.7518, device='cuda:0', grad_fn=<NllLossBackward>)
[3202] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[3203] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[3204] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[3205] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[3206] Train Loss:    tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[3207] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3208] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[3209] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[3210] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[3211] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[3212] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3213] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[3214] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[3215] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3216] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[3217] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[3218] Train Loss:    tensor(0.7574, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3219] Train Loss:   tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[3220] Train Loss:   tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[3221] Train Loss:   tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[3222] Train Loss:   tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3223] Train Loss:   tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[3224] Train Loss:   tensor(0.7309, device='cuda:0', grad_fn=<NllLossBackward>)
[3225] Train Loss:   tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[3226] Train Loss:   tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[3227] Train Loss:   tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[3228] Train Loss:   tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[3229] Train Loss:   tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[3230] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[3231] Train Loss:   tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3232] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3233] Train Loss:   tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[3234] Train Loss:   tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[3235] Train Loss:   tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[3236] Train Loss:   tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[3237] Train Loss:   tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[3238] Train Loss:   tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[3239] Train Loss:   tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[3240] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[3241] Train Loss:   tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[3242] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[3243] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[3244] Train Loss:   tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3245] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3246] Train Loss:   tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[3247] Train Loss:   tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[3248] Train Loss:   tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[3249] Train Loss:   tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[3250] Train Loss:   tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[3251] Train Loss:   tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[3252] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[3253] Train Loss:   tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[3254] Train Loss:   tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[3255] Train Loss:   tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[3256] Train Loss:   tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[3257] Train Loss:   tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[3258] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3259] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[3260] Train Loss:   tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[3261] Train Loss:   tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[3262] Train Loss:   tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[3263] Train Loss:   tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3264] Train Loss:   tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[3265] Train Loss:   tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[3266] Train Loss:   tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3267] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[3268] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[3269] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[3270] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3271] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[3272] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[3273] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[3274] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[3275] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[3276] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3277] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[3278] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[3279] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3280] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[3281] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[3282] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[3283] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[3284] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[3285] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[3286] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[3287] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3288] Train Loss:    tensor(0.7517, device='cuda:0', grad_fn=<NllLossBackward>)
[3289] Train Loss:    tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[3290] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[3291] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[3292] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[3293] Train Loss:    tensor(0.7393, device='cuda:0', grad_fn=<NllLossBackward>)
[3294] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[3295] Train Loss:    tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[3296] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[3297] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3298] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[3299] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[3300] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3301] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[3302] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[3303] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[3304] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[3305] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3306] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3307] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[3308] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3309] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[3310] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3311] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[3312] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[3313] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[3314] Train Loss:    tensor(0.7455, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3315] Train Loss:   tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[3316] Train Loss:   tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3317] Train Loss:   tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[3318] Train Loss:   tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[3319] Train Loss:   tensor(0.7639, device='cuda:0', grad_fn=<NllLossBackward>)
[3320] Train Loss:   tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[3321] Train Loss:   tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[3322] Train Loss:   tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[3323] Train Loss:   tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[3324] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3325] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[3326] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[3327] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3328] Train Loss:   tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[3329] Train Loss:   tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[3330] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[3331] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[3332] Train Loss:   tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[3333] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[3334] Train Loss:   tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[3335] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[3336] Train Loss:   tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[3337] Train Loss:   tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[3338] Train Loss:   tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[3339] Train Loss:   tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[3340] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[3341] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[3342] Train Loss:   tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3343] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[3344] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[3345] Train Loss:   tensor(0.7636, device='cuda:0', grad_fn=<NllLossBackward>)
[3346] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[3347] Train Loss:   tensor(0.6468, device='cuda:0', grad_fn=<NllLossBackward>)
[3348] Train Loss:   tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[3349] Train Loss:   tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[3350] Train Loss:   tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[3351] Train Loss:   tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3352] Train Loss:   tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[3353] Train Loss:   tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[3354] Train Loss:   tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[3355] Train Loss:   tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[3356] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3357] Train Loss:   tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[3358] Train Loss:   tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[3359] Train Loss:   tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3360] Train Loss:   tensor(0.7721, device='cuda:0', grad_fn=<NllLossBackward>)
[3361] Train Loss:   tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3362] Train Loss:   tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3363] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[3364] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[3365] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[3366] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[3367] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[3368] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[3369] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[3370] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[3371] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3372] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[3373] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[3374] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[3375] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[3376] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[3377] Train Loss:    tensor(0.6712, device='cuda:0', grad_fn=<NllLossBackward>)
[3378] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[3379] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[3380] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[3381] Train Loss:    tensor(0.7425, device='cuda:0', grad_fn=<NllLossBackward>)
[3382] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3383] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[3384] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[3385] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
[3386] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[3387] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3388] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[3389] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3390] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[3391] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[3392] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3393] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[3394] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[3395] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[3396] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[3397] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[3398] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[3399] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[3400] Train Loss:    tensor(0.7538, device='cuda:0', grad_fn=<NllLossBackward>)
[3401] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3402] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3403] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[3404] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[3405] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3406] Train Loss:    tensor(0.6668, device='cuda:0', grad_fn=<NllLossBackward>)
[3407] Train Loss:    tensor(0.6701, device='cuda:0', grad_fn=<NllLossBackward>)
[3408] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[3409] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[3410] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3411] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[3412] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[3413] Train Loss:    tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[3414] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[3415] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[3416] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[3417] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[3418] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[3419] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[3420] Train Loss:    tensor(0.7445, device='cuda:0', grad_fn=<NllLossBackward>)
[3421] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[3422] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3423] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[3424] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[3425] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[3426] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[3427] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[3428] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[3429] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3430] Train Loss:    tensor(0.7566, device='cuda:0', grad_fn=<NllLossBackward>)
[3431] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[3432] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[3433] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3434] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[3435] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3436] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[3437] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[3438] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[3439] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[3440] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[3441] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[3442] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[3443] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[3444] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[3445] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[3446] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[3447] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3448] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[3449] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[3450] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3451] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[3452] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[3453] Train Loss:    tensor(0.7426, device='cuda:0', grad_fn=<NllLossBackward>)
[3454] Train Loss:    tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[3455] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[3456] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[3457] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[3458] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3459] Train Loss:   tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3460] Train Loss:   tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[3461] Train Loss:   tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[3462] Train Loss:   tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[3463] Train Loss:   tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[3464] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[3465] Train Loss:   tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[3466] Train Loss:   tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3467] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[3468] Train Loss:   tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[3469] Train Loss:   tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[3470] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[3471] Train Loss:   tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[3472] Train Loss:   tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[3473] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[3474] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[3475] Train Loss:   tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[3476] Train Loss:   tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[3477] Train Loss:   tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[3478] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[3479] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[3480] Train Loss:   tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[3481] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[3482] Train Loss:   tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[3483] Train Loss:   tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[3484] Train Loss:   tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[3485] Train Loss:   tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[3486] Train Loss:   tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[3487] Train Loss:   tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[3488] Train Loss:   tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[3489] Train Loss:   tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[3490] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3491] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[3492] Train Loss:   tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[3493] Train Loss:   tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[3494] Train Loss:   tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3495] Train Loss:   tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3496] Train Loss:   tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[3497] Train Loss:   tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[3498] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[3499] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[3500] Train Loss:   tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[3501] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[3502] Train Loss:   tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[3503] Train Loss:   tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[3504] Train Loss:   tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[3505] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[3506] Train Loss:   tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3507] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[3508] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[3509] Train Loss:    tensor(0.7476, device='cuda:0', grad_fn=<NllLossBackward>)
[3510] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[3511] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[3512] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3513] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[3514] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[3515] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3516] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[3517] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[3518] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[3519] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[3520] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[3521] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[3522] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[3523] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[3524] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[3525] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[3526] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[3527] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[3528] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[3529] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[3530] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[3531] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3532] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[3533] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[3534] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3535] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[3536] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[3537] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[3538] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[3539] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[3540] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[3541] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[3542] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[3543] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3544] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[3545] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[3546] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[3547] Train Loss:    tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
[3548] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[3549] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[3550] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[3551] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3552] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[3553] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[3554] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3555] Train Loss:   tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[3556] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[3557] Train Loss:   tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[3558] Train Loss:   tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[3559] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[3560] Train Loss:   tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[3561] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3562] Train Loss:   tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3563] Train Loss:   tensor(0.7509, device='cuda:0', grad_fn=<NllLossBackward>)
[3564] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[3565] Train Loss:   tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[3566] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[3567] Train Loss:   tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[3568] Train Loss:   tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[3569] Train Loss:   tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[3570] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[3571] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3572] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3573] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[3574] Train Loss:   tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[3575] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[3576] Train Loss:   tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[3577] Train Loss:   tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[3578] Train Loss:   tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[3579] Train Loss:   tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[3580] Train Loss:   tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[3581] Train Loss:   tensor(0.7521, device='cuda:0', grad_fn=<NllLossBackward>)
[3582] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3583] Train Loss:   tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[3584] Train Loss:   tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[3585] Train Loss:   tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[3586] Train Loss:   tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[3587] Train Loss:   tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[3588] Train Loss:   tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3589] Train Loss:   tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[3590] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[3591] Train Loss:   tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[3592] Train Loss:   tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[3593] Train Loss:   tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[3594] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[3595] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[3596] Train Loss:   tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3597] Train Loss:   tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[3598] Train Loss:   tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[3599] Train Loss:   tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3600] Train Loss:   tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3601] Train Loss:   tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[3602] Train Loss:   tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3603] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[3604] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[3605] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3606] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[3607] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[3608] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[3609] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[3610] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[3611] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[3612] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[3613] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[3614] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[3615] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3616] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[3617] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[3618] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[3619] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3620] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[3621] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[3622] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[3623] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3624] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[3625] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[3626] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[3627] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[3628] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[3629] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[3630] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3631] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[3632] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[3633] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[3634] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3635] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[3636] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[3637] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[3638] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[3639] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[3640] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[3641] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3642] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[3643] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[3644] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[3645] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[3646] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[3647] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[3648] Train Loss:    tensor(0.7399, device='cuda:0', grad_fn=<NllLossBackward>)
[3649] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3650] Train Loss:    tensor(0.7415, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3651] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[3652] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[3653] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[3654] Train Loss:    tensor(0.7433, device='cuda:0', grad_fn=<NllLossBackward>)
[3655] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[3656] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[3657] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[3658] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[3659] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3660] Train Loss:    tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[3661] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[3662] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[3663] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3664] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[3665] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[3666] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[3667] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[3668] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[3669] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3670] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[3671] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[3672] Train Loss:    tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[3673] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[3674] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[3675] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[3676] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[3677] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3678] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[3679] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[3680] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[3681] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[3682] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[3683] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[3684] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[3685] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[3686] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[3687] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[3688] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[3689] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[3690] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[3691] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[3692] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[3693] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[3694] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[3695] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[3696] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[3697] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[3698] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3699] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[3700] Train Loss:   tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3701] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3702] Train Loss:   tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[3703] Train Loss:   tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[3704] Train Loss:   tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[3705] Train Loss:   tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[3706] Train Loss:   tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[3707] Train Loss:   tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3708] Train Loss:   tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[3709] Train Loss:   tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[3710] Train Loss:   tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[3711] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3712] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3713] Train Loss:   tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[3714] Train Loss:   tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[3715] Train Loss:   tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[3716] Train Loss:   tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[3717] Train Loss:   tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[3718] Train Loss:   tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[3719] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[3720] Train Loss:   tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[3721] Train Loss:   tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[3722] Train Loss:   tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[3723] Train Loss:   tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[3724] Train Loss:   tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[3725] Train Loss:   tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3726] Train Loss:   tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[3727] Train Loss:   tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[3728] Train Loss:   tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[3729] Train Loss:   tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[3730] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[3731] Train Loss:   tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[3732] Train Loss:   tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[3733] Train Loss:   tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[3734] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[3735] Train Loss:   tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[3736] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[3737] Train Loss:   tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[3738] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[3739] Train Loss:   tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[3740] Train Loss:   tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[3741] Train Loss:   tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3742] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3743] Train Loss:   tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[3744] Train Loss:   tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[3745] Train Loss:   tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[3746] Train Loss:   tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3747] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[3748] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3749] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[3750] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[3751] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[3752] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[3753] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[3754] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[3755] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[3756] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[3757] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[3758] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[3759] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[3760] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[3761] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[3762] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[3763] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[3764] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[3765] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[3766] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[3767] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[3768] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[3769] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3770] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[3771] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[3772] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[3773] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[3774] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3775] Train Loss:    tensor(0.6505, device='cuda:0', grad_fn=<NllLossBackward>)
[3776] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[3777] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[3778] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[3779] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[3780] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[3781] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[3782] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[3783] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[3784] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[3785] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[3786] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[3787] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[3788] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[3789] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3790] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[3791] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[3792] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3793] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[3794] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3795] Train Loss:   tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[3796] Train Loss:   tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[3797] Train Loss:   tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[3798] Train Loss:   tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[3799] Train Loss:   tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[3800] Train Loss:   tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[3801] Train Loss:   tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[3802] Train Loss:   tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[3803] Train Loss:   tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[3804] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[3805] Train Loss:   tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[3806] Train Loss:   tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[3807] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3808] Train Loss:   tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[3809] Train Loss:   tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[3810] Train Loss:   tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[3811] Train Loss:   tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[3812] Train Loss:   tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[3813] Train Loss:   tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3814] Train Loss:   tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[3815] Train Loss:   tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[3816] Train Loss:   tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[3817] Train Loss:   tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[3818] Train Loss:   tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[3819] Train Loss:   tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[3820] Train Loss:   tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[3821] Train Loss:   tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[3822] Train Loss:   tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[3823] Train Loss:   tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[3824] Train Loss:   tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[3825] Train Loss:   tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[3826] Train Loss:   tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3827] Train Loss:   tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[3828] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3829] Train Loss:   tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[3830] Train Loss:   tensor(0.6590, device='cuda:0', grad_fn=<NllLossBackward>)
[3831] Train Loss:   tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[3832] Train Loss:   tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3833] Train Loss:   tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[3834] Train Loss:   tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[3835] Train Loss:   tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[3836] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[3837] Train Loss:   tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[3838] Train Loss:   tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[3839] Train Loss:   tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[3840] Train Loss:   tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[3841] Train Loss:   tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[3842] Train Loss:   tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3843] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[3844] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[3845] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[3846] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[3847] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3848] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3849] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[3850] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[3851] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[3852] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[3853] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[3854] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[3855] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[3856] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[3857] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3858] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[3859] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[3860] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3861] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[3862] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[3863] Train Loss:    tensor(0.7547, device='cuda:0', grad_fn=<NllLossBackward>)
[3864] Train Loss:    tensor(0.7387, device='cuda:0', grad_fn=<NllLossBackward>)
[3865] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[3866] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[3867] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3868] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3869] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[3870] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3871] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[3872] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[3873] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[3874] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[3875] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3876] Train Loss:    tensor(0.7586, device='cuda:0', grad_fn=<NllLossBackward>)
[3877] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[3878] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[3879] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[3880] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3881] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[3882] Train Loss:    tensor(0.7475, device='cuda:0', grad_fn=<NllLossBackward>)
[3883] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[3884] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[3885] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[3886] Train Loss:    tensor(0.7604, device='cuda:0', grad_fn=<NllLossBackward>)
[3887] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[3888] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[3889] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[3890] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3891] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[3892] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3893] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[3894] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[3895] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[3896] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3897] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[3898] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[3899] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[3900] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[3901] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[3902] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[3903] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[3904] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[3905] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[3906] Train Loss:    tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[3907] Train Loss:    tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[3908] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[3909] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[3910] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[3911] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[3912] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[3913] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[3914] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[3915] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[3916] Train Loss:    tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[3917] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[3918] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[3919] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[3920] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3921] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[3922] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[3923] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[3924] Train Loss:    tensor(0.7405, device='cuda:0', grad_fn=<NllLossBackward>)
[3925] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[3926] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[3927] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[3928] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[3929] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[3930] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[3931] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[3932] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3933] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[3934] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[3935] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[3936] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[3937] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3938] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3939] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[3940] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3941] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[3942] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[3943] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[3944] Train Loss:    tensor(0.6646, device='cuda:0', grad_fn=<NllLossBackward>)
[3945] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[3946] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[3947] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[3948] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[3949] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[3950] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3951] Train Loss:    tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[3952] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[3953] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[3954] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[3955] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[3956] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[3957] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[3958] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[3959] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3960] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[3961] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[3962] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[3963] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[3964] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3965] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[3966] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[3967] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[3968] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[3969] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[3970] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[3971] Train Loss:    tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[3972] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[3973] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[3974] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[3975] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[3976] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[3977] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[3978] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3979] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[3980] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3981] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[3982] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[3983] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[3984] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3985] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[3986] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3987] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[3988] Train Loss:    tensor(0.7424, device='cuda:0', grad_fn=<NllLossBackward>)
[3989] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[3990] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[3991] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[3992] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[3993] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3994] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[3995] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[3996] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[3997] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3998] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3999] Train Loss:    tensor(0.7436, device='cuda:0', grad_fn=<NllLossBackward>)
[4000] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[4001] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[4002] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[4003] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[4004] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[4005] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[4006] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[4007] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[4008] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[4009] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[4010] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[4011] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[4012] Train Loss:    tensor(0.7531, device='cuda:0', grad_fn=<NllLossBackward>)
[4013] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[4014] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[4015] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[4016] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[4017] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[4018] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[4019] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[4020] Train Loss:    tensor(0.6627, device='cuda:0', grad_fn=<NllLossBackward>)
[4021] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[4022] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[4023] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[4024] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[4025] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[4026] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[4027] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[4028] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[4029] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[4030] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[4031] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[4032] Train Loss:    tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[4033] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[4034] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4035] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[4036] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[4037] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[4038] Train Loss:    tensor(0.7453, device='cuda:0', grad_fn=<NllLossBackward>)
[4039] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[4040] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[4041] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[4042] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[4043] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[4044] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[4045] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[4046] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[4047] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[4048] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[4049] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[4050] Train Loss:    tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[4051] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[4052] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[4053] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[4054] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[4055] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[4056] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[4057] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[4058] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[4059] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[4060] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[4061] Train Loss:    tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[4062] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[4063] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[4064] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[4065] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[4066] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[4067] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[4068] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[4069] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[4070] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[4071] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[4072] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[4073] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[4074] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[4075] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[4076] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[4077] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[4078] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[4079] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[4080] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[4081] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[4082] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4083] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[4084] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[4085] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[4086] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[4087] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[4088] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[4089] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[4090] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[4091] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[4092] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[4093] Train Loss:    tensor(0.7466, device='cuda:0', grad_fn=<NllLossBackward>)
[4094] Train Loss:    tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[4095] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[4096] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[4097] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[4098] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[4099] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[4100] Train Loss:    tensor(0.7524, device='cuda:0', grad_fn=<NllLossBackward>)
[4101] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[4102] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[4103] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[4104] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[4105] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[4106] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[4107] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[4108] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[4109] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[4110] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[4111] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[4112] Train Loss:    tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[4113] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[4114] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[4115] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[4116] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[4117] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[4118] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[4119] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[4120] Train Loss:    tensor(0.6658, device='cuda:0', grad_fn=<NllLossBackward>)
[4121] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[4122] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[4123] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[4124] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[4125] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[4126] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[4127] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[4128] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[4129] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[4130] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4131] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[4132] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[4133] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[4134] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[4135] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[4136] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[4137] Train Loss:    tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[4138] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[4139] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[4140] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[4141] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[4142] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[4143] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[4144] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[4145] Train Loss:    tensor(0.7666, device='cuda:0', grad_fn=<NllLossBackward>)
[4146] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[4147] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[4148] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[4149] Train Loss:    tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[4150] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[4151] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[4152] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[4153] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[4154] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[4155] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[4156] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[4157] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[4158] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[4159] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[4160] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[4161] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[4162] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[4163] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[4164] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[4165] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[4166] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[4167] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[4168] Train Loss:    tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[4169] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[4170] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[4171] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[4172] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[4173] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[4174] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[4175] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[4176] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[4177] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[4178] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4179] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[4180] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[4181] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[4182] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[4183] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[4184] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[4185] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[4186] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[4187] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[4188] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[4189] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[4190] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[4191] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[4192] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[4193] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[4194] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[4195] Train Loss:    tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[4196] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[4197] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[4198] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[4199] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[4200] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[4201] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[4202] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[4203] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[4204] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[4205] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[4206] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[4207] Train Loss:    tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[4208] Train Loss:    tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[4209] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[4210] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[4211] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[4212] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[4213] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[4214] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[4215] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[4216] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[4217] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[4218] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[4219] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[4220] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[4221] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[4222] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[4223] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[4224] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[4225] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[4226] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4227] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[4228] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[4229] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[4230] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[4231] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[4232] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[4233] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[4234] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[4235] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[4236] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[4237] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[4238] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[4239] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[4240] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[4241] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[4242] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[4243] Train Loss:    tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[4244] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[4245] Train Loss:    tensor(0.7450, device='cuda:0', grad_fn=<NllLossBackward>)
[4246] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[4247] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[4248] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[4249] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[4250] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[4251] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[4252] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[4253] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[4254] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[4255] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[4256] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[4257] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[4258] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[4259] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[4260] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[4261] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[4262] Train Loss:    tensor(0.6381, device='cuda:0', grad_fn=<NllLossBackward>)
[4263] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[4264] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[4265] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[4266] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[4267] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[4268] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[4269] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[4270] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[4271] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[4272] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[4273] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[4274] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4275] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[4276] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[4277] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[4278] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[4279] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[4280] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[4281] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[4282] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[4283] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[4284] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[4285] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[4286] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[4287] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[4288] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[4289] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[4290] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[4291] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[4292] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[4293] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[4294] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[4295] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[4296] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[4297] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[4298] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[4299] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[4300] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[4301] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[4302] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[4303] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[4304] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[4305] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[4306] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[4307] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[4308] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[4309] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[4310] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[4311] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[4312] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[4313] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[4314] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[4315] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[4316] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[4317] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[4318] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[4319] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[4320] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[4321] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[4322] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4323] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[4324] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[4325] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[4326] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[4327] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[4328] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[4329] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[4330] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[4331] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[4332] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[4333] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[4334] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[4335] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[4336] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[4337] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[4338] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[4339] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[4340] Train Loss:    tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[4341] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[4342] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[4343] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[4344] Train Loss:    tensor(0.6712, device='cuda:0', grad_fn=<NllLossBackward>)
[4345] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[4346] Train Loss:    tensor(0.7447, device='cuda:0', grad_fn=<NllLossBackward>)
[4347] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[4348] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[4349] Train Loss:    tensor(0.6706, device='cuda:0', grad_fn=<NllLossBackward>)
[4350] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[4351] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[4352] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[4353] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[4354] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[4355] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[4356] Train Loss:    tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[4357] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[4358] Train Loss:    tensor(0.7383, device='cuda:0', grad_fn=<NllLossBackward>)
[4359] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[4360] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[4361] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[4362] Train Loss:    tensor(0.7581, device='cuda:0', grad_fn=<NllLossBackward>)
[4363] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[4364] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[4365] Train Loss:    tensor(0.7461, device='cuda:0', grad_fn=<NllLossBackward>)
[4366] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[4367] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[4368] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[4369] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[4370] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4371] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[4372] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[4373] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[4374] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[4375] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[4376] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[4377] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[4378] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[4379] Train Loss:    tensor(0.7420, device='cuda:0', grad_fn=<NllLossBackward>)
[4380] Train Loss:    tensor(0.6647, device='cuda:0', grad_fn=<NllLossBackward>)
[4381] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[4382] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[4383] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[4384] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[4385] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[4386] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[4387] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[4388] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[4389] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[4390] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[4391] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[4392] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[4393] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[4394] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[4395] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[4396] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[4397] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[4398] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[4399] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[4400] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[4401] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[4402] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[4403] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[4404] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[4405] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[4406] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[4407] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[4408] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[4409] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[4410] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[4411] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[4412] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[4413] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[4414] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[4415] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[4416] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[4417] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[4418] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4419] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[4420] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[4421] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[4422] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[4423] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[4424] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[4425] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[4426] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[4427] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[4428] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[4429] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[4430] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[4431] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[4432] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[4433] Train Loss:    tensor(0.7565, device='cuda:0', grad_fn=<NllLossBackward>)
[4434] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[4435] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[4436] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[4437] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[4438] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[4439] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[4440] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[4441] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[4442] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[4443] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[4444] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[4445] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[4446] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[4447] Train Loss:    tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[4448] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[4449] Train Loss:    tensor(0.7385, device='cuda:0', grad_fn=<NllLossBackward>)
[4450] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[4451] Train Loss:    tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[4452] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[4453] Train Loss:    tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[4454] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[4455] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[4456] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[4457] Train Loss:    tensor(0.6671, device='cuda:0', grad_fn=<NllLossBackward>)
[4458] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[4459] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[4460] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[4461] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[4462] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[4463] Train Loss:    tensor(0.6564, device='cuda:0', grad_fn=<NllLossBackward>)
[4464] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[4465] Train Loss:    tensor(0.7406, device='cuda:0', grad_fn=<NllLossBackward>)
[4466] Train Loss:    tensor(0.7570, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4467] Train Loss:  tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[4468] Train Loss:  tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[4469] Train Loss:  tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[4470] Train Loss:  tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[4471] Train Loss:  tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[4472] Train Loss:  tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[4473] Train Loss:  tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[4474] Train Loss:  tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[4475] Train Loss:  tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[4476] Train Loss:  tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[4477] Train Loss:  tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[4478] Train Loss:  tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[4479] Train Loss:  tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[4480] Train Loss:  tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[4481] Train Loss:  tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[4482] Train Loss:  tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[4483] Train Loss:  tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[4484] Train Loss:  tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
[4485] Train Loss:  tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[4486] Train Loss:  tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[4487] Train Loss:  tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[4488] Train Loss:  tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[4489] Train Loss:  tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[4490] Train Loss:  tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[4491] Train Loss:  tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[4492] Train Loss:  tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[4493] Train Loss:  tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[4494] Train Loss:  tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[4495] Train Loss:  tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[4496] Train Loss:  tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[4497] Train Loss:  tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[4498] Train Loss:  tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[4499] Train Loss:  tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[4500] Train Loss:  tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
[4501] Train Loss:  tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[4502] Train Loss:  tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[4503] Train Loss:  tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[4504] Train Loss:  tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[4505] Train Loss:  tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[4506] Train Loss:  tensor(0.7469, device='cuda:0', grad_fn=<NllLossBackward>)
[4507] Train Loss:  tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[4508] Train Loss:  tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[4509] Train Loss:  tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[4510] Train Loss:  tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[4511] Train Loss:  tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
Training Loss:  0.7097978640664765 Time:  2506.319940328598 s
Training Accuaracy:  78.11539562035959 %
[0] Valid Loss:  tensor(0.7129, device='cuda:0')
```

```
[1] Valid Loss:  tensor(0.8820, device='cuda:0')
[2] Valid Loss:  tensor(0.6796, device='cuda:0')
[3] Valid Loss:  tensor(0.6170, device='cuda:0')
[4] Valid Loss:  tensor(0.7084, device='cuda:0')
[5] Valid Loss:  tensor(0.7181, device='cuda:0')
[6] Valid Loss:  tensor(0.7258, device='cuda:0')
[7] Valid Loss:  tensor(0.8595, device='cuda:0')
[8] Valid Loss:  tensor(0.8097, device='cuda:0')
[9] Valid Loss:  tensor(0.8145, device='cuda:0')
[10] Valid Loss:  tensor(0.7066, device='cuda:0')
[11] Valid Loss:  tensor(0.6792, device='cuda:0')
[12] Valid Loss:  tensor(1.1323, device='cuda:0')
[13] Valid Loss:  tensor(0.9650, device='cuda:0')
[14] Valid Loss:  tensor(0.6216, device='cuda:0')
[15] Valid Loss:  tensor(0.7410, device='cuda:0')
[16] Valid Loss:  tensor(0.6435, device='cuda:0')
[17] Valid Loss:  tensor(0.7944, device='cuda:0')
[18] Valid Loss:  tensor(0.6849, device='cuda:0')
[19] Valid Loss:  tensor(0.6277, device='cuda:0')
[20] Valid Loss:  tensor(0.6910, device='cuda:0')
[21] Valid Loss:  tensor(0.5895, device='cuda:0')
[22] Valid Loss:  tensor(0.5716, device='cuda:0')
[23] Valid Loss:  tensor(0.8225, device='cuda:0')
[24] Valid Loss:  tensor(0.7097, device='cuda:0')
[25] Valid Loss:  tensor(0.9711, device='cuda:0')
[26] Valid Loss:  tensor(0.5523, device='cuda:0')
[27] Valid Loss:  tensor(0.8937, device='cuda:0')
[28] Valid Loss:  tensor(0.6630, device='cuda:0')
[29] Valid Loss:  tensor(0.6907, device='cuda:0')
[30] Valid Loss:  tensor(0.6360, device='cuda:0')
[31] Valid Loss:  tensor(0.7194, device='cuda:0')
[32] Valid Loss:  tensor(0.8404, device='cuda:0')
[33] Valid Loss:  tensor(0.7950, device='cuda:0')
[34] Valid Loss:  tensor(0.7794, device='cuda:0')
[35] Valid Loss:  tensor(0.8584, device='cuda:0')
[36] Valid Loss:  tensor(0.6737, device='cuda:0')
[37] Valid Loss:  tensor(1.0251, device='cuda:0')
[38] Valid Loss:  tensor(0.8117, device='cuda:0')
[39] Valid Loss:  tensor(0.6633, device='cuda:0')
[40] Valid Loss:  tensor(0.7770, device='cuda:0')
[41] Valid Loss:  tensor(0.8671, device='cuda:0')
[42] Valid Loss:  tensor(0.7810, device='cuda:0')
[43] Valid Loss:  tensor(0.5690, device='cuda:0')
[44] Valid Loss:  tensor(0.6839, device='cuda:0')
[45] Valid Loss:  tensor(0.5821, device='cuda:0')
[46] Valid Loss:  tensor(0.8858, device='cuda:0')
[47] Valid Loss:  tensor(0.8197, device='cuda:0')
[48] Valid Loss:  tensor(0.6875, device='cuda:0')
```

```
[49] Valid Loss:   tensor(0.7628, device='cuda:0')
[50] Valid Loss:   tensor(0.6941, device='cuda:0')
[51] Valid Loss:   tensor(0.6597, device='cuda:0')
[52] Valid Loss:   tensor(0.5947, device='cuda:0')
[53] Valid Loss:   tensor(0.9182, device='cuda:0')
[54] Valid Loss:   tensor(0.6119, device='cuda:0')
[55] Valid Loss:   tensor(0.5576, device='cuda:0')
[56] Valid Loss:   tensor(0.8520, device='cuda:0')
[57] Valid Loss:   tensor(0.8184, device='cuda:0')
[58] Valid Loss:   tensor(0.7692, device='cuda:0')
[59] Valid Loss:   tensor(0.6623, device='cuda:0')
[60] Valid Loss:   tensor(1.0372, device='cuda:0')
[61] Valid Loss:   tensor(0.6971, device='cuda:0')
[62] Valid Loss:   tensor(0.8840, device='cuda:0')
[63] Valid Loss:   tensor(0.6637, device='cuda:0')
[64] Valid Loss:   tensor(0.7399, device='cuda:0')
[65] Valid Loss:   tensor(0.8805, device='cuda:0')
[66] Valid Loss:   tensor(0.7204, device='cuda:0')
[67] Valid Loss:   tensor(0.7515, device='cuda:0')
[68] Valid Loss:   tensor(0.7299, device='cuda:0')
[69] Valid Loss:   tensor(0.5743, device='cuda:0')
[70] Valid Loss:   tensor(0.6026, device='cuda:0')
[71] Valid Loss:   tensor(0.8278, device='cuda:0')
[72] Valid Loss:   tensor(0.7571, device='cuda:0')
[73] Valid Loss:   tensor(0.6833, device='cuda:0')
[74] Valid Loss:   tensor(0.7730, device='cuda:0')
[75] Valid Loss:   tensor(0.8478, device='cuda:0')
[76] Valid Loss:   tensor(0.8034, device='cuda:0')
[77] Valid Loss:   tensor(0.7116, device='cuda:0')
[78] Valid Loss:   tensor(0.7710, device='cuda:0')
[79] Valid Loss:   tensor(0.6332, device='cuda:0')
[80] Valid Loss:   tensor(0.7368, device='cuda:0')
[81] Valid Loss:   tensor(0.6290, device='cuda:0')
[82] Valid Loss:   tensor(0.8782, device='cuda:0')
[83] Valid Loss:   tensor(0.8053, device='cuda:0')
[84] Valid Loss:   tensor(0.6138, device='cuda:0')
[85] Valid Loss:   tensor(0.7377, device='cuda:0')
[86] Valid Loss:   tensor(0.7529, device='cuda:0')
[87] Valid Loss:   tensor(0.8048, device='cuda:0')
[88] Valid Loss:   tensor(0.6073, device='cuda:0')
[89] Valid Loss:   tensor(0.8124, device='cuda:0')
[90] Valid Loss:   tensor(0.9414, device='cuda:0')
[91] Valid Loss:   tensor(0.7346, device='cuda:0')
[92] Valid Loss:   tensor(0.7816, device='cuda:0')
[93] Valid Loss:   tensor(0.7978, device='cuda:0')
[94] Valid Loss:   tensor(0.8541, device='cuda:0')
[95] Valid Loss:   tensor(0.7122, device='cuda:0')
[96] Valid Loss:   tensor(0.8614, device='cuda:0')
```

```
[97] Valid Loss:  tensor(0.8089, device='cuda:0')
[98] Valid Loss:  tensor(0.8063, device='cuda:0')
[99] Valid Loss:  tensor(0.6269, device='cuda:0')
[100] Valid Loss:  tensor(0.7515, device='cuda:0')
[101] Valid Loss:  tensor(0.7331, device='cuda:0')
[102] Valid Loss:  tensor(0.9294, device='cuda:0')
[103] Valid Loss:  tensor(0.8514, device='cuda:0')
[104] Valid Loss:  tensor(0.6675, device='cuda:0')
[105] Valid Loss:  tensor(0.6859, device='cuda:0')
[106] Valid Loss:  tensor(0.6040, device='cuda:0')
[107] Valid Loss:  tensor(0.7689, device='cuda:0')
[108] Valid Loss:  tensor(0.8737, device='cuda:0')
[109] Valid Loss:  tensor(0.7576, device='cuda:0')
[110] Valid Loss:  tensor(0.7907, device='cuda:0')
[111] Valid Loss:  tensor(0.7645, device='cuda:0')
[112] Valid Loss:  tensor(0.8099, device='cuda:0')
[113] Valid Loss:  tensor(0.7687, device='cuda:0')
[114] Valid Loss:  tensor(0.7348, device='cuda:0')
[115] Valid Loss:  tensor(0.8284, device='cuda:0')
[116] Valid Loss:  tensor(0.8315, device='cuda:0')
[117] Valid Loss:  tensor(0.8417, device='cuda:0')
[118] Valid Loss:  tensor(0.6648, device='cuda:0')
[119] Valid Loss:  tensor(0.6072, device='cuda:0')
[120] Valid Loss:  tensor(1.0010, device='cuda:0')
[121] Valid Loss:  tensor(0.6897, device='cuda:0')
[122] Valid Loss:  tensor(0.7306, device='cuda:0')
[123] Valid Loss:  tensor(0.6929, device='cuda:0')
[124] Valid Loss:  tensor(0.9089, device='cuda:0')
[125] Valid Loss:  tensor(0.9452, device='cuda:0')
[126] Valid Loss:  tensor(0.5613, device='cuda:0')
[127] Valid Loss:  tensor(0.6260, device='cuda:0')
[128] Valid Loss:  tensor(0.8570, device='cuda:0')
[129] Valid Loss:  tensor(0.6982, device='cuda:0')
[130] Valid Loss:  tensor(0.8229, device='cuda:0')
[131] Valid Loss:  tensor(0.6840, device='cuda:0')
[132] Valid Loss:  tensor(0.6996, device='cuda:0')
[133] Valid Loss:  tensor(0.7743, device='cuda:0')
[134] Valid Loss:  tensor(0.6683, device='cuda:0')
[135] Valid Loss:  tensor(0.5992, device='cuda:0')
[136] Valid Loss:  tensor(0.6178, device='cuda:0')
[137] Valid Loss:  tensor(0.9748, device='cuda:0')
[138] Valid Loss:  tensor(0.8377, device='cuda:0')
[139] Valid Loss:  tensor(1.0550, device='cuda:0')
[140] Valid Loss:  tensor(0.8269, device='cuda:0')
[141] Valid Loss:  tensor(0.6483, device='cuda:0')
[142] Valid Loss:  tensor(0.9310, device='cuda:0')
[143] Valid Loss:  tensor(0.6278, device='cuda:0')
[144] Valid Loss:  tensor(0.7613, device='cuda:0')
```

```
[145] Valid Loss:  tensor(0.9401, device='cuda:0')
[146] Valid Loss:  tensor(0.6792, device='cuda:0')
[147] Valid Loss:  tensor(0.8113, device='cuda:0')
[148] Valid Loss:  tensor(0.6841, device='cuda:0')
[149] Valid Loss:  tensor(0.7708, device='cuda:0')
[150] Valid Loss:  tensor(0.7496, device='cuda:0')
[151] Valid Loss:  tensor(0.7433, device='cuda:0')
[152] Valid Loss:  tensor(0.6678, device='cuda:0')
[153] Valid Loss:  tensor(0.6694, device='cuda:0')
[154] Valid Loss:  tensor(0.7090, device='cuda:0')
[155] Valid Loss:  tensor(0.7540, device='cuda:0')
[156] Valid Loss:  tensor(0.8256, device='cuda:0')
[157] Valid Loss:  tensor(0.6495, device='cuda:0')
[158] Valid Loss:  tensor(0.7181, device='cuda:0')
[159] Valid Loss:  tensor(0.7319, device='cuda:0')
[160] Valid Loss:  tensor(0.6061, device='cuda:0')
[161] Valid Loss:  tensor(0.8238, device='cuda:0')
[162] Valid Loss:  tensor(0.7829, device='cuda:0')
[163] Valid Loss:  tensor(0.6482, device='cuda:0')
[164] Valid Loss:  tensor(0.9561, device='cuda:0')
[165] Valid Loss:  tensor(0.7169, device='cuda:0')
[166] Valid Loss:  tensor(0.7179, device='cuda:0')
[167] Valid Loss:  tensor(0.5936, device='cuda:0')
[168] Valid Loss:  tensor(0.6923, device='cuda:0')
[169] Valid Loss:  tensor(0.6887, device='cuda:0')
[170] Valid Loss:  tensor(0.6874, device='cuda:0')
[171] Valid Loss:  tensor(0.5946, device='cuda:0')
[172] Valid Loss:  tensor(0.7728, device='cuda:0')
[173] Valid Loss:  tensor(0.6346, device='cuda:0')
[174] Valid Loss:  tensor(0.7515, device='cuda:0')
[175] Valid Loss:  tensor(0.6638, device='cuda:0')
[176] Valid Loss:  tensor(0.6022, device='cuda:0')
[177] Valid Loss:  tensor(0.7464, device='cuda:0')
[178] Valid Loss:  tensor(0.7693, device='cuda:0')
[179] Valid Loss:  tensor(0.7837, device='cuda:0')
[180] Valid Loss:  tensor(0.8104, device='cuda:0')
[181] Valid Loss:  tensor(0.7655, device='cuda:0')
[182] Valid Loss:  tensor(0.7305, device='cuda:0')
[183] Valid Loss:  tensor(0.7450, device='cuda:0')
[184] Valid Loss:  tensor(0.9090, device='cuda:0')
[185] Valid Loss:  tensor(0.8282, device='cuda:0')
[186] Valid Loss:  tensor(0.7015, device='cuda:0')
[187] Valid Loss:  tensor(0.8174, device='cuda:0')
[188] Valid Loss:  tensor(0.8953, device='cuda:0')
[189] Valid Loss:  tensor(0.7347, device='cuda:0')
[190] Valid Loss:  tensor(0.7422, device='cuda:0')
[191] Valid Loss:  tensor(0.6155, device='cuda:0')
[192] Valid Loss:  tensor(1.0107, device='cuda:0')
```

```
[193] Valid Loss:  tensor(0.7477, device='cuda:0')
[194] Valid Loss:  tensor(0.8383, device='cuda:0')
[195] Valid Loss:  tensor(0.8540, device='cuda:0')
[196] Valid Loss:  tensor(0.6076, device='cuda:0')
[197] Valid Loss:  tensor(0.7773, device='cuda:0')
[198] Valid Loss:  tensor(0.7475, device='cuda:0')
[199] Valid Loss:  tensor(0.7875, device='cuda:0')
[200] Valid Loss:  tensor(0.7295, device='cuda:0')
[201] Valid Loss:  tensor(0.7068, device='cuda:0')
[202] Valid Loss:  tensor(0.6748, device='cuda:0')
[203] Valid Loss:  tensor(0.8930, device='cuda:0')
[204] Valid Loss:  tensor(0.8559, device='cuda:0')
[205] Valid Loss:  tensor(0.8506, device='cuda:0')
[206] Valid Loss:  tensor(0.9606, device='cuda:0')
[207] Valid Loss:  tensor(0.7768, device='cuda:0')
[208] Valid Loss:  tensor(0.6103, device='cuda:0')
[209] Valid Loss:  tensor(0.7259, device='cuda:0')
[210] Valid Loss:  tensor(1.1503, device='cuda:0')
[211] Valid Loss:  tensor(0.7681, device='cuda:0')
[212] Valid Loss:  tensor(0.6218, device='cuda:0')
[213] Valid Loss:  tensor(0.7071, device='cuda:0')
[214] Valid Loss:  tensor(0.7100, device='cuda:0')
[215] Valid Loss:  tensor(0.8278, device='cuda:0')
[216] Valid Loss:  tensor(0.7627, device='cuda:0')
[217] Valid Loss:  tensor(0.7182, device='cuda:0')
[218] Valid Loss:  tensor(0.5892, device='cuda:0')
[219] Valid Loss:  tensor(0.6973, device='cuda:0')
[220] Valid Loss:  tensor(0.6763, device='cuda:0')
[221] Valid Loss:  tensor(0.7676, device='cuda:0')
[222] Valid Loss:  tensor(0.5743, device='cuda:0')
[223] Valid Loss:  tensor(0.7005, device='cuda:0')
[224] Valid Loss:  tensor(0.8116, device='cuda:0')
[225] Valid Loss:  tensor(0.6055, device='cuda:0')
[226] Valid Loss:  tensor(0.7445, device='cuda:0')
[227] Valid Loss:  tensor(0.6188, device='cuda:0')
[228] Valid Loss:  tensor(0.8211, device='cuda:0')
[229] Valid Loss:  tensor(0.7017, device='cuda:0')
[230] Valid Loss:  tensor(0.7930, device='cuda:0')
[231] Valid Loss:  tensor(0.6880, device='cuda:0')
[232] Valid Loss:  tensor(0.6113, device='cuda:0')
[233] Valid Loss:  tensor(0.6964, device='cuda:0')
[234] Valid Loss:  tensor(0.5923, device='cuda:0')
[235] Valid Loss:  tensor(0.6273, device='cuda:0')
[236] Valid Loss:  tensor(0.7713, device='cuda:0')
[237] Valid Loss:  tensor(0.7907, device='cuda:0')
[238] Valid Loss:  tensor(0.5458, device='cuda:0')
[239] Valid Loss:  tensor(0.8100, device='cuda:0')
[240] Valid Loss:  tensor(0.7273, device='cuda:0')
```

```
[241] Valid Loss:  tensor(0.6359, device='cuda:0')
[242] Valid Loss:  tensor(0.5221, device='cuda:0')
[243] Valid Loss:  tensor(1.1629, device='cuda:0')
[244] Valid Loss:  tensor(0.7872, device='cuda:0')
[245] Valid Loss:  tensor(0.7007, device='cuda:0')
[246] Valid Loss:  tensor(0.6551, device='cuda:0')
[247] Valid Loss:  tensor(0.6946, device='cuda:0')
[248] Valid Loss:  tensor(0.9364, device='cuda:0')
[249] Valid Loss:  tensor(0.7032, device='cuda:0')
[250] Valid Loss:  tensor(0.7781, device='cuda:0')
[251] Valid Loss:  tensor(0.8023, device='cuda:0')
[252] Valid Loss:  tensor(0.6217, device='cuda:0')
[253] Valid Loss:  tensor(0.6345, device='cuda:0')
[254] Valid Loss:  tensor(0.6507, device='cuda:0')
[255] Valid Loss:  tensor(0.7867, device='cuda:0')
[256] Valid Loss:  tensor(0.8220, device='cuda:0')
[257] Valid Loss:  tensor(0.7141, device='cuda:0')
[258] Valid Loss:  tensor(0.8529, device='cuda:0')
[259] Valid Loss:  tensor(0.7423, device='cuda:0')
[260] Valid Loss:  tensor(0.9226, device='cuda:0')
[261] Valid Loss:  tensor(0.7653, device='cuda:0')
[262] Valid Loss:  tensor(0.8994, device='cuda:0')
[263] Valid Loss:  tensor(0.8763, device='cuda:0')
[264] Valid Loss:  tensor(0.8418, device='cuda:0')
[265] Valid Loss:  tensor(0.8034, device='cuda:0')
[266] Valid Loss:  tensor(0.6868, device='cuda:0')
[267] Valid Loss:  tensor(0.7311, device='cuda:0')
[268] Valid Loss:  tensor(0.7168, device='cuda:0')
[269] Valid Loss:  tensor(0.6912, device='cuda:0')
[270] Valid Loss:  tensor(0.6366, device='cuda:0')
[271] Valid Loss:  tensor(0.9568, device='cuda:0')
[272] Valid Loss:  tensor(0.7330, device='cuda:0')
[273] Valid Loss:  tensor(0.8741, device='cuda:0')
[274] Valid Loss:  tensor(0.6989, device='cuda:0')
[275] Valid Loss:  tensor(0.7485, device='cuda:0')
[276] Valid Loss:  tensor(0.6580, device='cuda:0')
[277] Valid Loss:  tensor(0.6331, device='cuda:0')
[278] Valid Loss:  tensor(0.9338, device='cuda:0')
[279] Valid Loss:  tensor(0.7969, device='cuda:0')
[280] Valid Loss:  tensor(0.8955, device='cuda:0')
[281] Valid Loss:  tensor(0.6584, device='cuda:0')
[282] Valid Loss:  tensor(0.7643, device='cuda:0')
[283] Valid Loss:  tensor(0.8970, device='cuda:0')
[284] Valid Loss:  tensor(0.6787, device='cuda:0')
[285] Valid Loss:  tensor(0.5744, device='cuda:0')
[286] Valid Loss:  tensor(0.8129, device='cuda:0')
[287] Valid Loss:  tensor(0.9342, device='cuda:0')
[288] Valid Loss:  tensor(0.5763, device='cuda:0')
```

```
[289] Valid Loss:   tensor(0.7710, device='cuda:0')
[290] Valid Loss:   tensor(0.5962, device='cuda:0')
[291] Valid Loss:   tensor(0.7318, device='cuda:0')
[292] Valid Loss:   tensor(0.8613, device='cuda:0')
[293] Valid Loss:   tensor(0.6540, device='cuda:0')
[294] Valid Loss:   tensor(0.8802, device='cuda:0')
[295] Valid Loss:   tensor(0.6691, device='cuda:0')
[296] Valid Loss:   tensor(0.7622, device='cuda:0')
[297] Valid Loss:   tensor(0.8112, device='cuda:0')
[298] Valid Loss:   tensor(0.6697, device='cuda:0')
[299] Valid Loss:   tensor(0.7790, device='cuda:0')
[300] Valid Loss:   tensor(0.9216, device='cuda:0')
[301] Valid Loss:   tensor(0.6000, device='cuda:0')
[302] Valid Loss:   tensor(0.5685, device='cuda:0')
[303] Valid Loss:   tensor(0.6828, device='cuda:0')
[304] Valid Loss:   tensor(0.6557, device='cuda:0')
[305] Valid Loss:   tensor(0.7389, device='cuda:0')
[306] Valid Loss:   tensor(0.7854, device='cuda:0')
[307] Valid Loss:   tensor(0.5952, device='cuda:0')
[308] Valid Loss:   tensor(0.6425, device='cuda:0')
[309] Valid Loss:   tensor(0.6773, device='cuda:0')
[310] Valid Loss:   tensor(0.7155, device='cuda:0')
[311] Valid Loss:   tensor(0.6272, device='cuda:0')
[312] Valid Loss:   tensor(0.6060, device='cuda:0')
[313] Valid Loss:   tensor(0.7059, device='cuda:0')
[314] Valid Loss:   tensor(0.7250, device='cuda:0')
[315] Valid Loss:   tensor(1.0374, device='cuda:0')
[316] Valid Loss:   tensor(0.8660, device='cuda:0')
[317] Valid Loss:   tensor(0.7168, device='cuda:0')
[318] Valid Loss:   tensor(0.6720, device='cuda:0')
[319] Valid Loss:   tensor(0.6461, device='cuda:0')
[320] Valid Loss:   tensor(0.6723, device='cuda:0')
[321] Valid Loss:   tensor(0.6825, device='cuda:0')
[322] Valid Loss:   tensor(0.6424, device='cuda:0')
[323] Valid Loss:   tensor(0.8218, device='cuda:0')
[324] Valid Loss:   tensor(0.7510, device='cuda:0')
[325] Valid Loss:   tensor(0.6292, device='cuda:0')
[326] Valid Loss:   tensor(0.7253, device='cuda:0')
[327] Valid Loss:   tensor(0.4743, device='cuda:0')
[328] Valid Loss:   tensor(0.6768, device='cuda:0')
[329] Valid Loss:   tensor(0.8024, device='cuda:0')
[330] Valid Loss:   tensor(0.5954, device='cuda:0')
[331] Valid Loss:   tensor(0.9218, device='cuda:0')
[332] Valid Loss:   tensor(0.6073, device='cuda:0')
[333] Valid Loss:   tensor(0.5950, device='cuda:0')
[334] Valid Loss:   tensor(0.6072, device='cuda:0')
[335] Valid Loss:   tensor(0.7085, device='cuda:0')
[336] Valid Loss:   tensor(0.7928, device='cuda:0')
```

```
[337] Valid Loss:  tensor(0.8614, device='cuda:0')
[338] Valid Loss:  tensor(0.7536, device='cuda:0')
[339] Valid Loss:  tensor(0.7171, device='cuda:0')
[340] Valid Loss:  tensor(0.7397, device='cuda:0')
[341] Valid Loss:  tensor(0.8616, device='cuda:0')
[342] Valid Loss:  tensor(0.9265, device='cuda:0')
[343] Valid Loss:  tensor(0.7763, device='cuda:0')
[344] Valid Loss:  tensor(0.6491, device='cuda:0')
[345] Valid Loss:  tensor(0.5604, device='cuda:0')
[346] Valid Loss:  tensor(0.7236, device='cuda:0')
[347] Valid Loss:  tensor(0.8407, device='cuda:0')
[348] Valid Loss:  tensor(0.7101, device='cuda:0')
[349] Valid Loss:  tensor(0.6767, device='cuda:0')
[350] Valid Loss:  tensor(0.7065, device='cuda:0')
[351] Valid Loss:  tensor(0.8212, device='cuda:0')
[352] Valid Loss:  tensor(0.6517, device='cuda:0')
[353] Valid Loss:  tensor(0.7413, device='cuda:0')
[354] Valid Loss:  tensor(0.9311, device='cuda:0')
[355] Valid Loss:  tensor(0.5307, device='cuda:0')
[356] Valid Loss:  tensor(0.6777, device='cuda:0')
[357] Valid Loss:  tensor(0.8207, device='cuda:0')
[358] Valid Loss:  tensor(0.6756, device='cuda:0')
[359] Valid Loss:  tensor(0.8117, device='cuda:0')
[360] Valid Loss:  tensor(0.7045, device='cuda:0')
[361] Valid Loss:  tensor(0.7577, device='cuda:0')
[362] Valid Loss:  tensor(0.7172, device='cuda:0')
[363] Valid Loss:  tensor(0.7941, device='cuda:0')
[364] Valid Loss:  tensor(0.9224, device='cuda:0')
[365] Valid Loss:  tensor(0.6933, device='cuda:0')
[366] Valid Loss:  tensor(0.6821, device='cuda:0')
[367] Valid Loss:  tensor(0.7750, device='cuda:0')
[368] Valid Loss:  tensor(0.8443, device='cuda:0')
[369] Valid Loss:  tensor(0.7258, device='cuda:0')
[370] Valid Loss:  tensor(0.6281, device='cuda:0')
[371] Valid Loss:  tensor(0.6500, device='cuda:0')
[372] Valid Loss:  tensor(0.8075, device='cuda:0')
[373] Valid Loss:  tensor(0.5464, device='cuda:0')
[374] Valid Loss:  tensor(0.7131, device='cuda:0')
[375] Valid Loss:  tensor(0.8062, device='cuda:0')
[376] Valid Loss:  tensor(0.6171, device='cuda:0')
[377] Valid Loss:  tensor(0.9106, device='cuda:0')
[378] Valid Loss:  tensor(0.6496, device='cuda:0')
[379] Valid Loss:  tensor(0.8229, device='cuda:0')
[380] Valid Loss:  tensor(1.0602, device='cuda:0')
Testing Loss:  0.7459756139069404
Testing Accuaracy:  77.05946607468489 %
Epoch Number:  8
[0] Train Loss:  tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1] Train Loss:  tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[2] Train Loss:  tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[3] Train Loss:  tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[4] Train Loss:  tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[5] Train Loss:  tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[6] Train Loss:  tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[7] Train Loss:  tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[8] Train Loss:  tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[9] Train Loss:  tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[10] Train Loss:  tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[11] Train Loss:  tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[12] Train Loss:  tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[13] Train Loss:  tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[14] Train Loss:  tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[15] Train Loss:  tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[16] Train Loss:  tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[17] Train Loss:  tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[18] Train Loss:  tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[19] Train Loss:  tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[20] Train Loss:  tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[21] Train Loss:  tensor(0.6592, device='cuda:0', grad_fn=<NllLossBackward>)
[22] Train Loss:  tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[23] Train Loss:  tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[24] Train Loss:  tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[25] Train Loss:  tensor(0.6650, device='cuda:0', grad_fn=<NllLossBackward>)
[26] Train Loss:  tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[27] Train Loss:  tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[28] Train Loss:  tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[29] Train Loss:  tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[30] Train Loss:  tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[31] Train Loss:  tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[32] Train Loss:  tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[33] Train Loss:  tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[34] Train Loss:  tensor(0.6684, device='cuda:0', grad_fn=<NllLossBackward>)
[35] Train Loss:  tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[36] Train Loss:  tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[37] Train Loss:  tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[38] Train Loss:  tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[39] Train Loss:  tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[40] Train Loss:  tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[41] Train Loss:  tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[42] Train Loss:  tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[43] Train Loss:  tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[44] Train Loss:  tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[45] Train Loss:  tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[46] Train Loss:  tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[47] Train Loss:  tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[48] Train Loss:  tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[49] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[50] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[51] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[52] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[53] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[54] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[55] Train Loss:    tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[56] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[57] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[58] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[59] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[60] Train Loss:    tensor(0.6573, device='cuda:0', grad_fn=<NllLossBackward>)
[61] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[62] Train Loss:    tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[63] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[64] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[65] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[66] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[67] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[68] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[69] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[70] Train Loss:    tensor(0.7432, device='cuda:0', grad_fn=<NllLossBackward>)
[71] Train Loss:    tensor(0.6598, device='cuda:0', grad_fn=<NllLossBackward>)
[72] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[73] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[74] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[75] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[76] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[77] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[78] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[79] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[80] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[81] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[82] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[83] Train Loss:    tensor(0.6739, device='cuda:0', grad_fn=<NllLossBackward>)
[84] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[85] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[86] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[87] Train Loss:    tensor(0.6538, device='cuda:0', grad_fn=<NllLossBackward>)
[88] Train Loss:    tensor(0.6540, device='cuda:0', grad_fn=<NllLossBackward>)
[89] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[90] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[91] Train Loss:    tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[92] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[93] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[94] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[95] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[96] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[97] Train Loss:  tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[98] Train Loss:  tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[99] Train Loss:  tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[100] Train Loss:  tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[101] Train Loss:  tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[102] Train Loss:  tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[103] Train Loss:  tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[104] Train Loss:  tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[105] Train Loss:  tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[106] Train Loss:  tensor(0.6637, device='cuda:0', grad_fn=<NllLossBackward>)
[107] Train Loss:  tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[108] Train Loss:  tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[109] Train Loss:  tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[110] Train Loss:  tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[111] Train Loss:  tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[112] Train Loss:  tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[113] Train Loss:  tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[114] Train Loss:  tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[115] Train Loss:  tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[116] Train Loss:  tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[117] Train Loss:  tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[118] Train Loss:  tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[119] Train Loss:  tensor(0.6631, device='cuda:0', grad_fn=<NllLossBackward>)
[120] Train Loss:  tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[121] Train Loss:  tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[122] Train Loss:  tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[123] Train Loss:  tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[124] Train Loss:  tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[125] Train Loss:  tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[126] Train Loss:  tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[127] Train Loss:  tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[128] Train Loss:  tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[129] Train Loss:  tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[130] Train Loss:  tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[131] Train Loss:  tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[132] Train Loss:  tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[133] Train Loss:  tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[134] Train Loss:  tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[135] Train Loss:  tensor(0.6620, device='cuda:0', grad_fn=<NllLossBackward>)
[136] Train Loss:  tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[137] Train Loss:  tensor(0.6625, device='cuda:0', grad_fn=<NllLossBackward>)
[138] Train Loss:  tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[139] Train Loss:  tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[140] Train Loss:  tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[141] Train Loss:  tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[142] Train Loss:  tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[143] Train Loss:  tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[144] Train Loss:  tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[145] Train Loss:   tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[146] Train Loss:   tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[147] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[148] Train Loss:   tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[149] Train Loss:   tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[150] Train Loss:   tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[151] Train Loss:   tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[152] Train Loss:   tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[153] Train Loss:   tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[154] Train Loss:   tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[155] Train Loss:   tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[156] Train Loss:   tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[157] Train Loss:   tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[158] Train Loss:   tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[159] Train Loss:   tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[160] Train Loss:   tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[161] Train Loss:   tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[162] Train Loss:   tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[163] Train Loss:   tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[164] Train Loss:   tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[165] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[166] Train Loss:   tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[167] Train Loss:   tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[168] Train Loss:   tensor(0.6645, device='cuda:0', grad_fn=<NllLossBackward>)
[169] Train Loss:   tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[170] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[171] Train Loss:   tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[172] Train Loss:   tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[173] Train Loss:   tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[174] Train Loss:   tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[175] Train Loss:   tensor(0.6527, device='cuda:0', grad_fn=<NllLossBackward>)
[176] Train Loss:   tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[177] Train Loss:   tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[178] Train Loss:   tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[179] Train Loss:   tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[180] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[181] Train Loss:   tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[182] Train Loss:   tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[183] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[184] Train Loss:   tensor(0.6632, device='cuda:0', grad_fn=<NllLossBackward>)
[185] Train Loss:   tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[186] Train Loss:   tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[187] Train Loss:   tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[188] Train Loss:   tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[189] Train Loss:   tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[190] Train Loss:   tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[191] Train Loss:   tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[192] Train Loss:   tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[193] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[194] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[195] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[196] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[197] Train Loss:    tensor(0.6729, device='cuda:0', grad_fn=<NllLossBackward>)
[198] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[199] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[200] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[201] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[202] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[203] Train Loss:    tensor(0.7372, device='cuda:0', grad_fn=<NllLossBackward>)
[204] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[205] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[206] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[207] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[208] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[209] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[210] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[211] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[212] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[213] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[214] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[215] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[216] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[217] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[218] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[219] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[220] Train Loss:    tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[221] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[222] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[223] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[224] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[225] Train Loss:    tensor(0.6679, device='cuda:0', grad_fn=<NllLossBackward>)
[226] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[227] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[228] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[229] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[230] Train Loss:    tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[231] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[232] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[233] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[234] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[235] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[236] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[237] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[238] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[239] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[240] Train Loss:    tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[241] Train Loss:   tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[242] Train Loss:   tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[243] Train Loss:   tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[244] Train Loss:   tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[245] Train Loss:   tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[246] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[247] Train Loss:   tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[248] Train Loss:   tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[249] Train Loss:   tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[250] Train Loss:   tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[251] Train Loss:   tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[252] Train Loss:   tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[253] Train Loss:   tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[254] Train Loss:   tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[255] Train Loss:   tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[256] Train Loss:   tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[257] Train Loss:   tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[258] Train Loss:   tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[259] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[260] Train Loss:   tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[261] Train Loss:   tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[262] Train Loss:   tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[263] Train Loss:   tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[264] Train Loss:   tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[265] Train Loss:   tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[266] Train Loss:   tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[267] Train Loss:   tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[268] Train Loss:   tensor(0.6586, device='cuda:0', grad_fn=<NllLossBackward>)
[269] Train Loss:   tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[270] Train Loss:   tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[271] Train Loss:   tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[272] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[273] Train Loss:   tensor(0.6669, device='cuda:0', grad_fn=<NllLossBackward>)
[274] Train Loss:   tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[275] Train Loss:   tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[276] Train Loss:   tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[277] Train Loss:   tensor(0.6504, device='cuda:0', grad_fn=<NllLossBackward>)
[278] Train Loss:   tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[279] Train Loss:   tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[280] Train Loss:   tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[281] Train Loss:   tensor(0.7456, device='cuda:0', grad_fn=<NllLossBackward>)
[282] Train Loss:   tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[283] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[284] Train Loss:   tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[285] Train Loss:   tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[286] Train Loss:   tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[287] Train Loss:   tensor(0.6612, device='cuda:0', grad_fn=<NllLossBackward>)
[288] Train Loss:   tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[289] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[290] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[291] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[292] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[293] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[294] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[295] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[296] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[297] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[298] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[299] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[300] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[301] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[302] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[303] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[304] Train Loss:    tensor(0.6683, device='cuda:0', grad_fn=<NllLossBackward>)
[305] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[306] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[307] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[308] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[309] Train Loss:    tensor(0.6704, device='cuda:0', grad_fn=<NllLossBackward>)
[310] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[311] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[312] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[313] Train Loss:    tensor(0.6607, device='cuda:0', grad_fn=<NllLossBackward>)
[314] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[315] Train Loss:    tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[316] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[317] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[318] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[319] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[320] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[321] Train Loss:    tensor(0.6647, device='cuda:0', grad_fn=<NllLossBackward>)
[322] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[323] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[324] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[325] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[326] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[327] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[328] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[329] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[330] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[331] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[332] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[333] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[334] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[335] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[336] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[337] Train Loss:   tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[338] Train Loss:   tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[339] Train Loss:   tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[340] Train Loss:   tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[341] Train Loss:   tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[342] Train Loss:   tensor(0.6659, device='cuda:0', grad_fn=<NllLossBackward>)
[343] Train Loss:   tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[344] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[345] Train Loss:   tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[346] Train Loss:   tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[347] Train Loss:   tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[348] Train Loss:   tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[349] Train Loss:   tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[350] Train Loss:   tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[351] Train Loss:   tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[352] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[353] Train Loss:   tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[354] Train Loss:   tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[355] Train Loss:   tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[356] Train Loss:   tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[357] Train Loss:   tensor(0.6589, device='cuda:0', grad_fn=<NllLossBackward>)
[358] Train Loss:   tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[359] Train Loss:   tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[360] Train Loss:   tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[361] Train Loss:   tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[362] Train Loss:   tensor(0.6695, device='cuda:0', grad_fn=<NllLossBackward>)
[363] Train Loss:   tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[364] Train Loss:   tensor(0.6644, device='cuda:0', grad_fn=<NllLossBackward>)
[365] Train Loss:   tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[366] Train Loss:   tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[367] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[368] Train Loss:   tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[369] Train Loss:   tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[370] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[371] Train Loss:   tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[372] Train Loss:   tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[373] Train Loss:   tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[374] Train Loss:   tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[375] Train Loss:   tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[376] Train Loss:   tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[377] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[378] Train Loss:   tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[379] Train Loss:   tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[380] Train Loss:   tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[381] Train Loss:   tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[382] Train Loss:   tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[383] Train Loss:   tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[384] Train Loss:   tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[385] Train Loss:   tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[386] Train Loss:   tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[387] Train Loss:   tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[388] Train Loss:   tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[389] Train Loss:   tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[390] Train Loss:   tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[391] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[392] Train Loss:   tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[393] Train Loss:   tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[394] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[395] Train Loss:   tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[396] Train Loss:   tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[397] Train Loss:   tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[398] Train Loss:   tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[399] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[400] Train Loss:   tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[401] Train Loss:   tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[402] Train Loss:   tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[403] Train Loss:   tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[404] Train Loss:   tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[405] Train Loss:   tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[406] Train Loss:   tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[407] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[408] Train Loss:   tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[409] Train Loss:   tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[410] Train Loss:   tensor(0.6606, device='cuda:0', grad_fn=<NllLossBackward>)
[411] Train Loss:   tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[412] Train Loss:   tensor(0.7560, device='cuda:0', grad_fn=<NllLossBackward>)
[413] Train Loss:   tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[414] Train Loss:   tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[415] Train Loss:   tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[416] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[417] Train Loss:   tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[418] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[419] Train Loss:   tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[420] Train Loss:   tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[421] Train Loss:   tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[422] Train Loss:   tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[423] Train Loss:   tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[424] Train Loss:   tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[425] Train Loss:   tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[426] Train Loss:   tensor(0.7448, device='cuda:0', grad_fn=<NllLossBackward>)
[427] Train Loss:   tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[428] Train Loss:   tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[429] Train Loss:   tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[430] Train Loss:   tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[431] Train Loss:   tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[432] Train Loss:   tensor(0.7351, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[433] Train Loss:   tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[434] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[435] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[436] Train Loss:   tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[437] Train Loss:   tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[438] Train Loss:   tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[439] Train Loss:   tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[440] Train Loss:   tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[441] Train Loss:   tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[442] Train Loss:   tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[443] Train Loss:   tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[444] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[445] Train Loss:   tensor(0.6682, device='cuda:0', grad_fn=<NllLossBackward>)
[446] Train Loss:   tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[447] Train Loss:   tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[448] Train Loss:   tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[449] Train Loss:   tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[450] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[451] Train Loss:   tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[452] Train Loss:   tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[453] Train Loss:   tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[454] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[455] Train Loss:   tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[456] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[457] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[458] Train Loss:   tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[459] Train Loss:   tensor(0.6665, device='cuda:0', grad_fn=<NllLossBackward>)
[460] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[461] Train Loss:   tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[462] Train Loss:   tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[463] Train Loss:   tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[464] Train Loss:   tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[465] Train Loss:   tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[466] Train Loss:   tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[467] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[468] Train Loss:   tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[469] Train Loss:   tensor(0.6691, device='cuda:0', grad_fn=<NllLossBackward>)
[470] Train Loss:   tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[471] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[472] Train Loss:   tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[473] Train Loss:   tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[474] Train Loss:   tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[475] Train Loss:   tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[476] Train Loss:   tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[477] Train Loss:   tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[478] Train Loss:   tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[479] Train Loss:   tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[480] Train Loss:   tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[481] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[482] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[483] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[484] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[485] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[486] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[487] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[488] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[489] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[490] Train Loss:    tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[491] Train Loss:    tensor(0.6691, device='cuda:0', grad_fn=<NllLossBackward>)
[492] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[493] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[494] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[495] Train Loss:    tensor(0.6577, device='cuda:0', grad_fn=<NllLossBackward>)
[496] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[497] Train Loss:    tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[498] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[499] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[500] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[501] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[502] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[503] Train Loss:    tensor(0.6516, device='cuda:0', grad_fn=<NllLossBackward>)
[504] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[505] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[506] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[507] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[508] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[509] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[510] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[511] Train Loss:    tensor(0.7377, device='cuda:0', grad_fn=<NllLossBackward>)
[512] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[513] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[514] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[515] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[516] Train Loss:    tensor(0.6618, device='cuda:0', grad_fn=<NllLossBackward>)
[517] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[518] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[519] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[520] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[521] Train Loss:    tensor(0.6533, device='cuda:0', grad_fn=<NllLossBackward>)
[522] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[523] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[524] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[525] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[526] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[527] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[528] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[529] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[530] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[531] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[532] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[533] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[534] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[535] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[536] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[537] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[538] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[539] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[540] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[541] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[542] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[543] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[544] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[545] Train Loss:    tensor(0.7410, device='cuda:0', grad_fn=<NllLossBackward>)
[546] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[547] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[548] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[549] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[550] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[551] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[552] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[553] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[554] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[555] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[556] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[557] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[558] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[559] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[560] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[561] Train Loss:    tensor(0.6622, device='cuda:0', grad_fn=<NllLossBackward>)
[562] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[563] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[564] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[565] Train Loss:    tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
[566] Train Loss:    tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[567] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[568] Train Loss:    tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[569] Train Loss:    tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[570] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[571] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[572] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[573] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[574] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[575] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[576] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[577] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[578] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[579] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[580] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[581] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[582] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[583] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[584] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[585] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[586] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[587] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[588] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[589] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[590] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[591] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[592] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[593] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[594] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[595] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[596] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[597] Train Loss:    tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[598] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[599] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[600] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[601] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[602] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[603] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[604] Train Loss:    tensor(0.6690, device='cuda:0', grad_fn=<NllLossBackward>)
[605] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[606] Train Loss:    tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[607] Train Loss:    tensor(0.7437, device='cuda:0', grad_fn=<NllLossBackward>)
[608] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[609] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[610] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[611] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[612] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[613] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[614] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[615] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[616] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[617] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[618] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[619] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[620] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[621] Train Loss:    tensor(0.6670, device='cuda:0', grad_fn=<NllLossBackward>)
[622] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[623] Train Loss:    tensor(0.6368, device='cuda:0', grad_fn=<NllLossBackward>)
[624] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[625] Train Loss:   tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[626] Train Loss:   tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[627] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[628] Train Loss:   tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[629] Train Loss:   tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[630] Train Loss:   tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[631] Train Loss:   tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[632] Train Loss:   tensor(0.6637, device='cuda:0', grad_fn=<NllLossBackward>)
[633] Train Loss:   tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[634] Train Loss:   tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[635] Train Loss:   tensor(0.6647, device='cuda:0', grad_fn=<NllLossBackward>)
[636] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[637] Train Loss:   tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[638] Train Loss:   tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[639] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[640] Train Loss:   tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[641] Train Loss:   tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[642] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[643] Train Loss:   tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[644] Train Loss:   tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[645] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[646] Train Loss:   tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[647] Train Loss:   tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[648] Train Loss:   tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[649] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[650] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[651] Train Loss:   tensor(0.6544, device='cuda:0', grad_fn=<NllLossBackward>)
[652] Train Loss:   tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[653] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[654] Train Loss:   tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[655] Train Loss:   tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[656] Train Loss:   tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[657] Train Loss:   tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[658] Train Loss:   tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[659] Train Loss:   tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[660] Train Loss:   tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[661] Train Loss:   tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[662] Train Loss:   tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[663] Train Loss:   tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[664] Train Loss:   tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[665] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[666] Train Loss:   tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[667] Train Loss:   tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[668] Train Loss:   tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[669] Train Loss:   tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[670] Train Loss:   tensor(0.6637, device='cuda:0', grad_fn=<NllLossBackward>)
[671] Train Loss:   tensor(0.6587, device='cuda:0', grad_fn=<NllLossBackward>)
[672] Train Loss:   tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[673] Train Loss:  tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[674] Train Loss:  tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[675] Train Loss:  tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[676] Train Loss:  tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[677] Train Loss:  tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[678] Train Loss:  tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[679] Train Loss:  tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[680] Train Loss:  tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[681] Train Loss:  tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[682] Train Loss:  tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[683] Train Loss:  tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[684] Train Loss:  tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[685] Train Loss:  tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[686] Train Loss:  tensor(0.6594, device='cuda:0', grad_fn=<NllLossBackward>)
[687] Train Loss:  tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[688] Train Loss:  tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[689] Train Loss:  tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[690] Train Loss:  tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[691] Train Loss:  tensor(0.7549, device='cuda:0', grad_fn=<NllLossBackward>)
[692] Train Loss:  tensor(0.6513, device='cuda:0', grad_fn=<NllLossBackward>)
[693] Train Loss:  tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[694] Train Loss:  tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[695] Train Loss:  tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[696] Train Loss:  tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[697] Train Loss:  tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[698] Train Loss:  tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[699] Train Loss:  tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[700] Train Loss:  tensor(0.6600, device='cuda:0', grad_fn=<NllLossBackward>)
[701] Train Loss:  tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[702] Train Loss:  tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[703] Train Loss:  tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[704] Train Loss:  tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[705] Train Loss:  tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[706] Train Loss:  tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[707] Train Loss:  tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[708] Train Loss:  tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[709] Train Loss:  tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[710] Train Loss:  tensor(0.7429, device='cuda:0', grad_fn=<NllLossBackward>)
[711] Train Loss:  tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[712] Train Loss:  tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[713] Train Loss:  tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[714] Train Loss:  tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[715] Train Loss:  tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[716] Train Loss:  tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[717] Train Loss:  tensor(0.7264, device='cuda:0', grad_fn=<NllLossBackward>)
[718] Train Loss:  tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[719] Train Loss:  tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[720] Train Loss:  tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[721] Train Loss:    tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[722] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[723] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[724] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[725] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[726] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[727] Train Loss:    tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[728] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[729] Train Loss:    tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[730] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[731] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[732] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[733] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[734] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[735] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[736] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[737] Train Loss:    tensor(0.6748, device='cuda:0', grad_fn=<NllLossBackward>)
[738] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[739] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[740] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[741] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[742] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[743] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[744] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[745] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[746] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[747] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[748] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[749] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[750] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[751] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[752] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[753] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[754] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[755] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[756] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[757] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[758] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[759] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[760] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[761] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[762] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[763] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[764] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[765] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[766] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[767] Train Loss:    tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[768] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[769] Train Loss:   tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[770] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[771] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[772] Train Loss:   tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[773] Train Loss:   tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[774] Train Loss:   tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[775] Train Loss:   tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[776] Train Loss:   tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[777] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[778] Train Loss:   tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[779] Train Loss:   tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[780] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[781] Train Loss:   tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[782] Train Loss:   tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[783] Train Loss:   tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[784] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[785] Train Loss:   tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[786] Train Loss:   tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[787] Train Loss:   tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[788] Train Loss:   tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[789] Train Loss:   tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[790] Train Loss:   tensor(0.6683, device='cuda:0', grad_fn=<NllLossBackward>)
[791] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[792] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[793] Train Loss:   tensor(0.6497, device='cuda:0', grad_fn=<NllLossBackward>)
[794] Train Loss:   tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[795] Train Loss:   tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[796] Train Loss:   tensor(0.6468, device='cuda:0', grad_fn=<NllLossBackward>)
[797] Train Loss:   tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[798] Train Loss:   tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[799] Train Loss:   tensor(0.6658, device='cuda:0', grad_fn=<NllLossBackward>)
[800] Train Loss:   tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[801] Train Loss:   tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[802] Train Loss:   tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[803] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[804] Train Loss:   tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[805] Train Loss:   tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[806] Train Loss:   tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[807] Train Loss:   tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[808] Train Loss:   tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[809] Train Loss:   tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[810] Train Loss:   tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[811] Train Loss:   tensor(0.6629, device='cuda:0', grad_fn=<NllLossBackward>)
[812] Train Loss:   tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[813] Train Loss:   tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[814] Train Loss:   tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[815] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[816] Train Loss:   tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[817] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[818] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[819] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[820] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[821] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[822] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[823] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[824] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[825] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[826] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[827] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[828] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[829] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[830] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[831] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[832] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[833] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[834] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[835] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[836] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[837] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[838] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[839] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[840] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[841] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[842] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[843] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[844] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[845] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[846] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[847] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[848] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[849] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[850] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[851] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[852] Train Loss:    tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[853] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[854] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[855] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[856] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[857] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[858] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[859] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[860] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[861] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[862] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[863] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[864] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[865] Train Loss:   tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[866] Train Loss:   tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[867] Train Loss:   tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[868] Train Loss:   tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[869] Train Loss:   tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[870] Train Loss:   tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[871] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[872] Train Loss:   tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[873] Train Loss:   tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[874] Train Loss:   tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[875] Train Loss:   tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[876] Train Loss:   tensor(0.6733, device='cuda:0', grad_fn=<NllLossBackward>)
[877] Train Loss:   tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[878] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[879] Train Loss:   tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[880] Train Loss:   tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[881] Train Loss:   tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[882] Train Loss:   tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[883] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[884] Train Loss:   tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[885] Train Loss:   tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[886] Train Loss:   tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[887] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[888] Train Loss:   tensor(0.6441, device='cuda:0', grad_fn=<NllLossBackward>)
[889] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[890] Train Loss:   tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[891] Train Loss:   tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[892] Train Loss:   tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[893] Train Loss:   tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[894] Train Loss:   tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[895] Train Loss:   tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[896] Train Loss:   tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[897] Train Loss:   tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[898] Train Loss:   tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[899] Train Loss:   tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[900] Train Loss:   tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[901] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[902] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[903] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[904] Train Loss:   tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[905] Train Loss:   tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[906] Train Loss:   tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[907] Train Loss:   tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[908] Train Loss:   tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[909] Train Loss:   tensor(0.6685, device='cuda:0', grad_fn=<NllLossBackward>)
[910] Train Loss:   tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[911] Train Loss:   tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[912] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[913] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[914] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[915] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[916] Train Loss:    tensor(0.6563, device='cuda:0', grad_fn=<NllLossBackward>)
[917] Train Loss:    tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[918] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[919] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[920] Train Loss:    tensor(0.6627, device='cuda:0', grad_fn=<NllLossBackward>)
[921] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[922] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[923] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[924] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[925] Train Loss:    tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[926] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[927] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[928] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[929] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[930] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[931] Train Loss:    tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[932] Train Loss:    tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[933] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[934] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[935] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[936] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[937] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[938] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[939] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[940] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[941] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[942] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[943] Train Loss:    tensor(0.6628, device='cuda:0', grad_fn=<NllLossBackward>)
[944] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[945] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[946] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[947] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[948] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[949] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[950] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[951] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[952] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[953] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[954] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[955] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[956] Train Loss:    tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[957] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[958] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[959] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[960] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[961]  Train Loss:  tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[962]  Train Loss:  tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[963]  Train Loss:  tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[964]  Train Loss:  tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[965]  Train Loss:  tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[966]  Train Loss:  tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[967]  Train Loss:  tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[968]  Train Loss:  tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[969]  Train Loss:  tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[970]  Train Loss:  tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[971]  Train Loss:  tensor(0.7584, device='cuda:0', grad_fn=<NllLossBackward>)
[972]  Train Loss:  tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[973]  Train Loss:  tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[974]  Train Loss:  tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[975]  Train Loss:  tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[976]  Train Loss:  tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[977]  Train Loss:  tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[978]  Train Loss:  tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[979]  Train Loss:  tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[980]  Train Loss:  tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[981]  Train Loss:  tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[982]  Train Loss:  tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[983]  Train Loss:  tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[984]  Train Loss:  tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[985]  Train Loss:  tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[986]  Train Loss:  tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[987]  Train Loss:  tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[988]  Train Loss:  tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[989]  Train Loss:  tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[990]  Train Loss:  tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[991]  Train Loss:  tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[992]  Train Loss:  tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[993]  Train Loss:  tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[994]  Train Loss:  tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[995]  Train Loss:  tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[996]  Train Loss:  tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[997]  Train Loss:  tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[998]  Train Loss:  tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[999]  Train Loss:  tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[1000]  Train Loss:  tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[1001]  Train Loss:  tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1002]  Train Loss:  tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1003]  Train Loss:  tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[1004]  Train Loss:  tensor(0.7490, device='cuda:0', grad_fn=<NllLossBackward>)
[1005]  Train Loss:  tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[1006]  Train Loss:  tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[1007]  Train Loss:  tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[1008]  Train Loss:  tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1009] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[1010] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[1011] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[1012] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[1013] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[1014] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[1015] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1016] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[1017] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[1018] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[1019] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[1020] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[1021] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1022] Train Loss:    tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[1023] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[1024] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[1025] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[1026] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[1027] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1028] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[1029] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[1030] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[1031] Train Loss:    tensor(0.6522, device='cuda:0', grad_fn=<NllLossBackward>)
[1032] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[1033] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[1034] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[1035] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[1036] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[1037] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[1038] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[1039] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[1040] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[1041] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
[1042] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[1043] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[1044] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[1045] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[1046] Train Loss:    tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[1047] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[1048] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[1049] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[1050] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[1051] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[1052] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[1053] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[1054] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[1055] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1056] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1057] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[1058] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[1059] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[1060] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[1061] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[1062] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[1063] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[1064] Train Loss:    tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[1065] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1066] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[1067] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[1068] Train Loss:    tensor(0.6620, device='cuda:0', grad_fn=<NllLossBackward>)
[1069] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[1070] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[1071] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[1072] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[1073] Train Loss:    tensor(0.6614, device='cuda:0', grad_fn=<NllLossBackward>)
[1074] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[1075] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[1076] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[1077] Train Loss:    tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[1078] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[1079] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[1080] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1081] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[1082] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[1083] Train Loss:    tensor(0.7472, device='cuda:0', grad_fn=<NllLossBackward>)
[1084] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[1085] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[1086] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[1087] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[1088] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[1089] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[1090] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[1091] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[1092] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[1093] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1094] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[1095] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1096] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[1097] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[1098] Train Loss:    tensor(0.7484, device='cuda:0', grad_fn=<NllLossBackward>)
[1099] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[1100] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[1101] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[1102] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[1103] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[1104] Train Loss:    tensor(0.6565, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1105] Train Loss:   tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[1106] Train Loss:   tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[1107] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[1108] Train Loss:   tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[1109] Train Loss:   tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[1110] Train Loss:   tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[1111] Train Loss:   tensor(0.6255, device='cuda:0', grad_fn=<NllLossBackward>)
[1112] Train Loss:   tensor(0.6636, device='cuda:0', grad_fn=<NllLossBackward>)
[1113] Train Loss:   tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[1114] Train Loss:   tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[1115] Train Loss:   tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[1116] Train Loss:   tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[1117] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[1118] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1119] Train Loss:   tensor(0.6664, device='cuda:0', grad_fn=<NllLossBackward>)
[1120] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[1121] Train Loss:   tensor(0.7630, device='cuda:0', grad_fn=<NllLossBackward>)
[1122] Train Loss:   tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[1123] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[1124] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[1125] Train Loss:   tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[1126] Train Loss:   tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1127] Train Loss:   tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[1128] Train Loss:   tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[1129] Train Loss:   tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[1130] Train Loss:   tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[1131] Train Loss:   tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[1132] Train Loss:   tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[1133] Train Loss:   tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[1134] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[1135] Train Loss:   tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[1136] Train Loss:   tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[1137] Train Loss:   tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[1138] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[1139] Train Loss:   tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1140] Train Loss:   tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[1141] Train Loss:   tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[1142] Train Loss:   tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[1143] Train Loss:   tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[1144] Train Loss:   tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[1145] Train Loss:   tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[1146] Train Loss:   tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[1147] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1148] Train Loss:   tensor(0.6719, device='cuda:0', grad_fn=<NllLossBackward>)
[1149] Train Loss:   tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[1150] Train Loss:   tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[1151] Train Loss:   tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[1152] Train Loss:   tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1153] Train Loss:    tensor(0.6588, device='cuda:0', grad_fn=<NllLossBackward>)
[1154] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[1155] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[1156] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[1157] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[1158] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[1159] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[1160] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[1161] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1162] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[1163] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[1164] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[1165] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[1166] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[1167] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[1168] Train Loss:    tensor(0.6543, device='cuda:0', grad_fn=<NllLossBackward>)
[1169] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[1170] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[1171] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[1172] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[1173] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[1174] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[1175] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[1176] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[1177] Train Loss:    tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[1178] Train Loss:    tensor(0.6602, device='cuda:0', grad_fn=<NllLossBackward>)
[1179] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[1180] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[1181] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[1182] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[1183] Train Loss:    tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[1184] Train Loss:    tensor(0.6632, device='cuda:0', grad_fn=<NllLossBackward>)
[1185] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[1186] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[1187] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[1188] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[1189] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[1190] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[1191] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[1192] Train Loss:    tensor(0.6610, device='cuda:0', grad_fn=<NllLossBackward>)
[1193] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[1194] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[1195] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[1196] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[1197] Train Loss:    tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[1198] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[1199] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[1200] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1201] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[1202] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[1203] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[1204] Train Loss:    tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[1205] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[1206] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[1207] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[1208] Train Loss:    tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[1209] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[1210] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[1211] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[1212] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1213] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[1214] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[1215] Train Loss:    tensor(0.6709, device='cuda:0', grad_fn=<NllLossBackward>)
[1216] Train Loss:    tensor(0.7443, device='cuda:0', grad_fn=<NllLossBackward>)
[1217] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[1218] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[1219] Train Loss:    tensor(0.6586, device='cuda:0', grad_fn=<NllLossBackward>)
[1220] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[1221] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[1222] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[1223] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[1224] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[1225] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[1226] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[1227] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[1228] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[1229] Train Loss:    tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
[1230] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[1231] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1232] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[1233] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[1234] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[1235] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1236] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[1237] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[1238] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[1239] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[1240] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[1241] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[1242] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[1243] Train Loss:    tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[1244] Train Loss:    tensor(0.6671, device='cuda:0', grad_fn=<NllLossBackward>)
[1245] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[1246] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[1247] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[1248] Train Loss:    tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1249] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1250] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[1251] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[1252] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[1253] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[1254] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[1255] Train Loss:    tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[1256] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[1257] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[1258] Train Loss:    tensor(0.6656, device='cuda:0', grad_fn=<NllLossBackward>)
[1259] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[1260] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[1261] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[1262] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[1263] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[1264] Train Loss:    tensor(0.6623, device='cuda:0', grad_fn=<NllLossBackward>)
[1265] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[1266] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[1267] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[1268] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[1269] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[1270] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1271] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[1272] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[1273] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[1274] Train Loss:    tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[1275] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[1276] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1277] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[1278] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[1279] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[1280] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[1281] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[1282] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[1283] Train Loss:    tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[1284] Train Loss:    tensor(0.6672, device='cuda:0', grad_fn=<NllLossBackward>)
[1285] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[1286] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[1287] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[1288] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[1289] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[1290] Train Loss:    tensor(0.6587, device='cuda:0', grad_fn=<NllLossBackward>)
[1291] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1292] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[1293] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[1294] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[1295] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[1296] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1297] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[1298] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1299] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[1300] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[1301] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[1302] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[1303] Train Loss:    tensor(0.6660, device='cuda:0', grad_fn=<NllLossBackward>)
[1304] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[1305] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[1306] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[1307] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[1308] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[1309] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[1310] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[1311] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[1312] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[1313] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[1314] Train Loss:    tensor(0.6733, device='cuda:0', grad_fn=<NllLossBackward>)
[1315] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[1316] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[1317] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[1318] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[1319] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[1320] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[1321] Train Loss:    tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[1322] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[1323] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[1324] Train Loss:    tensor(0.6696, device='cuda:0', grad_fn=<NllLossBackward>)
[1325] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[1326] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[1327] Train Loss:    tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[1328] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[1329] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[1330] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[1331] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[1332] Train Loss:    tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[1333] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[1334] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[1335] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[1336] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[1337] Train Loss:    tensor(0.6776, device='cuda:0', grad_fn=<NllLossBackward>)
[1338] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[1339] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[1340] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[1341] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1342] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[1343] Train Loss:    tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[1344] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1345] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[1346] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[1347] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[1348] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[1349] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[1350] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[1351] Train Loss:    tensor(0.6557, device='cuda:0', grad_fn=<NllLossBackward>)
[1352] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[1353] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[1354] Train Loss:    tensor(0.6594, device='cuda:0', grad_fn=<NllLossBackward>)
[1355] Train Loss:    tensor(0.6638, device='cuda:0', grad_fn=<NllLossBackward>)
[1356] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[1357] Train Loss:    tensor(0.6612, device='cuda:0', grad_fn=<NllLossBackward>)
[1358] Train Loss:    tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[1359] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[1360] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[1361] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[1362] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[1363] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[1364] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1365] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[1366] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[1367] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[1368] Train Loss:    tensor(0.6694, device='cuda:0', grad_fn=<NllLossBackward>)
[1369] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[1370] Train Loss:    tensor(0.6605, device='cuda:0', grad_fn=<NllLossBackward>)
[1371] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[1372] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1373] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[1374] Train Loss:    tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[1375] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[1376] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[1377] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[1378] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[1379] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[1380] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[1381] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1382] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[1383] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[1384] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[1385] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[1386] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[1387] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[1388] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[1389] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[1390] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[1391] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[1392] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1393] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[1394] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
[1395] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1396] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[1397] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[1398] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[1399] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[1400] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[1401] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[1402] Train Loss:    tensor(0.6514, device='cuda:0', grad_fn=<NllLossBackward>)
[1403] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[1404] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[1405] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[1406] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1407] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[1408] Train Loss:    tensor(0.6549, device='cuda:0', grad_fn=<NllLossBackward>)
[1409] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[1410] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[1411] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[1412] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[1413] Train Loss:    tensor(0.7245, device='cuda:0', grad_fn=<NllLossBackward>)
[1414] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[1415] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[1416] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[1417] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[1418] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1419] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[1420] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[1421] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[1422] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[1423] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[1424] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[1425] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[1426] Train Loss:    tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[1427] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[1428] Train Loss:    tensor(0.6563, device='cuda:0', grad_fn=<NllLossBackward>)
[1429] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[1430] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[1431] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[1432] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[1433] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[1434] Train Loss:    tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[1435] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[1436] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[1437] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[1438] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[1439] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[1440] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1441] Train Loss:   tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1442] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[1443] Train Loss:   tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[1444] Train Loss:   tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[1445] Train Loss:   tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[1446] Train Loss:   tensor(0.7345, device='cuda:0', grad_fn=<NllLossBackward>)
[1447] Train Loss:   tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1448] Train Loss:   tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[1449] Train Loss:   tensor(0.6712, device='cuda:0', grad_fn=<NllLossBackward>)
[1450] Train Loss:   tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[1451] Train Loss:   tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[1452] Train Loss:   tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1453] Train Loss:   tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[1454] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[1455] Train Loss:   tensor(0.7339, device='cuda:0', grad_fn=<NllLossBackward>)
[1456] Train Loss:   tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[1457] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1458] Train Loss:   tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[1459] Train Loss:   tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[1460] Train Loss:   tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[1461] Train Loss:   tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[1462] Train Loss:   tensor(0.6579, device='cuda:0', grad_fn=<NllLossBackward>)
[1463] Train Loss:   tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[1464] Train Loss:   tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[1465] Train Loss:   tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[1466] Train Loss:   tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[1467] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1468] Train Loss:   tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[1469] Train Loss:   tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[1470] Train Loss:   tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[1471] Train Loss:   tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1472] Train Loss:   tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[1473] Train Loss:   tensor(0.6683, device='cuda:0', grad_fn=<NllLossBackward>)
[1474] Train Loss:   tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[1475] Train Loss:   tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[1476] Train Loss:   tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[1477] Train Loss:   tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[1478] Train Loss:   tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[1479] Train Loss:   tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[1480] Train Loss:   tensor(0.6503, device='cuda:0', grad_fn=<NllLossBackward>)
[1481] Train Loss:   tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[1482] Train Loss:   tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[1483] Train Loss:   tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[1484] Train Loss:   tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[1485] Train Loss:   tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[1486] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[1487] Train Loss:   tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[1488] Train Loss:   tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1489] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[1490] Train Loss:    tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[1491] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1492] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1493] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[1494] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[1495] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[1496] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[1497] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[1498] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[1499] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[1500] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[1501] Train Loss:    tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[1502] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[1503] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[1504] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[1505] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[1506] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[1507] Train Loss:    tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[1508] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[1509] Train Loss:    tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[1510] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1511] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[1512] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[1513] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[1514] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[1515] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[1516] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[1517] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[1518] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[1519] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[1520] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[1521] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[1522] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[1523] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[1524] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[1525] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[1526] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1527] Train Loss:    tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[1528] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[1529] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[1530] Train Loss:    tensor(0.6546, device='cuda:0', grad_fn=<NllLossBackward>)
[1531] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[1532] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[1533] Train Loss:    tensor(0.6659, device='cuda:0', grad_fn=<NllLossBackward>)
[1534] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[1535] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[1536] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1537] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[1538] Train Loss:    tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[1539] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[1540] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[1541] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[1542] Train Loss:    tensor(0.7334, device='cuda:0', grad_fn=<NllLossBackward>)
[1543] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1544] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[1545] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[1546] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[1547] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[1548] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[1549] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[1550] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[1551] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[1552] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[1553] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[1554] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[1555] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[1556] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[1557] Train Loss:    tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[1558] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[1559] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[1560] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[1561] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[1562] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[1563] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[1564] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1565] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[1566] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[1567] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[1568] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[1569] Train Loss:    tensor(0.6691, device='cuda:0', grad_fn=<NllLossBackward>)
[1570] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[1571] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[1572] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[1573] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[1574] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[1575] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[1576] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[1577] Train Loss:    tensor(0.7347, device='cuda:0', grad_fn=<NllLossBackward>)
[1578] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[1579] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[1580] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[1581] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[1582] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[1583] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[1584] Train Loss:    tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1585] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[1586] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[1587] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[1588] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[1589] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[1590] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[1591] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[1592] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[1593] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[1594] Train Loss:    tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[1595] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[1596] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[1597] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[1598] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[1599] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1600] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[1601] Train Loss:    tensor(0.7336, device='cuda:0', grad_fn=<NllLossBackward>)
[1602] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[1603] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[1604] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[1605] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[1606] Train Loss:    tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[1607] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[1608] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1609] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[1610] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[1611] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[1612] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[1613] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[1614] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[1615] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[1616] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[1617] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[1618] Train Loss:    tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[1619] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1620] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[1621] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1622] Train Loss:    tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[1623] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[1624] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[1625] Train Loss:    tensor(0.6628, device='cuda:0', grad_fn=<NllLossBackward>)
[1626] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[1627] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[1628] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[1629] Train Loss:    tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[1630] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[1631] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[1632] Train Loss:    tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1633] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[1634] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[1635] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[1636] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[1637] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[1638] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[1639] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[1640] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[1641] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[1642] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[1643] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[1644] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[1645] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1646] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[1647] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[1648] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1649] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[1650] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[1651] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[1652] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[1653] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[1654] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[1655] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[1656] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[1657] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[1658] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1659] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[1660] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[1661] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[1662] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[1663] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[1664] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[1665] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[1666] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1667] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[1668] Train Loss:    tensor(0.6733, device='cuda:0', grad_fn=<NllLossBackward>)
[1669] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1670] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[1671] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[1672] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1673] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1674] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[1675] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[1676] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[1677] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[1678] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[1679] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[1680] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1681] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[1682] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[1683] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[1684] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[1685] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[1686] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[1687] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[1688] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[1689] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[1690] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[1691] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[1692] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[1693] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[1694] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[1695] Train Loss:    tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[1696] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[1697] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[1698] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[1699] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[1700] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[1701] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[1702] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[1703] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[1704] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[1705] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[1706] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[1707] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[1708] Train Loss:    tensor(0.7367, device='cuda:0', grad_fn=<NllLossBackward>)
[1709] Train Loss:    tensor(0.6591, device='cuda:0', grad_fn=<NllLossBackward>)
[1710] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[1711] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[1712] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[1713] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[1714] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[1715] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[1716] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[1717] Train Loss:    tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[1718] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1719] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[1720] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[1721] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[1722] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[1723] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1724] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[1725] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1726] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[1727] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[1728] Train Loss:    tensor(0.6632, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1729] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[1730] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[1731] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[1732] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1733] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[1734] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[1735] Train Loss:    tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[1736] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1737] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[1738] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1739] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[1740] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[1741] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[1742] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[1743] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1744] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[1745] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[1746] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[1747] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[1748] Train Loss:    tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[1749] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[1750] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[1751] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[1752] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[1753] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[1754] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[1755] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[1756] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[1757] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[1758] Train Loss:    tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[1759] Train Loss:    tensor(0.7271, device='cuda:0', grad_fn=<NllLossBackward>)
[1760] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[1761] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[1762] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[1763] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[1764] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[1765] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[1766] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[1767] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[1768] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[1769] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[1770] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[1771] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[1772] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[1773] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[1774] Train Loss:    tensor(0.6635, device='cuda:0', grad_fn=<NllLossBackward>)
[1775] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[1776] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1777] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[1778] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[1779] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[1780] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[1781] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[1782] Train Loss:    tensor(0.6598, device='cuda:0', grad_fn=<NllLossBackward>)
[1783] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[1784] Train Loss:    tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[1785] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[1786] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[1787] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[1788] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[1789] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[1790] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[1791] Train Loss:    tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[1792] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[1793] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[1794] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[1795] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[1796] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[1797] Train Loss:    tensor(0.6612, device='cuda:0', grad_fn=<NllLossBackward>)
[1798] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[1799] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1800] Train Loss:    tensor(0.6592, device='cuda:0', grad_fn=<NllLossBackward>)
[1801] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[1802] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[1803] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1804] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1805] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1806] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[1807] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[1808] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1809] Train Loss:    tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[1810] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[1811] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[1812] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[1813] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[1814] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[1815] Train Loss:    tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[1816] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[1817] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[1818] Train Loss:    tensor(0.6776, device='cuda:0', grad_fn=<NllLossBackward>)
[1819] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[1820] Train Loss:    tensor(0.7417, device='cuda:0', grad_fn=<NllLossBackward>)
[1821] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1822] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[1823] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[1824] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1825] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[1826] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[1827] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[1828] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[1829] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[1830] Train Loss:    tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[1831] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[1832] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[1833] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[1834] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[1835] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[1836] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[1837] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[1838] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[1839] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[1840] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[1841] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[1842] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[1843] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[1844] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[1845] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[1846] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[1847] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[1848] Train Loss:    tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[1849] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1850] Train Loss:    tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[1851] Train Loss:    tensor(0.6709, device='cuda:0', grad_fn=<NllLossBackward>)
[1852] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[1853] Train Loss:    tensor(0.6722, device='cuda:0', grad_fn=<NllLossBackward>)
[1854] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[1855] Train Loss:    tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[1856] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[1857] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[1858] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[1859] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[1860] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[1861] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[1862] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[1863] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[1864] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[1865] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[1866] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[1867] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1868] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[1869] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[1870] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[1871] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[1872] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1873] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[1874] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[1875] Train Loss:    tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[1876] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[1877] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[1878] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[1879] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[1880] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[1881] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[1882] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[1883] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[1884] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[1885] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[1886] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[1887] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[1888] Train Loss:    tensor(0.6679, device='cuda:0', grad_fn=<NllLossBackward>)
[1889] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[1890] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[1891] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[1892] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[1893] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1894] Train Loss:    tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
[1895] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[1896] Train Loss:    tensor(0.6688, device='cuda:0', grad_fn=<NllLossBackward>)
[1897] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1898] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[1899] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[1900] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[1901] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[1902] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[1903] Train Loss:    tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[1904] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[1905] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[1906] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[1907] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[1908] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1909] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[1910] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[1911] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[1912] Train Loss:    tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[1913] Train Loss:    tensor(0.6665, device='cuda:0', grad_fn=<NllLossBackward>)
[1914] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[1915] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1916] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[1917] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[1918] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[1919] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[1920] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1921] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[1922] Train Loss:    tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[1923] Train Loss:    tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[1924] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[1925] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[1926] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[1927] Train Loss:    tensor(0.6504, device='cuda:0', grad_fn=<NllLossBackward>)
[1928] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[1929] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[1930] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[1931] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[1932] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[1933] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[1934] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[1935] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[1936] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[1937] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[1938] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[1939] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[1940] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[1941] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[1942] Train Loss:    tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[1943] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[1944] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1945] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[1946] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[1947] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1948] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1949] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[1950] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[1951] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[1952] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[1953] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[1954] Train Loss:    tensor(0.6598, device='cuda:0', grad_fn=<NllLossBackward>)
[1955] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[1956] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[1957] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[1958] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[1959] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[1960] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[1961] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[1962] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[1963] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[1964] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[1965] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[1966] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[1967] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[1968] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1969] Train Loss:   tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[1970] Train Loss:   tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[1971] Train Loss:   tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[1972] Train Loss:   tensor(0.6628, device='cuda:0', grad_fn=<NllLossBackward>)
[1973] Train Loss:   tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[1974] Train Loss:   tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[1975] Train Loss:   tensor(0.7354, device='cuda:0', grad_fn=<NllLossBackward>)
[1976] Train Loss:   tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[1977] Train Loss:   tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[1978] Train Loss:   tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[1979] Train Loss:   tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[1980] Train Loss:   tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[1981] Train Loss:   tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[1982] Train Loss:   tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[1983] Train Loss:   tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[1984] Train Loss:   tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1985] Train Loss:   tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1986] Train Loss:   tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[1987] Train Loss:   tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[1988] Train Loss:   tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[1989] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[1990] Train Loss:   tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[1991] Train Loss:   tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[1992] Train Loss:   tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[1993] Train Loss:   tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[1994] Train Loss:   tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1995] Train Loss:   tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[1996] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[1997] Train Loss:   tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[1998] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[1999] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[2000] Train Loss:   tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[2001] Train Loss:   tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[2002] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[2003] Train Loss:   tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[2004] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[2005] Train Loss:   tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[2006] Train Loss:   tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[2007] Train Loss:   tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[2008] Train Loss:   tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[2009] Train Loss:   tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[2010] Train Loss:   tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
[2011] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[2012] Train Loss:   tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[2013] Train Loss:   tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[2014] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[2015] Train Loss:   tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[2016] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2017] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[2018] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[2019] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[2020] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[2021] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[2022] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[2023] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[2024] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[2025] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[2026] Train Loss:    tensor(0.6627, device='cuda:0', grad_fn=<NllLossBackward>)
[2027] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2028] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[2029] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2030] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[2031] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[2032] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[2033] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[2034] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[2035] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[2036] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[2037] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[2038] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[2039] Train Loss:    tensor(0.6606, device='cuda:0', grad_fn=<NllLossBackward>)
[2040] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[2041] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[2042] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[2043] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[2044] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[2045] Train Loss:    tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[2046] Train Loss:    tensor(0.7403, device='cuda:0', grad_fn=<NllLossBackward>)
[2047] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[2048] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2049] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[2050] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[2051] Train Loss:    tensor(0.6587, device='cuda:0', grad_fn=<NllLossBackward>)
[2052] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2053] Train Loss:    tensor(0.7397, device='cuda:0', grad_fn=<NllLossBackward>)
[2054] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[2055] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[2056] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[2057] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[2058] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[2059] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2060] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[2061] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[2062] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2063] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[2064] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2065] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[2066] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[2067] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[2068] Train Loss:    tensor(0.7282, device='cuda:0', grad_fn=<NllLossBackward>)
[2069] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[2070] Train Loss:    tensor(0.7462, device='cuda:0', grad_fn=<NllLossBackward>)
[2071] Train Loss:    tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[2072] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[2073] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[2074] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[2075] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[2076] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[2077] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[2078] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[2079] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[2080] Train Loss:    tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[2081] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[2082] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[2083] Train Loss:    tensor(0.6712, device='cuda:0', grad_fn=<NllLossBackward>)
[2084] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2085] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[2086] Train Loss:    tensor(0.7286, device='cuda:0', grad_fn=<NllLossBackward>)
[2087] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[2088] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[2089] Train Loss:    tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[2090] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[2091] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[2092] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[2093] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[2094] Train Loss:    tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[2095] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[2096] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[2097] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[2098] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[2099] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[2100] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[2101] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[2102] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[2103] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[2104] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[2105] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2106] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[2107] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[2108] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2109] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[2110] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[2111] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[2112] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2113] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[2114] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[2115] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[2116] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[2117] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
[2118] Train Loss:    tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[2119] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[2120] Train Loss:    tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[2121] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2122] Train Loss:    tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[2123] Train Loss:    tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[2124] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[2125] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2126] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[2127] Train Loss:    tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[2128] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[2129] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2130] Train Loss:    tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[2131] Train Loss:    tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[2132] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[2133] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[2134] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2135] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[2136] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[2137] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[2138] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[2139] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[2140] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[2141] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[2142] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[2143] Train Loss:    tensor(0.7486, device='cuda:0', grad_fn=<NllLossBackward>)
[2144] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[2145] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[2146] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[2147] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[2148] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[2149] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2150] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2151] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[2152] Train Loss:    tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[2153] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[2154] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[2155] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[2156] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[2157] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[2158] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[2159] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[2160] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2161] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[2162] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[2163] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[2164] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2165] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[2166] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[2167] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[2168] Train Loss:    tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[2169] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[2170] Train Loss:    tensor(0.6628, device='cuda:0', grad_fn=<NllLossBackward>)
[2171] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[2172] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[2173] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[2174] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[2175] Train Loss:    tensor(0.6624, device='cuda:0', grad_fn=<NllLossBackward>)
[2176] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[2177] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[2178] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[2179] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[2180] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[2181] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[2182] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[2183] Train Loss:    tensor(0.6647, device='cuda:0', grad_fn=<NllLossBackward>)
[2184] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[2185] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[2186] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[2187] Train Loss:    tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[2188] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2189] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[2190] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[2191] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[2192] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[2193] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[2194] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2195] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2196] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[2197] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[2198] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[2199] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[2200] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[2201] Train Loss:    tensor(0.6630, device='cuda:0', grad_fn=<NllLossBackward>)
[2202] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[2203] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[2204] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[2205] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[2206] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[2207] Train Loss:    tensor(0.6729, device='cuda:0', grad_fn=<NllLossBackward>)
[2208] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2209] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2210] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[2211] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[2212] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[2213] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[2214] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[2215] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[2216] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2217] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2218] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[2219] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[2220] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[2221] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[2222] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2223] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[2224] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[2225] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[2226] Train Loss:    tensor(0.6558, device='cuda:0', grad_fn=<NllLossBackward>)
[2227] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[2228] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[2229] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[2230] Train Loss:    tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[2231] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[2232] Train Loss:    tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[2233] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[2234] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2235] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[2236] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[2237] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2238] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[2239] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[2240] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[2241] Train Loss:    tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[2242] Train Loss:    tensor(0.7699, device='cuda:0', grad_fn=<NllLossBackward>)
[2243] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[2244] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[2245] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[2246] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[2247] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2248] Train Loss:    tensor(0.6686, device='cuda:0', grad_fn=<NllLossBackward>)
[2249] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[2250] Train Loss:    tensor(0.6564, device='cuda:0', grad_fn=<NllLossBackward>)
[2251] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[2252] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[2253] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[2254] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[2255] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[2256] Train Loss:    tensor(0.6705, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2257] Train Loss:   tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[2258] Train Loss:   tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[2259] Train Loss:   tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[2260] Train Loss:   tensor(0.7414, device='cuda:0', grad_fn=<NllLossBackward>)
[2261] Train Loss:   tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[2262] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[2263] Train Loss:   tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[2264] Train Loss:   tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[2265] Train Loss:   tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[2266] Train Loss:   tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[2267] Train Loss:   tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[2268] Train Loss:   tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[2269] Train Loss:   tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[2270] Train Loss:   tensor(0.6602, device='cuda:0', grad_fn=<NllLossBackward>)
[2271] Train Loss:   tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[2272] Train Loss:   tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[2273] Train Loss:   tensor(0.6619, device='cuda:0', grad_fn=<NllLossBackward>)
[2274] Train Loss:   tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[2275] Train Loss:   tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[2276] Train Loss:   tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[2277] Train Loss:   tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[2278] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[2279] Train Loss:   tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[2280] Train Loss:   tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[2281] Train Loss:   tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[2282] Train Loss:   tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[2283] Train Loss:   tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[2284] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[2285] Train Loss:   tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[2286] Train Loss:   tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[2287] Train Loss:   tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[2288] Train Loss:   tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[2289] Train Loss:   tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[2290] Train Loss:   tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[2291] Train Loss:   tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[2292] Train Loss:   tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[2293] Train Loss:   tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[2294] Train Loss:   tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[2295] Train Loss:   tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[2296] Train Loss:   tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2297] Train Loss:   tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[2298] Train Loss:   tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[2299] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[2300] Train Loss:   tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[2301] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[2302] Train Loss:   tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[2303] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[2304] Train Loss:   tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2305] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[2306] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[2307] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[2308] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[2309] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[2310] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2311] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[2312] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[2313] Train Loss:    tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[2314] Train Loss:    tensor(0.6655, device='cuda:0', grad_fn=<NllLossBackward>)
[2315] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[2316] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[2317] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[2318] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[2319] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[2320] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[2321] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[2322] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2323] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[2324] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[2325] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2326] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[2327] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[2328] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[2329] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[2330] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[2331] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[2332] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[2333] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[2334] Train Loss:    tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[2335] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[2336] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[2337] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[2338] Train Loss:    tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[2339] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[2340] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[2341] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[2342] Train Loss:    tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[2343] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2344] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[2345] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[2346] Train Loss:    tensor(0.6430, device='cuda:0', grad_fn=<NllLossBackward>)
[2347] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[2348] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[2349] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[2350] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[2351] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[2352] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2353] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[2354] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[2355] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[2356] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[2357] Train Loss:    tensor(0.6483, device='cuda:0', grad_fn=<NllLossBackward>)
[2358] Train Loss:    tensor(0.6636, device='cuda:0', grad_fn=<NllLossBackward>)
[2359] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[2360] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[2361] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2362] Train Loss:    tensor(0.7307, device='cuda:0', grad_fn=<NllLossBackward>)
[2363] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[2364] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[2365] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[2366] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[2367] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[2368] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[2369] Train Loss:    tensor(0.6704, device='cuda:0', grad_fn=<NllLossBackward>)
[2370] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[2371] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[2372] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[2373] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[2374] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[2375] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[2376] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[2377] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[2378] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[2379] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[2380] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[2381] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2382] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[2383] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[2384] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[2385] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[2386] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[2387] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[2388] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[2389] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[2390] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[2391] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[2392] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[2393] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[2394] Train Loss:    tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[2395] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[2396] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[2397] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2398] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[2399] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[2400] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2401] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[2402] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[2403] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[2404] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[2405] Train Loss:    tensor(0.6611, device='cuda:0', grad_fn=<NllLossBackward>)
[2406] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[2407] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[2408] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[2409] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[2410] Train Loss:    tensor(0.6637, device='cuda:0', grad_fn=<NllLossBackward>)
[2411] Train Loss:    tensor(0.6557, device='cuda:0', grad_fn=<NllLossBackward>)
[2412] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[2413] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[2414] Train Loss:    tensor(0.7491, device='cuda:0', grad_fn=<NllLossBackward>)
[2415] Train Loss:    tensor(0.6683, device='cuda:0', grad_fn=<NllLossBackward>)
[2416] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[2417] Train Loss:    tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[2418] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[2419] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[2420] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[2421] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[2422] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[2423] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[2424] Train Loss:    tensor(0.6583, device='cuda:0', grad_fn=<NllLossBackward>)
[2425] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[2426] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[2427] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[2428] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[2429] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[2430] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[2431] Train Loss:    tensor(0.6642, device='cuda:0', grad_fn=<NllLossBackward>)
[2432] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[2433] Train Loss:    tensor(0.6552, device='cuda:0', grad_fn=<NllLossBackward>)
[2434] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[2435] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[2436] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[2437] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[2438] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[2439] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[2440] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[2441] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[2442] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[2443] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[2444] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[2445] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[2446] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[2447] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[2448] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2449] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[2450] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2451] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[2452] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[2453] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[2454] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[2455] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[2456] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[2457] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[2458] Train Loss:    tensor(0.6709, device='cuda:0', grad_fn=<NllLossBackward>)
[2459] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[2460] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2461] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[2462] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[2463] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[2464] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[2465] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[2466] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[2467] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2468] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[2469] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[2470] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[2471] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[2472] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2473] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[2474] Train Loss:    tensor(0.6661, device='cuda:0', grad_fn=<NllLossBackward>)
[2475] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[2476] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[2477] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[2478] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[2479] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[2480] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[2481] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[2482] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[2483] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[2484] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[2485] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2486] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2487] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[2488] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[2489] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[2490] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[2491] Train Loss:    tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[2492] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[2493] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2494] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[2495] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[2496] Train Loss:    tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2497] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[2498] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[2499] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[2500] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[2501] Train Loss:    tensor(0.6543, device='cuda:0', grad_fn=<NllLossBackward>)
[2502] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[2503] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[2504] Train Loss:    tensor(0.6709, device='cuda:0', grad_fn=<NllLossBackward>)
[2505] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[2506] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[2507] Train Loss:    tensor(0.6615, device='cuda:0', grad_fn=<NllLossBackward>)
[2508] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[2509] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[2510] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[2511] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[2512] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[2513] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[2514] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[2515] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[2516] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[2517] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[2518] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[2519] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[2520] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[2521] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2522] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[2523] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[2524] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2525] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[2526] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[2527] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[2528] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2529] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[2530] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[2531] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[2532] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[2533] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2534] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[2535] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[2536] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[2537] Train Loss:    tensor(0.6665, device='cuda:0', grad_fn=<NllLossBackward>)
[2538] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[2539] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[2540] Train Loss:    tensor(0.7390, device='cuda:0', grad_fn=<NllLossBackward>)
[2541] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[2542] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2543] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2544] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2545] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[2546] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2547] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[2548] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[2549] Train Loss:    tensor(0.6733, device='cuda:0', grad_fn=<NllLossBackward>)
[2550] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[2551] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[2552] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[2553] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[2554] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[2555] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[2556] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2557] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[2558] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[2559] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[2560] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[2561] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[2562] Train Loss:    tensor(0.6576, device='cuda:0', grad_fn=<NllLossBackward>)
[2563] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[2564] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[2565] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[2566] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[2567] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[2568] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[2569] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[2570] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[2571] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[2572] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[2573] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[2574] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[2575] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[2576] Train Loss:    tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[2577] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[2578] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[2579] Train Loss:    tensor(0.7328, device='cuda:0', grad_fn=<NllLossBackward>)
[2580] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2581] Train Loss:    tensor(0.7381, device='cuda:0', grad_fn=<NllLossBackward>)
[2582] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[2583] Train Loss:    tensor(0.7588, device='cuda:0', grad_fn=<NllLossBackward>)
[2584] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2585] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[2586] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[2587] Train Loss:    tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[2588] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[2589] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[2590] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[2591] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[2592] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2593] Train Loss:   tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[2594] Train Loss:   tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[2595] Train Loss:   tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[2596] Train Loss:   tensor(0.6653, device='cuda:0', grad_fn=<NllLossBackward>)
[2597] Train Loss:   tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[2598] Train Loss:   tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[2599] Train Loss:   tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[2600] Train Loss:   tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[2601] Train Loss:   tensor(0.6545, device='cuda:0', grad_fn=<NllLossBackward>)
[2602] Train Loss:   tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2603] Train Loss:   tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2604] Train Loss:   tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[2605] Train Loss:   tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[2606] Train Loss:   tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[2607] Train Loss:   tensor(0.6661, device='cuda:0', grad_fn=<NllLossBackward>)
[2608] Train Loss:   tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[2609] Train Loss:   tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[2610] Train Loss:   tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[2611] Train Loss:   tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[2612] Train Loss:   tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[2613] Train Loss:   tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[2614] Train Loss:   tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[2615] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[2616] Train Loss:   tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[2617] Train Loss:   tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[2618] Train Loss:   tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[2619] Train Loss:   tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[2620] Train Loss:   tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[2621] Train Loss:   tensor(0.6695, device='cuda:0', grad_fn=<NllLossBackward>)
[2622] Train Loss:   tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[2623] Train Loss:   tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[2624] Train Loss:   tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[2625] Train Loss:   tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[2626] Train Loss:   tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[2627] Train Loss:   tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[2628] Train Loss:   tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[2629] Train Loss:   tensor(0.6669, device='cuda:0', grad_fn=<NllLossBackward>)
[2630] Train Loss:   tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[2631] Train Loss:   tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[2632] Train Loss:   tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[2633] Train Loss:   tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2634] Train Loss:   tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[2635] Train Loss:   tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[2636] Train Loss:   tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[2637] Train Loss:   tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[2638] Train Loss:   tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[2639] Train Loss:   tensor(0.7305, device='cuda:0', grad_fn=<NllLossBackward>)
[2640] Train Loss:   tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2641] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[2642] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[2643] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[2644] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[2645] Train Loss:    tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[2646] Train Loss:    tensor(0.6722, device='cuda:0', grad_fn=<NllLossBackward>)
[2647] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[2648] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[2649] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[2650] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[2651] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[2652] Train Loss:    tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[2653] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[2654] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[2655] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[2656] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[2657] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[2658] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[2659] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[2660] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[2661] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[2662] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[2663] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[2664] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2665] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[2666] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[2667] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[2668] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[2669] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[2670] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[2671] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[2672] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[2673] Train Loss:    tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[2674] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[2675] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[2676] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[2677] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[2678] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[2679] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[2680] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[2681] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[2682] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[2683] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[2684] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[2685] Train Loss:    tensor(0.6592, device='cuda:0', grad_fn=<NllLossBackward>)
[2686] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[2687] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[2688] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2689] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[2690] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[2691] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[2692] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2693] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[2694] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[2695] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[2696] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[2697] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[2698] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[2699] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[2700] Train Loss:    tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[2701] Train Loss:    tensor(0.6583, device='cuda:0', grad_fn=<NllLossBackward>)
[2702] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[2703] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[2704] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[2705] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[2706] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[2707] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[2708] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[2709] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[2710] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[2711] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2712] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[2713] Train Loss:    tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[2714] Train Loss:    tensor(0.6602, device='cuda:0', grad_fn=<NllLossBackward>)
[2715] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[2716] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[2717] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[2718] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[2719] Train Loss:    tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[2720] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2721] Train Loss:    tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
[2722] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2723] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[2724] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[2725] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[2726] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[2727] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[2728] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[2729] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[2730] Train Loss:    tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[2731] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[2732] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[2733] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[2734] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[2735] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2736] Train Loss:    tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2737] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[2738] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[2739] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[2740] Train Loss:    tensor(0.7298, device='cuda:0', grad_fn=<NllLossBackward>)
[2741] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[2742] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[2743] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[2744] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[2745] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[2746] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[2747] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[2748] Train Loss:    tensor(0.6590, device='cuda:0', grad_fn=<NllLossBackward>)
[2749] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[2750] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[2751] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[2752] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2753] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[2754] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[2755] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[2756] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2757] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[2758] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[2759] Train Loss:    tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[2760] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[2761] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[2762] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[2763] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[2764] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[2765] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[2766] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[2767] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[2768] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[2769] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[2770] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2771] Train Loss:    tensor(0.7373, device='cuda:0', grad_fn=<NllLossBackward>)
[2772] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[2773] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2774] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[2775] Train Loss:    tensor(0.6715, device='cuda:0', grad_fn=<NllLossBackward>)
[2776] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[2777] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[2778] Train Loss:    tensor(0.6429, device='cuda:0', grad_fn=<NllLossBackward>)
[2779] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[2780] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[2781] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[2782] Train Loss:    tensor(0.6572, device='cuda:0', grad_fn=<NllLossBackward>)
[2783] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2784] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2785] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[2786] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2787] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[2788] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[2789] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[2790] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[2791] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[2792] Train Loss:    tensor(0.7376, device='cuda:0', grad_fn=<NllLossBackward>)
[2793] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[2794] Train Loss:    tensor(0.7267, device='cuda:0', grad_fn=<NllLossBackward>)
[2795] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[2796] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[2797] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2798] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[2799] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[2800] Train Loss:    tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[2801] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2802] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[2803] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[2804] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[2805] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[2806] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[2807] Train Loss:    tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[2808] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[2809] Train Loss:    tensor(0.7318, device='cuda:0', grad_fn=<NllLossBackward>)
[2810] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[2811] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[2812] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[2813] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[2814] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2815] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[2816] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[2817] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[2818] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[2819] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[2820] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[2821] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[2822] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[2823] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[2824] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[2825] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[2826] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2827] Train Loss:    tensor(0.7321, device='cuda:0', grad_fn=<NllLossBackward>)
[2828] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2829] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[2830] Train Loss:    tensor(0.7500, device='cuda:0', grad_fn=<NllLossBackward>)
[2831] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[2832] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2833] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[2834] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[2835] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2836] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[2837] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[2838] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[2839] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[2840] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2841] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[2842] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[2843] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[2844] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[2845] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[2846] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[2847] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[2848] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2849] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[2850] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[2851] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[2852] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[2853] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[2854] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[2855] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[2856] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[2857] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[2858] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[2859] Train Loss:    tensor(0.7308, device='cuda:0', grad_fn=<NllLossBackward>)
[2860] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[2861] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[2862] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[2863] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[2864] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[2865] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2866] Train Loss:    tensor(0.6562, device='cuda:0', grad_fn=<NllLossBackward>)
[2867] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[2868] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[2869] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2870] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[2871] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[2872] Train Loss:    tensor(0.6715, device='cuda:0', grad_fn=<NllLossBackward>)
[2873] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[2874] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[2875] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[2876] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[2877] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[2878] Train Loss:    tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[2879] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[2880] Train Loss:    tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2881] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2882] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[2883] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[2884] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[2885] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[2886] Train Loss:    tensor(0.6332, device='cuda:0', grad_fn=<NllLossBackward>)
[2887] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[2888] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[2889] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[2890] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[2891] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[2892] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[2893] Train Loss:    tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[2894] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[2895] Train Loss:    tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[2896] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[2897] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[2898] Train Loss:    tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[2899] Train Loss:    tensor(0.7222, device='cuda:0', grad_fn=<NllLossBackward>)
[2900] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[2901] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[2902] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[2903] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[2904] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[2905] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[2906] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[2907] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2908] Train Loss:    tensor(0.6659, device='cuda:0', grad_fn=<NllLossBackward>)
[2909] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[2910] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[2911] Train Loss:    tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[2912] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[2913] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[2914] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2915] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[2916] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[2917] Train Loss:    tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[2918] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[2919] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[2920] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[2921] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[2922] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[2923] Train Loss:    tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[2924] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[2925] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[2926] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[2927] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[2928] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2929] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[2930] Train Loss:    tensor(0.6639, device='cuda:0', grad_fn=<NllLossBackward>)
[2931] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[2932] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[2933] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2934] Train Loss:    tensor(0.6618, device='cuda:0', grad_fn=<NllLossBackward>)
[2935] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[2936] Train Loss:    tensor(0.6739, device='cuda:0', grad_fn=<NllLossBackward>)
[2937] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[2938] Train Loss:    tensor(0.7442, device='cuda:0', grad_fn=<NllLossBackward>)
[2939] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[2940] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[2941] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[2942] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[2943] Train Loss:    tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[2944] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[2945] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[2946] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[2947] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2948] Train Loss:    tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[2949] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[2950] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[2951] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2952] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[2953] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[2954] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[2955] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[2956] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[2957] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2958] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[2959] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[2960] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[2961] Train Loss:    tensor(0.6695, device='cuda:0', grad_fn=<NllLossBackward>)
[2962] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[2963] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2964] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2965] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[2966] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[2967] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[2968] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2969] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[2970] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[2971] Train Loss:    tensor(0.6688, device='cuda:0', grad_fn=<NllLossBackward>)
[2972] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[2973] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[2974] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[2975] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[2976] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2977] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[2978] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[2979] Train Loss:    tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[2980] Train Loss:    tensor(0.7333, device='cuda:0', grad_fn=<NllLossBackward>)
[2981] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[2982] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[2983] Train Loss:    tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[2984] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2985] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[2986] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[2987] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[2988] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[2989] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[2990] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[2991] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[2992] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[2993] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[2994] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[2995] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[2996] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[2997] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[2998] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[2999] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[3000] Train Loss:    tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[3001] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[3002] Train Loss:    tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[3003] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[3004] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[3005] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[3006] Train Loss:    tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[3007] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[3008] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[3009] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[3010] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[3011] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[3012] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[3013] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[3014] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[3015] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[3016] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[3017] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[3018] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[3019] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[3020] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[3021] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[3022] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[3023] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[3024] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3025] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[3026] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[3027] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[3028] Train Loss:    tensor(0.6714, device='cuda:0', grad_fn=<NllLossBackward>)
[3029] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[3030] Train Loss:    tensor(0.7449, device='cuda:0', grad_fn=<NllLossBackward>)
[3031] Train Loss:    tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[3032] Train Loss:    tensor(0.6519, device='cuda:0', grad_fn=<NllLossBackward>)
[3033] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[3034] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[3035] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3036] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[3037] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[3038] Train Loss:    tensor(0.6568, device='cuda:0', grad_fn=<NllLossBackward>)
[3039] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[3040] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[3041] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[3042] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[3043] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[3044] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[3045] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3046] Train Loss:    tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[3047] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[3048] Train Loss:    tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[3049] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[3050] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
[3051] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[3052] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[3053] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[3054] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[3055] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[3056] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3057] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[3058] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[3059] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3060] Train Loss:    tensor(0.7562, device='cuda:0', grad_fn=<NllLossBackward>)
[3061] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3062] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3063] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[3064] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[3065] Train Loss:    tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[3066] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[3067] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[3068] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[3069] Train Loss:    tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[3070] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[3071] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[3072] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3073] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[3074] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[3075] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[3076] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[3077] Train Loss:    tensor(0.6597, device='cuda:0', grad_fn=<NllLossBackward>)
[3078] Train Loss:    tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[3079] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[3080] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[3081] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[3082] Train Loss:    tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[3083] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[3084] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3085] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3086] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[3087] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[3088] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[3089] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[3090] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[3091] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[3092] Train Loss:    tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[3093] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[3094] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[3095] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[3096] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[3097] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3098] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[3099] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3100] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[3101] Train Loss:    tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[3102] Train Loss:    tensor(0.7370, device='cuda:0', grad_fn=<NllLossBackward>)
[3103] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[3104] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[3105] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[3106] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[3107] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[3108] Train Loss:    tensor(0.6687, device='cuda:0', grad_fn=<NllLossBackward>)
[3109] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3110] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[3111] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3112] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[3113] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[3114] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[3115] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[3116] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[3117] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[3118] Train Loss:    tensor(0.6481, device='cuda:0', grad_fn=<NllLossBackward>)
[3119] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[3120] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3121] Train Loss:   tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[3122] Train Loss:   tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[3123] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[3124] Train Loss:   tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[3125] Train Loss:   tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[3126] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3127] Train Loss:   tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[3128] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[3129] Train Loss:   tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[3130] Train Loss:   tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3131] Train Loss:   tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3132] Train Loss:   tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[3133] Train Loss:   tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3134] Train Loss:   tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[3135] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[3136] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3137] Train Loss:   tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3138] Train Loss:   tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[3139] Train Loss:   tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[3140] Train Loss:   tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[3141] Train Loss:   tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3142] Train Loss:   tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[3143] Train Loss:   tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[3144] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[3145] Train Loss:   tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[3146] Train Loss:   tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[3147] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3148] Train Loss:   tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[3149] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3150] Train Loss:   tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[3151] Train Loss:   tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[3152] Train Loss:   tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[3153] Train Loss:   tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[3154] Train Loss:   tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[3155] Train Loss:   tensor(0.6671, device='cuda:0', grad_fn=<NllLossBackward>)
[3156] Train Loss:   tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[3157] Train Loss:   tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[3158] Train Loss:   tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[3159] Train Loss:   tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[3160] Train Loss:   tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[3161] Train Loss:   tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[3162] Train Loss:   tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[3163] Train Loss:   tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[3164] Train Loss:   tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3165] Train Loss:   tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[3166] Train Loss:   tensor(0.6712, device='cuda:0', grad_fn=<NllLossBackward>)
[3167] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[3168] Train Loss:   tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3169] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[3170] Train Loss:    tensor(0.6593, device='cuda:0', grad_fn=<NllLossBackward>)
[3171] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[3172] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3173] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3174] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[3175] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[3176] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[3177] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3178] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[3179] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[3180] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[3181] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[3182] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[3183] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[3184] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[3185] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[3186] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[3187] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[3188] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[3189] Train Loss:    tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[3190] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[3191] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[3192] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[3193] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[3194] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[3195] Train Loss:    tensor(0.6601, device='cuda:0', grad_fn=<NllLossBackward>)
[3196] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[3197] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[3198] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[3199] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[3200] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[3201] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[3202] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3203] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[3204] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[3205] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[3206] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[3207] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[3208] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[3209] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3210] Train Loss:    tensor(0.6649, device='cuda:0', grad_fn=<NllLossBackward>)
[3211] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[3212] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[3213] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[3214] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3215] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[3216] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3217] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[3218] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[3219] Train Loss:    tensor(0.6533, device='cuda:0', grad_fn=<NllLossBackward>)
[3220] Train Loss:    tensor(0.7324, device='cuda:0', grad_fn=<NllLossBackward>)
[3221] Train Loss:    tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[3222] Train Loss:    tensor(0.7342, device='cuda:0', grad_fn=<NllLossBackward>)
[3223] Train Loss:    tensor(0.6729, device='cuda:0', grad_fn=<NllLossBackward>)
[3224] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[3225] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[3226] Train Loss:    tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[3227] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[3228] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[3229] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[3230] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[3231] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[3232] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[3233] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[3234] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[3235] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[3236] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[3237] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[3238] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[3239] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[3240] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[3241] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[3242] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[3243] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3244] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[3245] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[3246] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[3247] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[3248] Train Loss:    tensor(0.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[3249] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[3250] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3251] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[3252] Train Loss:    tensor(0.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[3253] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[3254] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[3255] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[3256] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[3257] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[3258] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3259] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[3260] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[3261] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[3262] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[3263] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[3264] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3265] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[3266] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[3267] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[3268] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[3269] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[3270] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[3271] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[3272] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[3273] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3274] Train Loss:    tensor(0.7338, device='cuda:0', grad_fn=<NllLossBackward>)
[3275] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[3276] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[3277] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[3278] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[3279] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[3280] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[3281] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[3282] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[3283] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[3284] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[3285] Train Loss:    tensor(0.7326, device='cuda:0', grad_fn=<NllLossBackward>)
[3286] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[3287] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[3288] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[3289] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[3290] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[3291] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[3292] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[3293] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[3294] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[3295] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[3296] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[3297] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[3298] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[3299] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[3300] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3301] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[3302] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[3303] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[3304] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3305] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[3306] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[3307] Train Loss:    tensor(0.7331, device='cuda:0', grad_fn=<NllLossBackward>)
[3308] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[3309] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[3310] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[3311] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[3312] Train Loss:    tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
```

[3313] Train Loss:    tensor(0.7241, device='cuda:0', grad_fn=<NllLossBackward>)
[3314] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3315] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[3316] Train Loss:    tensor(0.7289, device='cuda:0', grad_fn=<NllLossBackward>)
[3317] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[3318] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[3319] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[3320] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[3321] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[3322] Train Loss:    tensor(0.6633, device='cuda:0', grad_fn=<NllLossBackward>)
[3323] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[3324] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[3325] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[3326] Train Loss:    tensor(0.7296, device='cuda:0', grad_fn=<NllLossBackward>)
[3327] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[3328] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[3329] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[3330] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[3331] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[3332] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[3333] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[3334] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[3335] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[3336] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[3337] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[3338] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3339] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[3340] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3341] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[3342] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[3343] Train Loss:    tensor(0.6722, device='cuda:0', grad_fn=<NllLossBackward>)
[3344] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[3345] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[3346] Train Loss:    tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[3347] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3348] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3349] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3350] Train Loss:    tensor(0.6660, device='cuda:0', grad_fn=<NllLossBackward>)
[3351] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[3352] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3353] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[3354] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[3355] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[3356] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[3357] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[3358] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[3359] Train Loss:    tensor(0.7337, device='cuda:0', grad_fn=<NllLossBackward>)
[3360] Train Loss:    tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)

```
[3361] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[3362] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3363] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[3364] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3365] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[3366] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[3367] Train Loss:    tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[3368] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[3369] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[3370] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[3371] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[3372] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[3373] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[3374] Train Loss:    tensor(0.7362, device='cuda:0', grad_fn=<NllLossBackward>)
[3375] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[3376] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[3377] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3378] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[3379] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[3380] Train Loss:    tensor(0.6663, device='cuda:0', grad_fn=<NllLossBackward>)
[3381] Train Loss:    tensor(0.7481, device='cuda:0', grad_fn=<NllLossBackward>)
[3382] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[3383] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[3384] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[3385] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[3386] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[3387] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[3388] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[3389] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[3390] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[3391] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3392] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[3393] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[3394] Train Loss:    tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[3395] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[3396] Train Loss:    tensor(0.7316, device='cuda:0', grad_fn=<NllLossBackward>)
[3397] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[3398] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[3399] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[3400] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[3401] Train Loss:    tensor(0.7465, device='cuda:0', grad_fn=<NllLossBackward>)
[3402] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[3403] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[3404] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[3405] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[3406] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3407] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[3408] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3409] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3410] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[3411] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[3412] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[3413] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[3414] Train Loss:    tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[3415] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[3416] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[3417] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[3418] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3419] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[3420] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[3421] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[3422] Train Loss:    tensor(0.6525, device='cuda:0', grad_fn=<NllLossBackward>)
[3423] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[3424] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[3425] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[3426] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3427] Train Loss:    tensor(0.7306, device='cuda:0', grad_fn=<NllLossBackward>)
[3428] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[3429] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[3430] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[3431] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[3432] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[3433] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[3434] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3435] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[3436] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3437] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[3438] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[3439] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[3440] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[3441] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[3442] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[3443] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[3444] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[3445] Train Loss:    tensor(0.7340, device='cuda:0', grad_fn=<NllLossBackward>)
[3446] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[3447] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[3448] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3449] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3450] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[3451] Train Loss:    tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[3452] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[3453] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[3454] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[3455] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[3456] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3457] Train Loss:   tensor(0.6595, device='cuda:0', grad_fn=<NllLossBackward>)
[3458] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[3459] Train Loss:   tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[3460] Train Loss:   tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[3461] Train Loss:   tensor(0.6375, device='cuda:0', grad_fn=<NllLossBackward>)
[3462] Train Loss:   tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[3463] Train Loss:   tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[3464] Train Loss:   tensor(0.6572, device='cuda:0', grad_fn=<NllLossBackward>)
[3465] Train Loss:   tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[3466] Train Loss:   tensor(0.6589, device='cuda:0', grad_fn=<NllLossBackward>)
[3467] Train Loss:   tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3468] Train Loss:   tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[3469] Train Loss:   tensor(0.6525, device='cuda:0', grad_fn=<NllLossBackward>)
[3470] Train Loss:   tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[3471] Train Loss:   tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[3472] Train Loss:   tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[3473] Train Loss:   tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[3474] Train Loss:   tensor(0.7386, device='cuda:0', grad_fn=<NllLossBackward>)
[3475] Train Loss:   tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[3476] Train Loss:   tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[3477] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[3478] Train Loss:   tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[3479] Train Loss:   tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[3480] Train Loss:   tensor(0.7198, device='cuda:0', grad_fn=<NllLossBackward>)
[3481] Train Loss:   tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[3482] Train Loss:   tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[3483] Train Loss:   tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[3484] Train Loss:   tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[3485] Train Loss:   tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[3486] Train Loss:   tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[3487] Train Loss:   tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[3488] Train Loss:   tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[3489] Train Loss:   tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[3490] Train Loss:   tensor(0.7219, device='cuda:0', grad_fn=<NllLossBackward>)
[3491] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[3492] Train Loss:   tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[3493] Train Loss:   tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[3494] Train Loss:   tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[3495] Train Loss:   tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[3496] Train Loss:   tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[3497] Train Loss:   tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[3498] Train Loss:   tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[3499] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[3500] Train Loss:   tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[3501] Train Loss:   tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[3502] Train Loss:   tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[3503] Train Loss:   tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3504] Train Loss:   tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3505] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3506] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[3507] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[3508] Train Loss:    tensor(0.7540, device='cuda:0', grad_fn=<NllLossBackward>)
[3509] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[3510] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[3511] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[3512] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[3513] Train Loss:    tensor(0.6646, device='cuda:0', grad_fn=<NllLossBackward>)
[3514] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3515] Train Loss:    tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[3516] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[3517] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[3518] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[3519] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[3520] Train Loss:    tensor(0.6617, device='cuda:0', grad_fn=<NllLossBackward>)
[3521] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3522] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[3523] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[3524] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3525] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3526] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[3527] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[3528] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[3529] Train Loss:    tensor(0.7499, device='cuda:0', grad_fn=<NllLossBackward>)
[3530] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3531] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3532] Train Loss:    tensor(0.6459, device='cuda:0', grad_fn=<NllLossBackward>)
[3533] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[3534] Train Loss:    tensor(0.6676, device='cuda:0', grad_fn=<NllLossBackward>)
[3535] Train Loss:    tensor(0.7388, device='cuda:0', grad_fn=<NllLossBackward>)
[3536] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[3537] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[3538] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[3539] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[3540] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[3541] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3542] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[3543] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[3544] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[3545] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[3546] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[3547] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3548] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[3549] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[3550] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3551] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3552] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3553] Train Loss:    tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[3554] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[3555] Train Loss:    tensor(0.7348, device='cuda:0', grad_fn=<NllLossBackward>)
[3556] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[3557] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[3558] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[3559] Train Loss:    tensor(0.7304, device='cuda:0', grad_fn=<NllLossBackward>)
[3560] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[3561] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[3562] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[3563] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[3564] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[3565] Train Loss:    tensor(0.6690, device='cuda:0', grad_fn=<NllLossBackward>)
[3566] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[3567] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3568] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[3569] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[3570] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[3571] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[3572] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[3573] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[3574] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[3575] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[3576] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[3577] Train Loss:    tensor(0.6601, device='cuda:0', grad_fn=<NllLossBackward>)
[3578] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[3579] Train Loss:    tensor(0.7378, device='cuda:0', grad_fn=<NllLossBackward>)
[3580] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[3581] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[3582] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[3583] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[3584] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[3585] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[3586] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[3587] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3588] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[3589] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[3590] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[3591] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[3592] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[3593] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[3594] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[3595] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[3596] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[3597] Train Loss:    tensor(0.6654, device='cuda:0', grad_fn=<NllLossBackward>)
[3598] Train Loss:    tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[3599] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[3600] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3601] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[3602] Train Loss:    tensor(0.6577, device='cuda:0', grad_fn=<NllLossBackward>)
[3603] Train Loss:    tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[3604] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[3605] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[3606] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[3607] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[3608] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[3609] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[3610] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[3611] Train Loss:    tensor(0.7231, device='cuda:0', grad_fn=<NllLossBackward>)
[3612] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[3613] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[3614] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[3615] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[3616] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[3617] Train Loss:    tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[3618] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[3619] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[3620] Train Loss:    tensor(0.6555, device='cuda:0', grad_fn=<NllLossBackward>)
[3621] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[3622] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[3623] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[3624] Train Loss:    tensor(0.6705, device='cuda:0', grad_fn=<NllLossBackward>)
[3625] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[3626] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[3627] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3628] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[3629] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3630] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[3631] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[3632] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[3633] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[3634] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[3635] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[3636] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[3637] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[3638] Train Loss:    tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[3639] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[3640] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[3641] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3642] Train Loss:    tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[3643] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[3644] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[3645] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[3646] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[3647] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[3648] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3649] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[3650] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[3651] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[3652] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3653] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3654] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[3655] Train Loss:    tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[3656] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3657] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[3658] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[3659] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[3660] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[3661] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[3662] Train Loss:    tensor(0.6668, device='cuda:0', grad_fn=<NllLossBackward>)
[3663] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[3664] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3665] Train Loss:    tensor(0.6656, device='cuda:0', grad_fn=<NllLossBackward>)
[3666] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[3667] Train Loss:    tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[3668] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[3669] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[3670] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[3671] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3672] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[3673] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3674] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[3675] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[3676] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[3677] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[3678] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3679] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[3680] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[3681] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[3682] Train Loss:    tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[3683] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[3684] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3685] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[3686] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[3687] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[3688] Train Loss:    tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[3689] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[3690] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[3691] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[3692] Train Loss:    tensor(0.7427, device='cuda:0', grad_fn=<NllLossBackward>)
[3693] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[3694] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[3695] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[3696] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3697] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[3698] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[3699] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[3700] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[3701] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[3702] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[3703] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[3704] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[3705] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[3706] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3707] Train Loss:    tensor(0.7350, device='cuda:0', grad_fn=<NllLossBackward>)
[3708] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[3709] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[3710] Train Loss:    tensor(0.6641, device='cuda:0', grad_fn=<NllLossBackward>)
[3711] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[3712] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[3713] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[3714] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[3715] Train Loss:    tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[3716] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[3717] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[3718] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[3719] Train Loss:    tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[3720] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[3721] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[3722] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3723] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[3724] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[3725] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[3726] Train Loss:    tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[3727] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[3728] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[3729] Train Loss:    tensor(0.7295, device='cuda:0', grad_fn=<NllLossBackward>)
[3730] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[3731] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[3732] Train Loss:    tensor(0.6638, device='cuda:0', grad_fn=<NllLossBackward>)
[3733] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[3734] Train Loss:    tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[3735] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3736] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[3737] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[3738] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3739] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3740] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[3741] Train Loss:    tensor(0.6611, device='cuda:0', grad_fn=<NllLossBackward>)
[3742] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[3743] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[3744] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3745] Train Loss:   tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[3746] Train Loss:   tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[3747] Train Loss:   tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[3748] Train Loss:   tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[3749] Train Loss:   tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[3750] Train Loss:   tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[3751] Train Loss:   tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[3752] Train Loss:   tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[3753] Train Loss:   tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[3754] Train Loss:   tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[3755] Train Loss:   tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[3756] Train Loss:   tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[3757] Train Loss:   tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[3758] Train Loss:   tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[3759] Train Loss:   tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3760] Train Loss:   tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[3761] Train Loss:   tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3762] Train Loss:   tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[3763] Train Loss:   tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[3764] Train Loss:   tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[3765] Train Loss:   tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[3766] Train Loss:   tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[3767] Train Loss:   tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[3768] Train Loss:   tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[3769] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3770] Train Loss:   tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[3771] Train Loss:   tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3772] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[3773] Train Loss:   tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[3774] Train Loss:   tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[3775] Train Loss:   tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[3776] Train Loss:   tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[3777] Train Loss:   tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3778] Train Loss:   tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3779] Train Loss:   tensor(0.6519, device='cuda:0', grad_fn=<NllLossBackward>)
[3780] Train Loss:   tensor(0.6715, device='cuda:0', grad_fn=<NllLossBackward>)
[3781] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[3782] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3783] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[3784] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3785] Train Loss:   tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[3786] Train Loss:   tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[3787] Train Loss:   tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[3788] Train Loss:   tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[3789] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3790] Train Loss:   tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[3791] Train Loss:   tensor(0.7300, device='cuda:0', grad_fn=<NllLossBackward>)
[3792] Train Loss:   tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3793] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[3794] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[3795] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[3796] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[3797] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[3798] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[3799] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[3800] Train Loss:    tensor(0.6748, device='cuda:0', grad_fn=<NllLossBackward>)
[3801] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3802] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3803] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[3804] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[3805] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[3806] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[3807] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3808] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[3809] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[3810] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[3811] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[3812] Train Loss:    tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[3813] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[3814] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[3815] Train Loss:    tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[3816] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[3817] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[3818] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[3819] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[3820] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[3821] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[3822] Train Loss:    tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[3823] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[3824] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[3825] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[3826] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[3827] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[3828] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[3829] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[3830] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[3831] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[3832] Train Loss:    tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[3833] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[3834] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[3835] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[3836] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[3837] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[3838] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3839] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3840] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3841] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3842] Train Loss:   tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[3843] Train Loss:   tensor(0.7287, device='cuda:0', grad_fn=<NllLossBackward>)
[3844] Train Loss:   tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[3845] Train Loss:   tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3846] Train Loss:   tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[3847] Train Loss:   tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[3848] Train Loss:   tensor(0.6706, device='cuda:0', grad_fn=<NllLossBackward>)
[3849] Train Loss:   tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[3850] Train Loss:   tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[3851] Train Loss:   tensor(0.7313, device='cuda:0', grad_fn=<NllLossBackward>)
[3852] Train Loss:   tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[3853] Train Loss:   tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[3854] Train Loss:   tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[3855] Train Loss:   tensor(0.7330, device='cuda:0', grad_fn=<NllLossBackward>)
[3856] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3857] Train Loss:   tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[3858] Train Loss:   tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[3859] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[3860] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[3861] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[3862] Train Loss:   tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[3863] Train Loss:   tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[3864] Train Loss:   tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[3865] Train Loss:   tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[3866] Train Loss:   tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[3867] Train Loss:   tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[3868] Train Loss:   tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[3869] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[3870] Train Loss:   tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[3871] Train Loss:   tensor(0.6623, device='cuda:0', grad_fn=<NllLossBackward>)
[3872] Train Loss:   tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[3873] Train Loss:   tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[3874] Train Loss:   tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[3875] Train Loss:   tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[3876] Train Loss:   tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[3877] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[3878] Train Loss:   tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3879] Train Loss:   tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[3880] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[3881] Train Loss:   tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[3882] Train Loss:   tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[3883] Train Loss:   tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[3884] Train Loss:   tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3885] Train Loss:   tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[3886] Train Loss:   tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[3887] Train Loss:   tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[3888] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3889] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[3890] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[3891] Train Loss:    tensor(0.7431, device='cuda:0', grad_fn=<NllLossBackward>)
[3892] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[3893] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[3894] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[3895] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3896] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[3897] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[3898] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[3899] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3900] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[3901] Train Loss:    tensor(0.7344, device='cuda:0', grad_fn=<NllLossBackward>)
[3902] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[3903] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[3904] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[3905] Train Loss:    tensor(0.6508, device='cuda:0', grad_fn=<NllLossBackward>)
[3906] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[3907] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[3908] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[3909] Train Loss:    tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[3910] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[3911] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[3912] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[3913] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[3914] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[3915] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[3916] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3917] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[3918] Train Loss:    tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[3919] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[3920] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[3921] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[3922] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[3923] Train Loss:    tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[3924] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[3925] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[3926] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[3927] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3928] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[3929] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[3930] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[3931] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[3932] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[3933] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[3934] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3935] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[3936] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3937] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[3938] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[3939] Train Loss:    tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[3940] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[3941] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[3942] Train Loss:    tensor(0.7480, device='cuda:0', grad_fn=<NllLossBackward>)
[3943] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[3944] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[3945] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[3946] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3947] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[3948] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[3949] Train Loss:    tensor(0.6650, device='cuda:0', grad_fn=<NllLossBackward>)
[3950] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[3951] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3952] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[3953] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[3954] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[3955] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[3956] Train Loss:    tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[3957] Train Loss:    tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[3958] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[3959] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[3960] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[3961] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[3962] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[3963] Train Loss:    tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[3964] Train Loss:    tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[3965] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[3966] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[3967] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[3968] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[3969] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[3970] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[3971] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[3972] Train Loss:    tensor(0.7358, device='cuda:0', grad_fn=<NllLossBackward>)
[3973] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3974] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[3975] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[3976] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[3977] Train Loss:    tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[3978] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[3979] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[3980] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3981] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[3982] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[3983] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[3984] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3985] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[3986] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[3987] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[3988] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3989] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[3990] Train Loss:    tensor(0.7278, device='cuda:0', grad_fn=<NllLossBackward>)
[3991] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3992] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[3993] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[3994] Train Loss:    tensor(0.6519, device='cuda:0', grad_fn=<NllLossBackward>)
[3995] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[3996] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[3997] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[3998] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[3999] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[4000] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[4001] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[4002] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[4003] Train Loss:    tensor(0.6591, device='cuda:0', grad_fn=<NllLossBackward>)
[4004] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[4005] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[4006] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[4007] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[4008] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[4009] Train Loss:    tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[4010] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[4011] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[4012] Train Loss:    tensor(0.6676, device='cuda:0', grad_fn=<NllLossBackward>)
[4013] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[4014] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[4015] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[4016] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[4017] Train Loss:    tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[4018] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[4019] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[4020] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[4021] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[4022] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[4023] Train Loss:    tensor(0.6627, device='cuda:0', grad_fn=<NllLossBackward>)
[4024] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[4025] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[4026] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[4027] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[4028] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[4029] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[4030] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[4031] Train Loss:    tensor(0.7416, device='cuda:0', grad_fn=<NllLossBackward>)
[4032] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4033] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[4034] Train Loss:    tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[4035] Train Loss:    tensor(0.7375, device='cuda:0', grad_fn=<NllLossBackward>)
[4036] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[4037] Train Loss:    tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[4038] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[4039] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[4040] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[4041] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[4042] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[4043] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[4044] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[4045] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[4046] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[4047] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[4048] Train Loss:    tensor(0.6699, device='cuda:0', grad_fn=<NllLossBackward>)
[4049] Train Loss:    tensor(0.7187, device='cuda:0', grad_fn=<NllLossBackward>)
[4050] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[4051] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[4052] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[4053] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[4054] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[4055] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[4056] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[4057] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[4058] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[4059] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[4060] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[4061] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[4062] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[4063] Train Loss:    tensor(0.7332, device='cuda:0', grad_fn=<NllLossBackward>)
[4064] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[4065] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[4066] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[4067] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[4068] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[4069] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[4070] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[4071] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[4072] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[4073] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[4074] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[4075] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[4076] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[4077] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[4078] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[4079] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[4080] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4081] Train Loss:   tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[4082] Train Loss:   tensor(0.7250, device='cuda:0', grad_fn=<NllLossBackward>)
[4083] Train Loss:   tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[4084] Train Loss:   tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[4085] Train Loss:   tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[4086] Train Loss:   tensor(0.7502, device='cuda:0', grad_fn=<NllLossBackward>)
[4087] Train Loss:   tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[4088] Train Loss:   tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[4089] Train Loss:   tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[4090] Train Loss:   tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[4091] Train Loss:   tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[4092] Train Loss:   tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[4093] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[4094] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[4095] Train Loss:   tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[4096] Train Loss:   tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[4097] Train Loss:   tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[4098] Train Loss:   tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[4099] Train Loss:   tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[4100] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[4101] Train Loss:   tensor(0.6692, device='cuda:0', grad_fn=<NllLossBackward>)
[4102] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[4103] Train Loss:   tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[4104] Train Loss:   tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[4105] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[4106] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[4107] Train Loss:   tensor(0.7371, device='cuda:0', grad_fn=<NllLossBackward>)
[4108] Train Loss:   tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[4109] Train Loss:   tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[4110] Train Loss:   tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[4111] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[4112] Train Loss:   tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[4113] Train Loss:   tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[4114] Train Loss:   tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[4115] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[4116] Train Loss:   tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[4117] Train Loss:   tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[4118] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[4119] Train Loss:   tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[4120] Train Loss:   tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[4121] Train Loss:   tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[4122] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[4123] Train Loss:   tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[4124] Train Loss:   tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[4125] Train Loss:   tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[4126] Train Loss:   tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[4127] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[4128] Train Loss:   tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4129] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[4130] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[4131] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[4132] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[4133] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[4134] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[4135] Train Loss:    tensor(0.7262, device='cuda:0', grad_fn=<NllLossBackward>)
[4136] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[4137] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[4138] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[4139] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[4140] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[4141] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[4142] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[4143] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[4144] Train Loss:    tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[4145] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[4146] Train Loss:    tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[4147] Train Loss:    tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[4148] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[4149] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[4150] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[4151] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[4152] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[4153] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[4154] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[4155] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[4156] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[4157] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[4158] Train Loss:    tensor(0.6645, device='cuda:0', grad_fn=<NllLossBackward>)
[4159] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[4160] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[4161] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[4162] Train Loss:    tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[4163] Train Loss:    tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[4164] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[4165] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[4166] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[4167] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[4168] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[4169] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[4170] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[4171] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[4172] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[4173] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[4174] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[4175] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[4176] Train Loss:    tensor(0.6605, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4177] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[4178] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[4179] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[4180] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[4181] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[4182] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[4183] Train Loss:    tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[4184] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[4185] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[4186] Train Loss:    tensor(0.6586, device='cuda:0', grad_fn=<NllLossBackward>)
[4187] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[4188] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[4189] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[4190] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[4191] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[4192] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[4193] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[4194] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[4195] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[4196] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[4197] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[4198] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[4199] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[4200] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[4201] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[4202] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[4203] Train Loss:    tensor(0.6471, device='cuda:0', grad_fn=<NllLossBackward>)
[4204] Train Loss:    tensor(0.7232, device='cuda:0', grad_fn=<NllLossBackward>)
[4205] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[4206] Train Loss:    tensor(0.7555, device='cuda:0', grad_fn=<NllLossBackward>)
[4207] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[4208] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[4209] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[4210] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[4211] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[4212] Train Loss:    tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[4213] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[4214] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[4215] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[4216] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[4217] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[4218] Train Loss:    tensor(0.7167, device='cuda:0', grad_fn=<NllLossBackward>)
[4219] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[4220] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[4221] Train Loss:    tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[4222] Train Loss:    tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
[4223] Train Loss:    tensor(0.6729, device='cuda:0', grad_fn=<NllLossBackward>)
[4224] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4225] Train Loss:    tensor(0.6644, device='cuda:0', grad_fn=<NllLossBackward>)
[4226] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[4227] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[4228] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[4229] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[4230] Train Loss:    tensor(0.6644, device='cuda:0', grad_fn=<NllLossBackward>)
[4231] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[4232] Train Loss:    tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[4233] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[4234] Train Loss:    tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[4235] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[4236] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[4237] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[4238] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[4239] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[4240] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[4241] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[4242] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[4243] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[4244] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[4245] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[4246] Train Loss:    tensor(0.6719, device='cuda:0', grad_fn=<NllLossBackward>)
[4247] Train Loss:    tensor(0.7451, device='cuda:0', grad_fn=<NllLossBackward>)
[4248] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[4249] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[4250] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[4251] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[4252] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[4253] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[4254] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[4255] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[4256] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[4257] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[4258] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[4259] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[4260] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[4261] Train Loss:    tensor(0.6695, device='cuda:0', grad_fn=<NllLossBackward>)
[4262] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[4263] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[4264] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[4265] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[4266] Train Loss:    tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[4267] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[4268] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[4269] Train Loss:    tensor(0.7434, device='cuda:0', grad_fn=<NllLossBackward>)
[4270] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[4271] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[4272] Train Loss:    tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4273] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[4274] Train Loss:    tensor(0.7301, device='cuda:0', grad_fn=<NllLossBackward>)
[4275] Train Loss:    tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[4276] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[4277] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[4278] Train Loss:    tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[4279] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[4280] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[4281] Train Loss:    tensor(0.6606, device='cuda:0', grad_fn=<NllLossBackward>)
[4282] Train Loss:    tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[4283] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[4284] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[4285] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[4286] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[4287] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[4288] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[4289] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[4290] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[4291] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[4292] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[4293] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[4294] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[4295] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[4296] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[4297] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[4298] Train Loss:    tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[4299] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[4300] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[4301] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[4302] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[4303] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[4304] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[4305] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[4306] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[4307] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[4308] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[4309] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[4310] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[4311] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[4312] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[4313] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[4314] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[4315] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[4316] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[4317] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[4318] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[4319] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[4320] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4321] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[4322] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[4323] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[4324] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[4325] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[4326] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[4327] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[4328] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[4329] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[4330] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[4331] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[4332] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[4333] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[4334] Train Loss:    tensor(0.6704, device='cuda:0', grad_fn=<NllLossBackward>)
[4335] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[4336] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[4337] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[4338] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[4339] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[4340] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[4341] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[4342] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[4343] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[4344] Train Loss:    tensor(0.6549, device='cuda:0', grad_fn=<NllLossBackward>)
[4345] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[4346] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[4347] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[4348] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[4349] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[4350] Train Loss:    tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[4351] Train Loss:    tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[4352] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[4353] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[4354] Train Loss:    tensor(0.6617, device='cuda:0', grad_fn=<NllLossBackward>)
[4355] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[4356] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[4357] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[4358] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[4359] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[4360] Train Loss:    tensor(0.7043, device='cuda:0', grad_fn=<NllLossBackward>)
[4361] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[4362] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[4363] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[4364] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[4365] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[4366] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[4367] Train Loss:    tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[4368] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4369] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[4370] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[4371] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[4372] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[4373] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[4374] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[4375] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[4376] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[4377] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[4378] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[4379] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[4380] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[4381] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[4382] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[4383] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[4384] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[4385] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[4386] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[4387] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[4388] Train Loss:    tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[4389] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[4390] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[4391] Train Loss:    tensor(0.7322, device='cuda:0', grad_fn=<NllLossBackward>)
[4392] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[4393] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[4394] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[4395] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[4396] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[4397] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[4398] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[4399] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[4400] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[4401] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[4402] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[4403] Train Loss:    tensor(0.6706, device='cuda:0', grad_fn=<NllLossBackward>)
[4404] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[4405] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[4406] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[4407] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[4408] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[4409] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[4410] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[4411] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[4412] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[4413] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[4414] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[4415] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[4416] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4417] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[4418] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[4419] Train Loss:    tensor(0.7411, device='cuda:0', grad_fn=<NllLossBackward>)
[4420] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[4421] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[4422] Train Loss:    tensor(0.6491, device='cuda:0', grad_fn=<NllLossBackward>)
[4423] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[4424] Train Loss:    tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[4425] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[4426] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[4427] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[4428] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[4429] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[4430] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[4431] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[4432] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[4433] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[4434] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[4435] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[4436] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[4437] Train Loss:    tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[4438] Train Loss:    tensor(0.6652, device='cuda:0', grad_fn=<NllLossBackward>)
[4439] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[4440] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[4441] Train Loss:    tensor(0.6668, device='cuda:0', grad_fn=<NllLossBackward>)
[4442] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[4443] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[4444] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[4445] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[4446] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[4447] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[4448] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[4449] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[4450] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[4451] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[4452] Train Loss:    tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[4453] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[4454] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[4455] Train Loss:    tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[4456] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[4457] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[4458] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[4459] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[4460] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[4461] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[4462] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[4463] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[4464] Train Loss:    tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4465] Train Loss:   tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[4466] Train Loss:   tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[4467] Train Loss:   tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[4468] Train Loss:   tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[4469] Train Loss:   tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[4470] Train Loss:   tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[4471] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[4472] Train Loss:   tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[4473] Train Loss:   tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[4474] Train Loss:   tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[4475] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[4476] Train Loss:   tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[4477] Train Loss:   tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[4478] Train Loss:   tensor(0.6600, device='cuda:0', grad_fn=<NllLossBackward>)
[4479] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[4480] Train Loss:   tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[4481] Train Loss:   tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[4482] Train Loss:   tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[4483] Train Loss:   tensor(0.6515, device='cuda:0', grad_fn=<NllLossBackward>)
[4484] Train Loss:   tensor(0.7422, device='cuda:0', grad_fn=<NllLossBackward>)
[4485] Train Loss:   tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[4486] Train Loss:   tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
[4487] Train Loss:   tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[4488] Train Loss:   tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[4489] Train Loss:   tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[4490] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[4491] Train Loss:   tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[4492] Train Loss:   tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[4493] Train Loss:   tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[4494] Train Loss:   tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[4495] Train Loss:   tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[4496] Train Loss:   tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[4497] Train Loss:   tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[4498] Train Loss:   tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[4499] Train Loss:   tensor(0.6716, device='cuda:0', grad_fn=<NllLossBackward>)
[4500] Train Loss:   tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[4501] Train Loss:   tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[4502] Train Loss:   tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[4503] Train Loss:   tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[4504] Train Loss:   tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[4505] Train Loss:   tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[4506] Train Loss:   tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[4507] Train Loss:   tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[4508] Train Loss:   tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[4509] Train Loss:   tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[4510] Train Loss:   tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[4511] Train Loss:   tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
Training Loss:   0.6987583677793014 Time:   2447.1590027809143 s
```

```
Training Accuaracy:   78.43709634550919 %
[0] Valid Loss:  tensor(0.7064, device='cuda:0')
[1] Valid Loss:  tensor(0.8799, device='cuda:0')
[2] Valid Loss:  tensor(0.6822, device='cuda:0')
[3] Valid Loss:  tensor(0.6154, device='cuda:0')
[4] Valid Loss:  tensor(0.6906, device='cuda:0')
[5] Valid Loss:  tensor(0.7017, device='cuda:0')
[6] Valid Loss:  tensor(0.7295, device='cuda:0')
[7] Valid Loss:  tensor(0.8357, device='cuda:0')
[8] Valid Loss:  tensor(0.8073, device='cuda:0')
[9] Valid Loss:  tensor(0.7933, device='cuda:0')
[10] Valid Loss:  tensor(0.6835, device='cuda:0')
[11] Valid Loss:  tensor(0.6482, device='cuda:0')
[12] Valid Loss:  tensor(1.1011, device='cuda:0')
[13] Valid Loss:  tensor(0.9695, device='cuda:0')
[14] Valid Loss:  tensor(0.6133, device='cuda:0')
[15] Valid Loss:  tensor(0.7393, device='cuda:0')
[16] Valid Loss:  tensor(0.6258, device='cuda:0')
[17] Valid Loss:  tensor(0.7839, device='cuda:0')
[18] Valid Loss:  tensor(0.6440, device='cuda:0')
[19] Valid Loss:  tensor(0.6196, device='cuda:0')
[20] Valid Loss:  tensor(0.6699, device='cuda:0')
[21] Valid Loss:  tensor(0.5742, device='cuda:0')
[22] Valid Loss:  tensor(0.5838, device='cuda:0')
[23] Valid Loss:  tensor(0.8273, device='cuda:0')
[24] Valid Loss:  tensor(0.7243, device='cuda:0')
[25] Valid Loss:  tensor(0.9316, device='cuda:0')
[26] Valid Loss:  tensor(0.5399, device='cuda:0')
[27] Valid Loss:  tensor(0.8737, device='cuda:0')
[28] Valid Loss:  tensor(0.6767, device='cuda:0')
[29] Valid Loss:  tensor(0.6905, device='cuda:0')
[30] Valid Loss:  tensor(0.6223, device='cuda:0')
[31] Valid Loss:  tensor(0.7234, device='cuda:0')
[32] Valid Loss:  tensor(0.8147, device='cuda:0')
[33] Valid Loss:  tensor(0.7987, device='cuda:0')
[34] Valid Loss:  tensor(0.7753, device='cuda:0')
[35] Valid Loss:  tensor(0.8695, device='cuda:0')
[36] Valid Loss:  tensor(0.6666, device='cuda:0')
[37] Valid Loss:  tensor(1.0254, device='cuda:0')
[38] Valid Loss:  tensor(0.8222, device='cuda:0')
[39] Valid Loss:  tensor(0.6520, device='cuda:0')
[40] Valid Loss:  tensor(0.7601, device='cuda:0')
[41] Valid Loss:  tensor(0.8319, device='cuda:0')
[42] Valid Loss:  tensor(0.7691, device='cuda:0')
[43] Valid Loss:  tensor(0.5424, device='cuda:0')
[44] Valid Loss:  tensor(0.6699, device='cuda:0')
[45] Valid Loss:  tensor(0.5661, device='cuda:0')
[46] Valid Loss:  tensor(0.8747, device='cuda:0')
```

```
[47] Valid Loss:    tensor(0.7925, device='cuda:0')
[48] Valid Loss:    tensor(0.6913, device='cuda:0')
[49] Valid Loss:    tensor(0.7606, device='cuda:0')
[50] Valid Loss:    tensor(0.6865, device='cuda:0')
[51] Valid Loss:    tensor(0.6467, device='cuda:0')
[52] Valid Loss:    tensor(0.5989, device='cuda:0')
[53] Valid Loss:    tensor(0.8912, device='cuda:0')
[54] Valid Loss:    tensor(0.6385, device='cuda:0')
[55] Valid Loss:    tensor(0.5566, device='cuda:0')
[56] Valid Loss:    tensor(0.8613, device='cuda:0')
[57] Valid Loss:    tensor(0.8073, device='cuda:0')
[58] Valid Loss:    tensor(0.7587, device='cuda:0')
[59] Valid Loss:    tensor(0.6663, device='cuda:0')
[60] Valid Loss:    tensor(1.0178, device='cuda:0')
[61] Valid Loss:    tensor(0.6968, device='cuda:0')
[62] Valid Loss:    tensor(0.8792, device='cuda:0')
[63] Valid Loss:    tensor(0.6610, device='cuda:0')
[64] Valid Loss:    tensor(0.7453, device='cuda:0')
[65] Valid Loss:    tensor(0.8698, device='cuda:0')
[66] Valid Loss:    tensor(0.7233, device='cuda:0')
[67] Valid Loss:    tensor(0.7299, device='cuda:0')
[68] Valid Loss:    tensor(0.7312, device='cuda:0')
[69] Valid Loss:    tensor(0.5610, device='cuda:0')
[70] Valid Loss:    tensor(0.6009, device='cuda:0')
[71] Valid Loss:    tensor(0.8113, device='cuda:0')
[72] Valid Loss:    tensor(0.7315, device='cuda:0')
[73] Valid Loss:    tensor(0.6712, device='cuda:0')
[74] Valid Loss:    tensor(0.7528, device='cuda:0')
[75] Valid Loss:    tensor(0.8414, device='cuda:0')
[76] Valid Loss:    tensor(0.7765, device='cuda:0')
[77] Valid Loss:    tensor(0.6855, device='cuda:0')
[78] Valid Loss:    tensor(0.7805, device='cuda:0')
[79] Valid Loss:    tensor(0.6294, device='cuda:0')
[80] Valid Loss:    tensor(0.7245, device='cuda:0')
[81] Valid Loss:    tensor(0.6239, device='cuda:0')
[82] Valid Loss:    tensor(0.8462, device='cuda:0')
[83] Valid Loss:    tensor(0.8059, device='cuda:0')
[84] Valid Loss:    tensor(0.6018, device='cuda:0')
[85] Valid Loss:    tensor(0.7318, device='cuda:0')
[86] Valid Loss:    tensor(0.7467, device='cuda:0')
[87] Valid Loss:    tensor(0.8055, device='cuda:0')
[88] Valid Loss:    tensor(0.5881, device='cuda:0')
[89] Valid Loss:    tensor(0.7814, device='cuda:0')
[90] Valid Loss:    tensor(0.9340, device='cuda:0')
[91] Valid Loss:    tensor(0.7364, device='cuda:0')
[92] Valid Loss:    tensor(0.7897, device='cuda:0')
[93] Valid Loss:    tensor(0.7876, device='cuda:0')
[94] Valid Loss:    tensor(0.8257, device='cuda:0')
```

```
[95] Valid Loss:  tensor(0.6851, device='cuda:0')
[96] Valid Loss:  tensor(0.8670, device='cuda:0')
[97] Valid Loss:  tensor(0.7939, device='cuda:0')
[98] Valid Loss:  tensor(0.8192, device='cuda:0')
[99] Valid Loss:  tensor(0.6155, device='cuda:0')
[100] Valid Loss:  tensor(0.7588, device='cuda:0')
[101] Valid Loss:  tensor(0.7338, device='cuda:0')
[102] Valid Loss:  tensor(0.8859, device='cuda:0')
[103] Valid Loss:  tensor(0.8498, device='cuda:0')
[104] Valid Loss:  tensor(0.6667, device='cuda:0')
[105] Valid Loss:  tensor(0.6565, device='cuda:0')
[106] Valid Loss:  tensor(0.6062, device='cuda:0')
[107] Valid Loss:  tensor(0.7514, device='cuda:0')
[108] Valid Loss:  tensor(0.8563, device='cuda:0')
[109] Valid Loss:  tensor(0.7583, device='cuda:0')
[110] Valid Loss:  tensor(0.7941, device='cuda:0')
[111] Valid Loss:  tensor(0.7779, device='cuda:0')
[112] Valid Loss:  tensor(0.8016, device='cuda:0')
[113] Valid Loss:  tensor(0.7718, device='cuda:0')
[114] Valid Loss:  tensor(0.7217, device='cuda:0')
[115] Valid Loss:  tensor(0.8259, device='cuda:0')
[116] Valid Loss:  tensor(0.8140, device='cuda:0')
[117] Valid Loss:  tensor(0.8361, device='cuda:0')
[118] Valid Loss:  tensor(0.6657, device='cuda:0')
[119] Valid Loss:  tensor(0.6058, device='cuda:0')
[120] Valid Loss:  tensor(0.9633, device='cuda:0')
[121] Valid Loss:  tensor(0.6739, device='cuda:0')
[122] Valid Loss:  tensor(0.7077, device='cuda:0')
[123] Valid Loss:  tensor(0.6577, device='cuda:0')
[124] Valid Loss:  tensor(0.8644, device='cuda:0')
[125] Valid Loss:  tensor(0.9381, device='cuda:0')
[126] Valid Loss:  tensor(0.5658, device='cuda:0')
[127] Valid Loss:  tensor(0.6185, device='cuda:0')
[128] Valid Loss:  tensor(0.8459, device='cuda:0')
[129] Valid Loss:  tensor(0.6875, device='cuda:0')
[130] Valid Loss:  tensor(0.8063, device='cuda:0')
[131] Valid Loss:  tensor(0.6726, device='cuda:0')
[132] Valid Loss:  tensor(0.7016, device='cuda:0')
[133] Valid Loss:  tensor(0.7613, device='cuda:0')
[134] Valid Loss:  tensor(0.6620, device='cuda:0')
[135] Valid Loss:  tensor(0.5777, device='cuda:0')
[136] Valid Loss:  tensor(0.6144, device='cuda:0')
[137] Valid Loss:  tensor(0.9522, device='cuda:0')
[138] Valid Loss:  tensor(0.8334, device='cuda:0')
[139] Valid Loss:  tensor(1.0241, device='cuda:0')
[140] Valid Loss:  tensor(0.7986, device='cuda:0')
[141] Valid Loss:  tensor(0.6409, device='cuda:0')
[142] Valid Loss:  tensor(0.9394, device='cuda:0')
```

```
[143] Valid Loss:   tensor(0.6338, device='cuda:0')
[144] Valid Loss:   tensor(0.7590, device='cuda:0')
[145] Valid Loss:   tensor(0.9432, device='cuda:0')
[146] Valid Loss:   tensor(0.6619, device='cuda:0')
[147] Valid Loss:   tensor(0.7829, device='cuda:0')
[148] Valid Loss:   tensor(0.6752, device='cuda:0')
[149] Valid Loss:   tensor(0.7542, device='cuda:0')
[150] Valid Loss:   tensor(0.7217, device='cuda:0')
[151] Valid Loss:   tensor(0.7482, device='cuda:0')
[152] Valid Loss:   tensor(0.6737, device='cuda:0')
[153] Valid Loss:   tensor(0.6581, device='cuda:0')
[154] Valid Loss:   tensor(0.6917, device='cuda:0')
[155] Valid Loss:   tensor(0.7652, device='cuda:0')
[156] Valid Loss:   tensor(0.8192, device='cuda:0')
[157] Valid Loss:   tensor(0.6396, device='cuda:0')
[158] Valid Loss:   tensor(0.7113, device='cuda:0')
[159] Valid Loss:   tensor(0.7183, device='cuda:0')
[160] Valid Loss:   tensor(0.5917, device='cuda:0')
[161] Valid Loss:   tensor(0.8288, device='cuda:0')
[162] Valid Loss:   tensor(0.7601, device='cuda:0')
[163] Valid Loss:   tensor(0.6282, device='cuda:0')
[164] Valid Loss:   tensor(0.9426, device='cuda:0')
[165] Valid Loss:   tensor(0.7091, device='cuda:0')
[166] Valid Loss:   tensor(0.7230, device='cuda:0')
[167] Valid Loss:   tensor(0.6040, device='cuda:0')
[168] Valid Loss:   tensor(0.6756, device='cuda:0')
[169] Valid Loss:   tensor(0.6704, device='cuda:0')
[170] Valid Loss:   tensor(0.7092, device='cuda:0')
[171] Valid Loss:   tensor(0.5862, device='cuda:0')
[172] Valid Loss:   tensor(0.7851, device='cuda:0')
[173] Valid Loss:   tensor(0.6156, device='cuda:0')
[174] Valid Loss:   tensor(0.7486, device='cuda:0')
[175] Valid Loss:   tensor(0.6338, device='cuda:0')
[176] Valid Loss:   tensor(0.5839, device='cuda:0')
[177] Valid Loss:   tensor(0.7612, device='cuda:0')
[178] Valid Loss:   tensor(0.7840, device='cuda:0')
[179] Valid Loss:   tensor(0.7841, device='cuda:0')
[180] Valid Loss:   tensor(0.8146, device='cuda:0')
[181] Valid Loss:   tensor(0.7771, device='cuda:0')
[182] Valid Loss:   tensor(0.7366, device='cuda:0')
[183] Valid Loss:   tensor(0.7410, device='cuda:0')
[184] Valid Loss:   tensor(0.8823, device='cuda:0')
[185] Valid Loss:   tensor(0.8187, device='cuda:0')
[186] Valid Loss:   tensor(0.6916, device='cuda:0')
[187] Valid Loss:   tensor(0.8313, device='cuda:0')
[188] Valid Loss:   tensor(0.8871, device='cuda:0')
[189] Valid Loss:   tensor(0.7285, device='cuda:0')
[190] Valid Loss:   tensor(0.7275, device='cuda:0')
```

```
[191] Valid Loss:   tensor(0.6042, device='cuda:0')
[192] Valid Loss:   tensor(1.0041, device='cuda:0')
[193] Valid Loss:   tensor(0.7637, device='cuda:0')
[194] Valid Loss:   tensor(0.8120, device='cuda:0')
[195] Valid Loss:   tensor(0.8477, device='cuda:0')
[196] Valid Loss:   tensor(0.5959, device='cuda:0')
[197] Valid Loss:   tensor(0.7750, device='cuda:0')
[198] Valid Loss:   tensor(0.7399, device='cuda:0')
[199] Valid Loss:   tensor(0.7844, device='cuda:0')
[200] Valid Loss:   tensor(0.7238, device='cuda:0')
[201] Valid Loss:   tensor(0.6946, device='cuda:0')
[202] Valid Loss:   tensor(0.6610, device='cuda:0')
[203] Valid Loss:   tensor(0.8783, device='cuda:0')
[204] Valid Loss:   tensor(0.8435, device='cuda:0')
[205] Valid Loss:   tensor(0.8350, device='cuda:0')
[206] Valid Loss:   tensor(0.9498, device='cuda:0')
[207] Valid Loss:   tensor(0.7761, device='cuda:0')
[208] Valid Loss:   tensor(0.6010, device='cuda:0')
[209] Valid Loss:   tensor(0.7154, device='cuda:0')
[210] Valid Loss:   tensor(1.1140, device='cuda:0')
[211] Valid Loss:   tensor(0.7540, device='cuda:0')
[212] Valid Loss:   tensor(0.6168, device='cuda:0')
[213] Valid Loss:   tensor(0.6951, device='cuda:0')
[214] Valid Loss:   tensor(0.7027, device='cuda:0')
[215] Valid Loss:   tensor(0.8169, device='cuda:0')
[216] Valid Loss:   tensor(0.7433, device='cuda:0')
[217] Valid Loss:   tensor(0.7230, device='cuda:0')
[218] Valid Loss:   tensor(0.5761, device='cuda:0')
[219] Valid Loss:   tensor(0.6856, device='cuda:0')
[220] Valid Loss:   tensor(0.6737, device='cuda:0')
[221] Valid Loss:   tensor(0.7519, device='cuda:0')
[222] Valid Loss:   tensor(0.5710, device='cuda:0')
[223] Valid Loss:   tensor(0.6944, device='cuda:0')
[224] Valid Loss:   tensor(0.7895, device='cuda:0')
[225] Valid Loss:   tensor(0.5988, device='cuda:0')
[226] Valid Loss:   tensor(0.7265, device='cuda:0')
[227] Valid Loss:   tensor(0.5957, device='cuda:0')
[228] Valid Loss:   tensor(0.8096, device='cuda:0')
[229] Valid Loss:   tensor(0.6779, device='cuda:0')
[230] Valid Loss:   tensor(0.7801, device='cuda:0')
[231] Valid Loss:   tensor(0.6729, device='cuda:0')
[232] Valid Loss:   tensor(0.5968, device='cuda:0')
[233] Valid Loss:   tensor(0.6741, device='cuda:0')
[234] Valid Loss:   tensor(0.5916, device='cuda:0')
[235] Valid Loss:   tensor(0.5992, device='cuda:0')
[236] Valid Loss:   tensor(0.7526, device='cuda:0')
[237] Valid Loss:   tensor(0.7931, device='cuda:0')
[238] Valid Loss:   tensor(0.5251, device='cuda:0')
```

```
[239] Valid Loss:   tensor(0.7909, device='cuda:0')
[240] Valid Loss:   tensor(0.6935, device='cuda:0')
[241] Valid Loss:   tensor(0.6351, device='cuda:0')
[242] Valid Loss:   tensor(0.5214, device='cuda:0')
[243] Valid Loss:   tensor(1.1594, device='cuda:0')
[244] Valid Loss:   tensor(0.7796, device='cuda:0')
[245] Valid Loss:   tensor(0.6854, device='cuda:0')
[246] Valid Loss:   tensor(0.6612, device='cuda:0')
[247] Valid Loss:   tensor(0.6797, device='cuda:0')
[248] Valid Loss:   tensor(0.9306, device='cuda:0')
[249] Valid Loss:   tensor(0.6873, device='cuda:0')
[250] Valid Loss:   tensor(0.7652, device='cuda:0')
[251] Valid Loss:   tensor(0.7943, device='cuda:0')
[252] Valid Loss:   tensor(0.5941, device='cuda:0')
[253] Valid Loss:   tensor(0.6324, device='cuda:0')
[254] Valid Loss:   tensor(0.6238, device='cuda:0')
[255] Valid Loss:   tensor(0.7604, device='cuda:0')
[256] Valid Loss:   tensor(0.8263, device='cuda:0')
[257] Valid Loss:   tensor(0.6710, device='cuda:0')
[258] Valid Loss:   tensor(0.8398, device='cuda:0')
[259] Valid Loss:   tensor(0.7239, device='cuda:0')
[260] Valid Loss:   tensor(0.9047, device='cuda:0')
[261] Valid Loss:   tensor(0.7505, device='cuda:0')
[262] Valid Loss:   tensor(0.8798, device='cuda:0')
[263] Valid Loss:   tensor(0.8623, device='cuda:0')
[264] Valid Loss:   tensor(0.8088, device='cuda:0')
[265] Valid Loss:   tensor(0.7976, device='cuda:0')
[266] Valid Loss:   tensor(0.6900, device='cuda:0')
[267] Valid Loss:   tensor(0.7318, device='cuda:0')
[268] Valid Loss:   tensor(0.6968, device='cuda:0')
[269] Valid Loss:   tensor(0.6962, device='cuda:0')
[270] Valid Loss:   tensor(0.6340, device='cuda:0')
[271] Valid Loss:   tensor(0.9507, device='cuda:0')
[272] Valid Loss:   tensor(0.7250, device='cuda:0')
[273] Valid Loss:   tensor(0.8767, device='cuda:0')
[274] Valid Loss:   tensor(0.6935, device='cuda:0')
[275] Valid Loss:   tensor(0.7313, device='cuda:0')
[276] Valid Loss:   tensor(0.6591, device='cuda:0')
[277] Valid Loss:   tensor(0.6299, device='cuda:0')
[278] Valid Loss:   tensor(0.9207, device='cuda:0')
[279] Valid Loss:   tensor(0.7925, device='cuda:0')
[280] Valid Loss:   tensor(0.8902, device='cuda:0')
[281] Valid Loss:   tensor(0.6503, device='cuda:0')
[282] Valid Loss:   tensor(0.7655, device='cuda:0')
[283] Valid Loss:   tensor(0.8817, device='cuda:0')
[284] Valid Loss:   tensor(0.6810, device='cuda:0')
[285] Valid Loss:   tensor(0.5808, device='cuda:0')
[286] Valid Loss:   tensor(0.7801, device='cuda:0')
```

```
[287] Valid Loss:   tensor(0.9540, device='cuda:0')
[288] Valid Loss:   tensor(0.5683, device='cuda:0')
[289] Valid Loss:   tensor(0.7495, device='cuda:0')
[290] Valid Loss:   tensor(0.5888, device='cuda:0')
[291] Valid Loss:   tensor(0.7181, device='cuda:0')
[292] Valid Loss:   tensor(0.8510, device='cuda:0')
[293] Valid Loss:   tensor(0.6285, device='cuda:0')
[294] Valid Loss:   tensor(0.8497, device='cuda:0')
[295] Valid Loss:   tensor(0.6703, device='cuda:0')
[296] Valid Loss:   tensor(0.7380, device='cuda:0')
[297] Valid Loss:   tensor(0.8055, device='cuda:0')
[298] Valid Loss:   tensor(0.6531, device='cuda:0')
[299] Valid Loss:   tensor(0.7697, device='cuda:0')
[300] Valid Loss:   tensor(0.9187, device='cuda:0')
[301] Valid Loss:   tensor(0.5921, device='cuda:0')
[302] Valid Loss:   tensor(0.5555, device='cuda:0')
[303] Valid Loss:   tensor(0.6893, device='cuda:0')
[304] Valid Loss:   tensor(0.6665, device='cuda:0')
[305] Valid Loss:   tensor(0.7397, device='cuda:0')
[306] Valid Loss:   tensor(0.7496, device='cuda:0')
[307] Valid Loss:   tensor(0.5916, device='cuda:0')
[308] Valid Loss:   tensor(0.6514, device='cuda:0')
[309] Valid Loss:   tensor(0.6554, device='cuda:0')
[310] Valid Loss:   tensor(0.6769, device='cuda:0')
[311] Valid Loss:   tensor(0.5958, device='cuda:0')
[312] Valid Loss:   tensor(0.5987, device='cuda:0')
[313] Valid Loss:   tensor(0.6991, device='cuda:0')
[314] Valid Loss:   tensor(0.7157, device='cuda:0')
[315] Valid Loss:   tensor(1.0131, device='cuda:0')
[316] Valid Loss:   tensor(0.8489, device='cuda:0')
[317] Valid Loss:   tensor(0.7010, device='cuda:0')
[318] Valid Loss:   tensor(0.6714, device='cuda:0')
[319] Valid Loss:   tensor(0.6521, device='cuda:0')
[320] Valid Loss:   tensor(0.6561, device='cuda:0')
[321] Valid Loss:   tensor(0.6699, device='cuda:0')
[322] Valid Loss:   tensor(0.6333, device='cuda:0')
[323] Valid Loss:   tensor(0.7995, device='cuda:0')
[324] Valid Loss:   tensor(0.7491, device='cuda:0')
[325] Valid Loss:   tensor(0.6165, device='cuda:0')
[326] Valid Loss:   tensor(0.7155, device='cuda:0')
[327] Valid Loss:   tensor(0.4648, device='cuda:0')
[328] Valid Loss:   tensor(0.6681, device='cuda:0')
[329] Valid Loss:   tensor(0.7937, device='cuda:0')
[330] Valid Loss:   tensor(0.5942, device='cuda:0')
[331] Valid Loss:   tensor(0.9184, device='cuda:0')
[332] Valid Loss:   tensor(0.5991, device='cuda:0')
[333] Valid Loss:   tensor(0.5907, device='cuda:0')
[334] Valid Loss:   tensor(0.6094, device='cuda:0')
```

```
[335] Valid Loss:  tensor(0.6890, device='cuda:0')
[336] Valid Loss:  tensor(0.7927, device='cuda:0')
[337] Valid Loss:  tensor(0.8763, device='cuda:0')
[338] Valid Loss:  tensor(0.7552, device='cuda:0')
[339] Valid Loss:  tensor(0.6965, device='cuda:0')
[340] Valid Loss:  tensor(0.7484, device='cuda:0')
[341] Valid Loss:  tensor(0.8460, device='cuda:0')
[342] Valid Loss:  tensor(0.9127, device='cuda:0')
[343] Valid Loss:  tensor(0.7437, device='cuda:0')
[344] Valid Loss:  tensor(0.6267, device='cuda:0')
[345] Valid Loss:  tensor(0.5553, device='cuda:0')
[346] Valid Loss:  tensor(0.7100, device='cuda:0')
[347] Valid Loss:  tensor(0.8414, device='cuda:0')
[348] Valid Loss:  tensor(0.7007, device='cuda:0')
[349] Valid Loss:  tensor(0.6780, device='cuda:0')
[350] Valid Loss:  tensor(0.6893, device='cuda:0')
[351] Valid Loss:  tensor(0.7974, device='cuda:0')
[352] Valid Loss:  tensor(0.6552, device='cuda:0')
[353] Valid Loss:  tensor(0.7368, device='cuda:0')
[354] Valid Loss:  tensor(0.9338, device='cuda:0')
[355] Valid Loss:  tensor(0.5174, device='cuda:0')
[356] Valid Loss:  tensor(0.6743, device='cuda:0')
[357] Valid Loss:  tensor(0.8027, device='cuda:0')
[358] Valid Loss:  tensor(0.6835, device='cuda:0')
[359] Valid Loss:  tensor(0.7831, device='cuda:0')
[360] Valid Loss:  tensor(0.6976, device='cuda:0')
[361] Valid Loss:  tensor(0.7246, device='cuda:0')
[362] Valid Loss:  tensor(0.7244, device='cuda:0')
[363] Valid Loss:  tensor(0.7640, device='cuda:0')
[364] Valid Loss:  tensor(0.9219, device='cuda:0')
[365] Valid Loss:  tensor(0.6875, device='cuda:0')
[366] Valid Loss:  tensor(0.6677, device='cuda:0')
[367] Valid Loss:  tensor(0.7688, device='cuda:0')
[368] Valid Loss:  tensor(0.8148, device='cuda:0')
[369] Valid Loss:  tensor(0.7378, device='cuda:0')
[370] Valid Loss:  tensor(0.6118, device='cuda:0')
[371] Valid Loss:  tensor(0.6257, device='cuda:0')
[372] Valid Loss:  tensor(0.7904, device='cuda:0')
[373] Valid Loss:  tensor(0.5416, device='cuda:0')
[374] Valid Loss:  tensor(0.7015, device='cuda:0')
[375] Valid Loss:  tensor(0.7999, device='cuda:0')
[376] Valid Loss:  tensor(0.6216, device='cuda:0')
[377] Valid Loss:  tensor(0.9198, device='cuda:0')
[378] Valid Loss:  tensor(0.6253, device='cuda:0')
[379] Valid Loss:  tensor(0.8058, device='cuda:0')
[380] Valid Loss:  tensor(1.0710, device='cuda:0')
Testing Loss:  0.7364098899789996
Testing Accuaracy:  77.30410113292842 %
```

```
Epoch Number:  9
[0] Train Loss:  tensor(0.6591, device='cuda:0', grad_fn=<NllLossBackward>)
[1] Train Loss:  tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[2] Train Loss:  tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[3] Train Loss:  tensor(0.6667, device='cuda:0', grad_fn=<NllLossBackward>)
[4] Train Loss:  tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[5] Train Loss:  tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[6] Train Loss:  tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[7] Train Loss:  tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[8] Train Loss:  tensor(0.6473, device='cuda:0', grad_fn=<NllLossBackward>)
[9] Train Loss:  tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[10] Train Loss:  tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[11] Train Loss:  tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[12] Train Loss:  tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[13] Train Loss:  tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[14] Train Loss:  tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[15] Train Loss:  tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[16] Train Loss:  tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[17] Train Loss:  tensor(0.6709, device='cuda:0', grad_fn=<NllLossBackward>)
[18] Train Loss:  tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[19] Train Loss:  tensor(0.6599, device='cuda:0', grad_fn=<NllLossBackward>)
[20] Train Loss:  tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
[21] Train Loss:  tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[22] Train Loss:  tensor(0.6605, device='cuda:0', grad_fn=<NllLossBackward>)
[23] Train Loss:  tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[24] Train Loss:  tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[25] Train Loss:  tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[26] Train Loss:  tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[27] Train Loss:  tensor(0.6589, device='cuda:0', grad_fn=<NllLossBackward>)
[28] Train Loss:  tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[29] Train Loss:  tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[30] Train Loss:  tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[31] Train Loss:  tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[32] Train Loss:  tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[33] Train Loss:  tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[34] Train Loss:  tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[35] Train Loss:  tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[36] Train Loss:  tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[37] Train Loss:  tensor(0.6439, device='cuda:0', grad_fn=<NllLossBackward>)
[38] Train Loss:  tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[39] Train Loss:  tensor(0.6724, device='cuda:0', grad_fn=<NllLossBackward>)
[40] Train Loss:  tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[41] Train Loss:  tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[42] Train Loss:  tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[43] Train Loss:  tensor(0.6716, device='cuda:0', grad_fn=<NllLossBackward>)
[44] Train Loss:  tensor(0.6628, device='cuda:0', grad_fn=<NllLossBackward>)
[45] Train Loss:  tensor(0.6692, device='cuda:0', grad_fn=<NllLossBackward>)
[46] Train Loss:  tensor(0.6500, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[47] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[48] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[49] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[50] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[51] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[52] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[53] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[54] Train Loss:    tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[55] Train Loss:    tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[56] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[57] Train Loss:    tensor(0.6682, device='cuda:0', grad_fn=<NllLossBackward>)
[58] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[59] Train Loss:    tensor(0.6636, device='cuda:0', grad_fn=<NllLossBackward>)
[60] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[61] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[62] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[63] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[64] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[65] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[66] Train Loss:    tensor(0.6621, device='cuda:0', grad_fn=<NllLossBackward>)
[67] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[68] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[69] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[70] Train Loss:    tensor(0.7382, device='cuda:0', grad_fn=<NllLossBackward>)
[71] Train Loss:    tensor(0.6532, device='cuda:0', grad_fn=<NllLossBackward>)
[72] Train Loss:    tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[73] Train Loss:    tensor(0.6669, device='cuda:0', grad_fn=<NllLossBackward>)
[74] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[75] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[76] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[77] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[78] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[79] Train Loss:    tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[80] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[81] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[82] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[83] Train Loss:    tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[84] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[85] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[86] Train Loss:    tensor(0.6692, device='cuda:0', grad_fn=<NllLossBackward>)
[87] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[88] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[89] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[90] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[91] Train Loss:    tensor(0.7221, device='cuda:0', grad_fn=<NllLossBackward>)
[92] Train Loss:    tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[93] Train Loss:    tensor(0.6619, device='cuda:0', grad_fn=<NllLossBackward>)
[94] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[95] Train Loss:   tensor(0.6655, device='cuda:0', grad_fn=<NllLossBackward>)
[96] Train Loss:   tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[97] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[98] Train Loss:   tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[99] Train Loss:   tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[100] Train Loss:  tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[101] Train Loss:  tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[102] Train Loss:  tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[103] Train Loss:  tensor(0.6483, device='cuda:0', grad_fn=<NllLossBackward>)
[104] Train Loss:  tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[105] Train Loss:  tensor(0.6465, device='cuda:0', grad_fn=<NllLossBackward>)
[106] Train Loss:  tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[107] Train Loss:  tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[108] Train Loss:  tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[109] Train Loss:  tensor(0.6470, device='cuda:0', grad_fn=<NllLossBackward>)
[110] Train Loss:  tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[111] Train Loss:  tensor(0.6640, device='cuda:0', grad_fn=<NllLossBackward>)
[112] Train Loss:  tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[113] Train Loss:  tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[114] Train Loss:  tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[115] Train Loss:  tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[116] Train Loss:  tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[117] Train Loss:  tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[118] Train Loss:  tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[119] Train Loss:  tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[120] Train Loss:  tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[121] Train Loss:  tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[122] Train Loss:  tensor(0.6678, device='cuda:0', grad_fn=<NllLossBackward>)
[123] Train Loss:  tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[124] Train Loss:  tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[125] Train Loss:  tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[126] Train Loss:  tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[127] Train Loss:  tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[128] Train Loss:  tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[129] Train Loss:  tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[130] Train Loss:  tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[131] Train Loss:  tensor(0.6492, device='cuda:0', grad_fn=<NllLossBackward>)
[132] Train Loss:  tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[133] Train Loss:  tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[134] Train Loss:  tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[135] Train Loss:  tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[136] Train Loss:  tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[137] Train Loss:  tensor(0.6557, device='cuda:0', grad_fn=<NllLossBackward>)
[138] Train Loss:  tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[139] Train Loss:  tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[140] Train Loss:  tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[141] Train Loss:  tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[142] Train Loss:  tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[143] Train Loss:   tensor(0.6592, device='cuda:0', grad_fn=<NllLossBackward>)
[144] Train Loss:   tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[145] Train Loss:   tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[146] Train Loss:   tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
[147] Train Loss:   tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[148] Train Loss:   tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[149] Train Loss:   tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[150] Train Loss:   tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[151] Train Loss:   tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[152] Train Loss:   tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[153] Train Loss:   tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[154] Train Loss:   tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[155] Train Loss:   tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[156] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[157] Train Loss:   tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[158] Train Loss:   tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[159] Train Loss:   tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[160] Train Loss:   tensor(0.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[161] Train Loss:   tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[162] Train Loss:   tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[163] Train Loss:   tensor(0.6562, device='cuda:0', grad_fn=<NllLossBackward>)
[164] Train Loss:   tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[165] Train Loss:   tensor(0.6467, device='cuda:0', grad_fn=<NllLossBackward>)
[166] Train Loss:   tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[167] Train Loss:   tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[168] Train Loss:   tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[169] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[170] Train Loss:   tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[171] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[172] Train Loss:   tensor(0.6714, device='cuda:0', grad_fn=<NllLossBackward>)
[173] Train Loss:   tensor(0.6533, device='cuda:0', grad_fn=<NllLossBackward>)
[174] Train Loss:   tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[175] Train Loss:   tensor(0.6556, device='cuda:0', grad_fn=<NllLossBackward>)
[176] Train Loss:   tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[177] Train Loss:   tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[178] Train Loss:   tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[179] Train Loss:   tensor(0.7389, device='cuda:0', grad_fn=<NllLossBackward>)
[180] Train Loss:   tensor(0.6649, device='cuda:0', grad_fn=<NllLossBackward>)
[181] Train Loss:   tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[182] Train Loss:   tensor(0.6624, device='cuda:0', grad_fn=<NllLossBackward>)
[183] Train Loss:   tensor(0.6619, device='cuda:0', grad_fn=<NllLossBackward>)
[184] Train Loss:   tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[185] Train Loss:   tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[186] Train Loss:   tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[187] Train Loss:   tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[188] Train Loss:   tensor(0.6572, device='cuda:0', grad_fn=<NllLossBackward>)
[189] Train Loss:   tensor(0.7255, device='cuda:0', grad_fn=<NllLossBackward>)
[190] Train Loss:   tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[191] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[192] Train Loss:    tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[193] Train Loss:    tensor(0.6481, device='cuda:0', grad_fn=<NllLossBackward>)
[194] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[195] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[196] Train Loss:    tensor(0.7294, device='cuda:0', grad_fn=<NllLossBackward>)
[197] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[198] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[199] Train Loss:    tensor(0.6535, device='cuda:0', grad_fn=<NllLossBackward>)
[200] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[201] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[202] Train Loss:    tensor(0.6667, device='cuda:0', grad_fn=<NllLossBackward>)
[203] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[204] Train Loss:    tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[205] Train Loss:    tensor(0.6597, device='cuda:0', grad_fn=<NllLossBackward>)
[206] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[207] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[208] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[209] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[210] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[211] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[212] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[213] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[214] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[215] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[216] Train Loss:    tensor(0.6625, device='cuda:0', grad_fn=<NllLossBackward>)
[217] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[218] Train Loss:    tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[219] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[220] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[221] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[222] Train Loss:    tensor(0.6621, device='cuda:0', grad_fn=<NllLossBackward>)
[223] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[224] Train Loss:    tensor(0.6450, device='cuda:0', grad_fn=<NllLossBackward>)
[225] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[226] Train Loss:    tensor(0.6418, device='cuda:0', grad_fn=<NllLossBackward>)
[227] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[228] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[229] Train Loss:    tensor(0.6570, device='cuda:0', grad_fn=<NllLossBackward>)
[230] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[231] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[232] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[233] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[234] Train Loss:    tensor(0.6568, device='cuda:0', grad_fn=<NllLossBackward>)
[235] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[236] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[237] Train Loss:    tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[238] Train Loss:    tensor(0.6687, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[239] Train Loss:   tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[240] Train Loss:   tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[241] Train Loss:   tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[242] Train Loss:   tensor(0.6567, device='cuda:0', grad_fn=<NllLossBackward>)
[243] Train Loss:   tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[244] Train Loss:   tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[245] Train Loss:   tensor(0.6625, device='cuda:0', grad_fn=<NllLossBackward>)
[246] Train Loss:   tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[247] Train Loss:   tensor(0.6277, device='cuda:0', grad_fn=<NllLossBackward>)
[248] Train Loss:   tensor(0.6724, device='cuda:0', grad_fn=<NllLossBackward>)
[249] Train Loss:   tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[250] Train Loss:   tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[251] Train Loss:   tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[252] Train Loss:   tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[253] Train Loss:   tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[254] Train Loss:   tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[255] Train Loss:   tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[256] Train Loss:   tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[257] Train Loss:   tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[258] Train Loss:   tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[259] Train Loss:   tensor(0.6460, device='cuda:0', grad_fn=<NllLossBackward>)
[260] Train Loss:   tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[261] Train Loss:   tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[262] Train Loss:   tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[263] Train Loss:   tensor(0.7320, device='cuda:0', grad_fn=<NllLossBackward>)
[264] Train Loss:   tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[265] Train Loss:   tensor(0.6486, device='cuda:0', grad_fn=<NllLossBackward>)
[266] Train Loss:   tensor(0.6353, device='cuda:0', grad_fn=<NllLossBackward>)
[267] Train Loss:   tensor(0.6654, device='cuda:0', grad_fn=<NllLossBackward>)
[268] Train Loss:   tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[269] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[270] Train Loss:   tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[271] Train Loss:   tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[272] Train Loss:   tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[273] Train Loss:   tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[274] Train Loss:   tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[275] Train Loss:   tensor(0.6621, device='cuda:0', grad_fn=<NllLossBackward>)
[276] Train Loss:   tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[277] Train Loss:   tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[278] Train Loss:   tensor(0.6671, device='cuda:0', grad_fn=<NllLossBackward>)
[279] Train Loss:   tensor(0.6655, device='cuda:0', grad_fn=<NllLossBackward>)
[280] Train Loss:   tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[281] Train Loss:   tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[282] Train Loss:   tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[283] Train Loss:   tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[284] Train Loss:   tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[285] Train Loss:   tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[286] Train Loss:   tensor(0.6508, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[287] Train Loss:   tensor(0.6583, device='cuda:0', grad_fn=<NllLossBackward>)
[288] Train Loss:   tensor(0.6714, device='cuda:0', grad_fn=<NllLossBackward>)
[289] Train Loss:   tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[290] Train Loss:   tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[291] Train Loss:   tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[292] Train Loss:   tensor(0.6739, device='cuda:0', grad_fn=<NllLossBackward>)
[293] Train Loss:   tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[294] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[295] Train Loss:   tensor(0.6577, device='cuda:0', grad_fn=<NllLossBackward>)
[296] Train Loss:   tensor(0.6658, device='cuda:0', grad_fn=<NllLossBackward>)
[297] Train Loss:   tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[298] Train Loss:   tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[299] Train Loss:   tensor(0.6682, device='cuda:0', grad_fn=<NllLossBackward>)
[300] Train Loss:   tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[301] Train Loss:   tensor(0.6696, device='cuda:0', grad_fn=<NllLossBackward>)
[302] Train Loss:   tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[303] Train Loss:   tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[304] Train Loss:   tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[305] Train Loss:   tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[306] Train Loss:   tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[307] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[308] Train Loss:   tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[309] Train Loss:   tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[310] Train Loss:   tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[311] Train Loss:   tensor(0.7263, device='cuda:0', grad_fn=<NllLossBackward>)
[312] Train Loss:   tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[313] Train Loss:   tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[314] Train Loss:   tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[315] Train Loss:   tensor(0.6492, device='cuda:0', grad_fn=<NllLossBackward>)
[316] Train Loss:   tensor(0.6613, device='cuda:0', grad_fn=<NllLossBackward>)
[317] Train Loss:   tensor(0.6685, device='cuda:0', grad_fn=<NllLossBackward>)
[318] Train Loss:   tensor(0.6375, device='cuda:0', grad_fn=<NllLossBackward>)
[319] Train Loss:   tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[320] Train Loss:   tensor(0.6719, device='cuda:0', grad_fn=<NllLossBackward>)
[321] Train Loss:   tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[322] Train Loss:   tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[323] Train Loss:   tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[324] Train Loss:   tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[325] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[326] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[327] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[328] Train Loss:   tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[329] Train Loss:   tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[330] Train Loss:   tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[331] Train Loss:   tensor(0.6694, device='cuda:0', grad_fn=<NllLossBackward>)
[332] Train Loss:   tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[333] Train Loss:   tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[334] Train Loss:   tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[335] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[336] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[337] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[338] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[339] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[340] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[341] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[342] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[343] Train Loss:    tensor(0.6480, device='cuda:0', grad_fn=<NllLossBackward>)
[344] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
[345] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[346] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[347] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[348] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[349] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[350] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[351] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[352] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[353] Train Loss:    tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
[354] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[355] Train Loss:    tensor(0.6641, device='cuda:0', grad_fn=<NllLossBackward>)
[356] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[357] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[358] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[359] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[360] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[361] Train Loss:    tensor(0.6339, device='cuda:0', grad_fn=<NllLossBackward>)
[362] Train Loss:    tensor(0.6687, device='cuda:0', grad_fn=<NllLossBackward>)
[363] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[364] Train Loss:    tensor(0.6545, device='cuda:0', grad_fn=<NllLossBackward>)
[365] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[366] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[367] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[368] Train Loss:    tensor(0.6729, device='cuda:0', grad_fn=<NllLossBackward>)
[369] Train Loss:    tensor(0.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[370] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[371] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[372] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[373] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[374] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[375] Train Loss:    tensor(0.7357, device='cuda:0', grad_fn=<NllLossBackward>)
[376] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[377] Train Loss:    tensor(0.6639, device='cuda:0', grad_fn=<NllLossBackward>)
[378] Train Loss:    tensor(0.6542, device='cuda:0', grad_fn=<NllLossBackward>)
[379] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[380] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[381] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[382] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[383] Train Loss:    tensor(0.6592, device='cuda:0', grad_fn=<NllLossBackward>)
[384] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[385] Train Loss:    tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[386] Train Loss:    tensor(0.6698, device='cuda:0', grad_fn=<NllLossBackward>)
[387] Train Loss:    tensor(0.6667, device='cuda:0', grad_fn=<NllLossBackward>)
[388] Train Loss:    tensor(0.7335, device='cuda:0', grad_fn=<NllLossBackward>)
[389] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[390] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[391] Train Loss:    tensor(0.6623, device='cuda:0', grad_fn=<NllLossBackward>)
[392] Train Loss:    tensor(0.6701, device='cuda:0', grad_fn=<NllLossBackward>)
[393] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[394] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[395] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[396] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[397] Train Loss:    tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[398] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[399] Train Loss:    tensor(0.6454, device='cuda:0', grad_fn=<NllLossBackward>)
[400] Train Loss:    tensor(0.6587, device='cuda:0', grad_fn=<NllLossBackward>)
[401] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[402] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[403] Train Loss:    tensor(0.6546, device='cuda:0', grad_fn=<NllLossBackward>)
[404] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[405] Train Loss:    tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
[406] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[407] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[408] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[409] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[410] Train Loss:    tensor(0.6698, device='cuda:0', grad_fn=<NllLossBackward>)
[411] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[412] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[413] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[414] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[415] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[416] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[417] Train Loss:    tensor(0.6554, device='cuda:0', grad_fn=<NllLossBackward>)
[418] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[419] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[420] Train Loss:    tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[421] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[422] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[423] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[424] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[425] Train Loss:    tensor(0.6682, device='cuda:0', grad_fn=<NllLossBackward>)
[426] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[427] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[428] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[429] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[430] Train Loss:    tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[431] Train Loss:   tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[432] Train Loss:   tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[433] Train Loss:   tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[434] Train Loss:   tensor(0.7285, device='cuda:0', grad_fn=<NllLossBackward>)
[435] Train Loss:   tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[436] Train Loss:   tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[437] Train Loss:   tensor(0.6682, device='cuda:0', grad_fn=<NllLossBackward>)
[438] Train Loss:   tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[439] Train Loss:   tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[440] Train Loss:   tensor(0.6516, device='cuda:0', grad_fn=<NllLossBackward>)
[441] Train Loss:   tensor(0.7302, device='cuda:0', grad_fn=<NllLossBackward>)
[442] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[443] Train Loss:   tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[444] Train Loss:   tensor(0.6563, device='cuda:0', grad_fn=<NllLossBackward>)
[445] Train Loss:   tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[446] Train Loss:   tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[447] Train Loss:   tensor(0.6521, device='cuda:0', grad_fn=<NllLossBackward>)
[448] Train Loss:   tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[449] Train Loss:   tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[450] Train Loss:   tensor(0.6709, device='cuda:0', grad_fn=<NllLossBackward>)
[451] Train Loss:   tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[452] Train Loss:   tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[453] Train Loss:   tensor(0.6581, device='cuda:0', grad_fn=<NllLossBackward>)
[454] Train Loss:   tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[455] Train Loss:   tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[456] Train Loss:   tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[457] Train Loss:   tensor(0.6645, device='cuda:0', grad_fn=<NllLossBackward>)
[458] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[459] Train Loss:   tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[460] Train Loss:   tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[461] Train Loss:   tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[462] Train Loss:   tensor(0.6580, device='cuda:0', grad_fn=<NllLossBackward>)
[463] Train Loss:   tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[464] Train Loss:   tensor(0.6621, device='cuda:0', grad_fn=<NllLossBackward>)
[465] Train Loss:   tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[466] Train Loss:   tensor(0.6672, device='cuda:0', grad_fn=<NllLossBackward>)
[467] Train Loss:   tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[468] Train Loss:   tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[469] Train Loss:   tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[470] Train Loss:   tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[471] Train Loss:   tensor(0.6714, device='cuda:0', grad_fn=<NllLossBackward>)
[472] Train Loss:   tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[473] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[474] Train Loss:   tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[475] Train Loss:   tensor(0.6719, device='cuda:0', grad_fn=<NllLossBackward>)
[476] Train Loss:   tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[477] Train Loss:   tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[478] Train Loss:   tensor(0.6739, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[479] Train Loss:   tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[480] Train Loss:   tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[481] Train Loss:   tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[482] Train Loss:   tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[483] Train Loss:   tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[484] Train Loss:   tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[485] Train Loss:   tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[486] Train Loss:   tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[487] Train Loss:   tensor(0.6686, device='cuda:0', grad_fn=<NllLossBackward>)
[488] Train Loss:   tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[489] Train Loss:   tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[490] Train Loss:   tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[491] Train Loss:   tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[492] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[493] Train Loss:   tensor(0.6603, device='cuda:0', grad_fn=<NllLossBackward>)
[494] Train Loss:   tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[495] Train Loss:   tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[496] Train Loss:   tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[497] Train Loss:   tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[498] Train Loss:   tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[499] Train Loss:   tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[500] Train Loss:   tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[501] Train Loss:   tensor(0.6682, device='cuda:0', grad_fn=<NllLossBackward>)
[502] Train Loss:   tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[503] Train Loss:   tensor(0.6629, device='cuda:0', grad_fn=<NllLossBackward>)
[504] Train Loss:   tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[505] Train Loss:   tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[506] Train Loss:   tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[507] Train Loss:   tensor(0.6606, device='cuda:0', grad_fn=<NllLossBackward>)
[508] Train Loss:   tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[509] Train Loss:   tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[510] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[511] Train Loss:   tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[512] Train Loss:   tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[513] Train Loss:   tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[514] Train Loss:   tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[515] Train Loss:   tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[516] Train Loss:   tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[517] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[518] Train Loss:   tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[519] Train Loss:   tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[520] Train Loss:   tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[521] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[522] Train Loss:   tensor(0.6525, device='cuda:0', grad_fn=<NllLossBackward>)
[523] Train Loss:   tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[524] Train Loss:   tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[525] Train Loss:   tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[526] Train Loss:   tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[527] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[528] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[529] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[530] Train Loss:    tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[531] Train Loss:    tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[532] Train Loss:    tensor(0.6699, device='cuda:0', grad_fn=<NllLossBackward>)
[533] Train Loss:    tensor(0.6580, device='cuda:0', grad_fn=<NllLossBackward>)
[534] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[535] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[536] Train Loss:    tensor(0.6581, device='cuda:0', grad_fn=<NllLossBackward>)
[537] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[538] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[539] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[540] Train Loss:    tensor(0.6701, device='cuda:0', grad_fn=<NllLossBackward>)
[541] Train Loss:    tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[542] Train Loss:    tensor(0.6639, device='cuda:0', grad_fn=<NllLossBackward>)
[543] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[544] Train Loss:    tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[545] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[546] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[547] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[548] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[549] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[550] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[551] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[552] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[553] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[554] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[555] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[556] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[557] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[558] Train Loss:    tensor(0.6590, device='cuda:0', grad_fn=<NllLossBackward>)
[559] Train Loss:    tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[560] Train Loss:    tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[561] Train Loss:    tensor(0.6638, device='cuda:0', grad_fn=<NllLossBackward>)
[562] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[563] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[564] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[565] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[566] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[567] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[568] Train Loss:    tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[569] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[570] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[571] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[572] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[573] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[574] Train Loss:    tensor(0.7205, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[575] Train Loss:   tensor(0.6547, device='cuda:0', grad_fn=<NllLossBackward>)
[576] Train Loss:   tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[577] Train Loss:   tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[578] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[579] Train Loss:   tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[580] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[581] Train Loss:   tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[582] Train Loss:   tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[583] Train Loss:   tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[584] Train Loss:   tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[585] Train Loss:   tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[586] Train Loss:   tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[587] Train Loss:   tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[588] Train Loss:   tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[589] Train Loss:   tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[590] Train Loss:   tensor(0.6714, device='cuda:0', grad_fn=<NllLossBackward>)
[591] Train Loss:   tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[592] Train Loss:   tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[593] Train Loss:   tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[594] Train Loss:   tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[595] Train Loss:   tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[596] Train Loss:   tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[597] Train Loss:   tensor(0.6704, device='cuda:0', grad_fn=<NllLossBackward>)
[598] Train Loss:   tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[599] Train Loss:   tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[600] Train Loss:   tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[601] Train Loss:   tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[602] Train Loss:   tensor(0.6610, device='cuda:0', grad_fn=<NllLossBackward>)
[603] Train Loss:   tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[604] Train Loss:   tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[605] Train Loss:   tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[606] Train Loss:   tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[607] Train Loss:   tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[608] Train Loss:   tensor(0.6613, device='cuda:0', grad_fn=<NllLossBackward>)
[609] Train Loss:   tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[610] Train Loss:   tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[611] Train Loss:   tensor(0.6570, device='cuda:0', grad_fn=<NllLossBackward>)
[612] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[613] Train Loss:   tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[614] Train Loss:   tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[615] Train Loss:   tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[616] Train Loss:   tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[617] Train Loss:   tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[618] Train Loss:   tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[619] Train Loss:   tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[620] Train Loss:   tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[621] Train Loss:   tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[622] Train Loss:   tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[623] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[624] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[625] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[626] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[627] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[628] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[629] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[630] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[631] Train Loss:    tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[632] Train Loss:    tensor(0.6616, device='cuda:0', grad_fn=<NllLossBackward>)
[633] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[634] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[635] Train Loss:    tensor(0.6692, device='cuda:0', grad_fn=<NllLossBackward>)
[636] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[637] Train Loss:    tensor(0.6496, device='cuda:0', grad_fn=<NllLossBackward>)
[638] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[639] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[640] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[641] Train Loss:    tensor(0.7117, device='cuda:0', grad_fn=<NllLossBackward>)
[642] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[643] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[644] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[645] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[646] Train Loss:    tensor(0.6500, device='cuda:0', grad_fn=<NllLossBackward>)
[647] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[648] Train Loss:    tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[649] Train Loss:    tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[650] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[651] Train Loss:    tensor(0.6676, device='cuda:0', grad_fn=<NllLossBackward>)
[652] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[653] Train Loss:    tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[654] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[655] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[656] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[657] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[658] Train Loss:    tensor(0.6401, device='cuda:0', grad_fn=<NllLossBackward>)
[659] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[660] Train Loss:    tensor(0.6561, device='cuda:0', grad_fn=<NllLossBackward>)
[661] Train Loss:    tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[662] Train Loss:    tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[663] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[664] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[665] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[666] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[667] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[668] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[669] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[670] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[671] Train Loss:   tensor(0.6483, device='cuda:0', grad_fn=<NllLossBackward>)
[672] Train Loss:   tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[673] Train Loss:   tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[674] Train Loss:   tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[675] Train Loss:   tensor(0.6676, device='cuda:0', grad_fn=<NllLossBackward>)
[676] Train Loss:   tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[677] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[678] Train Loss:   tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[679] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[680] Train Loss:   tensor(0.7319, device='cuda:0', grad_fn=<NllLossBackward>)
[681] Train Loss:   tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[682] Train Loss:   tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[683] Train Loss:   tensor(0.6273, device='cuda:0', grad_fn=<NllLossBackward>)
[684] Train Loss:   tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[685] Train Loss:   tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[686] Train Loss:   tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[687] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[688] Train Loss:   tensor(0.6687, device='cuda:0', grad_fn=<NllLossBackward>)
[689] Train Loss:   tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[690] Train Loss:   tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[691] Train Loss:   tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[692] Train Loss:   tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[693] Train Loss:   tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[694] Train Loss:   tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[695] Train Loss:   tensor(0.6525, device='cuda:0', grad_fn=<NllLossBackward>)
[696] Train Loss:   tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[697] Train Loss:   tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[698] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[699] Train Loss:   tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[700] Train Loss:   tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[701] Train Loss:   tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[702] Train Loss:   tensor(0.6578, device='cuda:0', grad_fn=<NllLossBackward>)
[703] Train Loss:   tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[704] Train Loss:   tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[705] Train Loss:   tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[706] Train Loss:   tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[707] Train Loss:   tensor(0.7280, device='cuda:0', grad_fn=<NllLossBackward>)
[708] Train Loss:   tensor(0.6535, device='cuda:0', grad_fn=<NllLossBackward>)
[709] Train Loss:   tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[710] Train Loss:   tensor(0.6663, device='cuda:0', grad_fn=<NllLossBackward>)
[711] Train Loss:   tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[712] Train Loss:   tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[713] Train Loss:   tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[714] Train Loss:   tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[715] Train Loss:   tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[716] Train Loss:   tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[717] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[718] Train Loss:   tensor(0.6549, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[719] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[720] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[721] Train Loss:    tensor(0.6608, device='cuda:0', grad_fn=<NllLossBackward>)
[722] Train Loss:    tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[723] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[724] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[725] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[726] Train Loss:    tensor(0.6705, device='cuda:0', grad_fn=<NllLossBackward>)
[727] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[728] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[729] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[730] Train Loss:    tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[731] Train Loss:    tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[732] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[733] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[734] Train Loss:    tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[735] Train Loss:    tensor(0.6504, device='cuda:0', grad_fn=<NllLossBackward>)
[736] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[737] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[738] Train Loss:    tensor(0.6691, device='cuda:0', grad_fn=<NllLossBackward>)
[739] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[740] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[741] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[742] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[743] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[744] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[745] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[746] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[747] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[748] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[749] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[750] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[751] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[752] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[753] Train Loss:    tensor(0.6654, device='cuda:0', grad_fn=<NllLossBackward>)
[754] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[755] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[756] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[757] Train Loss:    tensor(0.7310, device='cuda:0', grad_fn=<NllLossBackward>)
[758] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[759] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[760] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[761] Train Loss:    tensor(0.6701, device='cuda:0', grad_fn=<NllLossBackward>)
[762] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[763] Train Loss:    tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[764] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[765] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[766] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[767] Train Loss:   tensor(0.6776, device='cuda:0', grad_fn=<NllLossBackward>)
[768] Train Loss:   tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[769] Train Loss:   tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[770] Train Loss:   tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[771] Train Loss:   tensor(0.6545, device='cuda:0', grad_fn=<NllLossBackward>)
[772] Train Loss:   tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[773] Train Loss:   tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[774] Train Loss:   tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[775] Train Loss:   tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[776] Train Loss:   tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[777] Train Loss:   tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[778] Train Loss:   tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[779] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[780] Train Loss:   tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[781] Train Loss:   tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[782] Train Loss:   tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[783] Train Loss:   tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[784] Train Loss:   tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[785] Train Loss:   tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[786] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[787] Train Loss:   tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[788] Train Loss:   tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[789] Train Loss:   tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[790] Train Loss:   tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[791] Train Loss:   tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[792] Train Loss:   tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[793] Train Loss:   tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[794] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[795] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[796] Train Loss:   tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[797] Train Loss:   tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[798] Train Loss:   tensor(0.7402, device='cuda:0', grad_fn=<NllLossBackward>)
[799] Train Loss:   tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[800] Train Loss:   tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[801] Train Loss:   tensor(0.6502, device='cuda:0', grad_fn=<NllLossBackward>)
[802] Train Loss:   tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[803] Train Loss:   tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[804] Train Loss:   tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[805] Train Loss:   tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[806] Train Loss:   tensor(0.6311, device='cuda:0', grad_fn=<NllLossBackward>)
[807] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[808] Train Loss:   tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[809] Train Loss:   tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[810] Train Loss:   tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[811] Train Loss:   tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[812] Train Loss:   tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[813] Train Loss:   tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[814] Train Loss:   tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[815] Train Loss:   tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[816] Train Loss:   tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[817] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[818] Train Loss:   tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[819] Train Loss:   tensor(0.6704, device='cuda:0', grad_fn=<NllLossBackward>)
[820] Train Loss:   tensor(0.6433, device='cuda:0', grad_fn=<NllLossBackward>)
[821] Train Loss:   tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[822] Train Loss:   tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[823] Train Loss:   tensor(0.7329, device='cuda:0', grad_fn=<NllLossBackward>)
[824] Train Loss:   tensor(0.6739, device='cuda:0', grad_fn=<NllLossBackward>)
[825] Train Loss:   tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[826] Train Loss:   tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[827] Train Loss:   tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[828] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[829] Train Loss:   tensor(0.6701, device='cuda:0', grad_fn=<NllLossBackward>)
[830] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[831] Train Loss:   tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[832] Train Loss:   tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[833] Train Loss:   tensor(0.6519, device='cuda:0', grad_fn=<NllLossBackward>)
[834] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[835] Train Loss:   tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[836] Train Loss:   tensor(0.6624, device='cuda:0', grad_fn=<NllLossBackward>)
[837] Train Loss:   tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[838] Train Loss:   tensor(0.6660, device='cuda:0', grad_fn=<NllLossBackward>)
[839] Train Loss:   tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[840] Train Loss:   tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[841] Train Loss:   tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[842] Train Loss:   tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[843] Train Loss:   tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[844] Train Loss:   tensor(0.6623, device='cuda:0', grad_fn=<NllLossBackward>)
[845] Train Loss:   tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[846] Train Loss:   tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[847] Train Loss:   tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[848] Train Loss:   tensor(0.6646, device='cuda:0', grad_fn=<NllLossBackward>)
[849] Train Loss:   tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[850] Train Loss:   tensor(0.6602, device='cuda:0', grad_fn=<NllLossBackward>)
[851] Train Loss:   tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[852] Train Loss:   tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[853] Train Loss:   tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[854] Train Loss:   tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[855] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[856] Train Loss:   tensor(0.6535, device='cuda:0', grad_fn=<NllLossBackward>)
[857] Train Loss:   tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[858] Train Loss:   tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[859] Train Loss:   tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[860] Train Loss:   tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[861] Train Loss:   tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[862] Train Loss:   tensor(0.6644, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[863] Train Loss:   tensor(0.6683, device='cuda:0', grad_fn=<NllLossBackward>)
[864] Train Loss:   tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[865] Train Loss:   tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[866] Train Loss:   tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[867] Train Loss:   tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[868] Train Loss:   tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[869] Train Loss:   tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[870] Train Loss:   tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[871] Train Loss:   tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[872] Train Loss:   tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[873] Train Loss:   tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[874] Train Loss:   tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[875] Train Loss:   tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[876] Train Loss:   tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[877] Train Loss:   tensor(0.6484, device='cuda:0', grad_fn=<NllLossBackward>)
[878] Train Loss:   tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[879] Train Loss:   tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[880] Train Loss:   tensor(0.6691, device='cuda:0', grad_fn=<NllLossBackward>)
[881] Train Loss:   tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[882] Train Loss:   tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[883] Train Loss:   tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[884] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[885] Train Loss:   tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[886] Train Loss:   tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[887] Train Loss:   tensor(0.6378, device='cuda:0', grad_fn=<NllLossBackward>)
[888] Train Loss:   tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[889] Train Loss:   tensor(0.6739, device='cuda:0', grad_fn=<NllLossBackward>)
[890] Train Loss:   tensor(0.6661, device='cuda:0', grad_fn=<NllLossBackward>)
[891] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[892] Train Loss:   tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[893] Train Loss:   tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[894] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[895] Train Loss:   tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[896] Train Loss:   tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[897] Train Loss:   tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[898] Train Loss:   tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[899] Train Loss:   tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[900] Train Loss:   tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[901] Train Loss:   tensor(0.6716, device='cuda:0', grad_fn=<NllLossBackward>)
[902] Train Loss:   tensor(0.6502, device='cuda:0', grad_fn=<NllLossBackward>)
[903] Train Loss:   tensor(0.6550, device='cuda:0', grad_fn=<NllLossBackward>)
[904] Train Loss:   tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[905] Train Loss:   tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[906] Train Loss:   tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[907] Train Loss:   tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[908] Train Loss:   tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[909] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[910] Train Loss:   tensor(0.6654, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[911] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[912] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[913] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[914] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[915] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[916] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[917] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[918] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[919] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[920] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[921] Train Loss:    tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[922] Train Loss:    tensor(0.6630, device='cuda:0', grad_fn=<NllLossBackward>)
[923] Train Loss:    tensor(0.6668, device='cuda:0', grad_fn=<NllLossBackward>)
[924] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[925] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[926] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[927] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[928] Train Loss:    tensor(0.6593, device='cuda:0', grad_fn=<NllLossBackward>)
[929] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[930] Train Loss:    tensor(0.6590, device='cuda:0', grad_fn=<NllLossBackward>)
[931] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[932] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[933] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[934] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[935] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[936] Train Loss:    tensor(0.6716, device='cuda:0', grad_fn=<NllLossBackward>)
[937] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[938] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[939] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[940] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[941] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[942] Train Loss:    tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[943] Train Loss:    tensor(0.6527, device='cuda:0', grad_fn=<NllLossBackward>)
[944] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[945] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[946] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[947] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[948] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[949] Train Loss:    tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
[950] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[951] Train Loss:    tensor(0.6622, device='cuda:0', grad_fn=<NllLossBackward>)
[952] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[953] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[954] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[955] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[956] Train Loss:    tensor(0.6609, device='cuda:0', grad_fn=<NllLossBackward>)
[957] Train Loss:    tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[958] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[959] Train Loss:  tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[960] Train Loss:  tensor(0.6660, device='cuda:0', grad_fn=<NllLossBackward>)
[961] Train Loss:  tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[962] Train Loss:  tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[963] Train Loss:  tensor(0.6686, device='cuda:0', grad_fn=<NllLossBackward>)
[964] Train Loss:  tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[965] Train Loss:  tensor(0.6606, device='cuda:0', grad_fn=<NllLossBackward>)
[966] Train Loss:  tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[967] Train Loss:  tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[968] Train Loss:  tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[969] Train Loss:  tensor(0.6687, device='cuda:0', grad_fn=<NllLossBackward>)
[970] Train Loss:  tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[971] Train Loss:  tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[972] Train Loss:  tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[973] Train Loss:  tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[974] Train Loss:  tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[975] Train Loss:  tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[976] Train Loss:  tensor(0.6596, device='cuda:0', grad_fn=<NllLossBackward>)
[977] Train Loss:  tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[978] Train Loss:  tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[979] Train Loss:  tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[980] Train Loss:  tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[981] Train Loss:  tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[982] Train Loss:  tensor(0.6699, device='cuda:0', grad_fn=<NllLossBackward>)
[983] Train Loss:  tensor(0.6650, device='cuda:0', grad_fn=<NllLossBackward>)
[984] Train Loss:  tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[985] Train Loss:  tensor(0.6494, device='cuda:0', grad_fn=<NllLossBackward>)
[986] Train Loss:  tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[987] Train Loss:  tensor(0.6598, device='cuda:0', grad_fn=<NllLossBackward>)
[988] Train Loss:  tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[989] Train Loss:  tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[990] Train Loss:  tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[991] Train Loss:  tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[992] Train Loss:  tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[993] Train Loss:  tensor(0.6665, device='cuda:0', grad_fn=<NllLossBackward>)
[994] Train Loss:  tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[995] Train Loss:  tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[996] Train Loss:  tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[997] Train Loss:  tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[998] Train Loss:  tensor(0.6550, device='cuda:0', grad_fn=<NllLossBackward>)
[999] Train Loss:  tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[1000] Train Loss:  tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[1001] Train Loss:  tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[1002] Train Loss:  tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[1003] Train Loss:  tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[1004] Train Loss:  tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[1005] Train Loss:  tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[1006] Train Loss:  tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1007] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[1008] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[1009] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[1010] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[1011] Train Loss:    tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[1012] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[1013] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[1014] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[1015] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[1016] Train Loss:    tensor(0.6645, device='cuda:0', grad_fn=<NllLossBackward>)
[1017] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[1018] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[1019] Train Loss:    tensor(0.7210, device='cuda:0', grad_fn=<NllLossBackward>)
[1020] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[1021] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[1022] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[1023] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[1024] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[1025] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[1026] Train Loss:    tensor(0.6577, device='cuda:0', grad_fn=<NllLossBackward>)
[1027] Train Loss:    tensor(0.6639, device='cuda:0', grad_fn=<NllLossBackward>)
[1028] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[1029] Train Loss:    tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[1030] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[1031] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[1032] Train Loss:    tensor(0.6503, device='cuda:0', grad_fn=<NllLossBackward>)
[1033] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[1034] Train Loss:    tensor(0.7360, device='cuda:0', grad_fn=<NllLossBackward>)
[1035] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[1036] Train Loss:    tensor(0.6641, device='cuda:0', grad_fn=<NllLossBackward>)
[1037] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[1038] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[1039] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1040] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[1041] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[1042] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[1043] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[1044] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[1045] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[1046] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[1047] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[1048] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[1049] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[1050] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1051] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
[1052] Train Loss:    tensor(0.7408, device='cuda:0', grad_fn=<NllLossBackward>)
[1053] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[1054] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1055] Train Loss:   tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[1056] Train Loss:   tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[1057] Train Loss:   tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[1058] Train Loss:   tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[1059] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[1060] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1061] Train Loss:   tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[1062] Train Loss:   tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[1063] Train Loss:   tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[1064] Train Loss:   tensor(0.6607, device='cuda:0', grad_fn=<NllLossBackward>)
[1065] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[1066] Train Loss:   tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[1067] Train Loss:   tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[1068] Train Loss:   tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[1069] Train Loss:   tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[1070] Train Loss:   tensor(0.6630, device='cuda:0', grad_fn=<NllLossBackward>)
[1071] Train Loss:   tensor(0.6543, device='cuda:0', grad_fn=<NllLossBackward>)
[1072] Train Loss:   tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[1073] Train Loss:   tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[1074] Train Loss:   tensor(0.6627, device='cuda:0', grad_fn=<NllLossBackward>)
[1075] Train Loss:   tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[1076] Train Loss:   tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[1077] Train Loss:   tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[1078] Train Loss:   tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[1079] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[1080] Train Loss:   tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[1081] Train Loss:   tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1082] Train Loss:   tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[1083] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[1084] Train Loss:   tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[1085] Train Loss:   tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[1086] Train Loss:   tensor(0.6599, device='cuda:0', grad_fn=<NllLossBackward>)
[1087] Train Loss:   tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[1088] Train Loss:   tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1089] Train Loss:   tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[1090] Train Loss:   tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[1091] Train Loss:   tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[1092] Train Loss:   tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[1093] Train Loss:   tensor(0.6494, device='cuda:0', grad_fn=<NllLossBackward>)
[1094] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1095] Train Loss:   tensor(0.6385, device='cuda:0', grad_fn=<NllLossBackward>)
[1096] Train Loss:   tensor(0.6733, device='cuda:0', grad_fn=<NllLossBackward>)
[1097] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[1098] Train Loss:   tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[1099] Train Loss:   tensor(0.6957, device='cuda:0', grad_fn=<NllLossBackward>)
[1100] Train Loss:   tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[1101] Train Loss:   tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[1102] Train Loss:   tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1103] Train Loss:    tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[1104] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[1105] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[1106] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1107] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[1108] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[1109] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[1110] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[1111] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[1112] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[1113] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1114] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[1115] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[1116] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[1117] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[1118] Train Loss:    tensor(0.6686, device='cuda:0', grad_fn=<NllLossBackward>)
[1119] Train Loss:    tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[1120] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[1121] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[1122] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[1123] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[1124] Train Loss:    tensor(0.6721, device='cuda:0', grad_fn=<NllLossBackward>)
[1125] Train Loss:    tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[1126] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[1127] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[1128] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[1129] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[1130] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[1131] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[1132] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[1133] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[1134] Train Loss:    tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[1135] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[1136] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1137] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[1138] Train Loss:    tensor(0.6564, device='cuda:0', grad_fn=<NllLossBackward>)
[1139] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[1140] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[1141] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[1142] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1143] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[1144] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[1145] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[1146] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[1147] Train Loss:    tensor(0.7208, device='cuda:0', grad_fn=<NllLossBackward>)
[1148] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[1149] Train Loss:    tensor(0.6540, device='cuda:0', grad_fn=<NllLossBackward>)
[1150] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1151] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[1152] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[1153] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[1154] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[1155] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[1156] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[1157] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[1158] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[1159] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[1160] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[1161] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[1162] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[1163] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[1164] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[1165] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[1166] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[1167] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[1168] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[1169] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[1170] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[1171] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[1172] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1173] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1174] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[1175] Train Loss:    tensor(0.7225, device='cuda:0', grad_fn=<NllLossBackward>)
[1176] Train Loss:    tensor(0.6378, device='cuda:0', grad_fn=<NllLossBackward>)
[1177] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[1178] Train Loss:    tensor(0.6524, device='cuda:0', grad_fn=<NllLossBackward>)
[1179] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[1180] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[1181] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[1182] Train Loss:    tensor(0.6694, device='cuda:0', grad_fn=<NllLossBackward>)
[1183] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[1184] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[1185] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[1186] Train Loss:    tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[1187] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[1188] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[1189] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1190] Train Loss:    tensor(0.6678, device='cuda:0', grad_fn=<NllLossBackward>)
[1191] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[1192] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[1193] Train Loss:    tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[1194] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[1195] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[1196] Train Loss:    tensor(0.6599, device='cuda:0', grad_fn=<NllLossBackward>)
[1197] Train Loss:    tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[1198] Train Loss:    tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1199] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[1200] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[1201] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[1202] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1203] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[1204] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[1205] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[1206] Train Loss:    tensor(0.6555, device='cuda:0', grad_fn=<NllLossBackward>)
[1207] Train Loss:    tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[1208] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[1209] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[1210] Train Loss:    tensor(0.6581, device='cuda:0', grad_fn=<NllLossBackward>)
[1211] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[1212] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[1213] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[1214] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[1215] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[1216] Train Loss:    tensor(0.6701, device='cuda:0', grad_fn=<NllLossBackward>)
[1217] Train Loss:    tensor(0.6733, device='cuda:0', grad_fn=<NllLossBackward>)
[1218] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[1219] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[1220] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[1221] Train Loss:    tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[1222] Train Loss:    tensor(0.7317, device='cuda:0', grad_fn=<NllLossBackward>)
[1223] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[1224] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[1225] Train Loss:    tensor(0.6715, device='cuda:0', grad_fn=<NllLossBackward>)
[1226] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1227] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[1228] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1229] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[1230] Train Loss:    tensor(0.6526, device='cuda:0', grad_fn=<NllLossBackward>)
[1231] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[1232] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[1233] Train Loss:    tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[1234] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[1235] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[1236] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[1237] Train Loss:    tensor(0.6529, device='cuda:0', grad_fn=<NllLossBackward>)
[1238] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1239] Train Loss:    tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[1240] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[1241] Train Loss:    tensor(0.6551, device='cuda:0', grad_fn=<NllLossBackward>)
[1242] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[1243] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[1244] Train Loss:    tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[1245] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[1246] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1247] Train Loss:    tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[1248] Train Loss:    tensor(0.6566, device='cuda:0', grad_fn=<NllLossBackward>)
[1249] Train Loss:    tensor(0.6646, device='cuda:0', grad_fn=<NllLossBackward>)
[1250] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[1251] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1252] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[1253] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[1254] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[1255] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[1256] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[1257] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[1258] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[1259] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[1260] Train Loss:    tensor(0.6724, device='cuda:0', grad_fn=<NllLossBackward>)
[1261] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[1262] Train Loss:    tensor(0.6627, device='cuda:0', grad_fn=<NllLossBackward>)
[1263] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[1264] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[1265] Train Loss:    tensor(0.7248, device='cuda:0', grad_fn=<NllLossBackward>)
[1266] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[1267] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[1268] Train Loss:    tensor(0.7541, device='cuda:0', grad_fn=<NllLossBackward>)
[1269] Train Loss:    tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
[1270] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1271] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[1272] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[1273] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[1274] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[1275] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[1276] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[1277] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[1278] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[1279] Train Loss:    tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[1280] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[1281] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[1282] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[1283] Train Loss:    tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[1284] Train Loss:    tensor(0.6655, device='cuda:0', grad_fn=<NllLossBackward>)
[1285] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[1286] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[1287] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[1288] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[1289] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[1290] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[1291] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[1292] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[1293] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[1294] Train Loss:    tensor(0.6686, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1295] Train Loss:    tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[1296] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[1297] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[1298] Train Loss:    tensor(0.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[1299] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[1300] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[1301] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[1302] Train Loss:    tensor(0.6706, device='cuda:0', grad_fn=<NllLossBackward>)
[1303] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[1304] Train Loss:    tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[1305] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[1306] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[1307] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[1308] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[1309] Train Loss:    tensor(0.6678, device='cuda:0', grad_fn=<NllLossBackward>)
[1310] Train Loss:    tensor(0.7266, device='cuda:0', grad_fn=<NllLossBackward>)
[1311] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[1312] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[1313] Train Loss:    tensor(0.6667, device='cuda:0', grad_fn=<NllLossBackward>)
[1314] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[1315] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[1316] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[1317] Train Loss:    tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[1318] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[1319] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[1320] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[1321] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[1322] Train Loss:    tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[1323] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[1324] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[1325] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[1326] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[1327] Train Loss:    tensor(0.6715, device='cuda:0', grad_fn=<NllLossBackward>)
[1328] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[1329] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[1330] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[1331] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[1332] Train Loss:    tensor(0.6568, device='cuda:0', grad_fn=<NllLossBackward>)
[1333] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[1334] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[1335] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[1336] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[1337] Train Loss:    tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[1338] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[1339] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[1340] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[1341] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[1342] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1343] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1344] Train Loss:    tensor(0.6684, device='cuda:0', grad_fn=<NllLossBackward>)
[1345] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[1346] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[1347] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[1348] Train Loss:    tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[1349] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[1350] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[1351] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[1352] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[1353] Train Loss:    tensor(0.6653, device='cuda:0', grad_fn=<NllLossBackward>)
[1354] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[1355] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[1356] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[1357] Train Loss:    tensor(0.6578, device='cuda:0', grad_fn=<NllLossBackward>)
[1358] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[1359] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[1360] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[1361] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[1362] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[1363] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[1364] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[1365] Train Loss:    tensor(0.6664, device='cuda:0', grad_fn=<NllLossBackward>)
[1366] Train Loss:    tensor(0.7315, device='cuda:0', grad_fn=<NllLossBackward>)
[1367] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[1368] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[1369] Train Loss:    tensor(0.6714, device='cuda:0', grad_fn=<NllLossBackward>)
[1370] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1371] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[1372] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[1373] Train Loss:    tensor(0.6570, device='cuda:0', grad_fn=<NllLossBackward>)
[1374] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1375] Train Loss:    tensor(0.7105, device='cuda:0', grad_fn=<NllLossBackward>)
[1376] Train Loss:    tensor(0.6367, device='cuda:0', grad_fn=<NllLossBackward>)
[1377] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[1378] Train Loss:    tensor(0.7211, device='cuda:0', grad_fn=<NllLossBackward>)
[1379] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[1380] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[1381] Train Loss:    tensor(0.7276, device='cuda:0', grad_fn=<NllLossBackward>)
[1382] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[1383] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[1384] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[1385] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[1386] Train Loss:    tensor(0.7195, device='cuda:0', grad_fn=<NllLossBackward>)
[1387] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[1388] Train Loss:    tensor(0.6645, device='cuda:0', grad_fn=<NllLossBackward>)
[1389] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[1390] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1391] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[1392] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[1393] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[1394] Train Loss:    tensor(0.6270, device='cuda:0', grad_fn=<NllLossBackward>)
[1395] Train Loss:    tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[1396] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[1397] Train Loss:    tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[1398] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[1399] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[1400] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[1401] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[1402] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[1403] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[1404] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[1405] Train Loss:    tensor(0.7314, device='cuda:0', grad_fn=<NllLossBackward>)
[1406] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[1407] Train Loss:    tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[1408] Train Loss:    tensor(0.6722, device='cuda:0', grad_fn=<NllLossBackward>)
[1409] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[1410] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[1411] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[1412] Train Loss:    tensor(0.6671, device='cuda:0', grad_fn=<NllLossBackward>)
[1413] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[1414] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[1415] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1416] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[1417] Train Loss:    tensor(0.6479, device='cuda:0', grad_fn=<NllLossBackward>)
[1418] Train Loss:    tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[1419] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[1420] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[1421] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[1422] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[1423] Train Loss:    tensor(0.7394, device='cuda:0', grad_fn=<NllLossBackward>)
[1424] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[1425] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[1426] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[1427] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1428] Train Loss:    tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
[1429] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[1430] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[1431] Train Loss:    tensor(0.6462, device='cuda:0', grad_fn=<NllLossBackward>)
[1432] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[1433] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[1434] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[1435] Train Loss:    tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[1436] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[1437] Train Loss:    tensor(0.6656, device='cuda:0', grad_fn=<NllLossBackward>)
[1438] Train Loss:    tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1439] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[1440] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[1441] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[1442] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[1443] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1444] Train Loss:    tensor(0.6672, device='cuda:0', grad_fn=<NllLossBackward>)
[1445] Train Loss:    tensor(0.6435, device='cuda:0', grad_fn=<NllLossBackward>)
[1446] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[1447] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[1448] Train Loss:    tensor(0.6640, device='cuda:0', grad_fn=<NllLossBackward>)
[1449] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[1450] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1451] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[1452] Train Loss:    tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[1453] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[1454] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[1455] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[1456] Train Loss:    tensor(0.6637, device='cuda:0', grad_fn=<NllLossBackward>)
[1457] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[1458] Train Loss:    tensor(0.6676, device='cuda:0', grad_fn=<NllLossBackward>)
[1459] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[1460] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[1461] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[1462] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[1463] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[1464] Train Loss:    tensor(0.6513, device='cuda:0', grad_fn=<NllLossBackward>)
[1465] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[1466] Train Loss:    tensor(0.6670, device='cuda:0', grad_fn=<NllLossBackward>)
[1467] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1468] Train Loss:    tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[1469] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[1470] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[1471] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[1472] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[1473] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[1474] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[1475] Train Loss:    tensor(0.7341, device='cuda:0', grad_fn=<NllLossBackward>)
[1476] Train Loss:    tensor(0.6588, device='cuda:0', grad_fn=<NllLossBackward>)
[1477] Train Loss:    tensor(0.6452, device='cuda:0', grad_fn=<NllLossBackward>)
[1478] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[1479] Train Loss:    tensor(0.6669, device='cuda:0', grad_fn=<NllLossBackward>)
[1480] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[1481] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[1482] Train Loss:    tensor(0.6488, device='cuda:0', grad_fn=<NllLossBackward>)
[1483] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[1484] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[1485] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
[1486] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1487] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[1488] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[1489] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[1490] Train Loss:    tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[1491] Train Loss:    tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[1492] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[1493] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[1494] Train Loss:    tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[1495] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[1496] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1497] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[1498] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[1499] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[1500] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[1501] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[1502] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1503] Train Loss:    tensor(0.6593, device='cuda:0', grad_fn=<NllLossBackward>)
[1504] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[1505] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[1506] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[1507] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[1508] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[1509] Train Loss:    tensor(0.6575, device='cuda:0', grad_fn=<NllLossBackward>)
[1510] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1511] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[1512] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[1513] Train Loss:    tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[1514] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[1515] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[1516] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[1517] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[1518] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[1519] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1520] Train Loss:    tensor(0.6669, device='cuda:0', grad_fn=<NllLossBackward>)
[1521] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[1522] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[1523] Train Loss:    tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[1524] Train Loss:    tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[1525] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[1526] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[1527] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[1528] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[1529] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[1530] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[1531] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[1532] Train Loss:    tensor(0.6442, device='cuda:0', grad_fn=<NllLossBackward>)
[1533] Train Loss:    tensor(0.6649, device='cuda:0', grad_fn=<NllLossBackward>)
[1534] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1535] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[1536] Train Loss:    tensor(0.6543, device='cuda:0', grad_fn=<NllLossBackward>)
[1537] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[1538] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[1539] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[1540] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[1541] Train Loss:    tensor(0.6659, device='cuda:0', grad_fn=<NllLossBackward>)
[1542] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[1543] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[1544] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[1545] Train Loss:    tensor(0.6654, device='cuda:0', grad_fn=<NllLossBackward>)
[1546] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[1547] Train Loss:    tensor(0.6575, device='cuda:0', grad_fn=<NllLossBackward>)
[1548] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[1549] Train Loss:    tensor(0.6608, device='cuda:0', grad_fn=<NllLossBackward>)
[1550] Train Loss:    tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[1551] Train Loss:    tensor(0.6499, device='cuda:0', grad_fn=<NllLossBackward>)
[1552] Train Loss:    tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[1553] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[1554] Train Loss:    tensor(0.6652, device='cuda:0', grad_fn=<NllLossBackward>)
[1555] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[1556] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[1557] Train Loss:    tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[1558] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[1559] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[1560] Train Loss:    tensor(0.6466, device='cuda:0', grad_fn=<NllLossBackward>)
[1561] Train Loss:    tensor(0.6684, device='cuda:0', grad_fn=<NllLossBackward>)
[1562] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[1563] Train Loss:    tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[1564] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[1565] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[1566] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[1567] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[1568] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[1569] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[1570] Train Loss:    tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[1571] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[1572] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[1573] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[1574] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[1575] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[1576] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[1577] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[1578] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[1579] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[1580] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[1581] Train Loss:    tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[1582] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1583] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[1584] Train Loss:    tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[1585] Train Loss:    tensor(0.6676, device='cuda:0', grad_fn=<NllLossBackward>)
[1586] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[1587] Train Loss:    tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[1588] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[1589] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[1590] Train Loss:    tensor(0.7365, device='cuda:0', grad_fn=<NllLossBackward>)
[1591] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[1592] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[1593] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[1594] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[1595] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[1596] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[1597] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[1598] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[1599] Train Loss:    tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[1600] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[1601] Train Loss:    tensor(0.7190, device='cuda:0', grad_fn=<NllLossBackward>)
[1602] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[1603] Train Loss:    tensor(0.6776, device='cuda:0', grad_fn=<NllLossBackward>)
[1604] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[1605] Train Loss:    tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[1606] Train Loss:    tensor(0.7355, device='cuda:0', grad_fn=<NllLossBackward>)
[1607] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[1608] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[1609] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[1610] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[1611] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[1612] Train Loss:    tensor(0.6776, device='cuda:0', grad_fn=<NllLossBackward>)
[1613] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[1614] Train Loss:    tensor(0.6706, device='cuda:0', grad_fn=<NllLossBackward>)
[1615] Train Loss:    tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[1616] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[1617] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[1618] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[1619] Train Loss:    tensor(0.6694, device='cuda:0', grad_fn=<NllLossBackward>)
[1620] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[1621] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[1622] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[1623] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[1624] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[1625] Train Loss:    tensor(0.6706, device='cuda:0', grad_fn=<NllLossBackward>)
[1626] Train Loss:    tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[1627] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[1628] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[1629] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[1630] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1631] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[1632] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[1633] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[1634] Train Loss:    tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[1635] Train Loss:    tensor(0.6480, device='cuda:0', grad_fn=<NllLossBackward>)
[1636] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[1637] Train Loss:    tensor(0.6603, device='cuda:0', grad_fn=<NllLossBackward>)
[1638] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[1639] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[1640] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[1641] Train Loss:    tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
[1642] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[1643] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[1644] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[1645] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[1646] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[1647] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[1648] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[1649] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[1650] Train Loss:    tensor(0.6599, device='cuda:0', grad_fn=<NllLossBackward>)
[1651] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[1652] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[1653] Train Loss:    tensor(0.6856, device='cuda:0', grad_fn=<NllLossBackward>)
[1654] Train Loss:    tensor(0.6492, device='cuda:0', grad_fn=<NllLossBackward>)
[1655] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[1656] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[1657] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[1658] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[1659] Train Loss:    tensor(0.6415, device='cuda:0', grad_fn=<NllLossBackward>)
[1660] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[1661] Train Loss:    tensor(0.6699, device='cuda:0', grad_fn=<NllLossBackward>)
[1662] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[1663] Train Loss:    tensor(0.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[1664] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[1665] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[1666] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[1667] Train Loss:    tensor(0.6683, device='cuda:0', grad_fn=<NllLossBackward>)
[1668] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[1669] Train Loss:    tensor(0.7236, device='cuda:0', grad_fn=<NllLossBackward>)
[1670] Train Loss:    tensor(0.6724, device='cuda:0', grad_fn=<NllLossBackward>)
[1671] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[1672] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[1673] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[1674] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[1675] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[1676] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[1677] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[1678] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1679] Train Loss:    tensor(0.6706, device='cuda:0', grad_fn=<NllLossBackward>)
[1680] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1681] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[1682] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[1683] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[1684] Train Loss:    tensor(0.6654, device='cuda:0', grad_fn=<NllLossBackward>)
[1685] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[1686] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[1687] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[1688] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[1689] Train Loss:    tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[1690] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[1691] Train Loss:    tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[1692] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[1693] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[1694] Train Loss:    tensor(0.6566, device='cuda:0', grad_fn=<NllLossBackward>)
[1695] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[1696] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[1697] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[1698] Train Loss:    tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[1699] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[1700] Train Loss:    tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[1701] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[1702] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[1703] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[1704] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[1705] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[1706] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[1707] Train Loss:    tensor(0.6617, device='cuda:0', grad_fn=<NllLossBackward>)
[1708] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[1709] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[1710] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[1711] Train Loss:    tensor(0.6715, device='cuda:0', grad_fn=<NllLossBackward>)
[1712] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1713] Train Loss:    tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[1714] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[1715] Train Loss:    tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[1716] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[1717] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[1718] Train Loss:    tensor(0.6588, device='cuda:0', grad_fn=<NllLossBackward>)
[1719] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[1720] Train Loss:    tensor(0.6612, device='cuda:0', grad_fn=<NllLossBackward>)
[1721] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[1722] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[1723] Train Loss:    tensor(0.6632, device='cuda:0', grad_fn=<NllLossBackward>)
[1724] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[1725] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[1726] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1727] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[1728] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[1729] Train Loss:    tensor(0.6664, device='cuda:0', grad_fn=<NllLossBackward>)
[1730] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[1731] Train Loss:    tensor(0.6621, device='cuda:0', grad_fn=<NllLossBackward>)
[1732] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[1733] Train Loss:    tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[1734] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[1735] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[1736] Train Loss:    tensor(0.6590, device='cuda:0', grad_fn=<NllLossBackward>)
[1737] Train Loss:    tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[1738] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[1739] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[1740] Train Loss:    tensor(0.6527, device='cuda:0', grad_fn=<NllLossBackward>)
[1741] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[1742] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[1743] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[1744] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[1745] Train Loss:    tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[1746] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[1747] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[1748] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[1749] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[1750] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[1751] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[1752] Train Loss:    tensor(0.7145, device='cuda:0', grad_fn=<NllLossBackward>)
[1753] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[1754] Train Loss:    tensor(0.6703, device='cuda:0', grad_fn=<NllLossBackward>)
[1755] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[1756] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[1757] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[1758] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[1759] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[1760] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[1761] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[1762] Train Loss:    tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[1763] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[1764] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[1765] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[1766] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[1767] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[1768] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[1769] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[1770] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[1771] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[1772] Train Loss:    tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[1773] Train Loss:    tensor(0.6661, device='cuda:0', grad_fn=<NllLossBackward>)
[1774] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1775] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[1776] Train Loss:    tensor(0.6387, device='cuda:0', grad_fn=<NllLossBackward>)
[1777] Train Loss:    tensor(0.6997, device='cuda:0', grad_fn=<NllLossBackward>)
[1778] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[1779] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[1780] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[1781] Train Loss:    tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[1782] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[1783] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[1784] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[1785] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[1786] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[1787] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[1788] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[1789] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[1790] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[1791] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[1792] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[1793] Train Loss:    tensor(0.6631, device='cuda:0', grad_fn=<NllLossBackward>)
[1794] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[1795] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[1796] Train Loss:    tensor(0.6639, device='cuda:0', grad_fn=<NllLossBackward>)
[1797] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[1798] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[1799] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[1800] Train Loss:    tensor(0.6612, device='cuda:0', grad_fn=<NllLossBackward>)
[1801] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[1802] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[1803] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[1804] Train Loss:    tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[1805] Train Loss:    tensor(0.6552, device='cuda:0', grad_fn=<NllLossBackward>)
[1806] Train Loss:    tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[1807] Train Loss:    tensor(0.7169, device='cuda:0', grad_fn=<NllLossBackward>)
[1808] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[1809] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[1810] Train Loss:    tensor(0.6619, device='cuda:0', grad_fn=<NllLossBackward>)
[1811] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[1812] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[1813] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[1814] Train Loss:    tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[1815] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[1816] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[1817] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[1818] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[1819] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[1820] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[1821] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[1822] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1823] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[1824] Train Loss:    tensor(0.7446, device='cuda:0', grad_fn=<NllLossBackward>)
[1825] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[1826] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[1827] Train Loss:    tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[1828] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[1829] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[1830] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[1831] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[1832] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[1833] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[1834] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[1835] Train Loss:    tensor(0.6664, device='cuda:0', grad_fn=<NllLossBackward>)
[1836] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[1837] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[1838] Train Loss:    tensor(0.6609, device='cuda:0', grad_fn=<NllLossBackward>)
[1839] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[1840] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[1841] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[1842] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[1843] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[1844] Train Loss:    tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[1845] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[1846] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[1847] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[1848] Train Loss:    tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[1849] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[1850] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[1851] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[1852] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[1853] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[1854] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[1855] Train Loss:    tensor(0.6619, device='cuda:0', grad_fn=<NllLossBackward>)
[1856] Train Loss:    tensor(0.6653, device='cuda:0', grad_fn=<NllLossBackward>)
[1857] Train Loss:    tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[1858] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[1859] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[1860] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[1861] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[1862] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[1863] Train Loss:    tensor(0.6459, device='cuda:0', grad_fn=<NllLossBackward>)
[1864] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[1865] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[1866] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[1867] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[1868] Train Loss:    tensor(0.6494, device='cuda:0', grad_fn=<NllLossBackward>)
[1869] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[1870] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1871] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[1872] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[1873] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[1874] Train Loss:    tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[1875] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[1876] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[1877] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[1878] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[1879] Train Loss:    tensor(0.6596, device='cuda:0', grad_fn=<NllLossBackward>)
[1880] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[1881] Train Loss:    tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
[1882] Train Loss:    tensor(0.6661, device='cuda:0', grad_fn=<NllLossBackward>)
[1883] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1884] Train Loss:    tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
[1885] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
[1886] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[1887] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[1888] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[1889] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[1890] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[1891] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[1892] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[1893] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[1894] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[1895] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[1896] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[1897] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[1898] Train Loss:    tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[1899] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[1900] Train Loss:    tensor(0.6705, device='cuda:0', grad_fn=<NllLossBackward>)
[1901] Train Loss:    tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[1902] Train Loss:    tensor(0.6716, device='cuda:0', grad_fn=<NllLossBackward>)
[1903] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[1904] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[1905] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[1906] Train Loss:    tensor(0.6667, device='cuda:0', grad_fn=<NllLossBackward>)
[1907] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[1908] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[1909] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1910] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[1911] Train Loss:    tensor(0.6500, device='cuda:0', grad_fn=<NllLossBackward>)
[1912] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[1913] Train Loss:    tensor(0.6622, device='cuda:0', grad_fn=<NllLossBackward>)
[1914] Train Loss:    tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[1915] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[1916] Train Loss:    tensor(0.7292, device='cuda:0', grad_fn=<NllLossBackward>)
[1917] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[1918] Train Loss:    tensor(0.6502, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1919] Train Loss:    tensor(0.6637, device='cuda:0', grad_fn=<NllLossBackward>)
[1920] Train Loss:    tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[1921] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[1922] Train Loss:    tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[1923] Train Loss:    tensor(0.6724, device='cuda:0', grad_fn=<NllLossBackward>)
[1924] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[1925] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[1926] Train Loss:    tensor(0.6647, device='cuda:0', grad_fn=<NllLossBackward>)
[1927] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[1928] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[1929] Train Loss:    tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[1930] Train Loss:    tensor(0.6699, device='cuda:0', grad_fn=<NllLossBackward>)
[1931] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[1932] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[1933] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[1934] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[1935] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[1936] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[1937] Train Loss:    tensor(0.6600, device='cuda:0', grad_fn=<NllLossBackward>)
[1938] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[1939] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[1940] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1941] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[1942] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[1943] Train Loss:    tensor(0.6509, device='cuda:0', grad_fn=<NllLossBackward>)
[1944] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[1945] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[1946] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[1947] Train Loss:    tensor(0.6510, device='cuda:0', grad_fn=<NllLossBackward>)
[1948] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[1949] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[1950] Train Loss:    tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
[1951] Train Loss:    tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[1952] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[1953] Train Loss:    tensor(0.6776, device='cuda:0', grad_fn=<NllLossBackward>)
[1954] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[1955] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[1956] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[1957] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[1958] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[1959] Train Loss:    tensor(0.6503, device='cuda:0', grad_fn=<NllLossBackward>)
[1960] Train Loss:    tensor(0.6654, device='cuda:0', grad_fn=<NllLossBackward>)
[1961] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[1962] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[1963] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[1964] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[1965] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[1966] Train Loss:    tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[1967] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[1968] Train Loss:    tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[1969] Train Loss:    tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[1970] Train Loss:    tensor(0.6665, device='cuda:0', grad_fn=<NllLossBackward>)
[1971] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[1972] Train Loss:    tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[1973] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[1974] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[1975] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[1976] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[1977] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[1978] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[1979] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[1980] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[1981] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[1982] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[1983] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[1984] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[1985] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[1986] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[1987] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[1988] Train Loss:    tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[1989] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[1990] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[1991] Train Loss:    tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[1992] Train Loss:    tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[1993] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[1994] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[1995] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[1996] Train Loss:    tensor(0.6716, device='cuda:0', grad_fn=<NllLossBackward>)
[1997] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[1998] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[1999] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[2000] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2001] Train Loss:    tensor(0.6743, device='cuda:0', grad_fn=<NllLossBackward>)
[2002] Train Loss:    tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[2003] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[2004] Train Loss:    tensor(0.6682, device='cuda:0', grad_fn=<NllLossBackward>)
[2005] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[2006] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[2007] Train Loss:    tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[2008] Train Loss:    tensor(0.6598, device='cuda:0', grad_fn=<NllLossBackward>)
[2009] Train Loss:    tensor(0.6696, device='cuda:0', grad_fn=<NllLossBackward>)
[2010] Train Loss:    tensor(0.7489, device='cuda:0', grad_fn=<NllLossBackward>)
[2011] Train Loss:    tensor(0.7303, device='cuda:0', grad_fn=<NllLossBackward>)
[2012] Train Loss:    tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[2013] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[2014] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2015] Train Loss:    tensor(0.6568, device='cuda:0', grad_fn=<NllLossBackward>)
[2016] Train Loss:    tensor(0.6696, device='cuda:0', grad_fn=<NllLossBackward>)
[2017] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[2018] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[2019] Train Loss:    tensor(0.6629, device='cuda:0', grad_fn=<NllLossBackward>)
[2020] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[2021] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[2022] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[2023] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[2024] Train Loss:    tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[2025] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2026] Train Loss:    tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[2027] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[2028] Train Loss:    tensor(0.6704, device='cuda:0', grad_fn=<NllLossBackward>)
[2029] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[2030] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[2031] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[2032] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[2033] Train Loss:    tensor(0.6565, device='cuda:0', grad_fn=<NllLossBackward>)
[2034] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[2035] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[2036] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[2037] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[2038] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[2039] Train Loss:    tensor(0.6748, device='cuda:0', grad_fn=<NllLossBackward>)
[2040] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[2041] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[2042] Train Loss:    tensor(0.7418, device='cuda:0', grad_fn=<NllLossBackward>)
[2043] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[2044] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[2045] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[2046] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[2047] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[2048] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[2049] Train Loss:    tensor(0.6683, device='cuda:0', grad_fn=<NllLossBackward>)
[2050] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[2051] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[2052] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[2053] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[2054] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[2055] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2056] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[2057] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2058] Train Loss:    tensor(0.6548, device='cuda:0', grad_fn=<NllLossBackward>)
[2059] Train Loss:    tensor(0.6705, device='cuda:0', grad_fn=<NllLossBackward>)
[2060] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[2061] Train Loss:    tensor(0.6688, device='cuda:0', grad_fn=<NllLossBackward>)
[2062] Train Loss:    tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2063] Train Loss:    tensor(0.6705, device='cuda:0', grad_fn=<NllLossBackward>)
[2064] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[2065] Train Loss:    tensor(0.6638, device='cuda:0', grad_fn=<NllLossBackward>)
[2066] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[2067] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[2068] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[2069] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[2070] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[2071] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2072] Train Loss:    tensor(0.6687, device='cuda:0', grad_fn=<NllLossBackward>)
[2073] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[2074] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[2075] Train Loss:    tensor(0.6638, device='cuda:0', grad_fn=<NllLossBackward>)
[2076] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[2077] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[2078] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[2079] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[2080] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[2081] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[2082] Train Loss:    tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[2083] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[2084] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[2085] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[2086] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[2087] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[2088] Train Loss:    tensor(0.7152, device='cuda:0', grad_fn=<NllLossBackward>)
[2089] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[2090] Train Loss:    tensor(0.6544, device='cuda:0', grad_fn=<NllLossBackward>)
[2091] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[2092] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[2093] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[2094] Train Loss:    tensor(0.7251, device='cuda:0', grad_fn=<NllLossBackward>)
[2095] Train Loss:    tensor(0.6696, device='cuda:0', grad_fn=<NllLossBackward>)
[2096] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[2097] Train Loss:    tensor(0.6709, device='cuda:0', grad_fn=<NllLossBackward>)
[2098] Train Loss:    tensor(0.6582, device='cuda:0', grad_fn=<NllLossBackward>)
[2099] Train Loss:    tensor(0.6558, device='cuda:0', grad_fn=<NllLossBackward>)
[2100] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[2101] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[2102] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[2103] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[2104] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[2105] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[2106] Train Loss:    tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[2107] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[2108] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[2109] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[2110] Train Loss:    tensor(0.6716, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2111] Train Loss:    tensor(0.6580, device='cuda:0', grad_fn=<NllLossBackward>)
[2112] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[2113] Train Loss:    tensor(0.6635, device='cuda:0', grad_fn=<NllLossBackward>)
[2114] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[2115] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[2116] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[2117] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[2118] Train Loss:    tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[2119] Train Loss:    tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[2120] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[2121] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2122] Train Loss:    tensor(0.6406, device='cuda:0', grad_fn=<NllLossBackward>)
[2123] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[2124] Train Loss:    tensor(0.6656, device='cuda:0', grad_fn=<NllLossBackward>)
[2125] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[2126] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[2127] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[2128] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[2129] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[2130] Train Loss:    tensor(0.6669, device='cuda:0', grad_fn=<NllLossBackward>)
[2131] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[2132] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[2133] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2134] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[2135] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[2136] Train Loss:    tensor(0.6638, device='cuda:0', grad_fn=<NllLossBackward>)
[2137] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[2138] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[2139] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[2140] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[2141] Train Loss:    tensor(0.7460, device='cuda:0', grad_fn=<NllLossBackward>)
[2142] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[2143] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[2144] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[2145] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[2146] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[2147] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[2148] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
[2149] Train Loss:    tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[2150] Train Loss:    tensor(0.7396, device='cuda:0', grad_fn=<NllLossBackward>)
[2151] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[2152] Train Loss:    tensor(0.7196, device='cuda:0', grad_fn=<NllLossBackward>)
[2153] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2154] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[2155] Train Loss:    tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[2156] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[2157] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[2158] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2159] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[2160] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[2161] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[2162] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[2163] Train Loss:    tensor(0.6622, device='cuda:0', grad_fn=<NllLossBackward>)
[2164] Train Loss:    tensor(0.6722, device='cuda:0', grad_fn=<NllLossBackward>)
[2165] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[2166] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[2167] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[2168] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[2169] Train Loss:    tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[2170] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[2171] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[2172] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[2173] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[2174] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2175] Train Loss:    tensor(0.6509, device='cuda:0', grad_fn=<NllLossBackward>)
[2176] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[2177] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[2178] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[2179] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[2180] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[2181] Train Loss:    tensor(0.7193, device='cuda:0', grad_fn=<NllLossBackward>)
[2182] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[2183] Train Loss:    tensor(0.7107, device='cuda:0', grad_fn=<NllLossBackward>)
[2184] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[2185] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[2186] Train Loss:    tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[2187] Train Loss:    tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[2188] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[2189] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[2190] Train Loss:    tensor(0.7122, device='cuda:0', grad_fn=<NllLossBackward>)
[2191] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[2192] Train Loss:    tensor(0.7203, device='cuda:0', grad_fn=<NllLossBackward>)
[2193] Train Loss:    tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[2194] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[2195] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[2196] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[2197] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[2198] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[2199] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[2200] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[2201] Train Loss:    tensor(0.7213, device='cuda:0', grad_fn=<NllLossBackward>)
[2202] Train Loss:    tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[2203] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[2204] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[2205] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[2206] Train Loss:    tensor(0.7212, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2207] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[2208] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[2209] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[2210] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[2211] Train Loss:    tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[2212] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[2213] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[2214] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[2215] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[2216] Train Loss:    tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[2217] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[2218] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[2219] Train Loss:    tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[2220] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[2221] Train Loss:    tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[2222] Train Loss:    tensor(0.6624, device='cuda:0', grad_fn=<NllLossBackward>)
[2223] Train Loss:    tensor(0.7363, device='cuda:0', grad_fn=<NllLossBackward>)
[2224] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[2225] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[2226] Train Loss:    tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[2227] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[2228] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2229] Train Loss:    tensor(0.6460, device='cuda:0', grad_fn=<NllLossBackward>)
[2230] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[2231] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[2232] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[2233] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[2234] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[2235] Train Loss:    tensor(0.6592, device='cuda:0', grad_fn=<NllLossBackward>)
[2236] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[2237] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[2238] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[2239] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[2240] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[2241] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[2242] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[2243] Train Loss:    tensor(0.6655, device='cuda:0', grad_fn=<NllLossBackward>)
[2244] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[2245] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[2246] Train Loss:    tensor(0.6616, device='cuda:0', grad_fn=<NllLossBackward>)
[2247] Train Loss:    tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[2248] Train Loss:    tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[2249] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[2250] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[2251] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2252] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[2253] Train Loss:    tensor(0.7085, device='cuda:0', grad_fn=<NllLossBackward>)
[2254] Train Loss:    tensor(0.6523, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2255] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[2256] Train Loss:    tensor(0.6688, device='cuda:0', grad_fn=<NllLossBackward>)
[2257] Train Loss:    tensor(0.7233, device='cuda:0', grad_fn=<NllLossBackward>)
[2258] Train Loss:    tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[2259] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[2260] Train Loss:    tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[2261] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[2262] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[2263] Train Loss:    tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[2264] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[2265] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[2266] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[2267] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[2268] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[2269] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[2270] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[2271] Train Loss:    tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[2272] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[2273] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[2274] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[2275] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[2276] Train Loss:    tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[2277] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[2278] Train Loss:    tensor(0.7191, device='cuda:0', grad_fn=<NllLossBackward>)
[2279] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[2280] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[2281] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[2282] Train Loss:    tensor(0.6470, device='cuda:0', grad_fn=<NllLossBackward>)
[2283] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[2284] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[2285] Train Loss:    tensor(0.6819, device='cuda:0', grad_fn=<NllLossBackward>)
[2286] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[2287] Train Loss:    tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[2288] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[2289] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[2290] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[2291] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[2292] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[2293] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[2294] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[2295] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2296] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[2297] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[2298] Train Loss:    tensor(0.7092, device='cuda:0', grad_fn=<NllLossBackward>)
[2299] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[2300] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[2301] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2302] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2303] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[2304] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2305] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[2306] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[2307] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[2308] Train Loss:    tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[2309] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2310] Train Loss:    tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[2311] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[2312] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[2313] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
[2314] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[2315] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[2316] Train Loss:    tensor(0.7214, device='cuda:0', grad_fn=<NllLossBackward>)
[2317] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[2318] Train Loss:    tensor(0.7349, device='cuda:0', grad_fn=<NllLossBackward>)
[2319] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[2320] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[2321] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[2322] Train Loss:    tensor(0.6679, device='cuda:0', grad_fn=<NllLossBackward>)
[2323] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[2324] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[2325] Train Loss:    tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[2326] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[2327] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2328] Train Loss:    tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[2329] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[2330] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[2331] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[2332] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[2333] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[2334] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[2335] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[2336] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[2337] Train Loss:    tensor(0.7164, device='cuda:0', grad_fn=<NllLossBackward>)
[2338] Train Loss:    tensor(0.7080, device='cuda:0', grad_fn=<NllLossBackward>)
[2339] Train Loss:    tensor(0.6704, device='cuda:0', grad_fn=<NllLossBackward>)
[2340] Train Loss:    tensor(0.6484, device='cuda:0', grad_fn=<NllLossBackward>)
[2341] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[2342] Train Loss:    tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[2343] Train Loss:    tensor(0.6696, device='cuda:0', grad_fn=<NllLossBackward>)
[2344] Train Loss:    tensor(0.6409, device='cuda:0', grad_fn=<NllLossBackward>)
[2345] Train Loss:    tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[2346] Train Loss:    tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[2347] Train Loss:    tensor(0.6578, device='cuda:0', grad_fn=<NllLossBackward>)
[2348] Train Loss:    tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
[2349] Train Loss:    tensor(0.6672, device='cuda:0', grad_fn=<NllLossBackward>)
[2350] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2351] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[2352] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[2353] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[2354] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[2355] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[2356] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[2357] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[2358] Train Loss:    tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[2359] Train Loss:    tensor(0.6497, device='cuda:0', grad_fn=<NllLossBackward>)
[2360] Train Loss:    tensor(0.6679, device='cuda:0', grad_fn=<NllLossBackward>)
[2361] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[2362] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[2363] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[2364] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[2365] Train Loss:    tensor(0.6704, device='cuda:0', grad_fn=<NllLossBackward>)
[2366] Train Loss:    tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[2367] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[2368] Train Loss:    tensor(0.7200, device='cuda:0', grad_fn=<NllLossBackward>)
[2369] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[2370] Train Loss:    tensor(0.6638, device='cuda:0', grad_fn=<NllLossBackward>)
[2371] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[2372] Train Loss:    tensor(0.7458, device='cuda:0', grad_fn=<NllLossBackward>)
[2373] Train Loss:    tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
[2374] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[2375] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[2376] Train Loss:    tensor(0.6719, device='cuda:0', grad_fn=<NllLossBackward>)
[2377] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[2378] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[2379] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[2380] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[2381] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[2382] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2383] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[2384] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[2385] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[2386] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[2387] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2388] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[2389] Train Loss:    tensor(0.6616, device='cuda:0', grad_fn=<NllLossBackward>)
[2390] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2391] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[2392] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[2393] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[2394] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[2395] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[2396] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[2397] Train Loss:    tensor(0.7487, device='cuda:0', grad_fn=<NllLossBackward>)
[2398] Train Loss:    tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2399] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[2400] Train Loss:    tensor(0.6672, device='cuda:0', grad_fn=<NllLossBackward>)
[2401] Train Loss:    tensor(0.6664, device='cuda:0', grad_fn=<NllLossBackward>)
[2402] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[2403] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[2404] Train Loss:    tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[2405] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[2406] Train Loss:    tensor(0.6668, device='cuda:0', grad_fn=<NllLossBackward>)
[2407] Train Loss:    tensor(0.6685, device='cuda:0', grad_fn=<NllLossBackward>)
[2408] Train Loss:    tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[2409] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[2410] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[2411] Train Loss:    tensor(0.7016, device='cuda:0', grad_fn=<NllLossBackward>)
[2412] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2413] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[2414] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[2415] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[2416] Train Loss:    tensor(0.7050, device='cuda:0', grad_fn=<NllLossBackward>)
[2417] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[2418] Train Loss:    tensor(0.6739, device='cuda:0', grad_fn=<NllLossBackward>)
[2419] Train Loss:    tensor(0.6706, device='cuda:0', grad_fn=<NllLossBackward>)
[2420] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[2421] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2422] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[2423] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[2424] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[2425] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[2426] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[2427] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[2428] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[2429] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[2430] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[2431] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2432] Train Loss:    tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[2433] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[2434] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[2435] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[2436] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[2437] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[2438] Train Loss:    tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[2439] Train Loss:    tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[2440] Train Loss:    tensor(0.6742, device='cuda:0', grad_fn=<NllLossBackward>)
[2441] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[2442] Train Loss:    tensor(0.7060, device='cuda:0', grad_fn=<NllLossBackward>)
[2443] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[2444] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[2445] Train Loss:    tensor(0.7184, device='cuda:0', grad_fn=<NllLossBackward>)
[2446] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2447] Train Loss:    tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[2448] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[2449] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[2450] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[2451] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[2452] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[2453] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[2454] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[2455] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2456] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[2457] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[2458] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[2459] Train Loss:    tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[2460] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[2461] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[2462] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[2463] Train Loss:    tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[2464] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[2465] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[2466] Train Loss:    tensor(0.7413, device='cuda:0', grad_fn=<NllLossBackward>)
[2467] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[2468] Train Loss:    tensor(0.6356, device='cuda:0', grad_fn=<NllLossBackward>)
[2469] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[2470] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[2471] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[2472] Train Loss:    tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[2473] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[2474] Train Loss:    tensor(0.7015, device='cuda:0', grad_fn=<NllLossBackward>)
[2475] Train Loss:    tensor(0.6409, device='cuda:0', grad_fn=<NllLossBackward>)
[2476] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2477] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[2478] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[2479] Train Loss:    tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[2480] Train Loss:    tensor(0.6608, device='cuda:0', grad_fn=<NllLossBackward>)
[2481] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[2482] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[2483] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[2484] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[2485] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[2486] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[2487] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[2488] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[2489] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[2490] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[2491] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[2492] Train Loss:    tensor(0.7352, device='cuda:0', grad_fn=<NllLossBackward>)
[2493] Train Loss:    tensor(0.6617, device='cuda:0', grad_fn=<NllLossBackward>)
[2494] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2495] Train Loss:    tensor(0.6597, device='cuda:0', grad_fn=<NllLossBackward>)
[2496] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[2497] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[2498] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[2499] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[2500] Train Loss:    tensor(0.6599, device='cuda:0', grad_fn=<NllLossBackward>)
[2501] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[2502] Train Loss:    tensor(0.6650, device='cuda:0', grad_fn=<NllLossBackward>)
[2503] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[2504] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[2505] Train Loss:    tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[2506] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[2507] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[2508] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[2509] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[2510] Train Loss:    tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[2511] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[2512] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[2513] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[2514] Train Loss:    tensor(0.6501, device='cuda:0', grad_fn=<NllLossBackward>)
[2515] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[2516] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[2517] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[2518] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[2519] Train Loss:    tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[2520] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[2521] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[2522] Train Loss:    tensor(0.7130, device='cuda:0', grad_fn=<NllLossBackward>)
[2523] Train Loss:    tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
[2524] Train Loss:    tensor(0.6787, device='cuda:0', grad_fn=<NllLossBackward>)
[2525] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[2526] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[2527] Train Loss:    tensor(0.6392, device='cuda:0', grad_fn=<NllLossBackward>)
[2528] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[2529] Train Loss:    tensor(0.7148, device='cuda:0', grad_fn=<NllLossBackward>)
[2530] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[2531] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[2532] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[2533] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[2534] Train Loss:    tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[2535] Train Loss:    tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[2536] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[2537] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[2538] Train Loss:    tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[2539] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[2540] Train Loss:    tensor(0.6536, device='cuda:0', grad_fn=<NllLossBackward>)
[2541] Train Loss:    tensor(0.6998, device='cuda:0', grad_fn=<NllLossBackward>)
[2542] Train Loss:    tensor(0.6618, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2543] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[2544] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[2545] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[2546] Train Loss:    tensor(0.7124, device='cuda:0', grad_fn=<NllLossBackward>)
[2547] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[2548] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[2549] Train Loss:    tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[2550] Train Loss:    tensor(0.7100, device='cuda:0', grad_fn=<NllLossBackward>)
[2551] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[2552] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[2553] Train Loss:    tensor(0.6674, device='cuda:0', grad_fn=<NllLossBackward>)
[2554] Train Loss:    tensor(0.7401, device='cuda:0', grad_fn=<NllLossBackward>)
[2555] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[2556] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[2557] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[2558] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[2559] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[2560] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[2561] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[2562] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[2563] Train Loss:    tensor(0.6598, device='cuda:0', grad_fn=<NllLossBackward>)
[2564] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[2565] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[2566] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[2567] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[2568] Train Loss:    tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[2569] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[2570] Train Loss:    tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[2571] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[2572] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[2573] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[2574] Train Loss:    tensor(0.7412, device='cuda:0', grad_fn=<NllLossBackward>)
[2575] Train Loss:    tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[2576] Train Loss:    tensor(0.7160, device='cuda:0', grad_fn=<NllLossBackward>)
[2577] Train Loss:    tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[2578] Train Loss:    tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[2579] Train Loss:    tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[2580] Train Loss:    tensor(0.6649, device='cuda:0', grad_fn=<NllLossBackward>)
[2581] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[2582] Train Loss:    tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[2583] Train Loss:    tensor(0.7409, device='cuda:0', grad_fn=<NllLossBackward>)
[2584] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[2585] Train Loss:    tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[2586] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[2587] Train Loss:    tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[2588] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2589] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[2590] Train Loss:    tensor(0.6588, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2591] Train Loss:    tensor(0.6694, device='cuda:0', grad_fn=<NllLossBackward>)
[2592] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[2593] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2594] Train Loss:    tensor(0.7171, device='cuda:0', grad_fn=<NllLossBackward>)
[2595] Train Loss:    tensor(0.7272, device='cuda:0', grad_fn=<NllLossBackward>)
[2596] Train Loss:    tensor(0.6813, device='cuda:0', grad_fn=<NllLossBackward>)
[2597] Train Loss:    tensor(0.6722, device='cuda:0', grad_fn=<NllLossBackward>)
[2598] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[2599] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[2600] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[2601] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[2602] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[2603] Train Loss:    tensor(0.6669, device='cuda:0', grad_fn=<NllLossBackward>)
[2604] Train Loss:    tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[2605] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[2606] Train Loss:    tensor(0.6631, device='cuda:0', grad_fn=<NllLossBackward>)
[2607] Train Loss:    tensor(0.6660, device='cuda:0', grad_fn=<NllLossBackward>)
[2608] Train Loss:    tensor(0.6763, device='cuda:0', grad_fn=<NllLossBackward>)
[2609] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[2610] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[2611] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[2612] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[2613] Train Loss:    tensor(0.6661, device='cuda:0', grad_fn=<NllLossBackward>)
[2614] Train Loss:    tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[2615] Train Loss:    tensor(0.6512, device='cuda:0', grad_fn=<NllLossBackward>)
[2616] Train Loss:    tensor(0.6616, device='cuda:0', grad_fn=<NllLossBackward>)
[2617] Train Loss:    tensor(0.6528, device='cuda:0', grad_fn=<NllLossBackward>)
[2618] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2619] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[2620] Train Loss:    tensor(0.6602, device='cuda:0', grad_fn=<NllLossBackward>)
[2621] Train Loss:    tensor(0.6691, device='cuda:0', grad_fn=<NllLossBackward>)
[2622] Train Loss:    tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[2623] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[2624] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[2625] Train Loss:    tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[2626] Train Loss:    tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[2627] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[2628] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[2629] Train Loss:    tensor(0.6685, device='cuda:0', grad_fn=<NllLossBackward>)
[2630] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[2631] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[2632] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[2633] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[2634] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[2635] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[2636] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[2637] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[2638] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2639] Train Loss:    tensor(0.7346, device='cuda:0', grad_fn=<NllLossBackward>)
[2640] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2641] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[2642] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[2643] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[2644] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[2645] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2646] Train Loss:    tensor(0.7249, device='cuda:0', grad_fn=<NllLossBackward>)
[2647] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[2648] Train Loss:    tensor(0.6714, device='cuda:0', grad_fn=<NllLossBackward>)
[2649] Train Loss:    tensor(0.6647, device='cuda:0', grad_fn=<NllLossBackward>)
[2650] Train Loss:    tensor(0.6712, device='cuda:0', grad_fn=<NllLossBackward>)
[2651] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[2652] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[2653] Train Loss:    tensor(0.6650, device='cuda:0', grad_fn=<NllLossBackward>)
[2654] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2655] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[2656] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[2657] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[2658] Train Loss:    tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[2659] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[2660] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[2661] Train Loss:    tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[2662] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[2663] Train Loss:    tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[2664] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[2665] Train Loss:    tensor(0.6632, device='cuda:0', grad_fn=<NllLossBackward>)
[2666] Train Loss:    tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[2667] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[2668] Train Loss:    tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[2669] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[2670] Train Loss:    tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[2671] Train Loss:    tensor(0.6729, device='cuda:0', grad_fn=<NllLossBackward>)
[2672] Train Loss:    tensor(0.7175, device='cuda:0', grad_fn=<NllLossBackward>)
[2673] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[2674] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[2675] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[2676] Train Loss:    tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[2677] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[2678] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[2679] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[2680] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[2681] Train Loss:    tensor(0.7497, device='cuda:0', grad_fn=<NllLossBackward>)
[2682] Train Loss:    tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[2683] Train Loss:    tensor(0.6621, device='cuda:0', grad_fn=<NllLossBackward>)
[2684] Train Loss:    tensor(0.6612, device='cuda:0', grad_fn=<NllLossBackward>)
[2685] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[2686] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2687] Train Loss:    tensor(0.6545, device='cuda:0', grad_fn=<NllLossBackward>)
[2688] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[2689] Train Loss:    tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[2690] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[2691] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[2692] Train Loss:    tensor(0.6559, device='cuda:0', grad_fn=<NllLossBackward>)
[2693] Train Loss:    tensor(0.6622, device='cuda:0', grad_fn=<NllLossBackward>)
[2694] Train Loss:    tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[2695] Train Loss:    tensor(0.6545, device='cuda:0', grad_fn=<NllLossBackward>)
[2696] Train Loss:    tensor(0.7140, device='cuda:0', grad_fn=<NllLossBackward>)
[2697] Train Loss:    tensor(0.7277, device='cuda:0', grad_fn=<NllLossBackward>)
[2698] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[2699] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[2700] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2701] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2702] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[2703] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[2704] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[2705] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2706] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[2707] Train Loss:    tensor(0.7279, device='cuda:0', grad_fn=<NllLossBackward>)
[2708] Train Loss:    tensor(0.6715, device='cuda:0', grad_fn=<NllLossBackward>)
[2709] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[2710] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[2711] Train Loss:    tensor(0.6523, device='cuda:0', grad_fn=<NllLossBackward>)
[2712] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[2713] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[2714] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[2715] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[2716] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[2717] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[2718] Train Loss:    tensor(0.6616, device='cuda:0', grad_fn=<NllLossBackward>)
[2719] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[2720] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[2721] Train Loss:    tensor(0.6609, device='cuda:0', grad_fn=<NllLossBackward>)
[2722] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[2723] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[2724] Train Loss:    tensor(0.6629, device='cuda:0', grad_fn=<NllLossBackward>)
[2725] Train Loss:    tensor(0.6577, device='cuda:0', grad_fn=<NllLossBackward>)
[2726] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[2727] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[2728] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[2729] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[2730] Train Loss:    tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[2731] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[2732] Train Loss:    tensor(0.7108, device='cuda:0', grad_fn=<NllLossBackward>)
[2733] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[2734] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2735] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[2736] Train Loss:    tensor(0.7078, device='cuda:0', grad_fn=<NllLossBackward>)
[2737] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[2738] Train Loss:    tensor(0.6671, device='cuda:0', grad_fn=<NllLossBackward>)
[2739] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[2740] Train Loss:    tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[2741] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[2742] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[2743] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[2744] Train Loss:    tensor(0.7368, device='cuda:0', grad_fn=<NllLossBackward>)
[2745] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[2746] Train Loss:    tensor(0.6528, device='cuda:0', grad_fn=<NllLossBackward>)
[2747] Train Loss:    tensor(0.6578, device='cuda:0', grad_fn=<NllLossBackward>)
[2748] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[2749] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[2750] Train Loss:    tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[2751] Train Loss:    tensor(0.6719, device='cuda:0', grad_fn=<NllLossBackward>)
[2752] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[2753] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[2754] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[2755] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[2756] Train Loss:    tensor(0.6540, device='cuda:0', grad_fn=<NllLossBackward>)
[2757] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[2758] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[2759] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[2760] Train Loss:    tensor(0.6681, device='cuda:0', grad_fn=<NllLossBackward>)
[2761] Train Loss:    tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[2762] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[2763] Train Loss:    tensor(0.6714, device='cuda:0', grad_fn=<NllLossBackward>)
[2764] Train Loss:    tensor(0.6601, device='cuda:0', grad_fn=<NllLossBackward>)
[2765] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[2766] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[2767] Train Loss:    tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[2768] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[2769] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[2770] Train Loss:    tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[2771] Train Loss:    tensor(0.6574, device='cuda:0', grad_fn=<NllLossBackward>)
[2772] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[2773] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[2774] Train Loss:    tensor(0.7073, device='cuda:0', grad_fn=<NllLossBackward>)
[2775] Train Loss:    tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[2776] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[2777] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[2778] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[2779] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[2780] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[2781] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[2782] Train Loss:    tensor(0.6724, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2783] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[2784] Train Loss:    tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[2785] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[2786] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[2787] Train Loss:    tensor(0.7243, device='cuda:0', grad_fn=<NllLossBackward>)
[2788] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[2789] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[2790] Train Loss:    tensor(0.6831, device='cuda:0', grad_fn=<NllLossBackward>)
[2791] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[2792] Train Loss:    tensor(0.6462, device='cuda:0', grad_fn=<NllLossBackward>)
[2793] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[2794] Train Loss:    tensor(0.7201, device='cuda:0', grad_fn=<NllLossBackward>)
[2795] Train Loss:    tensor(0.6616, device='cuda:0', grad_fn=<NllLossBackward>)
[2796] Train Loss:    tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[2797] Train Loss:    tensor(0.7297, device='cuda:0', grad_fn=<NllLossBackward>)
[2798] Train Loss:    tensor(0.6667, device='cuda:0', grad_fn=<NllLossBackward>)
[2799] Train Loss:    tensor(0.7419, device='cuda:0', grad_fn=<NllLossBackward>)
[2800] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[2801] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[2802] Train Loss:    tensor(0.6690, device='cuda:0', grad_fn=<NllLossBackward>)
[2803] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[2804] Train Loss:    tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[2805] Train Loss:    tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[2806] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[2807] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[2808] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
[2809] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[2810] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[2811] Train Loss:    tensor(0.6489, device='cuda:0', grad_fn=<NllLossBackward>)
[2812] Train Loss:    tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[2813] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[2814] Train Loss:    tensor(0.6529, device='cuda:0', grad_fn=<NllLossBackward>)
[2815] Train Loss:    tensor(0.6676, device='cuda:0', grad_fn=<NllLossBackward>)
[2816] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[2817] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[2818] Train Loss:    tensor(0.6642, device='cuda:0', grad_fn=<NllLossBackward>)
[2819] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[2820] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[2821] Train Loss:    tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[2822] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[2823] Train Loss:    tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[2824] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[2825] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[2826] Train Loss:    tensor(0.6631, device='cuda:0', grad_fn=<NllLossBackward>)
[2827] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[2828] Train Loss:    tensor(0.6722, device='cuda:0', grad_fn=<NllLossBackward>)
[2829] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[2830] Train Loss:    tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2831] Train Loss:   tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[2832] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[2833] Train Loss:   tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[2834] Train Loss:   tensor(0.6706, device='cuda:0', grad_fn=<NllLossBackward>)
[2835] Train Loss:   tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[2836] Train Loss:   tensor(0.6381, device='cuda:0', grad_fn=<NllLossBackward>)
[2837] Train Loss:   tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[2838] Train Loss:   tensor(0.6601, device='cuda:0', grad_fn=<NllLossBackward>)
[2839] Train Loss:   tensor(0.6571, device='cuda:0', grad_fn=<NllLossBackward>)
[2840] Train Loss:   tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[2841] Train Loss:   tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[2842] Train Loss:   tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[2843] Train Loss:   tensor(0.6628, device='cuda:0', grad_fn=<NllLossBackward>)
[2844] Train Loss:   tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[2845] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[2846] Train Loss:   tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2847] Train Loss:   tensor(0.6579, device='cuda:0', grad_fn=<NllLossBackward>)
[2848] Train Loss:   tensor(0.7288, device='cuda:0', grad_fn=<NllLossBackward>)
[2849] Train Loss:   tensor(0.6483, device='cuda:0', grad_fn=<NllLossBackward>)
[2850] Train Loss:   tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[2851] Train Loss:   tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[2852] Train Loss:   tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[2853] Train Loss:   tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[2854] Train Loss:   tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[2855] Train Loss:   tensor(0.7238, device='cuda:0', grad_fn=<NllLossBackward>)
[2856] Train Loss:   tensor(0.6608, device='cuda:0', grad_fn=<NllLossBackward>)
[2857] Train Loss:   tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[2858] Train Loss:   tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[2859] Train Loss:   tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[2860] Train Loss:   tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[2861] Train Loss:   tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[2862] Train Loss:   tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[2863] Train Loss:   tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[2864] Train Loss:   tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[2865] Train Loss:   tensor(0.7374, device='cuda:0', grad_fn=<NllLossBackward>)
[2866] Train Loss:   tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[2867] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2868] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[2869] Train Loss:   tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[2870] Train Loss:   tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[2871] Train Loss:   tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[2872] Train Loss:   tensor(0.6598, device='cuda:0', grad_fn=<NllLossBackward>)
[2873] Train Loss:   tensor(0.6729, device='cuda:0', grad_fn=<NllLossBackward>)
[2874] Train Loss:   tensor(0.6650, device='cuda:0', grad_fn=<NllLossBackward>)
[2875] Train Loss:   tensor(0.6543, device='cuda:0', grad_fn=<NllLossBackward>)
[2876] Train Loss:   tensor(0.6752, device='cuda:0', grad_fn=<NllLossBackward>)
[2877] Train Loss:   tensor(0.7265, device='cuda:0', grad_fn=<NllLossBackward>)
[2878] Train Loss:   tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2879] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[2880] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[2881] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[2882] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[2883] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[2884] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[2885] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[2886] Train Loss:    tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[2887] Train Loss:    tensor(0.6584, device='cuda:0', grad_fn=<NllLossBackward>)
[2888] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[2889] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[2890] Train Loss:    tensor(0.6761, device='cuda:0', grad_fn=<NllLossBackward>)
[2891] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[2892] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[2893] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[2894] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[2895] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[2896] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[2897] Train Loss:    tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[2898] Train Loss:    tensor(0.6603, device='cuda:0', grad_fn=<NllLossBackward>)
[2899] Train Loss:    tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[2900] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[2901] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[2902] Train Loss:    tensor(0.6995, device='cuda:0', grad_fn=<NllLossBackward>)
[2903] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[2904] Train Loss:    tensor(0.6755, device='cuda:0', grad_fn=<NllLossBackward>)
[2905] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[2906] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[2907] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[2908] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[2909] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[2910] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[2911] Train Loss:    tensor(0.6701, device='cuda:0', grad_fn=<NllLossBackward>)
[2912] Train Loss:    tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[2913] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[2914] Train Loss:    tensor(0.6494, device='cuda:0', grad_fn=<NllLossBackward>)
[2915] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[2916] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[2917] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[2918] Train Loss:    tensor(0.6636, device='cuda:0', grad_fn=<NllLossBackward>)
[2919] Train Loss:    tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[2920] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[2921] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[2922] Train Loss:    tensor(0.6272, device='cuda:0', grad_fn=<NllLossBackward>)
[2923] Train Loss:    tensor(0.6501, device='cuda:0', grad_fn=<NllLossBackward>)
[2924] Train Loss:    tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[2925] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[2926] Train Loss:    tensor(0.7142, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2927] Train Loss:   tensor(0.7123, device='cuda:0', grad_fn=<NllLossBackward>)
[2928] Train Loss:   tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[2929] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[2930] Train Loss:   tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[2931] Train Loss:   tensor(0.7260, device='cuda:0', grad_fn=<NllLossBackward>)
[2932] Train Loss:   tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[2933] Train Loss:   tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[2934] Train Loss:   tensor(0.6566, device='cuda:0', grad_fn=<NllLossBackward>)
[2935] Train Loss:   tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[2936] Train Loss:   tensor(0.6628, device='cuda:0', grad_fn=<NllLossBackward>)
[2937] Train Loss:   tensor(0.6564, device='cuda:0', grad_fn=<NllLossBackward>)
[2938] Train Loss:   tensor(0.6506, device='cuda:0', grad_fn=<NllLossBackward>)
[2939] Train Loss:   tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[2940] Train Loss:   tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[2941] Train Loss:   tensor(0.6656, device='cuda:0', grad_fn=<NllLossBackward>)
[2942] Train Loss:   tensor(0.6582, device='cuda:0', grad_fn=<NllLossBackward>)
[2943] Train Loss:   tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[2944] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[2945] Train Loss:   tensor(0.7068, device='cuda:0', grad_fn=<NllLossBackward>)
[2946] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[2947] Train Loss:   tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[2948] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[2949] Train Loss:   tensor(0.7090, device='cuda:0', grad_fn=<NllLossBackward>)
[2950] Train Loss:   tensor(0.6615, device='cuda:0', grad_fn=<NllLossBackward>)
[2951] Train Loss:   tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[2952] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[2953] Train Loss:   tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[2954] Train Loss:   tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[2955] Train Loss:   tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[2956] Train Loss:   tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[2957] Train Loss:   tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[2958] Train Loss:   tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[2959] Train Loss:   tensor(0.7595, device='cuda:0', grad_fn=<NllLossBackward>)
[2960] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[2961] Train Loss:   tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[2962] Train Loss:   tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[2963] Train Loss:   tensor(0.6599, device='cuda:0', grad_fn=<NllLossBackward>)
[2964] Train Loss:   tensor(0.6512, device='cuda:0', grad_fn=<NllLossBackward>)
[2965] Train Loss:   tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[2966] Train Loss:   tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[2967] Train Loss:   tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[2968] Train Loss:   tensor(0.7275, device='cuda:0', grad_fn=<NllLossBackward>)
[2969] Train Loss:   tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[2970] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[2971] Train Loss:   tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[2972] Train Loss:   tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[2973] Train Loss:   tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[2974] Train Loss:   tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[2975] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[2976] Train Loss:    tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[2977] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[2978] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[2979] Train Loss:    tensor(0.6566, device='cuda:0', grad_fn=<NllLossBackward>)
[2980] Train Loss:    tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[2981] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[2982] Train Loss:    tensor(0.7119, device='cuda:0', grad_fn=<NllLossBackward>)
[2983] Train Loss:    tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[2984] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[2985] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[2986] Train Loss:    tensor(0.7244, device='cuda:0', grad_fn=<NllLossBackward>)
[2987] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[2988] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[2989] Train Loss:    tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[2990] Train Loss:    tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[2991] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[2992] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[2993] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[2994] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[2995] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[2996] Train Loss:    tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
[2997] Train Loss:    tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[2998] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[2999] Train Loss:    tensor(0.6614, device='cuda:0', grad_fn=<NllLossBackward>)
[3000] Train Loss:    tensor(0.6631, device='cuda:0', grad_fn=<NllLossBackward>)
[3001] Train Loss:    tensor(0.6739, device='cuda:0', grad_fn=<NllLossBackward>)
[3002] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[3003] Train Loss:    tensor(0.6692, device='cuda:0', grad_fn=<NllLossBackward>)
[3004] Train Loss:    tensor(0.6571, device='cuda:0', grad_fn=<NllLossBackward>)
[3005] Train Loss:    tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3006] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[3007] Train Loss:    tensor(0.6452, device='cuda:0', grad_fn=<NllLossBackward>)
[3008] Train Loss:    tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[3009] Train Loss:    tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[3010] Train Loss:    tensor(0.6501, device='cuda:0', grad_fn=<NllLossBackward>)
[3011] Train Loss:    tensor(0.6698, device='cuda:0', grad_fn=<NllLossBackward>)
[3012] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[3013] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[3014] Train Loss:    tensor(0.6520, device='cuda:0', grad_fn=<NllLossBackward>)
[3015] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[3016] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[3017] Train Loss:    tensor(0.6515, device='cuda:0', grad_fn=<NllLossBackward>)
[3018] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[3019] Train Loss:    tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[3020] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[3021] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[3022] Train Loss:    tensor(0.7194, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3023] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[3024] Train Loss:    tensor(0.6619, device='cuda:0', grad_fn=<NllLossBackward>)
[3025] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[3026] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[3027] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3028] Train Loss:    tensor(0.6430, device='cuda:0', grad_fn=<NllLossBackward>)
[3029] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3030] Train Loss:    tensor(0.7284, device='cuda:0', grad_fn=<NllLossBackward>)
[3031] Train Loss:    tensor(0.6628, device='cuda:0', grad_fn=<NllLossBackward>)
[3032] Train Loss:    tensor(0.6491, device='cuda:0', grad_fn=<NllLossBackward>)
[3033] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[3034] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3035] Train Loss:    tensor(0.6567, device='cuda:0', grad_fn=<NllLossBackward>)
[3036] Train Loss:    tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[3037] Train Loss:    tensor(0.6646, device='cuda:0', grad_fn=<NllLossBackward>)
[3038] Train Loss:    tensor(0.6962, device='cuda:0', grad_fn=<NllLossBackward>)
[3039] Train Loss:    tensor(0.7217, device='cuda:0', grad_fn=<NllLossBackward>)
[3040] Train Loss:    tensor(0.6595, device='cuda:0', grad_fn=<NllLossBackward>)
[3041] Train Loss:    tensor(0.6601, device='cuda:0', grad_fn=<NllLossBackward>)
[3042] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[3043] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[3044] Train Loss:    tensor(0.6583, device='cuda:0', grad_fn=<NllLossBackward>)
[3045] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[3046] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[3047] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3048] Train Loss:    tensor(0.7098, device='cuda:0', grad_fn=<NllLossBackward>)
[3049] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[3050] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3051] Train Loss:    tensor(0.7369, device='cuda:0', grad_fn=<NllLossBackward>)
[3052] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[3053] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[3054] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[3055] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[3056] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[3057] Train Loss:    tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[3058] Train Loss:    tensor(0.6816, device='cuda:0', grad_fn=<NllLossBackward>)
[3059] Train Loss:    tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[3060] Train Loss:    tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[3061] Train Loss:    tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[3062] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[3063] Train Loss:    tensor(0.7311, device='cuda:0', grad_fn=<NllLossBackward>)
[3064] Train Loss:    tensor(0.6553, device='cuda:0', grad_fn=<NllLossBackward>)
[3065] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[3066] Train Loss:    tensor(0.7113, device='cuda:0', grad_fn=<NllLossBackward>)
[3067] Train Loss:    tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[3068] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[3069] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[3070] Train Loss:    tensor(0.6588, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3071] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[3072] Train Loss:    tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[3073] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[3074] Train Loss:    tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[3075] Train Loss:    tensor(0.6695, device='cuda:0', grad_fn=<NllLossBackward>)
[3076] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[3077] Train Loss:    tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[3078] Train Loss:    tensor(0.7257, device='cuda:0', grad_fn=<NllLossBackward>)
[3079] Train Loss:    tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[3080] Train Loss:    tensor(0.6535, device='cuda:0', grad_fn=<NllLossBackward>)
[3081] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[3082] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[3083] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[3084] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[3085] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[3086] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[3087] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[3088] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[3089] Train Loss:    tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[3090] Train Loss:    tensor(0.6716, device='cuda:0', grad_fn=<NllLossBackward>)
[3091] Train Loss:    tensor(0.6498, device='cuda:0', grad_fn=<NllLossBackward>)
[3092] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[3093] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[3094] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[3095] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[3096] Train Loss:    tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[3097] Train Loss:    tensor(0.6679, device='cuda:0', grad_fn=<NllLossBackward>)
[3098] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[3099] Train Loss:    tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3100] Train Loss:    tensor(0.6649, device='cuda:0', grad_fn=<NllLossBackward>)
[3101] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[3102] Train Loss:    tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[3103] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[3104] Train Loss:    tensor(0.6740, device='cuda:0', grad_fn=<NllLossBackward>)
[3105] Train Loss:    tensor(0.6776, device='cuda:0', grad_fn=<NllLossBackward>)
[3106] Train Loss:    tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[3107] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[3108] Train Loss:    tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[3109] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3110] Train Loss:    tensor(0.6476, device='cuda:0', grad_fn=<NllLossBackward>)
[3111] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3112] Train Loss:    tensor(0.7163, device='cuda:0', grad_fn=<NllLossBackward>)
[3113] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[3114] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[3115] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[3116] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[3117] Train Loss:    tensor(0.6598, device='cuda:0', grad_fn=<NllLossBackward>)
[3118] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3119] Train Loss:   tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[3120] Train Loss:   tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[3121] Train Loss:   tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[3122] Train Loss:   tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3123] Train Loss:   tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[3124] Train Loss:   tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[3125] Train Loss:   tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[3126] Train Loss:   tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[3127] Train Loss:   tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[3128] Train Loss:   tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[3129] Train Loss:   tensor(0.6984, device='cuda:0', grad_fn=<NllLossBackward>)
[3130] Train Loss:   tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[3131] Train Loss:   tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[3132] Train Loss:   tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[3133] Train Loss:   tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[3134] Train Loss:   tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[3135] Train Loss:   tensor(0.6683, device='cuda:0', grad_fn=<NllLossBackward>)
[3136] Train Loss:   tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[3137] Train Loss:   tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[3138] Train Loss:   tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[3139] Train Loss:   tensor(0.6699, device='cuda:0', grad_fn=<NllLossBackward>)
[3140] Train Loss:   tensor(0.7012, device='cuda:0', grad_fn=<NllLossBackward>)
[3141] Train Loss:   tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[3142] Train Loss:   tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[3143] Train Loss:   tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[3144] Train Loss:   tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[3145] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3146] Train Loss:   tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[3147] Train Loss:   tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[3148] Train Loss:   tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[3149] Train Loss:   tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[3150] Train Loss:   tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[3151] Train Loss:   tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[3152] Train Loss:   tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3153] Train Loss:   tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[3154] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[3155] Train Loss:   tensor(0.6567, device='cuda:0', grad_fn=<NllLossBackward>)
[3156] Train Loss:   tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[3157] Train Loss:   tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[3158] Train Loss:   tensor(0.7135, device='cuda:0', grad_fn=<NllLossBackward>)
[3159] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[3160] Train Loss:   tensor(0.6471, device='cuda:0', grad_fn=<NllLossBackward>)
[3161] Train Loss:   tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[3162] Train Loss:   tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
[3163] Train Loss:   tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[3164] Train Loss:   tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[3165] Train Loss:   tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[3166] Train Loss:   tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3167] Train Loss:   tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[3168] Train Loss:   tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[3169] Train Loss:   tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[3170] Train Loss:   tensor(0.7179, device='cuda:0', grad_fn=<NllLossBackward>)
[3171] Train Loss:   tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[3172] Train Loss:   tensor(0.7441, device='cuda:0', grad_fn=<NllLossBackward>)
[3173] Train Loss:   tensor(0.7270, device='cuda:0', grad_fn=<NllLossBackward>)
[3174] Train Loss:   tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[3175] Train Loss:   tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[3176] Train Loss:   tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[3177] Train Loss:   tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[3178] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[3179] Train Loss:   tensor(0.6581, device='cuda:0', grad_fn=<NllLossBackward>)
[3180] Train Loss:   tensor(0.6553, device='cuda:0', grad_fn=<NllLossBackward>)
[3181] Train Loss:   tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[3182] Train Loss:   tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[3183] Train Loss:   tensor(0.7391, device='cuda:0', grad_fn=<NllLossBackward>)
[3184] Train Loss:   tensor(0.7611, device='cuda:0', grad_fn=<NllLossBackward>)
[3185] Train Loss:   tensor(0.6881, device='cuda:0', grad_fn=<NllLossBackward>)
[3186] Train Loss:   tensor(0.7153, device='cuda:0', grad_fn=<NllLossBackward>)
[3187] Train Loss:   tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[3188] Train Loss:   tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[3189] Train Loss:   tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[3190] Train Loss:   tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[3191] Train Loss:   tensor(0.7156, device='cuda:0', grad_fn=<NllLossBackward>)
[3192] Train Loss:   tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3193] Train Loss:   tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[3194] Train Loss:   tensor(0.6722, device='cuda:0', grad_fn=<NllLossBackward>)
[3195] Train Loss:   tensor(0.6747, device='cuda:0', grad_fn=<NllLossBackward>)
[3196] Train Loss:   tensor(0.7161, device='cuda:0', grad_fn=<NllLossBackward>)
[3197] Train Loss:   tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[3198] Train Loss:   tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[3199] Train Loss:   tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[3200] Train Loss:   tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[3201] Train Loss:   tensor(0.6653, device='cuda:0', grad_fn=<NllLossBackward>)
[3202] Train Loss:   tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[3203] Train Loss:   tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[3204] Train Loss:   tensor(0.6656, device='cuda:0', grad_fn=<NllLossBackward>)
[3205] Train Loss:   tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[3206] Train Loss:   tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3207] Train Loss:   tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3208] Train Loss:   tensor(0.7274, device='cuda:0', grad_fn=<NllLossBackward>)
[3209] Train Loss:   tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[3210] Train Loss:   tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[3211] Train Loss:   tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[3212] Train Loss:   tensor(0.6475, device='cuda:0', grad_fn=<NllLossBackward>)
[3213] Train Loss:   tensor(0.6671, device='cuda:0', grad_fn=<NllLossBackward>)
[3214] Train Loss:   tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3215] Train Loss:    tensor(0.6378, device='cuda:0', grad_fn=<NllLossBackward>)
[3216] Train Loss:    tensor(0.6620, device='cuda:0', grad_fn=<NllLossBackward>)
[3217] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[3218] Train Loss:    tensor(0.6434, device='cuda:0', grad_fn=<NllLossBackward>)
[3219] Train Loss:    tensor(0.7069, device='cuda:0', grad_fn=<NllLossBackward>)
[3220] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[3221] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[3222] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3223] Train Loss:    tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[3224] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[3225] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
[3226] Train Loss:    tensor(0.6662, device='cuda:0', grad_fn=<NllLossBackward>)
[3227] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3228] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[3229] Train Loss:    tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[3230] Train Loss:    tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[3231] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[3232] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[3233] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[3234] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[3235] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[3236] Train Loss:    tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[3237] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[3238] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3239] Train Loss:    tensor(0.6688, device='cuda:0', grad_fn=<NllLossBackward>)
[3240] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[3241] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3242] Train Loss:    tensor(0.6545, device='cuda:0', grad_fn=<NllLossBackward>)
[3243] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[3244] Train Loss:    tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[3245] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[3246] Train Loss:    tensor(0.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[3247] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3248] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[3249] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3250] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[3251] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[3252] Train Loss:    tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[3253] Train Loss:    tensor(0.6572, device='cuda:0', grad_fn=<NllLossBackward>)
[3254] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[3255] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[3256] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[3257] Train Loss:    tensor(0.6701, device='cuda:0', grad_fn=<NllLossBackward>)
[3258] Train Loss:    tensor(0.6613, device='cuda:0', grad_fn=<NllLossBackward>)
[3259] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3260] Train Loss:    tensor(0.7020, device='cuda:0', grad_fn=<NllLossBackward>)
[3261] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[3262] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3263] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[3264] Train Loss:    tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[3265] Train Loss:    tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[3266] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[3267] Train Loss:    tensor(0.6606, device='cuda:0', grad_fn=<NllLossBackward>)
[3268] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[3269] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3270] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[3271] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[3272] Train Loss:    tensor(0.6573, device='cuda:0', grad_fn=<NllLossBackward>)
[3273] Train Loss:    tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[3274] Train Loss:    tensor(0.6665, device='cuda:0', grad_fn=<NllLossBackward>)
[3275] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3276] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[3277] Train Loss:    tensor(0.7235, device='cuda:0', grad_fn=<NllLossBackward>)
[3278] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[3279] Train Loss:    tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[3280] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[3281] Train Loss:    tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[3282] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[3283] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[3284] Train Loss:    tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[3285] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[3286] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[3287] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[3288] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[3289] Train Loss:    tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[3290] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[3291] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[3292] Train Loss:    tensor(0.6638, device='cuda:0', grad_fn=<NllLossBackward>)
[3293] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[3294] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3295] Train Loss:    tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[3296] Train Loss:    tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[3297] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[3298] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[3299] Train Loss:    tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[3300] Train Loss:    tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[3301] Train Loss:    tensor(0.6609, device='cuda:0', grad_fn=<NllLossBackward>)
[3302] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3303] Train Loss:    tensor(0.6531, device='cuda:0', grad_fn=<NllLossBackward>)
[3304] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[3305] Train Loss:    tensor(0.7035, device='cuda:0', grad_fn=<NllLossBackward>)
[3306] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[3307] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[3308] Train Loss:    tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[3309] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[3310] Train Loss:    tensor(0.6515, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3311] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[3312] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[3313] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[3314] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3315] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[3316] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[3317] Train Loss:    tensor(0.6694, device='cuda:0', grad_fn=<NllLossBackward>)
[3318] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[3319] Train Loss:    tensor(0.7356, device='cuda:0', grad_fn=<NllLossBackward>)
[3320] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[3321] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[3322] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[3323] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[3324] Train Loss:    tensor(0.6597, device='cuda:0', grad_fn=<NllLossBackward>)
[3325] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3326] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[3327] Train Loss:    tensor(0.6621, device='cuda:0', grad_fn=<NllLossBackward>)
[3328] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[3329] Train Loss:    tensor(0.6544, device='cuda:0', grad_fn=<NllLossBackward>)
[3330] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[3331] Train Loss:    tensor(0.6516, device='cuda:0', grad_fn=<NllLossBackward>)
[3332] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[3333] Train Loss:    tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[3334] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[3335] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3336] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[3337] Train Loss:    tensor(0.7029, device='cuda:0', grad_fn=<NllLossBackward>)
[3338] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3339] Train Loss:    tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[3340] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[3341] Train Loss:    tensor(0.6503, device='cuda:0', grad_fn=<NllLossBackward>)
[3342] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[3343] Train Loss:    tensor(0.7121, device='cuda:0', grad_fn=<NllLossBackward>)
[3344] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[3345] Train Loss:    tensor(0.6682, device='cuda:0', grad_fn=<NllLossBackward>)
[3346] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[3347] Train Loss:    tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[3348] Train Loss:    tensor(0.6452, device='cuda:0', grad_fn=<NllLossBackward>)
[3349] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[3350] Train Loss:    tensor(0.6457, device='cuda:0', grad_fn=<NllLossBackward>)
[3351] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[3352] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[3353] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[3354] Train Loss:    tensor(0.6501, device='cuda:0', grad_fn=<NllLossBackward>)
[3355] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[3356] Train Loss:    tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[3357] Train Loss:    tensor(0.7104, device='cuda:0', grad_fn=<NllLossBackward>)
[3358] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3359] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[3360] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[3361] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[3362] Train Loss:    tensor(0.6842, device='cuda:0', grad_fn=<NllLossBackward>)
[3363] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[3364] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[3365] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3366] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[3367] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[3368] Train Loss:    tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[3369] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[3370] Train Loss:    tensor(0.6670, device='cuda:0', grad_fn=<NllLossBackward>)
[3371] Train Loss:    tensor(0.7004, device='cuda:0', grad_fn=<NllLossBackward>)
[3372] Train Loss:    tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[3373] Train Loss:    tensor(0.6798, device='cuda:0', grad_fn=<NllLossBackward>)
[3374] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[3375] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3376] Train Loss:    tensor(0.7017, device='cuda:0', grad_fn=<NllLossBackward>)
[3377] Train Loss:    tensor(0.6640, device='cuda:0', grad_fn=<NllLossBackward>)
[3378] Train Loss:    tensor(0.6548, device='cuda:0', grad_fn=<NllLossBackward>)
[3379] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[3380] Train Loss:    tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[3381] Train Loss:    tensor(0.6827, device='cuda:0', grad_fn=<NllLossBackward>)
[3382] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[3383] Train Loss:    tensor(0.6902, device='cuda:0', grad_fn=<NllLossBackward>)
[3384] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[3385] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3386] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[3387] Train Loss:    tensor(0.7052, device='cuda:0', grad_fn=<NllLossBackward>)
[3388] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[3389] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[3390] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[3391] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[3392] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[3393] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[3394] Train Loss:    tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
[3395] Train Loss:    tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[3396] Train Loss:    tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[3397] Train Loss:    tensor(0.6644, device='cuda:0', grad_fn=<NllLossBackward>)
[3398] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[3399] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[3400] Train Loss:    tensor(0.6525, device='cuda:0', grad_fn=<NllLossBackward>)
[3401] Train Loss:    tensor(0.6690, device='cuda:0', grad_fn=<NllLossBackward>)
[3402] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[3403] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[3404] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[3405] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[3406] Train Loss:    tensor(0.6562, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3407] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[3408] Train Loss:    tensor(0.6654, device='cuda:0', grad_fn=<NllLossBackward>)
[3409] Train Loss:    tensor(0.6660, device='cuda:0', grad_fn=<NllLossBackward>)
[3410] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[3411] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3412] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3413] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[3414] Train Loss:    tensor(0.6992, device='cuda:0', grad_fn=<NllLossBackward>)
[3415] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3416] Train Loss:    tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[3417] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3418] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[3419] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3420] Train Loss:    tensor(0.6577, device='cuda:0', grad_fn=<NllLossBackward>)
[3421] Train Loss:    tensor(0.6748, device='cuda:0', grad_fn=<NllLossBackward>)
[3422] Train Loss:    tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[3423] Train Loss:    tensor(0.6516, device='cuda:0', grad_fn=<NllLossBackward>)
[3424] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[3425] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[3426] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[3427] Train Loss:    tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
[3428] Train Loss:    tensor(0.6638, device='cuda:0', grad_fn=<NllLossBackward>)
[3429] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[3430] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[3431] Train Loss:    tensor(0.6954, device='cuda:0', grad_fn=<NllLossBackward>)
[3432] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[3433] Train Loss:    tensor(0.6539, device='cuda:0', grad_fn=<NllLossBackward>)
[3434] Train Loss:    tensor(0.7129, device='cuda:0', grad_fn=<NllLossBackward>)
[3435] Train Loss:    tensor(0.6669, device='cuda:0', grad_fn=<NllLossBackward>)
[3436] Train Loss:    tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[3437] Train Loss:    tensor(0.7172, device='cuda:0', grad_fn=<NllLossBackward>)
[3438] Train Loss:    tensor(0.6647, device='cuda:0', grad_fn=<NllLossBackward>)
[3439] Train Loss:    tensor(0.7404, device='cuda:0', grad_fn=<NllLossBackward>)
[3440] Train Loss:    tensor(0.7114, device='cuda:0', grad_fn=<NllLossBackward>)
[3441] Train Loss:    tensor(0.6669, device='cuda:0', grad_fn=<NllLossBackward>)
[3442] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[3443] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[3444] Train Loss:    tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[3445] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[3446] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[3447] Train Loss:    tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[3448] Train Loss:    tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[3449] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[3450] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[3451] Train Loss:    tensor(0.6696, device='cuda:0', grad_fn=<NllLossBackward>)
[3452] Train Loss:    tensor(0.7261, device='cuda:0', grad_fn=<NllLossBackward>)
[3453] Train Loss:    tensor(0.7293, device='cuda:0', grad_fn=<NllLossBackward>)
[3454] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3455] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[3456] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[3457] Train Loss:    tensor(0.7327, device='cuda:0', grad_fn=<NllLossBackward>)
[3458] Train Loss:    tensor(0.7133, device='cuda:0', grad_fn=<NllLossBackward>)
[3459] Train Loss:    tensor(0.7209, device='cuda:0', grad_fn=<NllLossBackward>)
[3460] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[3461] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[3462] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[3463] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[3464] Train Loss:    tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[3465] Train Loss:    tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[3466] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[3467] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[3468] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[3469] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[3470] Train Loss:    tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[3471] Train Loss:    tensor(0.6653, device='cuda:0', grad_fn=<NllLossBackward>)
[3472] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[3473] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[3474] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[3475] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3476] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[3477] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[3478] Train Loss:    tensor(0.6727, device='cuda:0', grad_fn=<NllLossBackward>)
[3479] Train Loss:    tensor(0.6473, device='cuda:0', grad_fn=<NllLossBackward>)
[3480] Train Loss:    tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[3481] Train Loss:    tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[3482] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[3483] Train Loss:    tensor(0.7242, device='cuda:0', grad_fn=<NllLossBackward>)
[3484] Train Loss:    tensor(0.6682, device='cuda:0', grad_fn=<NllLossBackward>)
[3485] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[3486] Train Loss:    tensor(0.6694, device='cuda:0', grad_fn=<NllLossBackward>)
[3487] Train Loss:    tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[3488] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[3489] Train Loss:    tensor(0.7151, device='cuda:0', grad_fn=<NllLossBackward>)
[3490] Train Loss:    tensor(0.7075, device='cuda:0', grad_fn=<NllLossBackward>)
[3491] Train Loss:    tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[3492] Train Loss:    tensor(0.6584, device='cuda:0', grad_fn=<NllLossBackward>)
[3493] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[3494] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3495] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[3496] Train Loss:    tensor(0.6836, device='cuda:0', grad_fn=<NllLossBackward>)
[3497] Train Loss:    tensor(0.6672, device='cuda:0', grad_fn=<NllLossBackward>)
[3498] Train Loss:    tensor(0.6398, device='cuda:0', grad_fn=<NllLossBackward>)
[3499] Train Loss:    tensor(0.6618, device='cuda:0', grad_fn=<NllLossBackward>)
[3500] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[3501] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[3502] Train Loss:    tensor(0.7207, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3503] Train Loss:   tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[3504] Train Loss:   tensor(0.6683, device='cuda:0', grad_fn=<NllLossBackward>)
[3505] Train Loss:   tensor(0.6671, device='cuda:0', grad_fn=<NllLossBackward>)
[3506] Train Loss:   tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[3507] Train Loss:   tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3508] Train Loss:   tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[3509] Train Loss:   tensor(0.6498, device='cuda:0', grad_fn=<NllLossBackward>)
[3510] Train Loss:   tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[3511] Train Loss:   tensor(0.6562, device='cuda:0', grad_fn=<NllLossBackward>)
[3512] Train Loss:   tensor(0.7125, device='cuda:0', grad_fn=<NllLossBackward>)
[3513] Train Loss:   tensor(0.7095, device='cuda:0', grad_fn=<NllLossBackward>)
[3514] Train Loss:   tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[3515] Train Loss:   tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3516] Train Loss:   tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[3517] Train Loss:   tensor(0.7013, device='cuda:0', grad_fn=<NllLossBackward>)
[3518] Train Loss:   tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[3519] Train Loss:   tensor(0.7173, device='cuda:0', grad_fn=<NllLossBackward>)
[3520] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[3521] Train Loss:   tensor(0.6695, device='cuda:0', grad_fn=<NllLossBackward>)
[3522] Train Loss:   tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[3523] Train Loss:   tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[3524] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[3525] Train Loss:   tensor(0.6622, device='cuda:0', grad_fn=<NllLossBackward>)
[3526] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[3527] Train Loss:   tensor(0.6616, device='cuda:0', grad_fn=<NllLossBackward>)
[3528] Train Loss:   tensor(0.7146, device='cuda:0', grad_fn=<NllLossBackward>)
[3529] Train Loss:   tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3530] Train Loss:   tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[3531] Train Loss:   tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[3532] Train Loss:   tensor(0.6675, device='cuda:0', grad_fn=<NllLossBackward>)
[3533] Train Loss:   tensor(0.6510, device='cuda:0', grad_fn=<NllLossBackward>)
[3534] Train Loss:   tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[3535] Train Loss:   tensor(0.6649, device='cuda:0', grad_fn=<NllLossBackward>)
[3536] Train Loss:   tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[3537] Train Loss:   tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[3538] Train Loss:   tensor(0.6698, device='cuda:0', grad_fn=<NllLossBackward>)
[3539] Train Loss:   tensor(0.6764, device='cuda:0', grad_fn=<NllLossBackward>)
[3540] Train Loss:   tensor(0.6606, device='cuda:0', grad_fn=<NllLossBackward>)
[3541] Train Loss:   tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[3542] Train Loss:   tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[3543] Train Loss:   tensor(0.6697, device='cuda:0', grad_fn=<NllLossBackward>)
[3544] Train Loss:   tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[3545] Train Loss:   tensor(0.6605, device='cuda:0', grad_fn=<NllLossBackward>)
[3546] Train Loss:   tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[3547] Train Loss:   tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[3548] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[3549] Train Loss:   tensor(0.6497, device='cuda:0', grad_fn=<NllLossBackward>)
[3550] Train Loss:   tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3551] Train Loss:    tensor(0.6860, device='cuda:0', grad_fn=<NllLossBackward>)
[3552] Train Loss:    tensor(0.6933, device='cuda:0', grad_fn=<NllLossBackward>)
[3553] Train Loss:    tensor(0.7234, device='cuda:0', grad_fn=<NllLossBackward>)
[3554] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[3555] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3556] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[3557] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[3558] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3559] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[3560] Train Loss:    tensor(0.6408, device='cuda:0', grad_fn=<NllLossBackward>)
[3561] Train Loss:    tensor(0.6645, device='cuda:0', grad_fn=<NllLossBackward>)
[3562] Train Loss:    tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[3563] Train Loss:    tensor(0.6879, device='cuda:0', grad_fn=<NllLossBackward>)
[3564] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[3565] Train Loss:    tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[3566] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[3567] Train Loss:    tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[3568] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[3569] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3570] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[3571] Train Loss:    tensor(0.7508, device='cuda:0', grad_fn=<NllLossBackward>)
[3572] Train Loss:    tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[3573] Train Loss:    tensor(0.6945, device='cuda:0', grad_fn=<NllLossBackward>)
[3574] Train Loss:    tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[3575] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3576] Train Loss:    tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[3577] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[3578] Train Loss:    tensor(0.7099, device='cuda:0', grad_fn=<NllLossBackward>)
[3579] Train Loss:    tensor(0.7364, device='cuda:0', grad_fn=<NllLossBackward>)
[3580] Train Loss:    tensor(0.6707, device='cuda:0', grad_fn=<NllLossBackward>)
[3581] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[3582] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[3583] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[3584] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[3585] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[3586] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[3587] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[3588] Train Loss:    tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[3589] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[3590] Train Loss:    tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[3591] Train Loss:    tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[3592] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3593] Train Loss:    tensor(0.6611, device='cuda:0', grad_fn=<NllLossBackward>)
[3594] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[3595] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[3596] Train Loss:    tensor(0.6737, device='cuda:0', grad_fn=<NllLossBackward>)
[3597] Train Loss:    tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[3598] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3599] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[3600] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[3601] Train Loss:    tensor(0.6527, device='cuda:0', grad_fn=<NllLossBackward>)
[3602] Train Loss:    tensor(0.7216, device='cuda:0', grad_fn=<NllLossBackward>)
[3603] Train Loss:    tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[3604] Train Loss:    tensor(0.6514, device='cuda:0', grad_fn=<NllLossBackward>)
[3605] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[3606] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[3607] Train Loss:    tensor(0.7273, device='cuda:0', grad_fn=<NllLossBackward>)
[3608] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[3609] Train Loss:    tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[3610] Train Loss:    tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[3611] Train Loss:    tensor(0.6627, device='cuda:0', grad_fn=<NllLossBackward>)
[3612] Train Loss:    tensor(0.6519, device='cuda:0', grad_fn=<NllLossBackward>)
[3613] Train Loss:    tensor(0.6886, device='cuda:0', grad_fn=<NllLossBackward>)
[3614] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[3615] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[3616] Train Loss:    tensor(0.6943, device='cuda:0', grad_fn=<NllLossBackward>)
[3617] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[3618] Train Loss:    tensor(0.6734, device='cuda:0', grad_fn=<NllLossBackward>)
[3619] Train Loss:    tensor(0.6942, device='cuda:0', grad_fn=<NllLossBackward>)
[3620] Train Loss:    tensor(0.6679, device='cuda:0', grad_fn=<NllLossBackward>)
[3621] Train Loss:    tensor(0.7166, device='cuda:0', grad_fn=<NllLossBackward>)
[3622] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[3623] Train Loss:    tensor(0.6726, device='cuda:0', grad_fn=<NllLossBackward>)
[3624] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[3625] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[3626] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3627] Train Loss:    tensor(0.6802, device='cuda:0', grad_fn=<NllLossBackward>)
[3628] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[3629] Train Loss:    tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[3630] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[3631] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[3632] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[3633] Train Loss:    tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[3634] Train Loss:    tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[3635] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[3636] Train Loss:    tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[3637] Train Loss:    tensor(0.6671, device='cuda:0', grad_fn=<NllLossBackward>)
[3638] Train Loss:    tensor(0.6864, device='cuda:0', grad_fn=<NllLossBackward>)
[3639] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[3640] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[3641] Train Loss:    tensor(0.6654, device='cuda:0', grad_fn=<NllLossBackward>)
[3642] Train Loss:    tensor(0.6640, device='cuda:0', grad_fn=<NllLossBackward>)
[3643] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[3644] Train Loss:    tensor(0.6922, device='cuda:0', grad_fn=<NllLossBackward>)
[3645] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[3646] Train Loss:    tensor(0.6735, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3647] Train Loss:    tensor(0.6509, device='cuda:0', grad_fn=<NllLossBackward>)
[3648] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[3649] Train Loss:    tensor(0.6724, device='cuda:0', grad_fn=<NllLossBackward>)
[3650] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3651] Train Loss:    tensor(0.7343, device='cuda:0', grad_fn=<NllLossBackward>)
[3652] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[3653] Train Loss:    tensor(0.7018, device='cuda:0', grad_fn=<NllLossBackward>)
[3654] Train Loss:    tensor(0.7204, device='cuda:0', grad_fn=<NllLossBackward>)
[3655] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[3656] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[3657] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[3658] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[3659] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3660] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[3661] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[3662] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[3663] Train Loss:    tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[3664] Train Loss:    tensor(0.6629, device='cuda:0', grad_fn=<NllLossBackward>)
[3665] Train Loss:    tensor(0.6684, device='cuda:0', grad_fn=<NllLossBackward>)
[3666] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3667] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[3668] Train Loss:    tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[3669] Train Loss:    tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[3670] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[3671] Train Loss:    tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[3672] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3673] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[3674] Train Loss:    tensor(0.6773, device='cuda:0', grad_fn=<NllLossBackward>)
[3675] Train Loss:    tensor(0.7283, device='cuda:0', grad_fn=<NllLossBackward>)
[3676] Train Loss:    tensor(0.6811, device='cuda:0', grad_fn=<NllLossBackward>)
[3677] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[3678] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[3679] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[3680] Train Loss:    tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[3681] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[3682] Train Loss:    tensor(0.6611, device='cuda:0', grad_fn=<NllLossBackward>)
[3683] Train Loss:    tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[3684] Train Loss:    tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[3685] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[3686] Train Loss:    tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[3687] Train Loss:    tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3688] Train Loss:    tensor(0.6714, device='cuda:0', grad_fn=<NllLossBackward>)
[3689] Train Loss:    tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[3690] Train Loss:    tensor(0.7229, device='cuda:0', grad_fn=<NllLossBackward>)
[3691] Train Loss:    tensor(0.7076, device='cuda:0', grad_fn=<NllLossBackward>)
[3692] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[3693] Train Loss:    tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[3694] Train Loss:    tensor(0.7094, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3695] Train Loss:    tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[3696] Train Loss:    tensor(0.6621, device='cuda:0', grad_fn=<NllLossBackward>)
[3697] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3698] Train Loss:    tensor(0.7040, device='cuda:0', grad_fn=<NllLossBackward>)
[3699] Train Loss:    tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[3700] Train Loss:    tensor(0.6730, device='cuda:0', grad_fn=<NllLossBackward>)
[3701] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3702] Train Loss:    tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[3703] Train Loss:    tensor(0.6510, device='cuda:0', grad_fn=<NllLossBackward>)
[3704] Train Loss:    tensor(0.7253, device='cuda:0', grad_fn=<NllLossBackward>)
[3705] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[3706] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[3707] Train Loss:    tensor(0.7323, device='cuda:0', grad_fn=<NllLossBackward>)
[3708] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[3709] Train Loss:    tensor(0.7118, device='cuda:0', grad_fn=<NllLossBackward>)
[3710] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[3711] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[3712] Train Loss:    tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[3713] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[3714] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3715] Train Loss:    tensor(0.6897, device='cuda:0', grad_fn=<NllLossBackward>)
[3716] Train Loss:    tensor(0.6776, device='cuda:0', grad_fn=<NllLossBackward>)
[3717] Train Loss:    tensor(0.7224, device='cuda:0', grad_fn=<NllLossBackward>)
[3718] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[3719] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[3720] Train Loss:    tensor(0.7066, device='cuda:0', grad_fn=<NllLossBackward>)
[3721] Train Loss:    tensor(0.6531, device='cuda:0', grad_fn=<NllLossBackward>)
[3722] Train Loss:    tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[3723] Train Loss:    tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[3724] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[3725] Train Loss:    tensor(0.7281, device='cuda:0', grad_fn=<NllLossBackward>)
[3726] Train Loss:    tensor(0.6578, device='cuda:0', grad_fn=<NllLossBackward>)
[3727] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[3728] Train Loss:    tensor(0.6812, device='cuda:0', grad_fn=<NllLossBackward>)
[3729] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[3730] Train Loss:    tensor(0.6642, device='cuda:0', grad_fn=<NllLossBackward>)
[3731] Train Loss:    tensor(0.6940, device='cuda:0', grad_fn=<NllLossBackward>)
[3732] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3733] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[3734] Train Loss:    tensor(0.6608, device='cuda:0', grad_fn=<NllLossBackward>)
[3735] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3736] Train Loss:    tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[3737] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3738] Train Loss:    tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[3739] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[3740] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[3741] Train Loss:    tensor(0.6661, device='cuda:0', grad_fn=<NllLossBackward>)
[3742] Train Loss:    tensor(0.6556, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3743] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[3744] Train Loss:    tensor(0.6480, device='cuda:0', grad_fn=<NllLossBackward>)
[3745] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[3746] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[3747] Train Loss:    tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[3748] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[3749] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[3750] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
[3751] Train Loss:    tensor(0.7290, device='cuda:0', grad_fn=<NllLossBackward>)
[3752] Train Loss:    tensor(0.7178, device='cuda:0', grad_fn=<NllLossBackward>)
[3753] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[3754] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[3755] Train Loss:    tensor(0.7170, device='cuda:0', grad_fn=<NllLossBackward>)
[3756] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[3757] Train Loss:    tensor(0.7165, device='cuda:0', grad_fn=<NllLossBackward>)
[3758] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3759] Train Loss:    tensor(0.6785, device='cuda:0', grad_fn=<NllLossBackward>)
[3760] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[3761] Train Loss:    tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[3762] Train Loss:    tensor(0.6535, device='cuda:0', grad_fn=<NllLossBackward>)
[3763] Train Loss:    tensor(0.7065, device='cuda:0', grad_fn=<NllLossBackward>)
[3764] Train Loss:    tensor(0.6967, device='cuda:0', grad_fn=<NllLossBackward>)
[3765] Train Loss:    tensor(0.7079, device='cuda:0', grad_fn=<NllLossBackward>)
[3766] Train Loss:    tensor(0.6815, device='cuda:0', grad_fn=<NllLossBackward>)
[3767] Train Loss:    tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[3768] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[3769] Train Loss:    tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[3770] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[3771] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[3772] Train Loss:    tensor(0.6724, device='cuda:0', grad_fn=<NllLossBackward>)
[3773] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[3774] Train Loss:    tensor(0.7024, device='cuda:0', grad_fn=<NllLossBackward>)
[3775] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
[3776] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3777] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[3778] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[3779] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[3780] Train Loss:    tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[3781] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[3782] Train Loss:    tensor(0.6425, device='cuda:0', grad_fn=<NllLossBackward>)
[3783] Train Loss:    tensor(0.7034, device='cuda:0', grad_fn=<NllLossBackward>)
[3784] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3785] Train Loss:    tensor(0.7601, device='cuda:0', grad_fn=<NllLossBackward>)
[3786] Train Loss:    tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[3787] Train Loss:    tensor(0.7258, device='cuda:0', grad_fn=<NllLossBackward>)
[3788] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[3789] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[3790] Train Loss:    tensor(0.7115, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3791] Train Loss:   tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[3792] Train Loss:   tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[3793] Train Loss:   tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[3794] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[3795] Train Loss:   tensor(0.6891, device='cuda:0', grad_fn=<NllLossBackward>)
[3796] Train Loss:   tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[3797] Train Loss:   tensor(0.6660, device='cuda:0', grad_fn=<NllLossBackward>)
[3798] Train Loss:   tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[3799] Train Loss:   tensor(0.6682, device='cuda:0', grad_fn=<NllLossBackward>)
[3800] Train Loss:   tensor(0.7007, device='cuda:0', grad_fn=<NllLossBackward>)
[3801] Train Loss:   tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3802] Train Loss:   tensor(0.6570, device='cuda:0', grad_fn=<NllLossBackward>)
[3803] Train Loss:   tensor(0.6882, device='cuda:0', grad_fn=<NllLossBackward>)
[3804] Train Loss:   tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[3805] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3806] Train Loss:   tensor(0.7112, device='cuda:0', grad_fn=<NllLossBackward>)
[3807] Train Loss:   tensor(0.7380, device='cuda:0', grad_fn=<NllLossBackward>)
[3808] Train Loss:   tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[3809] Train Loss:   tensor(0.7026, device='cuda:0', grad_fn=<NllLossBackward>)
[3810] Train Loss:   tensor(0.6754, device='cuda:0', grad_fn=<NllLossBackward>)
[3811] Train Loss:   tensor(0.7325, device='cuda:0', grad_fn=<NllLossBackward>)
[3812] Train Loss:   tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[3813] Train Loss:   tensor(0.6638, device='cuda:0', grad_fn=<NllLossBackward>)
[3814] Train Loss:   tensor(0.6765, device='cuda:0', grad_fn=<NllLossBackward>)
[3815] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[3816] Train Loss:   tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[3817] Train Loss:   tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[3818] Train Loss:   tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[3819] Train Loss:   tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[3820] Train Loss:   tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[3821] Train Loss:   tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[3822] Train Loss:   tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[3823] Train Loss:   tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[3824] Train Loss:   tensor(0.6923, device='cuda:0', grad_fn=<NllLossBackward>)
[3825] Train Loss:   tensor(0.6914, device='cuda:0', grad_fn=<NllLossBackward>)
[3826] Train Loss:   tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[3827] Train Loss:   tensor(0.7109, device='cuda:0', grad_fn=<NllLossBackward>)
[3828] Train Loss:   tensor(0.7155, device='cuda:0', grad_fn=<NllLossBackward>)
[3829] Train Loss:   tensor(0.7127, device='cuda:0', grad_fn=<NllLossBackward>)
[3830] Train Loss:   tensor(0.6642, device='cuda:0', grad_fn=<NllLossBackward>)
[3831] Train Loss:   tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[3832] Train Loss:   tensor(0.7008, device='cuda:0', grad_fn=<NllLossBackward>)
[3833] Train Loss:   tensor(0.7064, device='cuda:0', grad_fn=<NllLossBackward>)
[3834] Train Loss:   tensor(0.6976, device='cuda:0', grad_fn=<NllLossBackward>)
[3835] Train Loss:   tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[3836] Train Loss:   tensor(0.6899, device='cuda:0', grad_fn=<NllLossBackward>)
[3837] Train Loss:   tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3838] Train Loss:   tensor(0.7096, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3839] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[3840] Train Loss:    tensor(0.7102, device='cuda:0', grad_fn=<NllLossBackward>)
[3841] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[3842] Train Loss:    tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[3843] Train Loss:    tensor(0.6578, device='cuda:0', grad_fn=<NllLossBackward>)
[3844] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[3845] Train Loss:    tensor(0.6748, device='cuda:0', grad_fn=<NllLossBackward>)
[3846] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[3847] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[3848] Train Loss:    tensor(0.6649, device='cuda:0', grad_fn=<NllLossBackward>)
[3849] Train Loss:    tensor(0.6955, device='cuda:0', grad_fn=<NllLossBackward>)
[3850] Train Loss:    tensor(0.7054, device='cuda:0', grad_fn=<NllLossBackward>)
[3851] Train Loss:    tensor(0.7041, device='cuda:0', grad_fn=<NllLossBackward>)
[3852] Train Loss:    tensor(0.6624, device='cuda:0', grad_fn=<NllLossBackward>)
[3853] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[3854] Train Loss:    tensor(0.6814, device='cuda:0', grad_fn=<NllLossBackward>)
[3855] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3856] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[3857] Train Loss:    tensor(0.6514, device='cuda:0', grad_fn=<NllLossBackward>)
[3858] Train Loss:    tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[3859] Train Loss:    tensor(0.6612, device='cuda:0', grad_fn=<NllLossBackward>)
[3860] Train Loss:    tensor(0.7239, device='cuda:0', grad_fn=<NllLossBackward>)
[3861] Train Loss:    tensor(0.6789, device='cuda:0', grad_fn=<NllLossBackward>)
[3862] Train Loss:    tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[3863] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[3864] Train Loss:    tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[3865] Train Loss:    tensor(0.6912, device='cuda:0', grad_fn=<NllLossBackward>)
[3866] Train Loss:    tensor(0.7230, device='cuda:0', grad_fn=<NllLossBackward>)
[3867] Train Loss:    tensor(0.7183, device='cuda:0', grad_fn=<NllLossBackward>)
[3868] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[3869] Train Loss:    tensor(0.6575, device='cuda:0', grad_fn=<NllLossBackward>)
[3870] Train Loss:    tensor(0.6573, device='cuda:0', grad_fn=<NllLossBackward>)
[3871] Train Loss:    tensor(0.6875, device='cuda:0', grad_fn=<NllLossBackward>)
[3872] Train Loss:    tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3873] Train Loss:    tensor(0.6463, device='cuda:0', grad_fn=<NllLossBackward>)
[3874] Train Loss:    tensor(0.6878, device='cuda:0', grad_fn=<NllLossBackward>)
[3875] Train Loss:    tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[3876] Train Loss:    tensor(0.7392, device='cuda:0', grad_fn=<NllLossBackward>)
[3877] Train Loss:    tensor(0.6935, device='cuda:0', grad_fn=<NllLossBackward>)
[3878] Train Loss:    tensor(0.7189, device='cuda:0', grad_fn=<NllLossBackward>)
[3879] Train Loss:    tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[3880] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[3881] Train Loss:    tensor(0.6937, device='cuda:0', grad_fn=<NllLossBackward>)
[3882] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[3883] Train Loss:    tensor(0.6702, device='cuda:0', grad_fn=<NllLossBackward>)
[3884] Train Loss:    tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[3885] Train Loss:    tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[3886] Train Loss:    tensor(0.6619, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3887] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[3888] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[3889] Train Loss:    tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[3890] Train Loss:    tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[3891] Train Loss:    tensor(0.6650, device='cuda:0', grad_fn=<NllLossBackward>)
[3892] Train Loss:    tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[3893] Train Loss:    tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[3894] Train Loss:    tensor(0.6712, device='cuda:0', grad_fn=<NllLossBackward>)
[3895] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[3896] Train Loss:    tensor(0.6801, device='cuda:0', grad_fn=<NllLossBackward>)
[3897] Train Loss:    tensor(0.6840, device='cuda:0', grad_fn=<NllLossBackward>)
[3898] Train Loss:    tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[3899] Train Loss:    tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[3900] Train Loss:    tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[3901] Train Loss:    tensor(0.6867, device='cuda:0', grad_fn=<NllLossBackward>)
[3902] Train Loss:    tensor(0.6661, device='cuda:0', grad_fn=<NllLossBackward>)
[3903] Train Loss:    tensor(0.6874, device='cuda:0', grad_fn=<NllLossBackward>)
[3904] Train Loss:    tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[3905] Train Loss:    tensor(0.6762, device='cuda:0', grad_fn=<NllLossBackward>)
[3906] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[3907] Train Loss:    tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[3908] Train Loss:    tensor(0.6906, device='cuda:0', grad_fn=<NllLossBackward>)
[3909] Train Loss:    tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3910] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[3911] Train Loss:    tensor(0.7116, device='cuda:0', grad_fn=<NllLossBackward>)
[3912] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[3913] Train Loss:    tensor(0.7182, device='cuda:0', grad_fn=<NllLossBackward>)
[3914] Train Loss:    tensor(0.7027, device='cuda:0', grad_fn=<NllLossBackward>)
[3915] Train Loss:    tensor(0.6556, device='cuda:0', grad_fn=<NllLossBackward>)
[3916] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[3917] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3918] Train Loss:    tensor(0.7023, device='cuda:0', grad_fn=<NllLossBackward>)
[3919] Train Loss:    tensor(0.6823, device='cuda:0', grad_fn=<NllLossBackward>)
[3920] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[3921] Train Loss:    tensor(0.6700, device='cuda:0', grad_fn=<NllLossBackward>)
[3922] Train Loss:    tensor(0.7091, device='cuda:0', grad_fn=<NllLossBackward>)
[3923] Train Loss:    tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[3924] Train Loss:    tensor(0.6951, device='cuda:0', grad_fn=<NllLossBackward>)
[3925] Train Loss:    tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[3926] Train Loss:    tensor(0.6570, device='cuda:0', grad_fn=<NllLossBackward>)
[3927] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[3928] Train Loss:    tensor(0.6748, device='cuda:0', grad_fn=<NllLossBackward>)
[3929] Train Loss:    tensor(0.7031, device='cuda:0', grad_fn=<NllLossBackward>)
[3930] Train Loss:    tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[3931] Train Loss:    tensor(0.6659, device='cuda:0', grad_fn=<NllLossBackward>)
[3932] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[3933] Train Loss:    tensor(0.6994, device='cuda:0', grad_fn=<NllLossBackward>)
[3934] Train Loss:    tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3935] Train Loss:   tensor(0.6924, device='cuda:0', grad_fn=<NllLossBackward>)
[3936] Train Loss:   tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[3937] Train Loss:   tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[3938] Train Loss:   tensor(0.6839, device='cuda:0', grad_fn=<NllLossBackward>)
[3939] Train Loss:   tensor(0.7256, device='cuda:0', grad_fn=<NllLossBackward>)
[3940] Train Loss:   tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[3941] Train Loss:   tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[3942] Train Loss:   tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[3943] Train Loss:   tensor(0.6959, device='cuda:0', grad_fn=<NllLossBackward>)
[3944] Train Loss:   tensor(0.6679, device='cuda:0', grad_fn=<NllLossBackward>)
[3945] Train Loss:   tensor(0.6961, device='cuda:0', grad_fn=<NllLossBackward>)
[3946] Train Loss:   tensor(0.7061, device='cuda:0', grad_fn=<NllLossBackward>)
[3947] Train Loss:   tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[3948] Train Loss:   tensor(0.6941, device='cuda:0', grad_fn=<NllLossBackward>)
[3949] Train Loss:   tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[3950] Train Loss:   tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[3951] Train Loss:   tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[3952] Train Loss:   tensor(0.6884, device='cuda:0', grad_fn=<NllLossBackward>)
[3953] Train Loss:   tensor(0.6688, device='cuda:0', grad_fn=<NllLossBackward>)
[3954] Train Loss:   tensor(0.6596, device='cuda:0', grad_fn=<NllLossBackward>)
[3955] Train Loss:   tensor(0.7006, device='cuda:0', grad_fn=<NllLossBackward>)
[3956] Train Loss:   tensor(0.7081, device='cuda:0', grad_fn=<NllLossBackward>)
[3957] Train Loss:   tensor(0.6756, device='cuda:0', grad_fn=<NllLossBackward>)
[3958] Train Loss:   tensor(0.6791, device='cuda:0', grad_fn=<NllLossBackward>)
[3959] Train Loss:   tensor(0.6351, device='cuda:0', grad_fn=<NllLossBackward>)
[3960] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[3961] Train Loss:   tensor(0.6720, device='cuda:0', grad_fn=<NllLossBackward>)
[3962] Train Loss:   tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[3963] Train Loss:   tensor(0.6916, device='cuda:0', grad_fn=<NllLossBackward>)
[3964] Train Loss:   tensor(0.7138, device='cuda:0', grad_fn=<NllLossBackward>)
[3965] Train Loss:   tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[3966] Train Loss:   tensor(0.6893, device='cuda:0', grad_fn=<NllLossBackward>)
[3967] Train Loss:   tensor(0.6567, device='cuda:0', grad_fn=<NllLossBackward>)
[3968] Train Loss:   tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[3969] Train Loss:   tensor(0.6887, device='cuda:0', grad_fn=<NllLossBackward>)
[3970] Train Loss:   tensor(0.7188, device='cuda:0', grad_fn=<NllLossBackward>)
[3971] Train Loss:   tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[3972] Train Loss:   tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[3973] Train Loss:   tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[3974] Train Loss:   tensor(0.7128, device='cuda:0', grad_fn=<NllLossBackward>)
[3975] Train Loss:   tensor(0.6411, device='cuda:0', grad_fn=<NllLossBackward>)
[3976] Train Loss:   tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[3977] Train Loss:   tensor(0.6664, device='cuda:0', grad_fn=<NllLossBackward>)
[3978] Train Loss:   tensor(0.6623, device='cuda:0', grad_fn=<NllLossBackward>)
[3979] Train Loss:   tensor(0.7199, device='cuda:0', grad_fn=<NllLossBackward>)
[3980] Train Loss:   tensor(0.7002, device='cuda:0', grad_fn=<NllLossBackward>)
[3981] Train Loss:   tensor(0.7057, device='cuda:0', grad_fn=<NllLossBackward>)
[3982] Train Loss:   tensor(0.7070, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[3983] Train Loss:   tensor(0.7144, device='cuda:0', grad_fn=<NllLossBackward>)
[3984] Train Loss:   tensor(0.7089, device='cuda:0', grad_fn=<NllLossBackward>)
[3985] Train Loss:   tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[3986] Train Loss:   tensor(0.6663, device='cuda:0', grad_fn=<NllLossBackward>)
[3987] Train Loss:   tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[3988] Train Loss:   tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[3989] Train Loss:   tensor(0.6729, device='cuda:0', grad_fn=<NllLossBackward>)
[3990] Train Loss:   tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[3991] Train Loss:   tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[3992] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[3993] Train Loss:   tensor(0.6633, device='cuda:0', grad_fn=<NllLossBackward>)
[3994] Train Loss:   tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[3995] Train Loss:   tensor(0.7185, device='cuda:0', grad_fn=<NllLossBackward>)
[3996] Train Loss:   tensor(0.6582, device='cuda:0', grad_fn=<NllLossBackward>)
[3997] Train Loss:   tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[3998] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[3999] Train Loss:   tensor(0.6728, device='cuda:0', grad_fn=<NllLossBackward>)
[4000] Train Loss:   tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[4001] Train Loss:   tensor(0.6972, device='cuda:0', grad_fn=<NllLossBackward>)
[4002] Train Loss:   tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[4003] Train Loss:   tensor(0.6851, device='cuda:0', grad_fn=<NllLossBackward>)
[4004] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[4005] Train Loss:   tensor(0.6986, device='cuda:0', grad_fn=<NllLossBackward>)
[4006] Train Loss:   tensor(0.6537, device='cuda:0', grad_fn=<NllLossBackward>)
[4007] Train Loss:   tensor(0.7036, device='cuda:0', grad_fn=<NllLossBackward>)
[4008] Train Loss:   tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[4009] Train Loss:   tensor(0.7312, device='cuda:0', grad_fn=<NllLossBackward>)
[4010] Train Loss:   tensor(0.6644, device='cuda:0', grad_fn=<NllLossBackward>)
[4011] Train Loss:   tensor(0.7126, device='cuda:0', grad_fn=<NllLossBackward>)
[4012] Train Loss:   tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[4013] Train Loss:   tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[4014] Train Loss:   tensor(0.6852, device='cuda:0', grad_fn=<NllLossBackward>)
[4015] Train Loss:   tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[4016] Train Loss:   tensor(0.7047, device='cuda:0', grad_fn=<NllLossBackward>)
[4017] Train Loss:   tensor(0.7150, device='cuda:0', grad_fn=<NllLossBackward>)
[4018] Train Loss:   tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[4019] Train Loss:   tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[4020] Train Loss:   tensor(0.7088, device='cuda:0', grad_fn=<NllLossBackward>)
[4021] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[4022] Train Loss:   tensor(0.6946, device='cuda:0', grad_fn=<NllLossBackward>)
[4023] Train Loss:   tensor(0.6873, device='cuda:0', grad_fn=<NllLossBackward>)
[4024] Train Loss:   tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[4025] Train Loss:   tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[4026] Train Loss:   tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
[4027] Train Loss:   tensor(0.6846, device='cuda:0', grad_fn=<NllLossBackward>)
[4028] Train Loss:   tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[4029] Train Loss:   tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[4030] Train Loss:   tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4031] Train Loss:    tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[4032] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[4033] Train Loss:    tensor(0.7149, device='cuda:0', grad_fn=<NllLossBackward>)
[4034] Train Loss:    tensor(0.6658, device='cuda:0', grad_fn=<NllLossBackward>)
[4035] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[4036] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[4037] Train Loss:    tensor(0.6733, device='cuda:0', grad_fn=<NllLossBackward>)
[4038] Train Loss:    tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[4039] Train Loss:    tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[4040] Train Loss:    tensor(0.6770, device='cuda:0', grad_fn=<NllLossBackward>)
[4041] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[4042] Train Loss:    tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[4043] Train Loss:    tensor(0.6642, device='cuda:0', grad_fn=<NllLossBackward>)
[4044] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[4045] Train Loss:    tensor(0.6817, device='cuda:0', grad_fn=<NllLossBackward>)
[4046] Train Loss:    tensor(0.6648, device='cuda:0', grad_fn=<NllLossBackward>)
[4047] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
[4048] Train Loss:    tensor(0.7038, device='cuda:0', grad_fn=<NllLossBackward>)
[4049] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[4050] Train Loss:    tensor(0.6909, device='cuda:0', grad_fn=<NllLossBackward>)
[4051] Train Loss:    tensor(0.6918, device='cuda:0', grad_fn=<NllLossBackward>)
[4052] Train Loss:    tensor(0.6657, device='cuda:0', grad_fn=<NllLossBackward>)
[4053] Train Loss:    tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
[4054] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[4055] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[4056] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[4057] Train Loss:    tensor(0.6601, device='cuda:0', grad_fn=<NllLossBackward>)
[4058] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[4059] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[4060] Train Loss:    tensor(0.6958, device='cuda:0', grad_fn=<NllLossBackward>)
[4061] Train Loss:    tensor(0.7192, device='cuda:0', grad_fn=<NllLossBackward>)
[4062] Train Loss:    tensor(0.6784, device='cuda:0', grad_fn=<NllLossBackward>)
[4063] Train Loss:    tensor(0.6667, device='cuda:0', grad_fn=<NllLossBackward>)
[4064] Train Loss:    tensor(0.6844, device='cuda:0', grad_fn=<NllLossBackward>)
[4065] Train Loss:    tensor(0.6467, device='cuda:0', grad_fn=<NllLossBackward>)
[4066] Train Loss:    tensor(0.6971, device='cuda:0', grad_fn=<NllLossBackward>)
[4067] Train Loss:    tensor(0.6807, device='cuda:0', grad_fn=<NllLossBackward>)
[4068] Train Loss:    tensor(0.7132, device='cuda:0', grad_fn=<NllLossBackward>)
[4069] Train Loss:    tensor(0.6927, device='cuda:0', grad_fn=<NllLossBackward>)
[4070] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[4071] Train Loss:    tensor(0.6925, device='cuda:0', grad_fn=<NllLossBackward>)
[4072] Train Loss:    tensor(0.6693, device='cuda:0', grad_fn=<NllLossBackward>)
[4073] Train Loss:    tensor(0.7186, device='cuda:0', grad_fn=<NllLossBackward>)
[4074] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[4075] Train Loss:    tensor(0.7252, device='cuda:0', grad_fn=<NllLossBackward>)
[4076] Train Loss:    tensor(0.6667, device='cuda:0', grad_fn=<NllLossBackward>)
[4077] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[4078] Train Loss:    tensor(0.6863, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4079] Train Loss:    tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[4080] Train Loss:    tensor(0.6517, device='cuda:0', grad_fn=<NllLossBackward>)
[4081] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[4082] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[4083] Train Loss:    tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[4084] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
[4085] Train Loss:    tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[4086] Train Loss:    tensor(0.6969, device='cuda:0', grad_fn=<NllLossBackward>)
[4087] Train Loss:    tensor(0.7003, device='cuda:0', grad_fn=<NllLossBackward>)
[4088] Train Loss:    tensor(0.6904, device='cuda:0', grad_fn=<NllLossBackward>)
[4089] Train Loss:    tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[4090] Train Loss:    tensor(0.7101, device='cuda:0', grad_fn=<NllLossBackward>)
[4091] Train Loss:    tensor(0.7154, device='cuda:0', grad_fn=<NllLossBackward>)
[4092] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[4093] Train Loss:    tensor(0.7071, device='cuda:0', grad_fn=<NllLossBackward>)
[4094] Train Loss:    tensor(0.7136, device='cuda:0', grad_fn=<NllLossBackward>)
[4095] Train Loss:    tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[4096] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[4097] Train Loss:    tensor(0.6671, device='cuda:0', grad_fn=<NllLossBackward>)
[4098] Train Loss:    tensor(0.6790, device='cuda:0', grad_fn=<NllLossBackward>)
[4099] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[4100] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[4101] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[4102] Train Loss:    tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[4103] Train Loss:    tensor(0.6824, device='cuda:0', grad_fn=<NllLossBackward>)
[4104] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[4105] Train Loss:    tensor(0.6717, device='cuda:0', grad_fn=<NllLossBackward>)
[4106] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[4107] Train Loss:    tensor(0.7247, device='cuda:0', grad_fn=<NllLossBackward>)
[4108] Train Loss:    tensor(0.6582, device='cuda:0', grad_fn=<NllLossBackward>)
[4109] Train Loss:    tensor(0.6563, device='cuda:0', grad_fn=<NllLossBackward>)
[4110] Train Loss:    tensor(0.7045, device='cuda:0', grad_fn=<NllLossBackward>)
[4111] Train Loss:    tensor(0.6663, device='cuda:0', grad_fn=<NllLossBackward>)
[4112] Train Loss:    tensor(0.7046, device='cuda:0', grad_fn=<NllLossBackward>)
[4113] Train Loss:    tensor(0.6686, device='cuda:0', grad_fn=<NllLossBackward>)
[4114] Train Loss:    tensor(0.7053, device='cuda:0', grad_fn=<NllLossBackward>)
[4115] Train Loss:    tensor(0.6446, device='cuda:0', grad_fn=<NllLossBackward>)
[4116] Train Loss:    tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[4117] Train Loss:    tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[4118] Train Loss:    tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[4119] Train Loss:    tensor(0.6677, device='cuda:0', grad_fn=<NllLossBackward>)
[4120] Train Loss:    tensor(0.7049, device='cuda:0', grad_fn=<NllLossBackward>)
[4121] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[4122] Train Loss:    tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[4123] Train Loss:    tensor(0.6950, device='cuda:0', grad_fn=<NllLossBackward>)
[4124] Train Loss:    tensor(0.6584, device='cuda:0', grad_fn=<NllLossBackward>)
[4125] Train Loss:    tensor(0.6892, device='cuda:0', grad_fn=<NllLossBackward>)
[4126] Train Loss:    tensor(0.6932, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4127] Train Loss:   tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[4128] Train Loss:   tensor(0.6966, device='cuda:0', grad_fn=<NllLossBackward>)
[4129] Train Loss:   tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[4130] Train Loss:   tensor(0.7093, device='cuda:0', grad_fn=<NllLossBackward>)
[4131] Train Loss:   tensor(0.7227, device='cuda:0', grad_fn=<NllLossBackward>)
[4132] Train Loss:   tensor(0.6625, device='cuda:0', grad_fn=<NllLossBackward>)
[4133] Train Loss:   tensor(0.6511, device='cuda:0', grad_fn=<NllLossBackward>)
[4134] Train Loss:   tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[4135] Train Loss:   tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[4136] Train Loss:   tensor(0.6563, device='cuda:0', grad_fn=<NllLossBackward>)
[4137] Train Loss:   tensor(0.6804, device='cuda:0', grad_fn=<NllLossBackward>)
[4138] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[4139] Train Loss:   tensor(0.7044, device='cuda:0', grad_fn=<NllLossBackward>)
[4140] Train Loss:   tensor(0.7134, device='cuda:0', grad_fn=<NllLossBackward>)
[4141] Train Loss:   tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[4142] Train Loss:   tensor(0.6776, device='cuda:0', grad_fn=<NllLossBackward>)
[4143] Train Loss:   tensor(0.6676, device='cuda:0', grad_fn=<NllLossBackward>)
[4144] Train Loss:   tensor(0.6775, device='cuda:0', grad_fn=<NllLossBackward>)
[4145] Train Loss:   tensor(0.7218, device='cuda:0', grad_fn=<NllLossBackward>)
[4146] Train Loss:   tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[4147] Train Loss:   tensor(0.7206, device='cuda:0', grad_fn=<NllLossBackward>)
[4148] Train Loss:   tensor(0.6936, device='cuda:0', grad_fn=<NllLossBackward>)
[4149] Train Loss:   tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[4150] Train Loss:   tensor(0.6792, device='cuda:0', grad_fn=<NllLossBackward>)
[4151] Train Loss:   tensor(0.7011, device='cuda:0', grad_fn=<NllLossBackward>)
[4152] Train Loss:   tensor(0.7202, device='cuda:0', grad_fn=<NllLossBackward>)
[4153] Train Loss:   tensor(0.6759, device='cuda:0', grad_fn=<NllLossBackward>)
[4154] Train Loss:   tensor(0.7028, device='cuda:0', grad_fn=<NllLossBackward>)
[4155] Train Loss:   tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[4156] Train Loss:   tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[4157] Train Loss:   tensor(0.6780, device='cuda:0', grad_fn=<NllLossBackward>)
[4158] Train Loss:   tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[4159] Train Loss:   tensor(0.6829, device='cuda:0', grad_fn=<NllLossBackward>)
[4160] Train Loss:   tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[4161] Train Loss:   tensor(0.6928, device='cuda:0', grad_fn=<NllLossBackward>)
[4162] Train Loss:   tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[4163] Train Loss:   tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[4164] Train Loss:   tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[4165] Train Loss:   tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[4166] Train Loss:   tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[4167] Train Loss:   tensor(0.6628, device='cuda:0', grad_fn=<NllLossBackward>)
[4168] Train Loss:   tensor(0.7176, device='cuda:0', grad_fn=<NllLossBackward>)
[4169] Train Loss:   tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[4170] Train Loss:   tensor(0.6607, device='cuda:0', grad_fn=<NllLossBackward>)
[4171] Train Loss:   tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[4172] Train Loss:   tensor(0.6977, device='cuda:0', grad_fn=<NllLossBackward>)
[4173] Train Loss:   tensor(0.7147, device='cuda:0', grad_fn=<NllLossBackward>)
[4174] Train Loss:   tensor(0.7226, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4175] Train Loss:    tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[4176] Train Loss:    tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
[4177] Train Loss:    tensor(0.7055, device='cuda:0', grad_fn=<NllLossBackward>)
[4178] Train Loss:    tensor(0.6751, device='cuda:0', grad_fn=<NllLossBackward>)
[4179] Train Loss:    tensor(0.6853, device='cuda:0', grad_fn=<NllLossBackward>)
[4180] Train Loss:    tensor(0.6732, device='cuda:0', grad_fn=<NllLossBackward>)
[4181] Train Loss:    tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[4182] Train Loss:    tensor(0.6917, device='cuda:0', grad_fn=<NllLossBackward>)
[4183] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[4184] Train Loss:    tensor(0.6636, device='cuda:0', grad_fn=<NllLossBackward>)
[4185] Train Loss:    tensor(0.6979, device='cuda:0', grad_fn=<NllLossBackward>)
[4186] Train Loss:    tensor(0.6635, device='cuda:0', grad_fn=<NllLossBackward>)
[4187] Train Loss:    tensor(0.6919, device='cuda:0', grad_fn=<NllLossBackward>)
[4188] Train Loss:    tensor(0.7037, device='cuda:0', grad_fn=<NllLossBackward>)
[4189] Train Loss:    tensor(0.6680, device='cuda:0', grad_fn=<NllLossBackward>)
[4190] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[4191] Train Loss:    tensor(0.6639, device='cuda:0', grad_fn=<NllLossBackward>)
[4192] Train Loss:    tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[4193] Train Loss:    tensor(0.6768, device='cuda:0', grad_fn=<NllLossBackward>)
[4194] Train Loss:    tensor(0.7014, device='cuda:0', grad_fn=<NllLossBackward>)
[4195] Train Loss:    tensor(0.7056, device='cuda:0', grad_fn=<NllLossBackward>)
[4196] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[4197] Train Loss:    tensor(0.6828, device='cuda:0', grad_fn=<NllLossBackward>)
[4198] Train Loss:    tensor(0.7177, device='cuda:0', grad_fn=<NllLossBackward>)
[4199] Train Loss:    tensor(0.6862, device='cuda:0', grad_fn=<NllLossBackward>)
[4200] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[4201] Train Loss:    tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[4202] Train Loss:    tensor(0.7022, device='cuda:0', grad_fn=<NllLossBackward>)
[4203] Train Loss:    tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[4204] Train Loss:    tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[4205] Train Loss:    tensor(0.7039, device='cuda:0', grad_fn=<NllLossBackward>)
[4206] Train Loss:    tensor(0.6767, device='cuda:0', grad_fn=<NllLossBackward>)
[4207] Train Loss:    tensor(0.7254, device='cuda:0', grad_fn=<NllLossBackward>)
[4208] Train Loss:    tensor(0.6803, device='cuda:0', grad_fn=<NllLossBackward>)
[4209] Train Loss:    tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[4210] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[4211] Train Loss:    tensor(0.7062, device='cuda:0', grad_fn=<NllLossBackward>)
[4212] Train Loss:    tensor(0.6692, device='cuda:0', grad_fn=<NllLossBackward>)
[4213] Train Loss:    tensor(0.7087, device='cuda:0', grad_fn=<NllLossBackward>)
[4214] Train Loss:    tensor(0.7051, device='cuda:0', grad_fn=<NllLossBackward>)
[4215] Train Loss:    tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[4216] Train Loss:    tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[4217] Train Loss:    tensor(0.6830, device='cuda:0', grad_fn=<NllLossBackward>)
[4218] Train Loss:    tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[4219] Train Loss:    tensor(0.7268, device='cuda:0', grad_fn=<NllLossBackward>)
[4220] Train Loss:    tensor(0.6708, device='cuda:0', grad_fn=<NllLossBackward>)
[4221] Train Loss:    tensor(0.6774, device='cuda:0', grad_fn=<NllLossBackward>)
[4222] Train Loss:    tensor(0.7019, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4223] Train Loss:   tensor(0.7000, device='cuda:0', grad_fn=<NllLossBackward>)
[4224] Train Loss:   tensor(0.6903, device='cuda:0', grad_fn=<NllLossBackward>)
[4225] Train Loss:   tensor(0.7072, device='cuda:0', grad_fn=<NllLossBackward>)
[4226] Train Loss:   tensor(0.6652, device='cuda:0', grad_fn=<NllLossBackward>)
[4227] Train Loss:   tensor(0.6590, device='cuda:0', grad_fn=<NllLossBackward>)
[4228] Train Loss:   tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[4229] Train Loss:   tensor(0.7021, device='cuda:0', grad_fn=<NllLossBackward>)
[4230] Train Loss:   tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[4231] Train Loss:   tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[4232] Train Loss:   tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[4233] Train Loss:   tensor(0.6600, device='cuda:0', grad_fn=<NllLossBackward>)
[4234] Train Loss:   tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[4235] Train Loss:   tensor(0.6978, device='cuda:0', grad_fn=<NllLossBackward>)
[4236] Train Loss:   tensor(0.7506, device='cuda:0', grad_fn=<NllLossBackward>)
[4237] Train Loss:   tensor(0.6598, device='cuda:0', grad_fn=<NllLossBackward>)
[4238] Train Loss:   tensor(0.6980, device='cuda:0', grad_fn=<NllLossBackward>)
[4239] Train Loss:   tensor(0.6855, device='cuda:0', grad_fn=<NllLossBackward>)
[4240] Train Loss:   tensor(0.7141, device='cuda:0', grad_fn=<NllLossBackward>)
[4241] Train Loss:   tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[4242] Train Loss:   tensor(0.6938, device='cuda:0', grad_fn=<NllLossBackward>)
[4243] Train Loss:   tensor(0.6593, device='cuda:0', grad_fn=<NllLossBackward>)
[4244] Train Loss:   tensor(0.7237, device='cuda:0', grad_fn=<NllLossBackward>)
[4245] Train Loss:   tensor(0.6920, device='cuda:0', grad_fn=<NllLossBackward>)
[4246] Train Loss:   tensor(0.6866, device='cuda:0', grad_fn=<NllLossBackward>)
[4247] Train Loss:   tensor(0.7131, device='cuda:0', grad_fn=<NllLossBackward>)
[4248] Train Loss:   tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[4249] Train Loss:   tensor(0.6625, device='cuda:0', grad_fn=<NllLossBackward>)
[4250] Train Loss:   tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[4251] Train Loss:   tensor(0.6985, device='cuda:0', grad_fn=<NllLossBackward>)
[4252] Train Loss:   tensor(0.6385, device='cuda:0', grad_fn=<NllLossBackward>)
[4253] Train Loss:   tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[4254] Train Loss:   tensor(0.6479, device='cuda:0', grad_fn=<NllLossBackward>)
[4255] Train Loss:   tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[4256] Train Loss:   tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[4257] Train Loss:   tensor(0.7005, device='cuda:0', grad_fn=<NllLossBackward>)
[4258] Train Loss:   tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[4259] Train Loss:   tensor(0.6949, device='cuda:0', grad_fn=<NllLossBackward>)
[4260] Train Loss:   tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[4261] Train Loss:   tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[4262] Train Loss:   tensor(0.7220, device='cuda:0', grad_fn=<NllLossBackward>)
[4263] Train Loss:   tensor(0.6796, device='cuda:0', grad_fn=<NllLossBackward>)
[4264] Train Loss:   tensor(0.6861, device='cuda:0', grad_fn=<NllLossBackward>)
[4265] Train Loss:   tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[4266] Train Loss:   tensor(0.6684, device='cuda:0', grad_fn=<NllLossBackward>)
[4267] Train Loss:   tensor(0.6809, device='cuda:0', grad_fn=<NllLossBackward>)
[4268] Train Loss:   tensor(0.6990, device='cuda:0', grad_fn=<NllLossBackward>)
[4269] Train Loss:   tensor(0.7223, device='cuda:0', grad_fn=<NllLossBackward>)
[4270] Train Loss:   tensor(0.6869, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4271] Train Loss:   tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[4272] Train Loss:   tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[4273] Train Loss:   tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[4274] Train Loss:   tensor(0.7059, device='cuda:0', grad_fn=<NllLossBackward>)
[4275] Train Loss:   tensor(0.7030, device='cuda:0', grad_fn=<NllLossBackward>)
[4276] Train Loss:   tensor(0.6794, device='cuda:0', grad_fn=<NllLossBackward>)
[4277] Train Loss:   tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[4278] Train Loss:   tensor(0.6797, device='cuda:0', grad_fn=<NllLossBackward>)
[4279] Train Loss:   tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[4280] Train Loss:   tensor(0.6799, device='cuda:0', grad_fn=<NllLossBackward>)
[4281] Train Loss:   tensor(0.6849, device='cuda:0', grad_fn=<NllLossBackward>)
[4282] Train Loss:   tensor(0.6772, device='cuda:0', grad_fn=<NllLossBackward>)
[4283] Train Loss:   tensor(0.6744, device='cuda:0', grad_fn=<NllLossBackward>)
[4284] Train Loss:   tensor(0.6901, device='cuda:0', grad_fn=<NllLossBackward>)
[4285] Train Loss:   tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[4286] Train Loss:   tensor(0.6931, device='cuda:0', grad_fn=<NllLossBackward>)
[4287] Train Loss:   tensor(0.6800, device='cuda:0', grad_fn=<NllLossBackward>)
[4288] Train Loss:   tensor(0.6968, device='cuda:0', grad_fn=<NllLossBackward>)
[4289] Train Loss:   tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[4290] Train Loss:   tensor(0.6695, device='cuda:0', grad_fn=<NllLossBackward>)
[4291] Train Loss:   tensor(0.6854, device='cuda:0', grad_fn=<NllLossBackward>)
[4292] Train Loss:   tensor(0.7025, device='cuda:0', grad_fn=<NllLossBackward>)
[4293] Train Loss:   tensor(0.6885, device='cuda:0', grad_fn=<NllLossBackward>)
[4294] Train Loss:   tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[4295] Train Loss:   tensor(0.6847, device='cuda:0', grad_fn=<NllLossBackward>)
[4296] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[4297] Train Loss:   tensor(0.6838, device='cuda:0', grad_fn=<NllLossBackward>)
[4298] Train Loss:   tensor(0.6710, device='cuda:0', grad_fn=<NllLossBackward>)
[4299] Train Loss:   tensor(0.6921, device='cuda:0', grad_fn=<NllLossBackward>)
[4300] Train Loss:   tensor(0.6746, device='cuda:0', grad_fn=<NllLossBackward>)
[4301] Train Loss:   tensor(0.6779, device='cuda:0', grad_fn=<NllLossBackward>)
[4302] Train Loss:   tensor(0.6988, device='cuda:0', grad_fn=<NllLossBackward>)
[4303] Train Loss:   tensor(0.6336, device='cuda:0', grad_fn=<NllLossBackward>)
[4304] Train Loss:   tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[4305] Train Loss:   tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[4306] Train Loss:   tensor(0.7077, device='cuda:0', grad_fn=<NllLossBackward>)
[4307] Train Loss:   tensor(0.6689, device='cuda:0', grad_fn=<NllLossBackward>)
[4308] Train Loss:   tensor(0.6908, device='cuda:0', grad_fn=<NllLossBackward>)
[4309] Train Loss:   tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[4310] Train Loss:   tensor(0.6793, device='cuda:0', grad_fn=<NllLossBackward>)
[4311] Train Loss:   tensor(0.6843, device='cuda:0', grad_fn=<NllLossBackward>)
[4312] Train Loss:   tensor(0.7162, device='cuda:0', grad_fn=<NllLossBackward>)
[4313] Train Loss:   tensor(0.6694, device='cuda:0', grad_fn=<NllLossBackward>)
[4314] Train Loss:   tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[4315] Train Loss:   tensor(0.6644, device='cuda:0', grad_fn=<NllLossBackward>)
[4316] Train Loss:   tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[4317] Train Loss:   tensor(0.6989, device='cuda:0', grad_fn=<NllLossBackward>)
[4318] Train Loss:   tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4319] Train Loss:   tensor(0.7157, device='cuda:0', grad_fn=<NllLossBackward>)
[4320] Train Loss:   tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[4321] Train Loss:   tensor(0.6939, device='cuda:0', grad_fn=<NllLossBackward>)
[4322] Train Loss:   tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[4323] Train Loss:   tensor(0.6758, device='cuda:0', grad_fn=<NllLossBackward>)
[4324] Train Loss:   tensor(0.7269, device='cuda:0', grad_fn=<NllLossBackward>)
[4325] Train Loss:   tensor(0.6947, device='cuda:0', grad_fn=<NllLossBackward>)
[4326] Train Loss:   tensor(0.6845, device='cuda:0', grad_fn=<NllLossBackward>)
[4327] Train Loss:   tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[4328] Train Loss:   tensor(0.7009, device='cuda:0', grad_fn=<NllLossBackward>)
[4329] Train Loss:   tensor(0.6580, device='cuda:0', grad_fn=<NllLossBackward>)
[4330] Train Loss:   tensor(0.6888, device='cuda:0', grad_fn=<NllLossBackward>)
[4331] Train Loss:   tensor(0.6810, device='cuda:0', grad_fn=<NllLossBackward>)
[4332] Train Loss:   tensor(0.6783, device='cuda:0', grad_fn=<NllLossBackward>)
[4333] Train Loss:   tensor(0.6944, device='cuda:0', grad_fn=<NllLossBackward>)
[4334] Train Loss:   tensor(0.6659, device='cuda:0', grad_fn=<NllLossBackward>)
[4335] Train Loss:   tensor(0.7228, device='cuda:0', grad_fn=<NllLossBackward>)
[4336] Train Loss:   tensor(0.6580, device='cuda:0', grad_fn=<NllLossBackward>)
[4337] Train Loss:   tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[4338] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[4339] Train Loss:   tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[4340] Train Loss:   tensor(0.6991, device='cuda:0', grad_fn=<NllLossBackward>)
[4341] Train Loss:   tensor(0.6826, device='cuda:0', grad_fn=<NllLossBackward>)
[4342] Train Loss:   tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[4343] Train Loss:   tensor(0.6934, device='cuda:0', grad_fn=<NllLossBackward>)
[4344] Train Loss:   tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
[4345] Train Loss:   tensor(0.6987, device='cuda:0', grad_fn=<NllLossBackward>)
[4346] Train Loss:   tensor(0.6438, device='cuda:0', grad_fn=<NllLossBackward>)
[4347] Train Loss:   tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[4348] Train Loss:   tensor(0.6554, device='cuda:0', grad_fn=<NllLossBackward>)
[4349] Train Loss:   tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[4350] Train Loss:   tensor(0.6753, device='cuda:0', grad_fn=<NllLossBackward>)
[4351] Train Loss:   tensor(0.6666, device='cuda:0', grad_fn=<NllLossBackward>)
[4352] Train Loss:   tensor(0.6709, device='cuda:0', grad_fn=<NllLossBackward>)
[4353] Train Loss:   tensor(0.6634, device='cuda:0', grad_fn=<NllLossBackward>)
[4354] Train Loss:   tensor(0.7143, device='cuda:0', grad_fn=<NllLossBackward>)
[4355] Train Loss:   tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[4356] Train Loss:   tensor(0.7067, device='cuda:0', grad_fn=<NllLossBackward>)
[4357] Train Loss:   tensor(0.6929, device='cuda:0', grad_fn=<NllLossBackward>)
[4358] Train Loss:   tensor(0.6872, device='cuda:0', grad_fn=<NllLossBackward>)
[4359] Train Loss:   tensor(0.6883, device='cuda:0', grad_fn=<NllLossBackward>)
[4360] Train Loss:   tensor(0.6983, device='cuda:0', grad_fn=<NllLossBackward>)
[4361] Train Loss:   tensor(0.7103, device='cuda:0', grad_fn=<NllLossBackward>)
[4362] Train Loss:   tensor(0.7058, device='cuda:0', grad_fn=<NllLossBackward>)
[4363] Train Loss:   tensor(0.7033, device='cuda:0', grad_fn=<NllLossBackward>)
[4364] Train Loss:   tensor(0.6905, device='cuda:0', grad_fn=<NllLossBackward>)
[4365] Train Loss:   tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[4366] Train Loss:   tensor(0.6835, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4367] Train Loss:    tensor(0.6894, device='cuda:0', grad_fn=<NllLossBackward>)
[4368] Train Loss:    tensor(0.6865, device='cuda:0', grad_fn=<NllLossBackward>)
[4369] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[4370] Train Loss:    tensor(0.7048, device='cuda:0', grad_fn=<NllLossBackward>)
[4371] Train Loss:    tensor(0.6970, device='cuda:0', grad_fn=<NllLossBackward>)
[4372] Train Loss:    tensor(0.6760, device='cuda:0', grad_fn=<NllLossBackward>)
[4373] Train Loss:    tensor(0.7353, device='cuda:0', grad_fn=<NllLossBackward>)
[4374] Train Loss:    tensor(0.6741, device='cuda:0', grad_fn=<NllLossBackward>)
[4375] Train Loss:    tensor(0.6834, device='cuda:0', grad_fn=<NllLossBackward>)
[4376] Train Loss:    tensor(0.6880, device='cuda:0', grad_fn=<NllLossBackward>)
[4377] Train Loss:    tensor(0.7174, device='cuda:0', grad_fn=<NllLossBackward>)
[4378] Train Loss:    tensor(0.6777, device='cuda:0', grad_fn=<NllLossBackward>)
[4379] Train Loss:    tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[4380] Train Loss:    tensor(0.6766, device='cuda:0', grad_fn=<NllLossBackward>)
[4381] Train Loss:    tensor(0.7086, device='cuda:0', grad_fn=<NllLossBackward>)
[4382] Train Loss:    tensor(0.6825, device='cuda:0', grad_fn=<NllLossBackward>)
[4383] Train Loss:    tensor(0.6953, device='cuda:0', grad_fn=<NllLossBackward>)
[4384] Train Loss:    tensor(0.7240, device='cuda:0', grad_fn=<NllLossBackward>)
[4385] Train Loss:    tensor(0.6889, device='cuda:0', grad_fn=<NllLossBackward>)
[4386] Train Loss:    tensor(0.6963, device='cuda:0', grad_fn=<NllLossBackward>)
[4387] Train Loss:    tensor(0.6805, device='cuda:0', grad_fn=<NllLossBackward>)
[4388] Train Loss:    tensor(0.6974, device='cuda:0', grad_fn=<NllLossBackward>)
[4389] Train Loss:    tensor(0.6965, device='cuda:0', grad_fn=<NllLossBackward>)
[4390] Train Loss:    tensor(0.6656, device='cuda:0', grad_fn=<NllLossBackward>)
[4391] Train Loss:    tensor(0.7359, device='cuda:0', grad_fn=<NllLossBackward>)
[4392] Train Loss:    tensor(0.6982, device='cuda:0', grad_fn=<NllLossBackward>)
[4393] Train Loss:    tensor(0.6782, device='cuda:0', grad_fn=<NllLossBackward>)
[4394] Train Loss:    tensor(0.7137, device='cuda:0', grad_fn=<NllLossBackward>)
[4395] Train Loss:    tensor(0.7083, device='cuda:0', grad_fn=<NllLossBackward>)
[4396] Train Loss:    tensor(0.6857, device='cuda:0', grad_fn=<NllLossBackward>)
[4397] Train Loss:    tensor(0.6645, device='cuda:0', grad_fn=<NllLossBackward>)
[4398] Train Loss:    tensor(0.6858, device='cuda:0', grad_fn=<NllLossBackward>)
[4399] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
[4400] Train Loss:    tensor(0.6948, device='cuda:0', grad_fn=<NllLossBackward>)
[4401] Train Loss:    tensor(0.6848, device='cuda:0', grad_fn=<NllLossBackward>)
[4402] Train Loss:    tensor(0.6926, device='cuda:0', grad_fn=<NllLossBackward>)
[4403] Train Loss:    tensor(0.6778, device='cuda:0', grad_fn=<NllLossBackward>)
[4404] Train Loss:    tensor(0.6643, device='cuda:0', grad_fn=<NllLossBackward>)
[4405] Train Loss:    tensor(0.6517, device='cuda:0', grad_fn=<NllLossBackward>)
[4406] Train Loss:    tensor(0.6806, device='cuda:0', grad_fn=<NllLossBackward>)
[4407] Train Loss:    tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[4408] Train Loss:    tensor(0.7197, device='cuda:0', grad_fn=<NllLossBackward>)
[4409] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[4410] Train Loss:    tensor(0.6625, device='cuda:0', grad_fn=<NllLossBackward>)
[4411] Train Loss:    tensor(0.6749, device='cuda:0', grad_fn=<NllLossBackward>)
[4412] Train Loss:    tensor(0.6587, device='cuda:0', grad_fn=<NllLossBackward>)
[4413] Train Loss:    tensor(0.7168, device='cuda:0', grad_fn=<NllLossBackward>)
[4414] Train Loss:    tensor(0.6745, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4415] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[4416] Train Loss:    tensor(0.6841, device='cuda:0', grad_fn=<NllLossBackward>)
[4417] Train Loss:    tensor(0.6495, device='cuda:0', grad_fn=<NllLossBackward>)
[4418] Train Loss:    tensor(0.6833, device='cuda:0', grad_fn=<NllLossBackward>)
[4419] Train Loss:    tensor(0.7120, device='cuda:0', grad_fn=<NllLossBackward>)
[4420] Train Loss:    tensor(0.7082, device='cuda:0', grad_fn=<NllLossBackward>)
[4421] Train Loss:    tensor(0.6786, device='cuda:0', grad_fn=<NllLossBackward>)
[4422] Train Loss:    tensor(0.6627, device='cuda:0', grad_fn=<NllLossBackward>)
[4423] Train Loss:    tensor(0.6956, device='cuda:0', grad_fn=<NllLossBackward>)
[4424] Train Loss:    tensor(0.6501, device='cuda:0', grad_fn=<NllLossBackward>)
[4425] Train Loss:    tensor(0.6757, device='cuda:0', grad_fn=<NllLossBackward>)
[4426] Train Loss:    tensor(0.6981, device='cuda:0', grad_fn=<NllLossBackward>)
[4427] Train Loss:    tensor(0.6896, device='cuda:0', grad_fn=<NllLossBackward>)
[4428] Train Loss:    tensor(0.6975, device='cuda:0', grad_fn=<NllLossBackward>)
[4429] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[4430] Train Loss:    tensor(0.6898, device='cuda:0', grad_fn=<NllLossBackward>)
[4431] Train Loss:    tensor(0.6876, device='cuda:0', grad_fn=<NllLossBackward>)
[4432] Train Loss:    tensor(0.6651, device='cuda:0', grad_fn=<NllLossBackward>)
[4433] Train Loss:    tensor(0.6733, device='cuda:0', grad_fn=<NllLossBackward>)
[4434] Train Loss:    tensor(0.6822, device='cuda:0', grad_fn=<NllLossBackward>)
[4435] Train Loss:    tensor(0.6907, device='cuda:0', grad_fn=<NllLossBackward>)
[4436] Train Loss:    tensor(0.7158, device='cuda:0', grad_fn=<NllLossBackward>)
[4437] Train Loss:    tensor(0.6964, device='cuda:0', grad_fn=<NllLossBackward>)
[4438] Train Loss:    tensor(0.7291, device='cuda:0', grad_fn=<NllLossBackward>)
[4439] Train Loss:    tensor(0.6859, device='cuda:0', grad_fn=<NllLossBackward>)
[4440] Train Loss:    tensor(0.7180, device='cuda:0', grad_fn=<NllLossBackward>)
[4441] Train Loss:    tensor(0.6750, device='cuda:0', grad_fn=<NllLossBackward>)
[4442] Train Loss:    tensor(0.6910, device='cuda:0', grad_fn=<NllLossBackward>)
[4443] Train Loss:    tensor(0.6821, device='cuda:0', grad_fn=<NllLossBackward>)
[4444] Train Loss:    tensor(0.6837, device='cuda:0', grad_fn=<NllLossBackward>)
[4445] Train Loss:    tensor(0.7032, device='cuda:0', grad_fn=<NllLossBackward>)
[4446] Train Loss:    tensor(0.6913, device='cuda:0', grad_fn=<NllLossBackward>)
[4447] Train Loss:    tensor(0.6960, device='cuda:0', grad_fn=<NllLossBackward>)
[4448] Train Loss:    tensor(0.7259, device='cuda:0', grad_fn=<NllLossBackward>)
[4449] Train Loss:    tensor(0.6637, device='cuda:0', grad_fn=<NllLossBackward>)
[4450] Train Loss:    tensor(0.7106, device='cuda:0', grad_fn=<NllLossBackward>)
[4451] Train Loss:    tensor(0.7299, device='cuda:0', grad_fn=<NllLossBackward>)
[4452] Train Loss:    tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[4453] Train Loss:    tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[4454] Train Loss:    tensor(0.7042, device='cuda:0', grad_fn=<NllLossBackward>)
[4455] Train Loss:    tensor(0.6723, device='cuda:0', grad_fn=<NllLossBackward>)
[4456] Train Loss:    tensor(0.7111, device='cuda:0', grad_fn=<NllLossBackward>)
[4457] Train Loss:    tensor(0.6771, device='cuda:0', grad_fn=<NllLossBackward>)
[4458] Train Loss:    tensor(0.7139, device='cuda:0', grad_fn=<NllLossBackward>)
[4459] Train Loss:    tensor(0.6870, device='cuda:0', grad_fn=<NllLossBackward>)
[4460] Train Loss:    tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[4461] Train Loss:    tensor(0.6738, device='cuda:0', grad_fn=<NllLossBackward>)
[4462] Train Loss:    tensor(0.6820, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4463] Train Loss:   tensor(0.6725, device='cuda:0', grad_fn=<NllLossBackward>)
[4464] Train Loss:   tensor(0.6993, device='cuda:0', grad_fn=<NllLossBackward>)
[4465] Train Loss:   tensor(0.6713, device='cuda:0', grad_fn=<NllLossBackward>)
[4466] Train Loss:   tensor(0.6731, device='cuda:0', grad_fn=<NllLossBackward>)
[4467] Train Loss:   tensor(0.6911, device='cuda:0', grad_fn=<NllLossBackward>)
[4468] Train Loss:   tensor(0.7074, device='cuda:0', grad_fn=<NllLossBackward>)
[4469] Train Loss:   tensor(0.6718, device='cuda:0', grad_fn=<NllLossBackward>)
[4470] Train Loss:   tensor(0.6795, device='cuda:0', grad_fn=<NllLossBackward>)
[4471] Train Loss:   tensor(0.6915, device='cuda:0', grad_fn=<NllLossBackward>)
[4472] Train Loss:   tensor(0.7400, device='cuda:0', grad_fn=<NllLossBackward>)
[4473] Train Loss:   tensor(0.7159, device='cuda:0', grad_fn=<NllLossBackward>)
[4474] Train Loss:   tensor(0.6673, device='cuda:0', grad_fn=<NllLossBackward>)
[4475] Train Loss:   tensor(0.7010, device='cuda:0', grad_fn=<NllLossBackward>)
[4476] Train Loss:   tensor(0.7063, device='cuda:0', grad_fn=<NllLossBackward>)
[4477] Train Loss:   tensor(0.6999, device='cuda:0', grad_fn=<NllLossBackward>)
[4478] Train Loss:   tensor(0.6631, device='cuda:0', grad_fn=<NllLossBackward>)
[4479] Train Loss:   tensor(0.6930, device='cuda:0', grad_fn=<NllLossBackward>)
[4480] Train Loss:   tensor(0.7084, device='cuda:0', grad_fn=<NllLossBackward>)
[4481] Train Loss:   tensor(0.6622, device='cuda:0', grad_fn=<NllLossBackward>)
[4482] Train Loss:   tensor(0.6788, device='cuda:0', grad_fn=<NllLossBackward>)
[4483] Train Loss:   tensor(0.6818, device='cuda:0', grad_fn=<NllLossBackward>)
[4484] Train Loss:   tensor(0.7246, device='cuda:0', grad_fn=<NllLossBackward>)
[4485] Train Loss:   tensor(0.7001, device='cuda:0', grad_fn=<NllLossBackward>)
[4486] Train Loss:   tensor(0.6952, device='cuda:0', grad_fn=<NllLossBackward>)
[4487] Train Loss:   tensor(0.7181, device='cuda:0', grad_fn=<NllLossBackward>)
[4488] Train Loss:   tensor(0.7097, device='cuda:0', grad_fn=<NllLossBackward>)
[4489] Train Loss:   tensor(0.6619, device='cuda:0', grad_fn=<NllLossBackward>)
[4490] Train Loss:   tensor(0.6655, device='cuda:0', grad_fn=<NllLossBackward>)
[4491] Train Loss:   tensor(0.6808, device='cuda:0', grad_fn=<NllLossBackward>)
[4492] Train Loss:   tensor(0.6832, device='cuda:0', grad_fn=<NllLossBackward>)
[4493] Train Loss:   tensor(0.7215, device='cuda:0', grad_fn=<NllLossBackward>)
[4494] Train Loss:   tensor(0.6711, device='cuda:0', grad_fn=<NllLossBackward>)
[4495] Train Loss:   tensor(0.6877, device='cuda:0', grad_fn=<NllLossBackward>)
[4496] Train Loss:   tensor(0.6868, device='cuda:0', grad_fn=<NllLossBackward>)
[4497] Train Loss:   tensor(0.6769, device='cuda:0', grad_fn=<NllLossBackward>)
[4498] Train Loss:   tensor(0.6996, device='cuda:0', grad_fn=<NllLossBackward>)
[4499] Train Loss:   tensor(0.6781, device='cuda:0', grad_fn=<NllLossBackward>)
[4500] Train Loss:   tensor(0.6514, device='cuda:0', grad_fn=<NllLossBackward>)
[4501] Train Loss:   tensor(0.6871, device='cuda:0', grad_fn=<NllLossBackward>)
[4502] Train Loss:   tensor(0.6900, device='cuda:0', grad_fn=<NllLossBackward>)
[4503] Train Loss:   tensor(0.6890, device='cuda:0', grad_fn=<NllLossBackward>)
[4504] Train Loss:   tensor(0.6722, device='cuda:0', grad_fn=<NllLossBackward>)
[4505] Train Loss:   tensor(0.7110, device='cuda:0', grad_fn=<NllLossBackward>)
[4506] Train Loss:   tensor(0.6736, device='cuda:0', grad_fn=<NllLossBackward>)
[4507] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[4508] Train Loss:   tensor(0.6850, device='cuda:0', grad_fn=<NllLossBackward>)
[4509] Train Loss:   tensor(0.6895, device='cuda:0', grad_fn=<NllLossBackward>)
[4510] Train Loss:   tensor(0.6525, device='cuda:0', grad_fn=<NllLossBackward>)
```

```
[4511] Train Loss:  tensor(0.6973, device='cuda:0', grad_fn=<NllLossBackward>)
Training Loss:  0.6888509186262146 Time:  2517.247688293457 s
Training Accuaracy:  78.70972004342588 %
[0] Valid Loss:  tensor(0.6958, device='cuda:0')
[1] Valid Loss:  tensor(0.8811, device='cuda:0')
[2] Valid Loss:  tensor(0.6665, device='cuda:0')
[3] Valid Loss:  tensor(0.6099, device='cuda:0')
[4] Valid Loss:  tensor(0.7065, device='cuda:0')
[5] Valid Loss:  tensor(0.6996, device='cuda:0')
[6] Valid Loss:  tensor(0.7201, device='cuda:0')
[7] Valid Loss:  tensor(0.8241, device='cuda:0')
[8] Valid Loss:  tensor(0.7997, device='cuda:0')
[9] Valid Loss:  tensor(0.8065, device='cuda:0')
[10] Valid Loss:  tensor(0.6860, device='cuda:0')
[11] Valid Loss:  tensor(0.6595, device='cuda:0')
[12] Valid Loss:  tensor(1.1094, device='cuda:0')
[13] Valid Loss:  tensor(0.9812, device='cuda:0')
[14] Valid Loss:  tensor(0.6063, device='cuda:0')
[15] Valid Loss:  tensor(0.7393, device='cuda:0')
[16] Valid Loss:  tensor(0.6060, device='cuda:0')
[17] Valid Loss:  tensor(0.7679, device='cuda:0')
[18] Valid Loss:  tensor(0.6543, device='cuda:0')
[19] Valid Loss:  tensor(0.6165, device='cuda:0')
[20] Valid Loss:  tensor(0.6625, device='cuda:0')
[21] Valid Loss:  tensor(0.5785, device='cuda:0')
[22] Valid Loss:  tensor(0.5637, device='cuda:0')
[23] Valid Loss:  tensor(0.8206, device='cuda:0')
[24] Valid Loss:  tensor(0.6828, device='cuda:0')
[25] Valid Loss:  tensor(0.9639, device='cuda:0')
[26] Valid Loss:  tensor(0.5354, device='cuda:0')
[27] Valid Loss:  tensor(0.9110, device='cuda:0')
[28] Valid Loss:  tensor(0.6651, device='cuda:0')
[29] Valid Loss:  tensor(0.6799, device='cuda:0')
[30] Valid Loss:  tensor(0.6267, device='cuda:0')
[31] Valid Loss:  tensor(0.7337, device='cuda:0')
[32] Valid Loss:  tensor(0.8196, device='cuda:0')
[33] Valid Loss:  tensor(0.7911, device='cuda:0')
[34] Valid Loss:  tensor(0.7813, device='cuda:0')
[35] Valid Loss:  tensor(0.8689, device='cuda:0')
[36] Valid Loss:  tensor(0.6681, device='cuda:0')
[37] Valid Loss:  tensor(1.0375, device='cuda:0')
[38] Valid Loss:  tensor(0.8184, device='cuda:0')
[39] Valid Loss:  tensor(0.6456, device='cuda:0')
[40] Valid Loss:  tensor(0.7468, device='cuda:0')
[41] Valid Loss:  tensor(0.8375, device='cuda:0')
[42] Valid Loss:  tensor(0.7687, device='cuda:0')
[43] Valid Loss:  tensor(0.5613, device='cuda:0')
[44] Valid Loss:  tensor(0.6704, device='cuda:0')
```

```
[45] Valid Loss:   tensor(0.5887, device='cuda:0')
[46] Valid Loss:   tensor(0.8612, device='cuda:0')
[47] Valid Loss:   tensor(0.7859, device='cuda:0')
[48] Valid Loss:   tensor(0.6878, device='cuda:0')
[49] Valid Loss:   tensor(0.7612, device='cuda:0')
[50] Valid Loss:   tensor(0.6885, device='cuda:0')
[51] Valid Loss:   tensor(0.6490, device='cuda:0')
[52] Valid Loss:   tensor(0.5946, device='cuda:0')
[53] Valid Loss:   tensor(0.8930, device='cuda:0')
[54] Valid Loss:   tensor(0.6061, device='cuda:0')
[55] Valid Loss:   tensor(0.5536, device='cuda:0')
[56] Valid Loss:   tensor(0.8548, device='cuda:0')
[57] Valid Loss:   tensor(0.7855, device='cuda:0')
[58] Valid Loss:   tensor(0.7608, device='cuda:0')
[59] Valid Loss:   tensor(0.6571, device='cuda:0')
[60] Valid Loss:   tensor(1.0200, device='cuda:0')
[61] Valid Loss:   tensor(0.6872, device='cuda:0')
[62] Valid Loss:   tensor(0.8671, device='cuda:0')
[63] Valid Loss:   tensor(0.6684, device='cuda:0')
[64] Valid Loss:   tensor(0.7372, device='cuda:0')
[65] Valid Loss:   tensor(0.8626, device='cuda:0')
[66] Valid Loss:   tensor(0.7116, device='cuda:0')
[67] Valid Loss:   tensor(0.7413, device='cuda:0')
[68] Valid Loss:   tensor(0.7252, device='cuda:0')
[69] Valid Loss:   tensor(0.5526, device='cuda:0')
[70] Valid Loss:   tensor(0.5798, device='cuda:0')
[71] Valid Loss:   tensor(0.8419, device='cuda:0')
[72] Valid Loss:   tensor(0.7504, device='cuda:0')
[73] Valid Loss:   tensor(0.6566, device='cuda:0')
[74] Valid Loss:   tensor(0.7643, device='cuda:0')
[75] Valid Loss:   tensor(0.8485, device='cuda:0')
[76] Valid Loss:   tensor(0.7761, device='cuda:0')
[77] Valid Loss:   tensor(0.6849, device='cuda:0')
[78] Valid Loss:   tensor(0.7821, device='cuda:0')
[79] Valid Loss:   tensor(0.6262, device='cuda:0')
[80] Valid Loss:   tensor(0.7223, device='cuda:0')
[81] Valid Loss:   tensor(0.6268, device='cuda:0')
[82] Valid Loss:   tensor(0.8519, device='cuda:0')
[83] Valid Loss:   tensor(0.8162, device='cuda:0')
[84] Valid Loss:   tensor(0.5918, device='cuda:0')
[85] Valid Loss:   tensor(0.7242, device='cuda:0')
[86] Valid Loss:   tensor(0.7178, device='cuda:0')
[87] Valid Loss:   tensor(0.8070, device='cuda:0')
[88] Valid Loss:   tensor(0.5803, device='cuda:0')
[89] Valid Loss:   tensor(0.8066, device='cuda:0')
[90] Valid Loss:   tensor(0.9206, device='cuda:0')
[91] Valid Loss:   tensor(0.7286, device='cuda:0')
[92] Valid Loss:   tensor(0.7761, device='cuda:0')
```

```
[93] Valid Loss:  tensor(0.7692, device='cuda:0')
[94] Valid Loss:  tensor(0.8439, device='cuda:0')
[95] Valid Loss:  tensor(0.6844, device='cuda:0')
[96] Valid Loss:  tensor(0.8536, device='cuda:0')
[97] Valid Loss:  tensor(0.7938, device='cuda:0')
[98] Valid Loss:  tensor(0.8058, device='cuda:0')
[99] Valid Loss:  tensor(0.6076, device='cuda:0')
[100] Valid Loss:  tensor(0.7547, device='cuda:0')
[101] Valid Loss:  tensor(0.7257, device='cuda:0')
[102] Valid Loss:  tensor(0.9081, device='cuda:0')
[103] Valid Loss:  tensor(0.8569, device='cuda:0')
[104] Valid Loss:  tensor(0.6510, device='cuda:0')
[105] Valid Loss:  tensor(0.6652, device='cuda:0')
[106] Valid Loss:  tensor(0.6154, device='cuda:0')
[107] Valid Loss:  tensor(0.7473, device='cuda:0')
[108] Valid Loss:  tensor(0.8451, device='cuda:0')
[109] Valid Loss:  tensor(0.7520, device='cuda:0')
[110] Valid Loss:  tensor(0.7936, device='cuda:0')
[111] Valid Loss:  tensor(0.7646, device='cuda:0')
[112] Valid Loss:  tensor(0.8174, device='cuda:0')
[113] Valid Loss:  tensor(0.7562, device='cuda:0')
[114] Valid Loss:  tensor(0.6984, device='cuda:0')
[115] Valid Loss:  tensor(0.8262, device='cuda:0')
[116] Valid Loss:  tensor(0.8189, device='cuda:0')
[117] Valid Loss:  tensor(0.8393, device='cuda:0')
[118] Valid Loss:  tensor(0.6483, device='cuda:0')
[119] Valid Loss:  tensor(0.6004, device='cuda:0')
[120] Valid Loss:  tensor(0.9919, device='cuda:0')
[121] Valid Loss:  tensor(0.6678, device='cuda:0')
[122] Valid Loss:  tensor(0.7248, device='cuda:0')
[123] Valid Loss:  tensor(0.6563, device='cuda:0')
[124] Valid Loss:  tensor(0.8733, device='cuda:0')
[125] Valid Loss:  tensor(0.9398, device='cuda:0')
[126] Valid Loss:  tensor(0.5669, device='cuda:0')
[127] Valid Loss:  tensor(0.6124, device='cuda:0')
[128] Valid Loss:  tensor(0.8434, device='cuda:0')
[129] Valid Loss:  tensor(0.6708, device='cuda:0')
[130] Valid Loss:  tensor(0.8055, device='cuda:0')
[131] Valid Loss:  tensor(0.6861, device='cuda:0')
[132] Valid Loss:  tensor(0.6971, device='cuda:0')
[133] Valid Loss:  tensor(0.7574, device='cuda:0')
[134] Valid Loss:  tensor(0.6492, device='cuda:0')
[135] Valid Loss:  tensor(0.5753, device='cuda:0')
[136] Valid Loss:  tensor(0.6013, device='cuda:0')
[137] Valid Loss:  tensor(0.9703, device='cuda:0')
[138] Valid Loss:  tensor(0.8402, device='cuda:0')
[139] Valid Loss:  tensor(1.0228, device='cuda:0')
[140] Valid Loss:  tensor(0.8148, device='cuda:0')
```

```
[141] Valid Loss:  tensor(0.6208, device='cuda:0')
[142] Valid Loss:  tensor(0.9083, device='cuda:0')
[143] Valid Loss:  tensor(0.6401, device='cuda:0')
[144] Valid Loss:  tensor(0.7502, device='cuda:0')
[145] Valid Loss:  tensor(0.9429, device='cuda:0')
[146] Valid Loss:  tensor(0.6711, device='cuda:0')
[147] Valid Loss:  tensor(0.7916, device='cuda:0')
[148] Valid Loss:  tensor(0.6599, device='cuda:0')
[149] Valid Loss:  tensor(0.7689, device='cuda:0')
[150] Valid Loss:  tensor(0.7346, device='cuda:0')
[151] Valid Loss:  tensor(0.7267, device='cuda:0')
[152] Valid Loss:  tensor(0.6500, device='cuda:0')
[153] Valid Loss:  tensor(0.6629, device='cuda:0')
[154] Valid Loss:  tensor(0.6871, device='cuda:0')
[155] Valid Loss:  tensor(0.7598, device='cuda:0')
[156] Valid Loss:  tensor(0.8186, device='cuda:0')
[157] Valid Loss:  tensor(0.6114, device='cuda:0')
[158] Valid Loss:  tensor(0.7233, device='cuda:0')
[159] Valid Loss:  tensor(0.7107, device='cuda:0')
[160] Valid Loss:  tensor(0.5952, device='cuda:0')
[161] Valid Loss:  tensor(0.8297, device='cuda:0')
[162] Valid Loss:  tensor(0.7628, device='cuda:0')
[163] Valid Loss:  tensor(0.6248, device='cuda:0')
[164] Valid Loss:  tensor(0.9695, device='cuda:0')
[165] Valid Loss:  tensor(0.7034, device='cuda:0')
[166] Valid Loss:  tensor(0.7192, device='cuda:0')
[167] Valid Loss:  tensor(0.6056, device='cuda:0')
[168] Valid Loss:  tensor(0.6653, device='cuda:0')
[169] Valid Loss:  tensor(0.6478, device='cuda:0')
[170] Valid Loss:  tensor(0.6904, device='cuda:0')
[171] Valid Loss:  tensor(0.5760, device='cuda:0')
[172] Valid Loss:  tensor(0.7675, device='cuda:0')
[173] Valid Loss:  tensor(0.6356, device='cuda:0')
[174] Valid Loss:  tensor(0.7557, device='cuda:0')
[175] Valid Loss:  tensor(0.6408, device='cuda:0')
[176] Valid Loss:  tensor(0.5787, device='cuda:0')
[177] Valid Loss:  tensor(0.7521, device='cuda:0')
[178] Valid Loss:  tensor(0.7617, device='cuda:0')
[179] Valid Loss:  tensor(0.7745, device='cuda:0')
[180] Valid Loss:  tensor(0.8089, device='cuda:0')
[181] Valid Loss:  tensor(0.7775, device='cuda:0')
[182] Valid Loss:  tensor(0.7355, device='cuda:0')
[183] Valid Loss:  tensor(0.7375, device='cuda:0')
[184] Valid Loss:  tensor(0.8877, device='cuda:0')
[185] Valid Loss:  tensor(0.8100, device='cuda:0')
[186] Valid Loss:  tensor(0.7039, device='cuda:0')
[187] Valid Loss:  tensor(0.8219, device='cuda:0')
[188] Valid Loss:  tensor(0.8823, device='cuda:0')
```
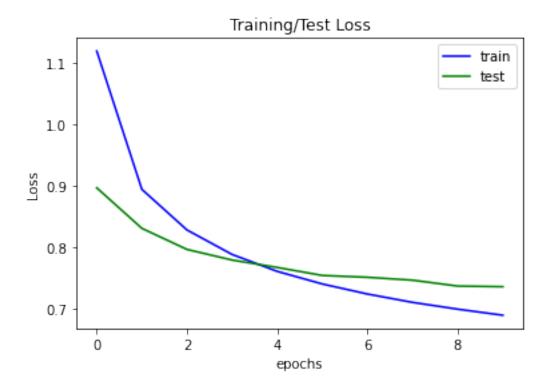
```
[189] Valid Loss:  tensor(0.7243, device='cuda:0')
[190] Valid Loss:  tensor(0.7311, device='cuda:0')
[191] Valid Loss:  tensor(0.6055, device='cuda:0')
[192] Valid Loss:  tensor(0.9992, device='cuda:0')
[193] Valid Loss:  tensor(0.7742, device='cuda:0')
[194] Valid Loss:  tensor(0.8225, device='cuda:0')
[195] Valid Loss:  tensor(0.8402, device='cuda:0')
[196] Valid Loss:  tensor(0.5966, device='cuda:0')
[197] Valid Loss:  tensor(0.7764, device='cuda:0')
[198] Valid Loss:  tensor(0.7369, device='cuda:0')
[199] Valid Loss:  tensor(0.7949, device='cuda:0')
[200] Valid Loss:  tensor(0.7301, device='cuda:0')
[201] Valid Loss:  tensor(0.6916, device='cuda:0')
[202] Valid Loss:  tensor(0.6529, device='cuda:0')
[203] Valid Loss:  tensor(0.8975, device='cuda:0')
[204] Valid Loss:  tensor(0.8286, device='cuda:0')
[205] Valid Loss:  tensor(0.8428, device='cuda:0')
[206] Valid Loss:  tensor(0.9414, device='cuda:0')
[207] Valid Loss:  tensor(0.7738, device='cuda:0')
[208] Valid Loss:  tensor(0.6033, device='cuda:0')
[209] Valid Loss:  tensor(0.7166, device='cuda:0')
[210] Valid Loss:  tensor(1.1225, device='cuda:0')
[211] Valid Loss:  tensor(0.7506, device='cuda:0')
[212] Valid Loss:  tensor(0.6011, device='cuda:0')
[213] Valid Loss:  tensor(0.7004, device='cuda:0')
[214] Valid Loss:  tensor(0.6956, device='cuda:0')
[215] Valid Loss:  tensor(0.8082, device='cuda:0')
[216] Valid Loss:  tensor(0.7474, device='cuda:0')
[217] Valid Loss:  tensor(0.7204, device='cuda:0')
[218] Valid Loss:  tensor(0.5834, device='cuda:0')
[219] Valid Loss:  tensor(0.6995, device='cuda:0')
[220] Valid Loss:  tensor(0.6658, device='cuda:0')
[221] Valid Loss:  tensor(0.7713, device='cuda:0')
[222] Valid Loss:  tensor(0.5844, device='cuda:0')
[223] Valid Loss:  tensor(0.6856, device='cuda:0')
[224] Valid Loss:  tensor(0.7982, device='cuda:0')
[225] Valid Loss:  tensor(0.5914, device='cuda:0')
[226] Valid Loss:  tensor(0.7393, device='cuda:0')
[227] Valid Loss:  tensor(0.6008, device='cuda:0')
[228] Valid Loss:  tensor(0.8104, device='cuda:0')
[229] Valid Loss:  tensor(0.6847, device='cuda:0')
[230] Valid Loss:  tensor(0.7898, device='cuda:0')
[231] Valid Loss:  tensor(0.6850, device='cuda:0')
[232] Valid Loss:  tensor(0.6003, device='cuda:0')
[233] Valid Loss:  tensor(0.6655, device='cuda:0')
[234] Valid Loss:  tensor(0.5892, device='cuda:0')
[235] Valid Loss:  tensor(0.6199, device='cuda:0')
[236] Valid Loss:  tensor(0.7438, device='cuda:0')
```

```
[237] Valid Loss:  tensor(0.7979, device='cuda:0')
[238] Valid Loss:  tensor(0.5175, device='cuda:0')
[239] Valid Loss:  tensor(0.7934, device='cuda:0')
[240] Valid Loss:  tensor(0.7022, device='cuda:0')
[241] Valid Loss:  tensor(0.6220, device='cuda:0')
[242] Valid Loss:  tensor(0.5204, device='cuda:0')
[243] Valid Loss:  tensor(1.1646, device='cuda:0')
[244] Valid Loss:  tensor(0.7650, device='cuda:0')
[245] Valid Loss:  tensor(0.6821, device='cuda:0')
[246] Valid Loss:  tensor(0.6377, device='cuda:0')
[247] Valid Loss:  tensor(0.6757, device='cuda:0')
[248] Valid Loss:  tensor(0.9341, device='cuda:0')
[249] Valid Loss:  tensor(0.6902, device='cuda:0')
[250] Valid Loss:  tensor(0.7705, device='cuda:0')
[251] Valid Loss:  tensor(0.8109, device='cuda:0')
[252] Valid Loss:  tensor(0.5873, device='cuda:0')
[253] Valid Loss:  tensor(0.6205, device='cuda:0')
[254] Valid Loss:  tensor(0.6240, device='cuda:0')
[255] Valid Loss:  tensor(0.7660, device='cuda:0')
[256] Valid Loss:  tensor(0.8026, device='cuda:0')
[257] Valid Loss:  tensor(0.6879, device='cuda:0')
[258] Valid Loss:  tensor(0.8634, device='cuda:0')
[259] Valid Loss:  tensor(0.7196, device='cuda:0')
[260] Valid Loss:  tensor(0.9207, device='cuda:0')
[261] Valid Loss:  tensor(0.7632, device='cuda:0')
[262] Valid Loss:  tensor(0.8826, device='cuda:0')
[263] Valid Loss:  tensor(0.8569, device='cuda:0')
[264] Valid Loss:  tensor(0.8096, device='cuda:0')
[265] Valid Loss:  tensor(0.8122, device='cuda:0')
[266] Valid Loss:  tensor(0.6831, device='cuda:0')
[267] Valid Loss:  tensor(0.7257, device='cuda:0')
[268] Valid Loss:  tensor(0.7024, device='cuda:0')
[269] Valid Loss:  tensor(0.6702, device='cuda:0')
[270] Valid Loss:  tensor(0.6310, device='cuda:0')
[271] Valid Loss:  tensor(0.9573, device='cuda:0')
[272] Valid Loss:  tensor(0.7234, device='cuda:0')
[273] Valid Loss:  tensor(0.8662, device='cuda:0')
[274] Valid Loss:  tensor(0.6886, device='cuda:0')
[275] Valid Loss:  tensor(0.7321, device='cuda:0')
[276] Valid Loss:  tensor(0.6435, device='cuda:0')
[277] Valid Loss:  tensor(0.6185, device='cuda:0')
[278] Valid Loss:  tensor(0.9004, device='cuda:0')
[279] Valid Loss:  tensor(0.7953, device='cuda:0')
[280] Valid Loss:  tensor(0.8850, device='cuda:0')
[281] Valid Loss:  tensor(0.6354, device='cuda:0')
[282] Valid Loss:  tensor(0.7763, device='cuda:0')
[283] Valid Loss:  tensor(0.8899, device='cuda:0')
[284] Valid Loss:  tensor(0.6722, device='cuda:0')
```

```
[285] Valid Loss:   tensor(0.5821, device='cuda:0')
[286] Valid Loss:   tensor(0.8023, device='cuda:0')
[287] Valid Loss:   tensor(0.9451, device='cuda:0')
[288] Valid Loss:   tensor(0.5606, device='cuda:0')
[289] Valid Loss:   tensor(0.7399, device='cuda:0')
[290] Valid Loss:   tensor(0.5877, device='cuda:0')
[291] Valid Loss:   tensor(0.7055, device='cuda:0')
[292] Valid Loss:   tensor(0.8359, device='cuda:0')
[293] Valid Loss:   tensor(0.6301, device='cuda:0')
[294] Valid Loss:   tensor(0.8665, device='cuda:0')
[295] Valid Loss:   tensor(0.6642, device='cuda:0')
[296] Valid Loss:   tensor(0.7324, device='cuda:0')
[297] Valid Loss:   tensor(0.8124, device='cuda:0')
[298] Valid Loss:   tensor(0.6688, device='cuda:0')
[299] Valid Loss:   tensor(0.7645, device='cuda:0')
[300] Valid Loss:   tensor(0.9159, device='cuda:0')
[301] Valid Loss:   tensor(0.5862, device='cuda:0')
[302] Valid Loss:   tensor(0.5620, device='cuda:0')
[303] Valid Loss:   tensor(0.6774, device='cuda:0')
[304] Valid Loss:   tensor(0.6463, device='cuda:0')
[305] Valid Loss:   tensor(0.7669, device='cuda:0')
[306] Valid Loss:   tensor(0.7515, device='cuda:0')
[307] Valid Loss:   tensor(0.5852, device='cuda:0')
[308] Valid Loss:   tensor(0.6391, device='cuda:0')
[309] Valid Loss:   tensor(0.6783, device='cuda:0')
[310] Valid Loss:   tensor(0.6903, device='cuda:0')
[311] Valid Loss:   tensor(0.6159, device='cuda:0')
[312] Valid Loss:   tensor(0.5946, device='cuda:0')
[313] Valid Loss:   tensor(0.6961, device='cuda:0')
[314] Valid Loss:   tensor(0.7238, device='cuda:0')
[315] Valid Loss:   tensor(1.0154, device='cuda:0')
[316] Valid Loss:   tensor(0.8448, device='cuda:0')
[317] Valid Loss:   tensor(0.7062, device='cuda:0')
[318] Valid Loss:   tensor(0.6636, device='cuda:0')
[319] Valid Loss:   tensor(0.6514, device='cuda:0')
[320] Valid Loss:   tensor(0.6660, device='cuda:0')
[321] Valid Loss:   tensor(0.6772, device='cuda:0')
[322] Valid Loss:   tensor(0.6320, device='cuda:0')
[323] Valid Loss:   tensor(0.7880, device='cuda:0')
[324] Valid Loss:   tensor(0.7513, device='cuda:0')
[325] Valid Loss:   tensor(0.6019, device='cuda:0')
[326] Valid Loss:   tensor(0.7205, device='cuda:0')
[327] Valid Loss:   tensor(0.4528, device='cuda:0')
[328] Valid Loss:   tensor(0.6599, device='cuda:0')
[329] Valid Loss:   tensor(0.7822, device='cuda:0')
[330] Valid Loss:   tensor(0.5994, device='cuda:0')
[331] Valid Loss:   tensor(0.9281, device='cuda:0')
[332] Valid Loss:   tensor(0.5904, device='cuda:0')
```
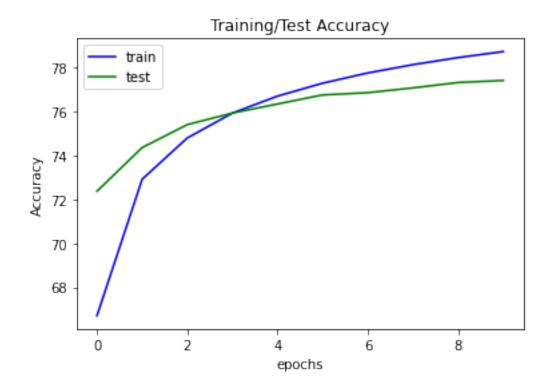
```
[333] Valid Loss:   tensor(0.5806, device='cuda:0')
[334] Valid Loss:   tensor(0.6111, device='cuda:0')
[335] Valid Loss:   tensor(0.7167, device='cuda:0')
[336] Valid Loss:   tensor(0.8092, device='cuda:0')
[337] Valid Loss:   tensor(0.8595, device='cuda:0')
[338] Valid Loss:   tensor(0.7553, device='cuda:0')
[339] Valid Loss:   tensor(0.6891, device='cuda:0')
[340] Valid Loss:   tensor(0.7609, device='cuda:0')
[341] Valid Loss:   tensor(0.8273, device='cuda:0')
[342] Valid Loss:   tensor(0.9081, device='cuda:0')
[343] Valid Loss:   tensor(0.7459, device='cuda:0')
[344] Valid Loss:   tensor(0.6488, device='cuda:0')
[345] Valid Loss:   tensor(0.5605, device='cuda:0')
[346] Valid Loss:   tensor(0.7139, device='cuda:0')
[347] Valid Loss:   tensor(0.8328, device='cuda:0')
[348] Valid Loss:   tensor(0.6914, device='cuda:0')
[349] Valid Loss:   tensor(0.6649, device='cuda:0')
[350] Valid Loss:   tensor(0.6924, device='cuda:0')
[351] Valid Loss:   tensor(0.7961, device='cuda:0')
[352] Valid Loss:   tensor(0.6316, device='cuda:0')
[353] Valid Loss:   tensor(0.7238, device='cuda:0')
[354] Valid Loss:   tensor(0.9339, device='cuda:0')
[355] Valid Loss:   tensor(0.5144, device='cuda:0')
[356] Valid Loss:   tensor(0.6744, device='cuda:0')
[357] Valid Loss:   tensor(0.8009, device='cuda:0')
[358] Valid Loss:   tensor(0.6786, device='cuda:0')
[359] Valid Loss:   tensor(0.7870, device='cuda:0')
[360] Valid Loss:   tensor(0.7079, device='cuda:0')
[361] Valid Loss:   tensor(0.7448, device='cuda:0')
[362] Valid Loss:   tensor(0.7179, device='cuda:0')
[363] Valid Loss:   tensor(0.7699, device='cuda:0')
[364] Valid Loss:   tensor(0.9254, device='cuda:0')
[365] Valid Loss:   tensor(0.6861, device='cuda:0')
[366] Valid Loss:   tensor(0.6662, device='cuda:0')
[367] Valid Loss:   tensor(0.7649, device='cuda:0')
[368] Valid Loss:   tensor(0.8272, device='cuda:0')
[369] Valid Loss:   tensor(0.7260, device='cuda:0')
[370] Valid Loss:   tensor(0.6007, device='cuda:0')
[371] Valid Loss:   tensor(0.6236, device='cuda:0')
[372] Valid Loss:   tensor(0.7917, device='cuda:0')
[373] Valid Loss:   tensor(0.5372, device='cuda:0')
[374] Valid Loss:   tensor(0.7127, device='cuda:0')
[375] Valid Loss:   tensor(0.7866, device='cuda:0')
[376] Valid Loss:   tensor(0.6266, device='cuda:0')
[377] Valid Loss:   tensor(0.9205, device='cuda:0')
[378] Valid Loss:   tensor(0.6171, device='cuda:0')
[379] Valid Loss:   tensor(0.8039, device='cuda:0')
[380] Valid Loss:   tensor(1.0983, device='cuda:0')
```
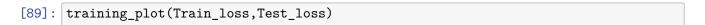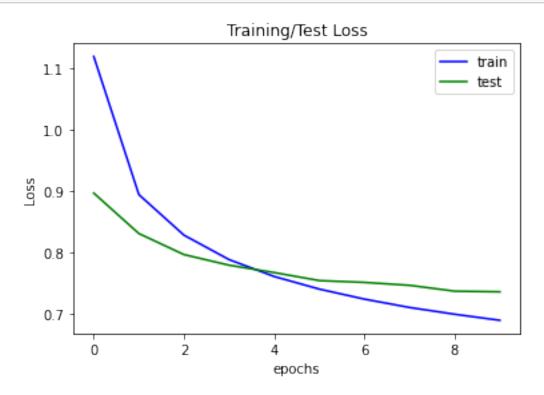
```
Testing Loss:   0.7353862481793081
Testing Accuaracy:   77.39704192983322 %
```



```
[88]: training_plot_acc(Train_acc,Test_acc)
```

Training/Test Accuracy

```
[89]: training_plot(Train_loss,Test_loss)
```



Training/Test Loss

```
[90]: model = torch.load(os.path.join(Path,'{}.pt'.format('model_V2_'+ str(9))))
```

```
[98]: test_loss, test_acc = val_model(model,val_loader,criterion=criterion)
```

```
[0] Valid Loss:  tensor(0.6859, device='cuda:0')
[1] Valid Loss:  tensor(0.8796, device='cuda:0')
[2] Valid Loss:  tensor(0.6652, device='cuda:0')
[3] Valid Loss:  tensor(0.6006, device='cuda:0')
[4] Valid Loss:  tensor(0.7383, device='cuda:0')
[5] Valid Loss:  tensor(0.6930, device='cuda:0')
[6] Valid Loss:  tensor(0.7131, device='cuda:0')
[7] Valid Loss:  tensor(0.8259, device='cuda:0')
[8] Valid Loss:  tensor(0.7578, device='cuda:0')
[9] Valid Loss:  tensor(0.8166, device='cuda:0')
[10] Valid Loss:  tensor(0.6720, device='cuda:0')
[11] Valid Loss:  tensor(0.6669, device='cuda:0')
[12] Valid Loss:  tensor(1.1379, device='cuda:0')
[13] Valid Loss:  tensor(0.9776, device='cuda:0')
[14] Valid Loss:  tensor(0.6125, device='cuda:0')
[15] Valid Loss:  tensor(0.7342, device='cuda:0')
[16] Valid Loss:  tensor(0.6075, device='cuda:0')
[17] Valid Loss:  tensor(0.7500, device='cuda:0')
[18] Valid Loss:  tensor(0.6714, device='cuda:0')
[19] Valid Loss:  tensor(0.6158, device='cuda:0')
[20] Valid Loss:  tensor(0.6852, device='cuda:0')
[21] Valid Loss:  tensor(0.5963, device='cuda:0')
[22] Valid Loss:  tensor(0.5806, device='cuda:0')
[23] Valid Loss:  tensor(0.8667, device='cuda:0')
[24] Valid Loss:  tensor(0.7177, device='cuda:0')
[25] Valid Loss:  tensor(0.9747, device='cuda:0')
[26] Valid Loss:  tensor(0.4976, device='cuda:0')
[27] Valid Loss:  tensor(0.8630, device='cuda:0')
[28] Valid Loss:  tensor(0.6632, device='cuda:0')
[29] Valid Loss:  tensor(0.7179, device='cuda:0')
[30] Valid Loss:  tensor(0.6315, device='cuda:0')
[31] Valid Loss:  tensor(0.7399, device='cuda:0')
[32] Valid Loss:  tensor(0.8654, device='cuda:0')
[33] Valid Loss:  tensor(0.8140, device='cuda:0')
[34] Valid Loss:  tensor(0.7857, device='cuda:0')
[35] Valid Loss:  tensor(0.8638, device='cuda:0')
[36] Valid Loss:  tensor(0.6564, device='cuda:0')
[37] Valid Loss:  tensor(1.0045, device='cuda:0')
[38] Valid Loss:  tensor(0.8157, device='cuda:0')
[39] Valid Loss:  tensor(0.6757, device='cuda:0')
[40] Valid Loss:  tensor(0.7651, device='cuda:0')
```

```
[41] Valid Loss:   tensor(0.9153, device='cuda:0')
[42] Valid Loss:   tensor(0.7586, device='cuda:0')
[43] Valid Loss:   tensor(0.5532, device='cuda:0')
[44] Valid Loss:   tensor(0.6601, device='cuda:0')
[45] Valid Loss:   tensor(0.5728, device='cuda:0')
[46] Valid Loss:   tensor(0.8986, device='cuda:0')
[47] Valid Loss:   tensor(0.7672, device='cuda:0')
[48] Valid Loss:   tensor(0.6443, device='cuda:0')
[49] Valid Loss:   tensor(0.7430, device='cuda:0')
[50] Valid Loss:   tensor(0.7337, device='cuda:0')
[51] Valid Loss:   tensor(0.6364, device='cuda:0')
[52] Valid Loss:   tensor(0.5964, device='cuda:0')
[53] Valid Loss:   tensor(0.8939, device='cuda:0')
[54] Valid Loss:   tensor(0.6344, device='cuda:0')
[55] Valid Loss:   tensor(0.5645, device='cuda:0')
[56] Valid Loss:   tensor(0.8452, device='cuda:0')
[57] Valid Loss:   tensor(0.8538, device='cuda:0')
[58] Valid Loss:   tensor(0.7665, device='cuda:0')
[59] Valid Loss:   tensor(0.6498, device='cuda:0')
[60] Valid Loss:   tensor(1.0518, device='cuda:0')
[61] Valid Loss:   tensor(0.6860, device='cuda:0')
[62] Valid Loss:   tensor(0.8480, device='cuda:0')
[63] Valid Loss:   tensor(0.6753, device='cuda:0')
[64] Valid Loss:   tensor(0.7420, device='cuda:0')
[65] Valid Loss:   tensor(0.8720, device='cuda:0')
[66] Valid Loss:   tensor(0.7027, device='cuda:0')
[67] Valid Loss:   tensor(0.7410, device='cuda:0')
[68] Valid Loss:   tensor(0.7141, device='cuda:0')
[69] Valid Loss:   tensor(0.5681, device='cuda:0')
[70] Valid Loss:   tensor(0.5910, device='cuda:0')
[71] Valid Loss:   tensor(0.8626, device='cuda:0')
[72] Valid Loss:   tensor(0.7487, device='cuda:0')
[73] Valid Loss:   tensor(0.6595, device='cuda:0')
[74] Valid Loss:   tensor(0.7901, device='cuda:0')
[75] Valid Loss:   tensor(0.8731, device='cuda:0')
[76] Valid Loss:   tensor(0.8105, device='cuda:0')
[77] Valid Loss:   tensor(0.6859, device='cuda:0')
[78] Valid Loss:   tensor(0.7714, device='cuda:0')
[79] Valid Loss:   tensor(0.6168, device='cuda:0')
[80] Valid Loss:   tensor(0.7394, device='cuda:0')
[81] Valid Loss:   tensor(0.6165, device='cuda:0')
[82] Valid Loss:   tensor(0.8603, device='cuda:0')
[83] Valid Loss:   tensor(0.8005, device='cuda:0')
[84] Valid Loss:   tensor(0.6183, device='cuda:0')
[85] Valid Loss:   tensor(0.7270, device='cuda:0')
[86] Valid Loss:   tensor(0.7616, device='cuda:0')
[87] Valid Loss:   tensor(0.8001, device='cuda:0')
[88] Valid Loss:   tensor(0.5494, device='cuda:0')
```

```
[89] Valid Loss:  tensor(0.7885, device='cuda:0')
[90] Valid Loss:  tensor(0.9244, device='cuda:0')
[91] Valid Loss:  tensor(0.7226, device='cuda:0')
[92] Valid Loss:  tensor(0.7449, device='cuda:0')
[93] Valid Loss:  tensor(0.7989, device='cuda:0')
[94] Valid Loss:  tensor(0.8324, device='cuda:0')
[95] Valid Loss:  tensor(0.7281, device='cuda:0')
[96] Valid Loss:  tensor(0.8468, device='cuda:0')
[97] Valid Loss:  tensor(0.7998, device='cuda:0')
[98] Valid Loss:  tensor(0.8245, device='cuda:0')
[99] Valid Loss:  tensor(0.6169, device='cuda:0')
[100] Valid Loss:  tensor(0.7158, device='cuda:0')
[101] Valid Loss:  tensor(0.7221, device='cuda:0')
[102] Valid Loss:  tensor(0.8442, device='cuda:0')
[103] Valid Loss:  tensor(0.9019, device='cuda:0')
[104] Valid Loss:  tensor(0.6512, device='cuda:0')
[105] Valid Loss:  tensor(0.6616, device='cuda:0')
[106] Valid Loss:  tensor(0.6159, device='cuda:0')
[107] Valid Loss:  tensor(0.7594, device='cuda:0')
[108] Valid Loss:  tensor(0.8381, device='cuda:0')
[109] Valid Loss:  tensor(0.7623, device='cuda:0')
[110] Valid Loss:  tensor(0.8327, device='cuda:0')
[111] Valid Loss:  tensor(0.7461, device='cuda:0')
[112] Valid Loss:  tensor(0.8213, device='cuda:0')
[113] Valid Loss:  tensor(0.7356, device='cuda:0')
[114] Valid Loss:  tensor(0.7310, device='cuda:0')
[115] Valid Loss:  tensor(0.8023, device='cuda:0')
[116] Valid Loss:  tensor(0.8143, device='cuda:0')
[117] Valid Loss:  tensor(0.8408, device='cuda:0')
[118] Valid Loss:  tensor(0.6678, device='cuda:0')
[119] Valid Loss:  tensor(0.6269, device='cuda:0')
[120] Valid Loss:  tensor(0.9712, device='cuda:0')
[121] Valid Loss:  tensor(0.6538, device='cuda:0')
[122] Valid Loss:  tensor(0.7248, device='cuda:0')
[123] Valid Loss:  tensor(0.6487, device='cuda:0')
[124] Valid Loss:  tensor(0.8830, device='cuda:0')
[125] Valid Loss:  tensor(0.9495, device='cuda:0')
[126] Valid Loss:  tensor(0.5673, device='cuda:0')
[127] Valid Loss:  tensor(0.6380, device='cuda:0')
[128] Valid Loss:  tensor(0.8147, device='cuda:0')
[129] Valid Loss:  tensor(0.6855, device='cuda:0')
[130] Valid Loss:  tensor(0.8199, device='cuda:0')
[131] Valid Loss:  tensor(0.6740, device='cuda:0')
[132] Valid Loss:  tensor(0.7087, device='cuda:0')
[133] Valid Loss:  tensor(0.7484, device='cuda:0')
[134] Valid Loss:  tensor(0.6671, device='cuda:0')
[135] Valid Loss:  tensor(0.5918, device='cuda:0')
[136] Valid Loss:  tensor(0.6331, device='cuda:0')
```

```
[137] Valid Loss:   tensor(1.0025, device='cuda:0')
[138] Valid Loss:   tensor(0.8493, device='cuda:0')
[139] Valid Loss:   tensor(1.0463, device='cuda:0')
[140] Valid Loss:   tensor(0.7854, device='cuda:0')
[141] Valid Loss:   tensor(0.6269, device='cuda:0')
[142] Valid Loss:   tensor(0.9535, device='cuda:0')
[143] Valid Loss:   tensor(0.6541, device='cuda:0')
[144] Valid Loss:   tensor(0.7489, device='cuda:0')
[145] Valid Loss:   tensor(0.9314, device='cuda:0')
[146] Valid Loss:   tensor(0.6658, device='cuda:0')
[147] Valid Loss:   tensor(0.7856, device='cuda:0')
[148] Valid Loss:   tensor(0.6699, device='cuda:0')
[149] Valid Loss:   tensor(0.7680, device='cuda:0')
[150] Valid Loss:   tensor(0.7609, device='cuda:0')
[151] Valid Loss:   tensor(0.7579, device='cuda:0')
[152] Valid Loss:   tensor(0.6549, device='cuda:0')
[153] Valid Loss:   tensor(0.6619, device='cuda:0')
[154] Valid Loss:   tensor(0.6919, device='cuda:0')
[155] Valid Loss:   tensor(0.7746, device='cuda:0')
[156] Valid Loss:   tensor(0.8329, device='cuda:0')
[157] Valid Loss:   tensor(0.6578, device='cuda:0')
[158] Valid Loss:   tensor(0.7193, device='cuda:0')
[159] Valid Loss:   tensor(0.7222, device='cuda:0')
[160] Valid Loss:   tensor(0.5980, device='cuda:0')
[161] Valid Loss:   tensor(0.8121, device='cuda:0')
[162] Valid Loss:   tensor(0.7743, device='cuda:0')
[163] Valid Loss:   tensor(0.6574, device='cuda:0')
[164] Valid Loss:   tensor(0.9269, device='cuda:0')
[165] Valid Loss:   tensor(0.7214, device='cuda:0')
[166] Valid Loss:   tensor(0.7098, device='cuda:0')
[167] Valid Loss:   tensor(0.6009, device='cuda:0')
[168] Valid Loss:   tensor(0.6840, device='cuda:0')
[169] Valid Loss:   tensor(0.6365, device='cuda:0')
[170] Valid Loss:   tensor(0.6687, device='cuda:0')
[171] Valid Loss:   tensor(0.5713, device='cuda:0')
[172] Valid Loss:   tensor(0.7843, device='cuda:0')
[173] Valid Loss:   tensor(0.6243, device='cuda:0')
[174] Valid Loss:   tensor(0.7587, device='cuda:0')
[175] Valid Loss:   tensor(0.6867, device='cuda:0')
[176] Valid Loss:   tensor(0.6014, device='cuda:0')
[177] Valid Loss:   tensor(0.7329, device='cuda:0')
[178] Valid Loss:   tensor(0.7966, device='cuda:0')
[179] Valid Loss:   tensor(0.8119, device='cuda:0')
[180] Valid Loss:   tensor(0.7953, device='cuda:0')
[181] Valid Loss:   tensor(0.7703, device='cuda:0')
[182] Valid Loss:   tensor(0.7223, device='cuda:0')
[183] Valid Loss:   tensor(0.7090, device='cuda:0')
[184] Valid Loss:   tensor(0.9061, device='cuda:0')
```

```
[185] Valid Loss:    tensor(0.8267, device='cuda:0')
[186] Valid Loss:    tensor(0.7092, device='cuda:0')
[187] Valid Loss:    tensor(0.8351, device='cuda:0')
[188] Valid Loss:    tensor(0.8636, device='cuda:0')
[189] Valid Loss:    tensor(0.7091, device='cuda:0')
[190] Valid Loss:    tensor(0.7710, device='cuda:0')
[191] Valid Loss:    tensor(0.6380, device='cuda:0')
[192] Valid Loss:    tensor(1.0217, device='cuda:0')
[193] Valid Loss:    tensor(0.7548, device='cuda:0')
[194] Valid Loss:    tensor(0.8072, device='cuda:0')
[195] Valid Loss:    tensor(0.8617, device='cuda:0')
[196] Valid Loss:    tensor(0.5909, device='cuda:0')
[197] Valid Loss:    tensor(0.7662, device='cuda:0')
[198] Valid Loss:    tensor(0.7321, device='cuda:0')
[199] Valid Loss:    tensor(0.7878, device='cuda:0')
[200] Valid Loss:    tensor(0.7194, device='cuda:0')
[201] Valid Loss:    tensor(0.6891, device='cuda:0')
[202] Valid Loss:    tensor(0.6555, device='cuda:0')
[203] Valid Loss:    tensor(0.8828, device='cuda:0')
[204] Valid Loss:    tensor(0.8536, device='cuda:0')
[205] Valid Loss:    tensor(0.8495, device='cuda:0')
[206] Valid Loss:    tensor(0.9197, device='cuda:0')
[207] Valid Loss:    tensor(0.7831, device='cuda:0')
[208] Valid Loss:    tensor(0.6102, device='cuda:0')
[209] Valid Loss:    tensor(0.7244, device='cuda:0')
[210] Valid Loss:    tensor(1.1404, device='cuda:0')
[211] Valid Loss:    tensor(0.7650, device='cuda:0')
[212] Valid Loss:    tensor(0.6129, device='cuda:0')
[213] Valid Loss:    tensor(0.6881, device='cuda:0')
[214] Valid Loss:    tensor(0.6841, device='cuda:0')
[215] Valid Loss:    tensor(0.7971, device='cuda:0')
[216] Valid Loss:    tensor(0.7537, device='cuda:0')
[217] Valid Loss:    tensor(0.7246, device='cuda:0')
[218] Valid Loss:    tensor(0.5863, device='cuda:0')
[219] Valid Loss:    tensor(0.7043, device='cuda:0')
[220] Valid Loss:    tensor(0.6671, device='cuda:0')
[221] Valid Loss:    tensor(0.7693, device='cuda:0')
[222] Valid Loss:    tensor(0.5835, device='cuda:0')
[223] Valid Loss:    tensor(0.6721, device='cuda:0')
[224] Valid Loss:    tensor(0.8178, device='cuda:0')
[225] Valid Loss:    tensor(0.5999, device='cuda:0')
[226] Valid Loss:    tensor(0.7514, device='cuda:0')
[227] Valid Loss:    tensor(0.6153, device='cuda:0')
[228] Valid Loss:    tensor(0.7921, device='cuda:0')
[229] Valid Loss:    tensor(0.6815, device='cuda:0')
[230] Valid Loss:    tensor(0.7582, device='cuda:0')
[231] Valid Loss:    tensor(0.6767, device='cuda:0')
[232] Valid Loss:    tensor(0.5798, device='cuda:0')
```

```
[233] Valid Loss:    tensor(0.6846, device='cuda:0')
[234] Valid Loss:    tensor(0.6100, device='cuda:0')
[235] Valid Loss:    tensor(0.5903, device='cuda:0')
[236] Valid Loss:    tensor(0.7245, device='cuda:0')
[237] Valid Loss:    tensor(0.8183, device='cuda:0')
[238] Valid Loss:    tensor(0.5477, device='cuda:0')
[239] Valid Loss:    tensor(0.8387, device='cuda:0')
[240] Valid Loss:    tensor(0.6930, device='cuda:0')
[241] Valid Loss:    tensor(0.6304, device='cuda:0')
[242] Valid Loss:    tensor(0.5454, device='cuda:0')
[243] Valid Loss:    tensor(1.1684, device='cuda:0')
[244] Valid Loss:    tensor(0.7904, device='cuda:0')
[245] Valid Loss:    tensor(0.6689, device='cuda:0')
[246] Valid Loss:    tensor(0.6740, device='cuda:0')
[247] Valid Loss:    tensor(0.6686, device='cuda:0')
[248] Valid Loss:    tensor(0.9386, device='cuda:0')
[249] Valid Loss:    tensor(0.7264, device='cuda:0')
[250] Valid Loss:    tensor(0.7733, device='cuda:0')
[251] Valid Loss:    tensor(0.8178, device='cuda:0')
[252] Valid Loss:    tensor(0.6102, device='cuda:0')
[253] Valid Loss:    tensor(0.6418, device='cuda:0')
[254] Valid Loss:    tensor(0.6423, device='cuda:0')
[255] Valid Loss:    tensor(0.7956, device='cuda:0')
[256] Valid Loss:    tensor(0.8101, device='cuda:0')
[257] Valid Loss:    tensor(0.6705, device='cuda:0')
[258] Valid Loss:    tensor(0.8271, device='cuda:0')
[259] Valid Loss:    tensor(0.7304, device='cuda:0')
[260] Valid Loss:    tensor(0.9256, device='cuda:0')
[261] Valid Loss:    tensor(0.7585, device='cuda:0')
[262] Valid Loss:    tensor(0.8931, device='cuda:0')
[263] Valid Loss:    tensor(0.8679, device='cuda:0')
[264] Valid Loss:    tensor(0.7925, device='cuda:0')
[265] Valid Loss:    tensor(0.7770, device='cuda:0')
[266] Valid Loss:    tensor(0.6949, device='cuda:0')
[267] Valid Loss:    tensor(0.7319, device='cuda:0')
[268] Valid Loss:    tensor(0.7334, device='cuda:0')
[269] Valid Loss:    tensor(0.6790, device='cuda:0')
[270] Valid Loss:    tensor(0.6464, device='cuda:0')
[271] Valid Loss:    tensor(1.0071, device='cuda:0')
[272] Valid Loss:    tensor(0.7136, device='cuda:0')
[273] Valid Loss:    tensor(0.8749, device='cuda:0')
[274] Valid Loss:    tensor(0.6948, device='cuda:0')
[275] Valid Loss:    tensor(0.7967, device='cuda:0')
[276] Valid Loss:    tensor(0.6398, device='cuda:0')
[277] Valid Loss:    tensor(0.6381, device='cuda:0')
[278] Valid Loss:    tensor(0.8979, device='cuda:0')
[279] Valid Loss:    tensor(0.7668, device='cuda:0')
[280] Valid Loss:    tensor(0.8587, device='cuda:0')
```

```
[281] Valid Loss:  tensor(0.6503, device='cuda:0')
[282] Valid Loss:  tensor(0.7938, device='cuda:0')
[283] Valid Loss:  tensor(0.9199, device='cuda:0')
[284] Valid Loss:  tensor(0.6817, device='cuda:0')
[285] Valid Loss:  tensor(0.5786, device='cuda:0')
[286] Valid Loss:  tensor(0.8067, device='cuda:0')
[287] Valid Loss:  tensor(0.9474, device='cuda:0')
[288] Valid Loss:  tensor(0.5705, device='cuda:0')
[289] Valid Loss:  tensor(0.7546, device='cuda:0')
[290] Valid Loss:  tensor(0.6010, device='cuda:0')
[291] Valid Loss:  tensor(0.7438, device='cuda:0')
[292] Valid Loss:  tensor(0.8417, device='cuda:0')
[293] Valid Loss:  tensor(0.6765, device='cuda:0')
[294] Valid Loss:  tensor(0.8801, device='cuda:0')
[295] Valid Loss:  tensor(0.6550, device='cuda:0')
[296] Valid Loss:  tensor(0.7549, device='cuda:0')
[297] Valid Loss:  tensor(0.8127, device='cuda:0')
[298] Valid Loss:  tensor(0.6672, device='cuda:0')
[299] Valid Loss:  tensor(0.8048, device='cuda:0')
[300] Valid Loss:  tensor(0.9199, device='cuda:0')
[301] Valid Loss:  tensor(0.5721, device='cuda:0')
[302] Valid Loss:  tensor(0.5656, device='cuda:0')
[303] Valid Loss:  tensor(0.6229, device='cuda:0')
[304] Valid Loss:  tensor(0.6570, device='cuda:0')
[305] Valid Loss:  tensor(0.7692, device='cuda:0')
[306] Valid Loss:  tensor(0.7955, device='cuda:0')
[307] Valid Loss:  tensor(0.6211, device='cuda:0')
[308] Valid Loss:  tensor(0.6200, device='cuda:0')
[309] Valid Loss:  tensor(0.6652, device='cuda:0')
[310] Valid Loss:  tensor(0.6987, device='cuda:0')
[311] Valid Loss:  tensor(0.6214, device='cuda:0')
[312] Valid Loss:  tensor(0.6195, device='cuda:0')
[313] Valid Loss:  tensor(0.6967, device='cuda:0')
[314] Valid Loss:  tensor(0.7353, device='cuda:0')
[315] Valid Loss:  tensor(0.9952, device='cuda:0')
[316] Valid Loss:  tensor(0.8238, device='cuda:0')
[317] Valid Loss:  tensor(0.6979, device='cuda:0')
[318] Valid Loss:  tensor(0.6718, device='cuda:0')
[319] Valid Loss:  tensor(0.6504, device='cuda:0')
[320] Valid Loss:  tensor(0.6609, device='cuda:0')
[321] Valid Loss:  tensor(0.6582, device='cuda:0')
[322] Valid Loss:  tensor(0.6378, device='cuda:0')
[323] Valid Loss:  tensor(0.8134, device='cuda:0')
[324] Valid Loss:  tensor(0.7659, device='cuda:0')
[325] Valid Loss:  tensor(0.6269, device='cuda:0')
[326] Valid Loss:  tensor(0.7054, device='cuda:0')
[327] Valid Loss:  tensor(0.4554, device='cuda:0')
[328] Valid Loss:  tensor(0.6469, device='cuda:0')
```

```
[329] Valid Loss:  tensor(0.8306, device='cuda:0')
[330] Valid Loss:  tensor(0.6265, device='cuda:0')
[331] Valid Loss:  tensor(0.9424, device='cuda:0')
[332] Valid Loss:  tensor(0.5868, device='cuda:0')
[333] Valid Loss:  tensor(0.5850, device='cuda:0')
[334] Valid Loss:  tensor(0.6086, device='cuda:0')
[335] Valid Loss:  tensor(0.6905, device='cuda:0')
[336] Valid Loss:  tensor(0.7930, device='cuda:0')
[337] Valid Loss:  tensor(0.8768, device='cuda:0')
[338] Valid Loss:  tensor(0.7642, device='cuda:0')
[339] Valid Loss:  tensor(0.6904, device='cuda:0')
[340] Valid Loss:  tensor(0.7531, device='cuda:0')
[341] Valid Loss:  tensor(0.8244, device='cuda:0')
[342] Valid Loss:  tensor(0.9318, device='cuda:0')
[343] Valid Loss:  tensor(0.7611, device='cuda:0')
[344] Valid Loss:  tensor(0.6460, device='cuda:0')
[345] Valid Loss:  tensor(0.5559, device='cuda:0')
[346] Valid Loss:  tensor(0.7063, device='cuda:0')
[347] Valid Loss:  tensor(0.8482, device='cuda:0')
[348] Valid Loss:  tensor(0.7094, device='cuda:0')
[349] Valid Loss:  tensor(0.6505, device='cuda:0')
[350] Valid Loss:  tensor(0.6947, device='cuda:0')
[351] Valid Loss:  tensor(0.8416, device='cuda:0')
[352] Valid Loss:  tensor(0.6438, device='cuda:0')
[353] Valid Loss:  tensor(0.7131, device='cuda:0')
[354] Valid Loss:  tensor(0.9194, device='cuda:0')
[355] Valid Loss:  tensor(0.5376, device='cuda:0')
[356] Valid Loss:  tensor(0.6708, device='cuda:0')
[357] Valid Loss:  tensor(0.8087, device='cuda:0')
[358] Valid Loss:  tensor(0.6900, device='cuda:0')
[359] Valid Loss:  tensor(0.8092, device='cuda:0')
[360] Valid Loss:  tensor(0.6848, device='cuda:0')
[361] Valid Loss:  tensor(0.7166, device='cuda:0')
[362] Valid Loss:  tensor(0.7184, device='cuda:0')
[363] Valid Loss:  tensor(0.7779, device='cuda:0')
[364] Valid Loss:  tensor(0.9048, device='cuda:0')
[365] Valid Loss:  tensor(0.6912, device='cuda:0')
[366] Valid Loss:  tensor(0.6681, device='cuda:0')
[367] Valid Loss:  tensor(0.7818, device='cuda:0')
[368] Valid Loss:  tensor(0.8540, device='cuda:0')
[369] Valid Loss:  tensor(0.7213, device='cuda:0')
[370] Valid Loss:  tensor(0.6229, device='cuda:0')
[371] Valid Loss:  tensor(0.6378, device='cuda:0')
[372] Valid Loss:  tensor(0.8244, device='cuda:0')
[373] Valid Loss:  tensor(0.5423, device='cuda:0')
[374] Valid Loss:  tensor(0.7034, device='cuda:0')
[375] Valid Loss:  tensor(0.8094, device='cuda:0')
[376] Valid Loss:  tensor(0.6217, device='cuda:0')
```

```
[377] Valid Loss:  tensor(0.9024, device='cuda:0')
[378] Valid Loss:  tensor(0.6351, device='cuda:0')
[379] Valid Loss:  tensor(0.8160, device='cuda:0')
[380] Valid Loss:  tensor(1.0755, device='cuda:0')
Testing Loss:  0.7401996966734958
Testing Accuaracy:  77.19743408996592 %
```

[ ]: