

Koç University, Deep Learning Course (COMP541)

Assignment 4: Graph Neural Networks and Transformers

In this assignment, you will implement the vanilla version of Graph Convolution Networks (GCN) [Kipf and Welling \(2016\)](#) and Graph Attention Networks (GAT) [Veličković, et al. \(2018\)](#), as well as a transformer for translation.

Background

Basics of GCN

Recall from the lectures, the goal of a GCN is to learn a function of signals features on a graph $G = (V, E)$, which takes as inputs:

1. the input features of each node, $x_i \in R^F$ (in matrix form: $X \in R^{|V| \times F}$)
2. some information about the graph structure, typically the adjacency matrix A

Each convolutional layer can be written as $H^{(l+1)} = f(H^{(l)}, A)$ for some function f .

The function f we are using for this assignment is in the form of

$f(H^{(l)}, A) = \sigma(\hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2} H^{(l)} W^{(l)})$, where $\hat{A} = A + I$ and \hat{D} is the diagonal node degree matrix ($D^{-1} \hat{A}$ normalizes \hat{A} such that all rows sum to one). Let

$\tilde{A} = \hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2}$. The GCN we will implement takes two convolution layers, $Z = f(X, A) = \text{softmax}(\tilde{A} \cdot \text{Dropout}(\text{ReLU}(\tilde{A} X W^{(0)})) \cdot W^{(1)})$

Basics of GAT

Graph Attention Network (GAT) is a novel convolution-style neural network. It operates on graph-structured data and leverages masked self-attentional layers. In this assignment, we will implement the graph attention layer.

Dataset

The dataset we used for this assignment is Cora ([Sen et al. \(2008\)](#)). Cora is one of standard citation network benchmark dataset (just like MNIST dataset for computer vision tasks). It consists of 2708 scientific publications and 5429 links. Each publication is classified into one of 7 classes. Each publication is described by a word vector (length 1433) that indicates the absence/presence of the corresponding word. This is used as the features of each node for our experiment. The task is to perform node classification (predict which class each node belongs to).

Google Colab Setup

```
In [1]: from google.colab import drive
drive.mount('/content/drive')

BASE = '/content/drive/MyDrive/comp541_assignment4_fall2025'
DIR = f'{BASE}/tr_to_en'

Mounted at /content/drive
```

```
In [2]: import os

print(os.listdir(BASE))
print(os.listdir(DIR))

['.DS_Store', 'cora.tgz', 'tr_to_en', '.venv', 'cora', 'comp541_assignment4_(1).ipynb']
['tr.dev', 'tr.train.r0.125', 'en.train.r0.125', 'tr.train.r0.5', 'en.dev', 'tr.train', 'en.test', 'en.train.r0.5', 'en.train', 'tr.train.r0.25', 'en.train.r0.25', 'tr.test']
```

Experiments

Experiments: Open GCN notebook on Colab and implement the following parts.

1. Implementation of Graph Convolution Layer: Complete the code for `GraphConvolution` Class
2. Implementation of Graph Convolution Network: Complete the code for `GCN` Class
3. Train your Graph Convolution Network: After implementing the required classes, now you can train your GCN. We want you to play with the architecture (such as the number of Graph Convolution Layers, usage of Dropout etc.) and the hyperparameters, and report your results in various settings.
4. Implementation of Graph Attention Layer: Complete the code for `GraphAttentionLayer` Class
5. Train your Graph Convolution Network: After implementing the required classes, now you can train your GAT. Again, you are need to play with the structure and the hyperparameters and conduct various experiments.
6. Compare your models: Compare the evaluation results for Vanilla GCN and GAT. Comment on the discrepancy in their performance (if any) and briefly explain why you think it's the case.

Download the Cora data

```
In [3]: !curl -O https://linqs-data.soe.ucsc.edu/public/lbc/cora.tgz
! tar -zxf cora.tgz
```

% Total rrent	% Received	% Xferd	Average Speed	Time	Time	Time	Cu
			Dload Upload	Total	Spent	Left	Sp
eed							
100 164k 100 164k 0 0 62124				0 0:00:02	0:00:02	--:--:--	6
2126							
cora/							
cora/README							
cora/cora.cites							
cora/cora.content							

import modules and set random seed

```
In [4]: import numpy as np
import scipy.sparse as sp
import torch
import pandas as pd
import math
import torch
import torch.nn as nn
import torch.nn.functional as F
import torch.optim as optim
import time
import matplotlib.pyplot as plt

seed = 0

np.random.seed(seed)
torch.manual_seed(seed)
torch.cuda.manual_seed(seed)
torch.cuda.manual_seed_all(seed)
device = torch.device('cuda:0' if torch.cuda.is_available() else 'cpu')
```

Loading and preprocessing the data

```
In [ ]: def encode_onehot(labels):
    # The classes must be sorted before encoding to enable static class e
    # In other words, make sure the first class always maps to index 0.
    classes = sorted(list(set(labels)))
    classes_dict = {c: np.identity(len(classes))[i, :] for i, c in
                    enumerate(classes)}
    labels_onehot = np.array(list(map(classes_dict.get, labels)),
                           dtype=np.int32)
    return labels_onehot

def load_data(path="cora/", dataset="cora", training_samples=140):
    """Load citation network dataset (cora only for now)"""
    print('Loading {} dataset...'.format(dataset))

    idx_features_labels = np.genfromtxt("{}{}.content".format(path, dataset),
                                        dtype=np.dtype(str))
    features = sp.csr_matrix(idx_features_labels[:, 1:-1], dtype=np.float)
    labels = encode_onehot(idx_features_labels[:, -1])

    # build graph
```

```

idx = np.array(idx_features_labels[:, 0], dtype=np.int32)
idx_map = {j: i for i, j in enumerate(idx)}
edges_unordered = np.genfromtxt("{}{}.cites".format(path, dataset),
                                 dtype=np.int32)
edges = np.array(list(map(idx_map.get, edges_unordered.flatten()))),
                           dtype=np.int32).reshape(edges_unordered.shape)
adj = sp.coo_matrix((np.ones(edges.shape[0]), (edges[:, 0], edges[:, 1]),
                     shape=(labels.shape[0], labels.shape[0])),
                     dtype=np.float32)

# build symmetric adjacency matrix
adj = adj + adj.T.multiply(adj.T > adj) - adj.multiply(adj.T > adj)

features = normalize(features)
adj = adj + sp.eye(adj.shape[0])
adj = normalize_adj(adj)

# Random indexes
idx_rand = torch.randperm(len(labels))
# Nodes for training
idx_train = idx_rand[:training_samples]
# Nodes for validation
idx_val = idx_rand[training_samples:]

adj = torch.FloatTensor(np.array(adj.todense()))
features = torch.FloatTensor(np.array(features.todense()))
labels = torch.LongTensor(np.where(labels)[1])

idx_train = torch.LongTensor(idx_train)
idx_val = torch.LongTensor(idx_val)

return adj, features, labels, idx_train, idx_val

def normalize_adj(mx):
    """Symmetric normalization"""
    rowsum = np.array(mx.sum(1))
    r_inv_sqrt = np.power(rowsum, -0.5).flatten()
    r_inv_sqrt[np.isinf(r_inv_sqrt)] = 0.
    r_mat_inv_sqrt = sp.diags(r_inv_sqrt)
    return mx.dot(r_mat_inv_sqrt).transpose().dot(r_mat_inv_sqrt)

def normalize(mx):
    """Row-normalize sparse matrix"""
    rowsum = np.array(mx.sum(1))
    r_inv = np.power(rowsum, -1).flatten()
    r_inv[np.isinf(r_inv)] = 0.
    r_mat_inv = sp.diags(r_inv)
    mx = r_mat_inv.dot(mx)
    return mx

def accuracy(output, labels):
    preds = output.max(1)[1].type_as(labels)
    correct = preds.eq(labels).double()
    correct = correct.sum()
    return correct / len(labels)

```

Check the data

```
In [ ]: adj, features, labels, idx_train, idx_val = load_data()
```

Loading cora dataset...

```
In [ ]: print(adj)
print(adj.shape)
```

```
tensor([[0.1667, 0.0000, 0.0000, ..., 0.0000, 0.0000, 0.0000],
        [0.0000, 0.5000, 0.0000, ..., 0.0000, 0.0000, 0.0000],
        [0.0000, 0.0000, 0.2000, ..., 0.0000, 0.0000, 0.0000],
        ...,
        [0.0000, 0.0000, 0.0000, ..., 0.2000, 0.0000, 0.0000],
        [0.0000, 0.0000, 0.0000, ..., 0.0000, 0.2000, 0.0000],
        [0.0000, 0.0000, 0.0000, ..., 0.0000, 0.0000, 0.2500]])
torch.Size([2708, 2708])
```

```
In [ ]: print(features)
print(features.shape)
```

```
tensor([[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., 0., 0.]])
torch.Size([2708, 1433])
```

```
In [ ]: print(labels)
print(labels.unique())
print(len(labels))
```

```
tensor([2, 5, 4, ..., 1, 0, 2])
tensor([0, 1, 2, 3, 4, 5, 6])
2708
```

```
In [ ]: print(len(idx_train))
print(len(idx_val))
```

```
140
2568
```

Part 1: Graph Attention Networks

Part 1.1: Graph attention layer

A GAT is made up of multiple such layers. In this section, you will implement a single graph attention layer. Similar to the `GraphConvolution()`, this

`GraphAttentionLayer()` module takes $\mathbf{h} = \{\overrightarrow{h}_1, \overrightarrow{h}_2, \dots, \overrightarrow{h}_N\}$ where $\overrightarrow{h}_i \in \mathbb{R}^F$ as input and outputs $\mathbf{h}' = \{\overrightarrow{h}'_1, \overrightarrow{h}'_2, \dots, \overrightarrow{h}'_N\}$, where $\overrightarrow{h}'_i \in \mathbb{R}^{F'}$. However, instead of weighing each neighbouring node based on the adjacency matrix, we will use self attention to learn the relative importance of each neighbouring node. Recall from HW4 where you are asked to write out the equation for single headed attention, here we will implement multi-headed attention, which involves the following steps:

The initial transformation

In GCN above, you have completed similar transformation. But here, we need to define a weight matrix and perform this transformation for each head: $\vec{s}_i^k = \mathbf{W}^k \vec{h}_i$. We will perform a single linear transformation and then split it up for each head later.

Note the input \vec{h} has shape `[n_nodes, in_features]` and \vec{s} has shape of `[n_nodes, n_heads * n_hidden]`. Remember to reshape \vec{s} has shape of `[n_nodes, n_heads, n_hidden]` for later uses. Note: set `bias=False` for this linear transformation.

attention score

We calculate these for each head k . Here for simplicity of the notation, we omit k in the following equations. The attention scores are defined as the follows:

$e_{ij} = a(\vec{\mathbf{W}}\vec{h}_i, \vec{\mathbf{W}}\vec{h}_j) = a(\vec{s}_i, \vec{s}_j)$, where e_{ij} is the attention score (importance) of node j to node i . We will have to calculate this for each head. a is the attention mechanism, that calculates the attention score. The paper concatenates \vec{s}_i, \vec{s}_j and does a linear transformation with a weight vector $\mathbf{a} \in \mathbb{R}^{2F'}$ followed by a LeakyReLU.

$$e_{ij} = \text{LeakyReLU}\left(\mathbf{a}^\top \left[\vec{s}_i \| \vec{s}_j\right]\right)$$

How to vectorize this? Some hints:

1. `tensor.repeat()` gives you $\{\vec{s}_1, \vec{s}_2, \dots, \vec{s}_N, \vec{s}_1, \vec{s}_2, \dots, \vec{s}_N, \dots\}$.
2. `tensor.repeat_interleave()` gives you $\{\vec{s}_1, \vec{s}_1, \dots, \vec{s}_1, \vec{s}_2, \vec{s}_2, \dots, \vec{s}_2, \dots\}$.
3. concatenate to get $\left[\vec{s}_i \| \vec{s}_j\right]$ for all pairs of i, j . Reshape $\vec{s}_i \| \vec{s}_j$ has shape of `[n_nodes, n_nodes, n_heads, 2 * n_hidden]`
4. apply the attention layer and non-linear activation function to get
 $e_{ij} = \text{LeakyReLU}\left(\mathbf{a}^\top \left[\vec{s}_i \| \vec{s}_j\right]\right)$, where \mathbf{a}^\top is a single linear transformation that maps from dimension `n_hidden * 2` to `1`. Note: set the `bias=False` for this linear transformation. \mathbf{a} is of shape `[n_nodes, n_nodes, n_heads, 1]`. Remove the last dimension `1` using `squeeze()`.

Perform softmax

First, we need to mask e_{ij} based on adjacency matrix. We only need to sum over the neighbouring nodes for the attention calculation. Set the elements in e_{ij} to $-\infty$ if there is no edge from i to j for the softmax calculation. We need to do this for all heads and the adjacency matrix is the same for each head. Use `tensor.masked_fill()` to mask e_{ij} based on adjacency matrix for all heads.

Hint: reshape the adjacency matrix to `[n_nodes, n_nodes, 1]` using `unsqueeze()`. Now we are ready to normalize attention scores (or coefficients)

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}$$

Apply dropout

Apply the dropout layer. (this step is easy)

Calculate final output for each head

$$\vec{h}_i'^k = \sum_{j \in \mathcal{N}_i} \alpha_{ij}^k \vec{s}_j^k$$

Concat or Mean

Finally we concateneate the transformed features: $\vec{h}_i' = \left\| \begin{array}{c} \vec{h}_i'^1 \\ \vdots \\ \vec{h}_i'^K \end{array} \right\|$. In the code, we only need to reshape the tensor to shape of `[n_nodes, n_heads * n_hidden]`. Note that if it is the final layer, then it doesn't make sense to do concatenation anymore. Instead, we sum over the `n_heads` dimension: $\vec{h}_i' = \frac{1}{K} \sum_{k=1}^K \vec{h}_i'^k$.

```
In [ ]: class GraphAttentionLayer(nn.Module):

    def __init__(self, in_features: int, out_features: int, n_heads: int,
                 is_concat: bool = True,
                 dropout: float = 0.6,
                 alpha: float = 0.2):
        """
        in_features: F, the number of input features per node
        out_features: F', the number of output features per node
        n_heads: K, the number of attention heads
        is_concat: whether the multi-head results should be concatenated
        dropout: the dropout probability
        alpha: the negative slope for leaky relu activation
        """
        super(GraphAttentionLayer, self).__init__()

        self.is_concat = is_concat
        self.n_heads = n_heads

        if is_concat:
            assert out_features % n_heads == 0
            self.n_hidden = out_features // n_heads
        else:
            self.n_hidden = out_features

        # TODO: initialize the following modules:
        # (1) self.W: Linear layer that transform the input feature before
        # You should NOT use for loops for the multiheaded implementation
        # (2) self.attention: Linear layer that compute the attention scores
        # (3) self.activation: Activation function (LeakyReLU with negative slope)
        # (4) self.softmax: Softmax function (what's the dim to compute the softmax)
```

```

# (5) self.dropout_layer: Dropout function(with ratio=dropout)
#####
# your code here #####
self.W = nn.Linear(in_features, n_heads * self.n_hidden, bias=False)
self.attention = nn.Linear(2 * self.n_hidden, 1, bias=False)
self.activation = nn.LeakyReLU(negative_slope=alpha)
self.softmax = nn.Softmax(dim=1)
self.dropout_layer = nn.Dropout(p=dropout)

#####

def forward(self, h: torch.Tensor, adj_mat: torch.Tensor):
    # Number of nodes
    n_nodes = h.shape[0]

    # TODO:
    # (1) calculate s = Wh and reshape it to [n_nodes, n_heads, n_hid
    #     (you can use tensor.view() function)
    # (2) get [s_i || s_j] using tensor.repeat(), repeat_interleave()
    # (3) apply the attention layer
    # (4) apply the activation layer (you will get the attention scor
    # (5) remove the last dimension 1 use tensor.squeeze()
    # (6) mask the attention score with the adjacency matrix (if ther
    #     note: check the dimensions of e and your adjacency matrix.
    # (7) apply softmax
    # (8) apply dropout_layer
    #####
    s = self.W(h)
    s = s.view(n_nodes, self.n_heads, self.n_hidden)
    s_i = s.repeat_interleave(n_nodes, dim=0)
    s_i = s_i.view(n_nodes, n_nodes, self.n_heads, self.n_hidden)
    s_j = s.repeat(n_nodes, 1, 1)
    s_j = s_j.view(n_nodes, n_nodes, self.n_heads, self.n_hidden)
    s_cat = torch.cat([s_i, s_j], dim=-1)
    e = self.attention(s_cat)
    e = self.activation(e)
    e = e.squeeze(-1)
    mask = adj_mat.unsqueeze(-1)
    e = e.masked_fill(mask == 0, float('-inf'))
    a = self.softmax(e)
    a = self.dropout_layer(a)

#####

# Summation
h_prime = torch.einsum('ijh,jhf->ihf', a, s) #[n_nodes, n_heads,

# TODO: Concat or Mean
# Concatenate the heads
if self.is_concat:
    #####
    h_prime = h_prime.reshape(n_nodes, self.n_heads * self.n_hidd
    #####
# Take the mean of the heads (for the last layer)
else:

```

```
#####
# Your code here #####
h_prime = h_prime.mean(dim=1)

#####
# Return the result #####
return h_prime
```

Part 1.2: Define GAT network

it's really similar to how we defined GCN. You can follow the paper to set the number of attention layers and the activation function

```
In [ ]: class GAT(nn.Module):

    def __init__(self, nfeat: int, n_hidden: int, n_classes: int, n_heads
                 ):
        """
        in_features: the number of features per node
        n_hidden: the number of features in the first graph attention layer
        n_classes: the number of classes
        n_heads: the number of heads in the graph attention layers
        dropout: the dropout probability
        alpha: the negative input slope for leaky ReLU of the attention layer
        """
        super().__init__()

        # TODO: Set your GraphAttentionLayers, activation function and pool
        self.dropout = dropout

        self.layer1 = GraphAttentionLayer(
            in_features=nfeat,
            out_features=n_hidden,
            n_heads=n_heads,
            is_concat=True,
            dropout=dropout,
            alpha=alpha
        )

        self.elu = nn.ELU()

        self.layer2 = GraphAttentionLayer(
            in_features=n_hidden,
            out_features=n_classes,
            n_heads=n_heads,
            is_concat=False,
            dropout=dropout,
            alpha=alpha
        )

        self.dropout_layer = nn.Dropout(p=dropout)

    def forward(self, x: torch.Tensor, adj_mat: torch.Tensor):
        """
        x: the features vectors
        adj_mat: the adjacency matrix
        """
        # TODO: implement the forward function
        x = self.dropout_layer(x)
```

```

x = self.layer1(x, adj_mat)
x = self.elu(x)
x = self.dropout_layer(x)
x = self.layer2(x, adj_mat)
x = F.log_softmax(x, dim=1)
return x

```

Part 1.3: Training GAT

Play with the structure and the experiment settings and report your results as stated.

```
In [ ]: # TODO: SET YOUR HYPERPARAMETERS HERE
args = {
    "training_samples": 140,
    "epochs": 1000,
    "lr": 0.005,
    "weight_decay": 0.0005,
    "dropout": 0.6,
    "n_heads": 8,
    "n_hidden": 8,
    "alpha": 0.2,
    "n_classes": 7,
}
```

```
In [ ]: # TODO: SET YOUR MODE AND OPTIMIZER HERE
model = GAT(
    nfeat=features.shape[1],
    n_hidden=args["n_hidden"],
    n_classes=args["n_classes"],
    n_heads=args["n_heads"],
    dropout=args["dropout"],
    alpha=args["alpha"]
).to(device)

optimizer = torch.optim.Adam(
    model.parameters(),
    lr=args["lr"],
    weight_decay=args["weight_decay"]
)

criterion = nn.CrossEntropyLoss()

adj, features, labels, idx_train, idx_val = load_data(training_samples=args["training_samples"], val_size=500)
adj, features, labels, idx_train, idx_val = adj.to(device), features.to(device), labels.to(device), idx_train.to(device), idx_val.to(device)
```

Loading cora dataset...

```
In [ ]: def train(epoch):
    model.train()
    optimizer.zero_grad()
    output = model(features, adj)
    train_loss = criterion(output[idx_train], labels[idx_train])
    train_acc = accuracy(output[idx_train], labels[idx_train])
    train_loss.backward()
    optimizer.step()

    model.eval()
```

```
with torch.no_grad():
    output_validation = model(features, adj)
    validation_loss = criterion(output_validation[idx_val], labels[idx_val])
    validation_acc = accuracy(output_validation[idx_val], labels[idx_val])

    print(f"Epoch {epoch+1}/{args['epochs']}, Train Loss: {train_loss.item():.4f}, Validation Loss: {validation_loss.item():.4f}, Validation Acc: {validation_acc.item():.4f}")

def test():
    model.eval()
    output = model(features, adj)
    test_loss = criterion(output[idx_val], labels[idx_val])
    test_acc = accuracy(output[idx_val], labels[idx_val])
    preds = output[idx_val].argmax(dim=1)

    print(f"Test Loss: {test_loss.item():.4f}, Test Acc: {test_acc.item():.4f}")
    return test_loss.item(), test_acc.item(), preds
```

```
In [ ]: # Train model
t_total = time.time()
for epoch in range(args["epochs"]):
    train(epoch)
print("Optimization Finished!")
print("Total time elapsed: {:.4f}s".format(time.time() - t_total))

# Testing
test()
```

Epoch 1/1000, Train Loss: 1.9460, Train Acc: 0.1929, Validation Loss: 1.9457, Validation Acc: 0.3092
Epoch 2/1000, Train Loss: 1.9456, Train Acc: 0.1500, Validation Loss: 1.9455, Validation Acc: 0.3820
Epoch 3/1000, Train Loss: 1.9454, Train Acc: 0.2214, Validation Loss: 1.9453, Validation Acc: 0.3851
Epoch 4/1000, Train Loss: 1.9449, Train Acc: 0.2786, Validation Loss: 1.9451, Validation Acc: 0.3843
Epoch 5/1000, Train Loss: 1.9449, Train Acc: 0.2786, Validation Loss: 1.9449, Validation Acc: 0.3773
Epoch 6/1000, Train Loss: 1.9442, Train Acc: 0.3500, Validation Loss: 1.9446, Validation Acc: 0.3610
Epoch 7/1000, Train Loss: 1.9440, Train Acc: 0.3071, Validation Loss: 1.9444, Validation Acc: 0.3462
Epoch 8/1000, Train Loss: 1.9435, Train Acc: 0.3857, Validation Loss: 1.9441, Validation Acc: 0.3396
Epoch 9/1000, Train Loss: 1.9432, Train Acc: 0.3429, Validation Loss: 1.9438, Validation Acc: 0.3361
Epoch 10/1000, Train Loss: 1.9429, Train Acc: 0.3357, Validation Loss: 1.9435, Validation Acc: 0.3326
Epoch 11/1000, Train Loss: 1.9425, Train Acc: 0.3929, Validation Loss: 1.9431, Validation Acc: 0.3306
Epoch 12/1000, Train Loss: 1.9412, Train Acc: 0.3357, Validation Loss: 1.9427, Validation Acc: 0.3271
Epoch 13/1000, Train Loss: 1.9403, Train Acc: 0.3857, Validation Loss: 1.9422, Validation Acc: 0.3248
Epoch 14/1000, Train Loss: 1.9400, Train Acc: 0.4286, Validation Loss: 1.9417, Validation Acc: 0.3213
Epoch 15/1000, Train Loss: 1.9397, Train Acc: 0.3714, Validation Loss: 1.9411, Validation Acc: 0.3181
Epoch 16/1000, Train Loss: 1.9398, Train Acc: 0.3786, Validation Loss: 1.9405, Validation Acc: 0.3154
Epoch 17/1000, Train Loss: 1.9379, Train Acc: 0.3929, Validation Loss: 1.9399, Validation Acc: 0.3127
Epoch 18/1000, Train Loss: 1.9373, Train Acc: 0.3786, Validation Loss: 1.9392, Validation Acc: 0.3111
Epoch 19/1000, Train Loss: 1.9369, Train Acc: 0.3929, Validation Loss: 1.9384, Validation Acc: 0.3107
Epoch 20/1000, Train Loss: 1.9358, Train Acc: 0.3571, Validation Loss: 1.9376, Validation Acc: 0.3084
Epoch 21/1000, Train Loss: 1.9349, Train Acc: 0.3929, Validation Loss: 1.9368, Validation Acc: 0.3053
Epoch 22/1000, Train Loss: 1.9307, Train Acc: 0.4000, Validation Loss: 1.9359, Validation Acc: 0.3045
Epoch 23/1000, Train Loss: 1.9313, Train Acc: 0.3571, Validation Loss: 1.9349, Validation Acc: 0.3033
Epoch 24/1000, Train Loss: 1.9316, Train Acc: 0.3571, Validation Loss: 1.9338, Validation Acc: 0.3030
Epoch 25/1000, Train Loss: 1.9296, Train Acc: 0.3500, Validation Loss: 1.9327, Validation Acc: 0.3030
Epoch 26/1000, Train Loss: 1.9279, Train Acc: 0.3714, Validation Loss: 1.9315, Validation Acc: 0.3026
Epoch 27/1000, Train Loss: 1.9267, Train Acc: 0.3643, Validation Loss: 1.9303, Validation Acc: 0.3026
Epoch 28/1000, Train Loss: 1.9233, Train Acc: 0.3286, Validation Loss: 1.9290, Validation Acc: 0.3026
Epoch 29/1000, Train Loss: 1.9237, Train Acc: 0.3714, Validation Loss: 1.9276, Validation Acc: 0.3026
Epoch 30/1000, Train Loss: 1.9218, Train Acc: 0.3643, Validation Loss: 1.9261, Validation Acc: 0.3026

Epoch 31/1000, Train Loss: 1.9185, Train Acc: 0.3857, Validation Loss: 1.9246, Validation Acc: 0.3033
Epoch 32/1000, Train Loss: 1.9198, Train Acc: 0.3643, Validation Loss: 1.9231, Validation Acc: 0.3033
Epoch 33/1000, Train Loss: 1.9160, Train Acc: 0.3786, Validation Loss: 1.9214, Validation Acc: 0.3037
Epoch 34/1000, Train Loss: 1.9067, Train Acc: 0.3714, Validation Loss: 1.9197, Validation Acc: 0.3037
Epoch 35/1000, Train Loss: 1.9151, Train Acc: 0.3571, Validation Loss: 1.9179, Validation Acc: 0.3041
Epoch 36/1000, Train Loss: 1.9088, Train Acc: 0.3429, Validation Loss: 1.9160, Validation Acc: 0.3041
Epoch 37/1000, Train Loss: 1.9068, Train Acc: 0.3714, Validation Loss: 1.9140, Validation Acc: 0.3041
Epoch 38/1000, Train Loss: 1.9030, Train Acc: 0.3714, Validation Loss: 1.9120, Validation Acc: 0.3045
Epoch 39/1000, Train Loss: 1.8928, Train Acc: 0.3929, Validation Loss: 1.9098, Validation Acc: 0.3045
Epoch 40/1000, Train Loss: 1.8949, Train Acc: 0.3857, Validation Loss: 1.9075, Validation Acc: 0.3045
Epoch 41/1000, Train Loss: 1.8943, Train Acc: 0.3857, Validation Loss: 1.9052, Validation Acc: 0.3045
Epoch 42/1000, Train Loss: 1.8895, Train Acc: 0.4000, Validation Loss: 1.9028, Validation Acc: 0.3045
Epoch 43/1000, Train Loss: 1.8907, Train Acc: 0.3714, Validation Loss: 1.9003, Validation Acc: 0.3045
Epoch 44/1000, Train Loss: 1.8809, Train Acc: 0.3714, Validation Loss: 1.8977, Validation Acc: 0.3045
Epoch 45/1000, Train Loss: 1.8795, Train Acc: 0.3786, Validation Loss: 1.8950, Validation Acc: 0.3045
Epoch 46/1000, Train Loss: 1.8764, Train Acc: 0.3786, Validation Loss: 1.8923, Validation Acc: 0.3045
Epoch 47/1000, Train Loss: 1.8646, Train Acc: 0.3929, Validation Loss: 1.8894, Validation Acc: 0.3045
Epoch 48/1000, Train Loss: 1.8676, Train Acc: 0.3857, Validation Loss: 1.8865, Validation Acc: 0.3045
Epoch 49/1000, Train Loss: 1.8514, Train Acc: 0.3714, Validation Loss: 1.8834, Validation Acc: 0.3045
Epoch 50/1000, Train Loss: 1.8543, Train Acc: 0.3786, Validation Loss: 1.8803, Validation Acc: 0.3041
Epoch 51/1000, Train Loss: 1.8480, Train Acc: 0.3286, Validation Loss: 1.8771, Validation Acc: 0.3041
Epoch 52/1000, Train Loss: 1.8298, Train Acc: 0.3857, Validation Loss: 1.8737, Validation Acc: 0.3041
Epoch 53/1000, Train Loss: 1.8246, Train Acc: 0.3500, Validation Loss: 1.8702, Validation Acc: 0.3041
Epoch 54/1000, Train Loss: 1.8458, Train Acc: 0.3357, Validation Loss: 1.8667, Validation Acc: 0.3041
Epoch 55/1000, Train Loss: 1.8297, Train Acc: 0.3714, Validation Loss: 1.8632, Validation Acc: 0.3041
Epoch 56/1000, Train Loss: 1.8371, Train Acc: 0.3500, Validation Loss: 1.8596, Validation Acc: 0.3041
Epoch 57/1000, Train Loss: 1.8243, Train Acc: 0.3571, Validation Loss: 1.8560, Validation Acc: 0.3037
Epoch 58/1000, Train Loss: 1.8053, Train Acc: 0.3714, Validation Loss: 1.8523, Validation Acc: 0.3037
Epoch 59/1000, Train Loss: 1.8097, Train Acc: 0.3571, Validation Loss: 1.8485, Validation Acc: 0.3037
Epoch 60/1000, Train Loss: 1.8097, Train Acc: 0.3786, Validation Loss: 1.8447, Validation Acc: 0.3037

Epoch 61/1000, Train Loss: 1.7990, Train Acc: 0.3857, Validation Loss: 1.8409, Validation Acc: 0.3037
Epoch 62/1000, Train Loss: 1.7941, Train Acc: 0.3714, Validation Loss: 1.8370, Validation Acc: 0.3037
Epoch 63/1000, Train Loss: 1.7948, Train Acc: 0.3571, Validation Loss: 1.8331, Validation Acc: 0.3041
Epoch 64/1000, Train Loss: 1.7782, Train Acc: 0.3786, Validation Loss: 1.8292, Validation Acc: 0.3041
Epoch 65/1000, Train Loss: 1.7840, Train Acc: 0.3857, Validation Loss: 1.8252, Validation Acc: 0.3045
Epoch 66/1000, Train Loss: 1.7712, Train Acc: 0.3714, Validation Loss: 1.8211, Validation Acc: 0.3049
Epoch 67/1000, Train Loss: 1.7601, Train Acc: 0.3571, Validation Loss: 1.8170, Validation Acc: 0.3061
Epoch 68/1000, Train Loss: 1.7402, Train Acc: 0.3571, Validation Loss: 1.8129, Validation Acc: 0.3076
Epoch 69/1000, Train Loss: 1.7708, Train Acc: 0.3429, Validation Loss: 1.8087, Validation Acc: 0.3076
Epoch 70/1000, Train Loss: 1.7435, Train Acc: 0.3500, Validation Loss: 1.8046, Validation Acc: 0.3080
Epoch 71/1000, Train Loss: 1.7274, Train Acc: 0.3714, Validation Loss: 1.8003, Validation Acc: 0.3080
Epoch 72/1000, Train Loss: 1.7013, Train Acc: 0.3500, Validation Loss: 1.7960, Validation Acc: 0.3080
Epoch 73/1000, Train Loss: 1.6803, Train Acc: 0.3571, Validation Loss: 1.7916, Validation Acc: 0.3092
Epoch 74/1000, Train Loss: 1.7189, Train Acc: 0.3786, Validation Loss: 1.7873, Validation Acc: 0.3092
Epoch 75/1000, Train Loss: 1.7076, Train Acc: 0.3714, Validation Loss: 1.7829, Validation Acc: 0.3092
Epoch 76/1000, Train Loss: 1.6995, Train Acc: 0.3500, Validation Loss: 1.7785, Validation Acc: 0.3092
Epoch 77/1000, Train Loss: 1.6947, Train Acc: 0.4000, Validation Loss: 1.7741, Validation Acc: 0.3096
Epoch 78/1000, Train Loss: 1.6695, Train Acc: 0.3857, Validation Loss: 1.7697, Validation Acc: 0.3100
Epoch 79/1000, Train Loss: 1.6560, Train Acc: 0.3929, Validation Loss: 1.7654, Validation Acc: 0.3100
Epoch 80/1000, Train Loss: 1.6713, Train Acc: 0.3786, Validation Loss: 1.7610, Validation Acc: 0.3100
Epoch 81/1000, Train Loss: 1.6649, Train Acc: 0.3786, Validation Loss: 1.7567, Validation Acc: 0.3100
Epoch 82/1000, Train Loss: 1.6341, Train Acc: 0.3786, Validation Loss: 1.7523, Validation Acc: 0.3104
Epoch 83/1000, Train Loss: 1.6426, Train Acc: 0.3571, Validation Loss: 1.7481, Validation Acc: 0.3111
Epoch 84/1000, Train Loss: 1.6199, Train Acc: 0.3500, Validation Loss: 1.7438, Validation Acc: 0.3119
Epoch 85/1000, Train Loss: 1.6198, Train Acc: 0.4071, Validation Loss: 1.7396, Validation Acc: 0.3143
Epoch 86/1000, Train Loss: 1.6472, Train Acc: 0.3857, Validation Loss: 1.7355, Validation Acc: 0.3174
Epoch 87/1000, Train Loss: 1.6112, Train Acc: 0.3857, Validation Loss: 1.7315, Validation Acc: 0.3181
Epoch 88/1000, Train Loss: 1.6246, Train Acc: 0.4000, Validation Loss: 1.7274, Validation Acc: 0.3189
Epoch 89/1000, Train Loss: 1.6144, Train Acc: 0.3786, Validation Loss: 1.7234, Validation Acc: 0.3189
Epoch 90/1000, Train Loss: 1.6507, Train Acc: 0.3357, Validation Loss: 1.7194, Validation Acc: 0.3197

Epoch 91/1000, Train Loss: 1.6131, Train Acc: 0.3857, Validation Loss: 1.7155, Validation Acc: 0.3217
Epoch 92/1000, Train Loss: 1.5437, Train Acc: 0.3857, Validation Loss: 1.7116, Validation Acc: 0.3232
Epoch 93/1000, Train Loss: 1.5565, Train Acc: 0.4429, Validation Loss: 1.7077, Validation Acc: 0.3252
Epoch 94/1000, Train Loss: 1.5903, Train Acc: 0.3857, Validation Loss: 1.7039, Validation Acc: 0.3290
Epoch 95/1000, Train Loss: 1.5581, Train Acc: 0.4357, Validation Loss: 1.7000, Validation Acc: 0.3333
Epoch 96/1000, Train Loss: 1.5633, Train Acc: 0.4143, Validation Loss: 1.6962, Validation Acc: 0.3357
Epoch 97/1000, Train Loss: 1.5507, Train Acc: 0.4143, Validation Loss: 1.6925, Validation Acc: 0.3372
Epoch 98/1000, Train Loss: 1.5260, Train Acc: 0.4143, Validation Loss: 1.6888, Validation Acc: 0.3431
Epoch 99/1000, Train Loss: 1.5466, Train Acc: 0.4143, Validation Loss: 1.6851, Validation Acc: 0.3481
Epoch 100/1000, Train Loss: 1.5596, Train Acc: 0.4071, Validation Loss: 1.6815, Validation Acc: 0.3532
Epoch 101/1000, Train Loss: 1.5864, Train Acc: 0.4357, Validation Loss: 1.6779, Validation Acc: 0.3590
Epoch 102/1000, Train Loss: 1.5275, Train Acc: 0.4000, Validation Loss: 1.6744, Validation Acc: 0.3625
Epoch 103/1000, Train Loss: 1.5193, Train Acc: 0.4214, Validation Loss: 1.6709, Validation Acc: 0.3695
Epoch 104/1000, Train Loss: 1.5041, Train Acc: 0.4357, Validation Loss: 1.6674, Validation Acc: 0.3731
Epoch 105/1000, Train Loss: 1.5220, Train Acc: 0.4286, Validation Loss: 1.6639, Validation Acc: 0.3812
Epoch 106/1000, Train Loss: 1.5273, Train Acc: 0.4571, Validation Loss: 1.6604, Validation Acc: 0.3886
Epoch 107/1000, Train Loss: 1.4626, Train Acc: 0.4857, Validation Loss: 1.6570, Validation Acc: 0.3960
Epoch 108/1000, Train Loss: 1.5047, Train Acc: 0.4571, Validation Loss: 1.6537, Validation Acc: 0.4026
Epoch 109/1000, Train Loss: 1.4787, Train Acc: 0.4786, Validation Loss: 1.6503, Validation Acc: 0.4081
Epoch 110/1000, Train Loss: 1.4702, Train Acc: 0.4714, Validation Loss: 1.6468, Validation Acc: 0.4178
Epoch 111/1000, Train Loss: 1.4983, Train Acc: 0.4643, Validation Loss: 1.6435, Validation Acc: 0.4268
Epoch 112/1000, Train Loss: 1.4973, Train Acc: 0.4143, Validation Loss: 1.6403, Validation Acc: 0.4357
Epoch 113/1000, Train Loss: 1.4907, Train Acc: 0.4929, Validation Loss: 1.6371, Validation Acc: 0.4400
Epoch 114/1000, Train Loss: 1.4912, Train Acc: 0.4000, Validation Loss: 1.6340, Validation Acc: 0.4509
Epoch 115/1000, Train Loss: 1.5377, Train Acc: 0.4643, Validation Loss: 1.6311, Validation Acc: 0.4614
Epoch 116/1000, Train Loss: 1.5007, Train Acc: 0.4857, Validation Loss: 1.6282, Validation Acc: 0.4716
Epoch 117/1000, Train Loss: 1.4644, Train Acc: 0.5071, Validation Loss: 1.6253, Validation Acc: 0.4790
Epoch 118/1000, Train Loss: 1.4270, Train Acc: 0.4929, Validation Loss: 1.6224, Validation Acc: 0.4899
Epoch 119/1000, Train Loss: 1.4666, Train Acc: 0.5071, Validation Loss: 1.6195, Validation Acc: 0.4977
Epoch 120/1000, Train Loss: 1.4874, Train Acc: 0.4857, Validation Loss: 1.6166, Validation Acc: 0.5070

Epoch 121/1000, Train Loss: 1.4519, Train Acc: 0.5214, Validation Loss: 1.6137, Validation Acc: 0.5148
Epoch 122/1000, Train Loss: 1.4110, Train Acc: 0.5000, Validation Loss: 1.6108, Validation Acc: 0.5202
Epoch 123/1000, Train Loss: 1.4182, Train Acc: 0.5143, Validation Loss: 1.6079, Validation Acc: 0.5269
Epoch 124/1000, Train Loss: 1.4905, Train Acc: 0.5500, Validation Loss: 1.6051, Validation Acc: 0.5350
Epoch 125/1000, Train Loss: 1.4309, Train Acc: 0.5643, Validation Loss: 1.6022, Validation Acc: 0.5393
Epoch 126/1000, Train Loss: 1.3999, Train Acc: 0.5714, Validation Loss: 1.5993, Validation Acc: 0.5471
Epoch 127/1000, Train Loss: 1.4398, Train Acc: 0.5571, Validation Loss: 1.5965, Validation Acc: 0.5522
Epoch 128/1000, Train Loss: 1.4335, Train Acc: 0.5429, Validation Loss: 1.5936, Validation Acc: 0.5553
Epoch 129/1000, Train Loss: 1.3627, Train Acc: 0.5071, Validation Loss: 1.5908, Validation Acc: 0.5615
Epoch 130/1000, Train Loss: 1.3614, Train Acc: 0.6000, Validation Loss: 1.5880, Validation Acc: 0.5674
Epoch 131/1000, Train Loss: 1.4060, Train Acc: 0.5714, Validation Loss: 1.5851, Validation Acc: 0.5748
Epoch 132/1000, Train Loss: 1.3885, Train Acc: 0.5786, Validation Loss: 1.5823, Validation Acc: 0.5798
Epoch 133/1000, Train Loss: 1.4146, Train Acc: 0.5643, Validation Loss: 1.5796, Validation Acc: 0.5833
Epoch 134/1000, Train Loss: 1.3659, Train Acc: 0.6000, Validation Loss: 1.5768, Validation Acc: 0.5907
Epoch 135/1000, Train Loss: 1.3804, Train Acc: 0.6143, Validation Loss: 1.5741, Validation Acc: 0.5954
Epoch 136/1000, Train Loss: 1.3602, Train Acc: 0.5929, Validation Loss: 1.5715, Validation Acc: 0.6012
Epoch 137/1000, Train Loss: 1.3668, Train Acc: 0.5929, Validation Loss: 1.5689, Validation Acc: 0.6044
Epoch 138/1000, Train Loss: 1.3715, Train Acc: 0.5786, Validation Loss: 1.5663, Validation Acc: 0.6086
Epoch 139/1000, Train Loss: 1.3843, Train Acc: 0.5857, Validation Loss: 1.5636, Validation Acc: 0.6125
Epoch 140/1000, Train Loss: 1.3591, Train Acc: 0.6143, Validation Loss: 1.5609, Validation Acc: 0.6149
Epoch 141/1000, Train Loss: 1.3543, Train Acc: 0.5571, Validation Loss: 1.5583, Validation Acc: 0.6192
Epoch 142/1000, Train Loss: 1.3924, Train Acc: 0.6000, Validation Loss: 1.5556, Validation Acc: 0.6219
Epoch 143/1000, Train Loss: 1.4179, Train Acc: 0.6286, Validation Loss: 1.5529, Validation Acc: 0.6254
Epoch 144/1000, Train Loss: 1.3602, Train Acc: 0.6571, Validation Loss: 1.5503, Validation Acc: 0.6285
Epoch 145/1000, Train Loss: 1.3193, Train Acc: 0.6286, Validation Loss: 1.5476, Validation Acc: 0.6308
Epoch 146/1000, Train Loss: 1.3348, Train Acc: 0.6214, Validation Loss: 1.5450, Validation Acc: 0.6336
Epoch 147/1000, Train Loss: 1.3551, Train Acc: 0.6000, Validation Loss: 1.5423, Validation Acc: 0.6375
Epoch 148/1000, Train Loss: 1.3470, Train Acc: 0.6286, Validation Loss: 1.5398, Validation Acc: 0.6402
Epoch 149/1000, Train Loss: 1.3217, Train Acc: 0.6000, Validation Loss: 1.5373, Validation Acc: 0.6445
Epoch 150/1000, Train Loss: 1.2831, Train Acc: 0.6643, Validation Loss: 1.5347, Validation Acc: 0.6499

Epoch 151/1000, Train Loss: 1.3221, Train Acc: 0.6357, Validation Loss: 1.5322, Validation Acc: 0.6515
Epoch 152/1000, Train Loss: 1.2656, Train Acc: 0.6571, Validation Loss: 1.5296, Validation Acc: 0.6542
Epoch 153/1000, Train Loss: 1.3536, Train Acc: 0.6000, Validation Loss: 1.5270, Validation Acc: 0.6562
Epoch 154/1000, Train Loss: 1.3505, Train Acc: 0.6571, Validation Loss: 1.5246, Validation Acc: 0.6569
Epoch 155/1000, Train Loss: 1.3540, Train Acc: 0.5857, Validation Loss: 1.5222, Validation Acc: 0.6589
Epoch 156/1000, Train Loss: 1.2967, Train Acc: 0.6214, Validation Loss: 1.5198, Validation Acc: 0.6597
Epoch 157/1000, Train Loss: 1.3277, Train Acc: 0.6143, Validation Loss: 1.5174, Validation Acc: 0.6612
Epoch 158/1000, Train Loss: 1.3222, Train Acc: 0.6714, Validation Loss: 1.5150, Validation Acc: 0.6632
Epoch 159/1000, Train Loss: 1.2926, Train Acc: 0.6286, Validation Loss: 1.5127, Validation Acc: 0.6643
Epoch 160/1000, Train Loss: 1.2630, Train Acc: 0.6929, Validation Loss: 1.5105, Validation Acc: 0.6663
Epoch 161/1000, Train Loss: 1.2841, Train Acc: 0.6429, Validation Loss: 1.5083, Validation Acc: 0.6678
Epoch 162/1000, Train Loss: 1.2855, Train Acc: 0.6071, Validation Loss: 1.5061, Validation Acc: 0.6690
Epoch 163/1000, Train Loss: 1.2534, Train Acc: 0.6429, Validation Loss: 1.5039, Validation Acc: 0.6694
Epoch 164/1000, Train Loss: 1.2882, Train Acc: 0.6143, Validation Loss: 1.5018, Validation Acc: 0.6698
Epoch 165/1000, Train Loss: 1.2693, Train Acc: 0.6214, Validation Loss: 1.4996, Validation Acc: 0.6717
Epoch 166/1000, Train Loss: 1.2499, Train Acc: 0.6286, Validation Loss: 1.4974, Validation Acc: 0.6733
Epoch 167/1000, Train Loss: 1.2822, Train Acc: 0.6286, Validation Loss: 1.4952, Validation Acc: 0.6764
Epoch 168/1000, Train Loss: 1.3030, Train Acc: 0.5929, Validation Loss: 1.4929, Validation Acc: 0.6764
Epoch 169/1000, Train Loss: 1.2932, Train Acc: 0.6214, Validation Loss: 1.4907, Validation Acc: 0.6772
Epoch 170/1000, Train Loss: 1.2498, Train Acc: 0.6643, Validation Loss: 1.4884, Validation Acc: 0.6772
Epoch 171/1000, Train Loss: 1.3043, Train Acc: 0.6857, Validation Loss: 1.4861, Validation Acc: 0.6787
Epoch 172/1000, Train Loss: 1.2198, Train Acc: 0.7286, Validation Loss: 1.4838, Validation Acc: 0.6799
Epoch 173/1000, Train Loss: 1.2971, Train Acc: 0.6214, Validation Loss: 1.4815, Validation Acc: 0.6811
Epoch 174/1000, Train Loss: 1.2901, Train Acc: 0.6429, Validation Loss: 1.4794, Validation Acc: 0.6830
Epoch 175/1000, Train Loss: 1.2348, Train Acc: 0.6500, Validation Loss: 1.4773, Validation Acc: 0.6830
Epoch 176/1000, Train Loss: 1.2497, Train Acc: 0.6929, Validation Loss: 1.4753, Validation Acc: 0.6846
Epoch 177/1000, Train Loss: 1.3177, Train Acc: 0.6714, Validation Loss: 1.4735, Validation Acc: 0.6857
Epoch 178/1000, Train Loss: 1.3060, Train Acc: 0.6643, Validation Loss: 1.4717, Validation Acc: 0.6869
Epoch 179/1000, Train Loss: 1.2200, Train Acc: 0.6857, Validation Loss: 1.4700, Validation Acc: 0.6873
Epoch 180/1000, Train Loss: 1.2921, Train Acc: 0.6714, Validation Loss: 1.4684, Validation Acc: 0.6873

Epoch 181/1000, Train Loss: 1.2679, Train Acc: 0.7000, Validation Loss: 1.4667, Validation Acc: 0.6881
Epoch 182/1000, Train Loss: 1.2796, Train Acc: 0.6714, Validation Loss: 1.4651, Validation Acc: 0.6885
Epoch 183/1000, Train Loss: 1.2206, Train Acc: 0.6714, Validation Loss: 1.4634, Validation Acc: 0.6900
Epoch 184/1000, Train Loss: 1.2455, Train Acc: 0.6429, Validation Loss: 1.4617, Validation Acc: 0.6908
Epoch 185/1000, Train Loss: 1.1946, Train Acc: 0.7000, Validation Loss: 1.4599, Validation Acc: 0.6912
Epoch 186/1000, Train Loss: 1.2221, Train Acc: 0.7071, Validation Loss: 1.4581, Validation Acc: 0.6928
Epoch 187/1000, Train Loss: 1.3028, Train Acc: 0.6714, Validation Loss: 1.4563, Validation Acc: 0.6935
Epoch 188/1000, Train Loss: 1.1738, Train Acc: 0.6500, Validation Loss: 1.4544, Validation Acc: 0.6935
Epoch 189/1000, Train Loss: 1.2250, Train Acc: 0.7071, Validation Loss: 1.4526, Validation Acc: 0.6943
Epoch 190/1000, Train Loss: 1.1962, Train Acc: 0.7214, Validation Loss: 1.4508, Validation Acc: 0.6967
Epoch 191/1000, Train Loss: 1.2243, Train Acc: 0.6786, Validation Loss: 1.4489, Validation Acc: 0.6970
Epoch 192/1000, Train Loss: 1.1755, Train Acc: 0.7071, Validation Loss: 1.4470, Validation Acc: 0.6974
Epoch 193/1000, Train Loss: 1.2577, Train Acc: 0.6714, Validation Loss: 1.4451, Validation Acc: 0.6982
Epoch 194/1000, Train Loss: 1.2014, Train Acc: 0.7071, Validation Loss: 1.4431, Validation Acc: 0.6978
Epoch 195/1000, Train Loss: 1.1921, Train Acc: 0.7571, Validation Loss: 1.4412, Validation Acc: 0.6990
Epoch 196/1000, Train Loss: 1.1969, Train Acc: 0.7286, Validation Loss: 1.4393, Validation Acc: 0.7009
Epoch 197/1000, Train Loss: 1.1819, Train Acc: 0.6786, Validation Loss: 1.4374, Validation Acc: 0.7025
Epoch 198/1000, Train Loss: 1.1466, Train Acc: 0.7143, Validation Loss: 1.4354, Validation Acc: 0.7033
Epoch 199/1000, Train Loss: 1.2141, Train Acc: 0.7429, Validation Loss: 1.4334, Validation Acc: 0.7040
Epoch 200/1000, Train Loss: 1.2257, Train Acc: 0.6786, Validation Loss: 1.4315, Validation Acc: 0.7040
Epoch 201/1000, Train Loss: 1.1937, Train Acc: 0.6714, Validation Loss: 1.4296, Validation Acc: 0.7040
Epoch 202/1000, Train Loss: 1.1693, Train Acc: 0.7357, Validation Loss: 1.4276, Validation Acc: 0.7048
Epoch 203/1000, Train Loss: 1.1619, Train Acc: 0.7214, Validation Loss: 1.4256, Validation Acc: 0.7056
Epoch 204/1000, Train Loss: 1.1618, Train Acc: 0.7143, Validation Loss: 1.4236, Validation Acc: 0.7076
Epoch 205/1000, Train Loss: 1.1325, Train Acc: 0.7714, Validation Loss: 1.4216, Validation Acc: 0.7079
Epoch 206/1000, Train Loss: 1.1446, Train Acc: 0.7071, Validation Loss: 1.4196, Validation Acc: 0.7083
Epoch 207/1000, Train Loss: 1.1706, Train Acc: 0.7000, Validation Loss: 1.4175, Validation Acc: 0.7087
Epoch 208/1000, Train Loss: 1.1432, Train Acc: 0.7357, Validation Loss: 1.4154, Validation Acc: 0.7083
Epoch 209/1000, Train Loss: 1.1647, Train Acc: 0.7357, Validation Loss: 1.4133, Validation Acc: 0.7087
Epoch 210/1000, Train Loss: 1.1285, Train Acc: 0.6857, Validation Loss: 1.4111, Validation Acc: 0.7083

Epoch 211/1000, Train Loss: 1.2142, Train Acc: 0.7286, Validation Loss: 1.4090, Validation Acc: 0.7079
Epoch 212/1000, Train Loss: 1.1796, Train Acc: 0.7071, Validation Loss: 1.4068, Validation Acc: 0.7087
Epoch 213/1000, Train Loss: 1.2299, Train Acc: 0.6857, Validation Loss: 1.4048, Validation Acc: 0.7095
Epoch 214/1000, Train Loss: 1.1740, Train Acc: 0.6429, Validation Loss: 1.4028, Validation Acc: 0.7083
Epoch 215/1000, Train Loss: 1.1853, Train Acc: 0.6786, Validation Loss: 1.4007, Validation Acc: 0.7083
Epoch 216/1000, Train Loss: 1.1538, Train Acc: 0.6643, Validation Loss: 1.3988, Validation Acc: 0.7087
Epoch 217/1000, Train Loss: 1.1955, Train Acc: 0.7143, Validation Loss: 1.3970, Validation Acc: 0.7087
Epoch 218/1000, Train Loss: 1.0973, Train Acc: 0.7429, Validation Loss: 1.3951, Validation Acc: 0.7095
Epoch 219/1000, Train Loss: 1.2542, Train Acc: 0.6357, Validation Loss: 1.3934, Validation Acc: 0.7103
Epoch 220/1000, Train Loss: 1.1475, Train Acc: 0.7286, Validation Loss: 1.3919, Validation Acc: 0.7114
Epoch 221/1000, Train Loss: 1.1774, Train Acc: 0.7143, Validation Loss: 1.3903, Validation Acc: 0.7114
Epoch 222/1000, Train Loss: 1.0969, Train Acc: 0.6929, Validation Loss: 1.3888, Validation Acc: 0.7118
Epoch 223/1000, Train Loss: 1.1584, Train Acc: 0.7071, Validation Loss: 1.3873, Validation Acc: 0.7122
Epoch 224/1000, Train Loss: 1.1524, Train Acc: 0.6714, Validation Loss: 1.3857, Validation Acc: 0.7122
Epoch 225/1000, Train Loss: 1.1595, Train Acc: 0.6857, Validation Loss: 1.3841, Validation Acc: 0.7126
Epoch 226/1000, Train Loss: 1.1258, Train Acc: 0.6857, Validation Loss: 1.3826, Validation Acc: 0.7126
Epoch 227/1000, Train Loss: 1.1258, Train Acc: 0.7143, Validation Loss: 1.3811, Validation Acc: 0.7130
Epoch 228/1000, Train Loss: 1.0747, Train Acc: 0.7429, Validation Loss: 1.3796, Validation Acc: 0.7142
Epoch 229/1000, Train Loss: 1.1561, Train Acc: 0.7000, Validation Loss: 1.3779, Validation Acc: 0.7146
Epoch 230/1000, Train Loss: 1.1664, Train Acc: 0.6929, Validation Loss: 1.3764, Validation Acc: 0.7138
Epoch 231/1000, Train Loss: 1.1217, Train Acc: 0.7357, Validation Loss: 1.3748, Validation Acc: 0.7146
Epoch 232/1000, Train Loss: 1.1013, Train Acc: 0.7357, Validation Loss: 1.3732, Validation Acc: 0.7142
Epoch 233/1000, Train Loss: 1.1097, Train Acc: 0.6714, Validation Loss: 1.3716, Validation Acc: 0.7146
Epoch 234/1000, Train Loss: 1.2066, Train Acc: 0.6786, Validation Loss: 1.3700, Validation Acc: 0.7146
Epoch 235/1000, Train Loss: 1.0590, Train Acc: 0.7357, Validation Loss: 1.3684, Validation Acc: 0.7146
Epoch 236/1000, Train Loss: 1.0895, Train Acc: 0.7286, Validation Loss: 1.3669, Validation Acc: 0.7134
Epoch 237/1000, Train Loss: 1.0710, Train Acc: 0.7500, Validation Loss: 1.3653, Validation Acc: 0.7130
Epoch 238/1000, Train Loss: 1.1347, Train Acc: 0.7143, Validation Loss: 1.3636, Validation Acc: 0.7126
Epoch 239/1000, Train Loss: 1.1196, Train Acc: 0.7214, Validation Loss: 1.3619, Validation Acc: 0.7126
Epoch 240/1000, Train Loss: 1.0631, Train Acc: 0.7143, Validation Loss: 1.3603, Validation Acc: 0.7126

Epoch 241/1000, Train Loss: 1.0980, Train Acc: 0.6929, Validation Loss: 1.3585, Validation Acc: 0.7111
Epoch 242/1000, Train Loss: 1.1083, Train Acc: 0.7071, Validation Loss: 1.3567, Validation Acc: 0.7107
Epoch 243/1000, Train Loss: 1.1606, Train Acc: 0.7214, Validation Loss: 1.3550, Validation Acc: 0.7111
Epoch 244/1000, Train Loss: 1.0693, Train Acc: 0.7000, Validation Loss: 1.3534, Validation Acc: 0.7114
Epoch 245/1000, Train Loss: 1.1167, Train Acc: 0.7286, Validation Loss: 1.3517, Validation Acc: 0.7107
Epoch 246/1000, Train Loss: 1.1152, Train Acc: 0.7071, Validation Loss: 1.3502, Validation Acc: 0.7111
Epoch 247/1000, Train Loss: 1.1405, Train Acc: 0.7000, Validation Loss: 1.3487, Validation Acc: 0.7118
Epoch 248/1000, Train Loss: 1.0875, Train Acc: 0.6929, Validation Loss: 1.3472, Validation Acc: 0.7114
Epoch 249/1000, Train Loss: 1.1234, Train Acc: 0.6929, Validation Loss: 1.3458, Validation Acc: 0.7111
Epoch 250/1000, Train Loss: 1.1179, Train Acc: 0.7500, Validation Loss: 1.3444, Validation Acc: 0.7114
Epoch 251/1000, Train Loss: 1.0648, Train Acc: 0.7357, Validation Loss: 1.3430, Validation Acc: 0.7111
Epoch 252/1000, Train Loss: 1.0393, Train Acc: 0.7857, Validation Loss: 1.3416, Validation Acc: 0.7114
Epoch 253/1000, Train Loss: 1.0875, Train Acc: 0.7643, Validation Loss: 1.3402, Validation Acc: 0.7118
Epoch 254/1000, Train Loss: 1.0986, Train Acc: 0.6786, Validation Loss: 1.3389, Validation Acc: 0.7118
Epoch 255/1000, Train Loss: 1.0657, Train Acc: 0.7286, Validation Loss: 1.3376, Validation Acc: 0.7130
Epoch 256/1000, Train Loss: 1.0839, Train Acc: 0.7714, Validation Loss: 1.3362, Validation Acc: 0.7142
Epoch 257/1000, Train Loss: 1.0460, Train Acc: 0.7429, Validation Loss: 1.3347, Validation Acc: 0.7142
Epoch 258/1000, Train Loss: 1.1335, Train Acc: 0.7286, Validation Loss: 1.3334, Validation Acc: 0.7157
Epoch 259/1000, Train Loss: 1.0282, Train Acc: 0.8214, Validation Loss: 1.3319, Validation Acc: 0.7161
Epoch 260/1000, Train Loss: 1.0930, Train Acc: 0.7286, Validation Loss: 1.3304, Validation Acc: 0.7161
Epoch 261/1000, Train Loss: 0.9952, Train Acc: 0.8143, Validation Loss: 1.3289, Validation Acc: 0.7161
Epoch 262/1000, Train Loss: 1.1058, Train Acc: 0.7071, Validation Loss: 1.3274, Validation Acc: 0.7165
Epoch 263/1000, Train Loss: 1.0576, Train Acc: 0.7286, Validation Loss: 1.3259, Validation Acc: 0.7173
Epoch 264/1000, Train Loss: 1.0873, Train Acc: 0.6786, Validation Loss: 1.3243, Validation Acc: 0.7177
Epoch 265/1000, Train Loss: 1.1146, Train Acc: 0.7214, Validation Loss: 1.3229, Validation Acc: 0.7181
Epoch 266/1000, Train Loss: 1.0868, Train Acc: 0.6929, Validation Loss: 1.3214, Validation Acc: 0.7185
Epoch 267/1000, Train Loss: 1.0567, Train Acc: 0.7143, Validation Loss: 1.3199, Validation Acc: 0.7188
Epoch 268/1000, Train Loss: 1.1029, Train Acc: 0.6929, Validation Loss: 1.3184, Validation Acc: 0.7200
Epoch 269/1000, Train Loss: 1.1221, Train Acc: 0.7000, Validation Loss: 1.3169, Validation Acc: 0.7204
Epoch 270/1000, Train Loss: 1.1021, Train Acc: 0.6643, Validation Loss: 1.3154, Validation Acc: 0.7196

Epoch 271/1000, Train Loss: 0.9691, Train Acc: 0.7929, Validation Loss: 1.3138, Validation Acc: 0.7196
Epoch 272/1000, Train Loss: 0.9864, Train Acc: 0.7714, Validation Loss: 1.3121, Validation Acc: 0.7200
Epoch 273/1000, Train Loss: 1.1618, Train Acc: 0.7143, Validation Loss: 1.3106, Validation Acc: 0.7212
Epoch 274/1000, Train Loss: 1.0851, Train Acc: 0.6643, Validation Loss: 1.3092, Validation Acc: 0.7216
Epoch 275/1000, Train Loss: 1.1005, Train Acc: 0.6857, Validation Loss: 1.3079, Validation Acc: 0.7220
Epoch 276/1000, Train Loss: 1.0947, Train Acc: 0.7214, Validation Loss: 1.3066, Validation Acc: 0.7224
Epoch 277/1000, Train Loss: 1.0377, Train Acc: 0.7571, Validation Loss: 1.3053, Validation Acc: 0.7227
Epoch 278/1000, Train Loss: 1.0551, Train Acc: 0.7214, Validation Loss: 1.3042, Validation Acc: 0.7227
Epoch 279/1000, Train Loss: 1.0765, Train Acc: 0.6857, Validation Loss: 1.3030, Validation Acc: 0.7235
Epoch 280/1000, Train Loss: 1.0067, Train Acc: 0.7571, Validation Loss: 1.3019, Validation Acc: 0.7231
Epoch 281/1000, Train Loss: 1.0108, Train Acc: 0.7286, Validation Loss: 1.3009, Validation Acc: 0.7235
Epoch 282/1000, Train Loss: 1.1150, Train Acc: 0.7357, Validation Loss: 1.2999, Validation Acc: 0.7235
Epoch 283/1000, Train Loss: 1.0098, Train Acc: 0.7071, Validation Loss: 1.2989, Validation Acc: 0.7243
Epoch 284/1000, Train Loss: 1.0391, Train Acc: 0.7143, Validation Loss: 1.2979, Validation Acc: 0.7255
Epoch 285/1000, Train Loss: 1.0272, Train Acc: 0.7357, Validation Loss: 1.2969, Validation Acc: 0.7262
Epoch 286/1000, Train Loss: 1.0248, Train Acc: 0.7214, Validation Loss: 1.2959, Validation Acc: 0.7262
Epoch 287/1000, Train Loss: 0.9730, Train Acc: 0.7500, Validation Loss: 1.2948, Validation Acc: 0.7270
Epoch 288/1000, Train Loss: 1.0620, Train Acc: 0.7786, Validation Loss: 1.2937, Validation Acc: 0.7274
Epoch 289/1000, Train Loss: 0.9425, Train Acc: 0.7643, Validation Loss: 1.2926, Validation Acc: 0.7286
Epoch 290/1000, Train Loss: 0.9549, Train Acc: 0.7643, Validation Loss: 1.2915, Validation Acc: 0.7290
Epoch 291/1000, Train Loss: 0.9975, Train Acc: 0.8071, Validation Loss: 1.2904, Validation Acc: 0.7301
Epoch 292/1000, Train Loss: 1.1003, Train Acc: 0.6857, Validation Loss: 1.2893, Validation Acc: 0.7309
Epoch 293/1000, Train Loss: 0.9849, Train Acc: 0.7714, Validation Loss: 1.2880, Validation Acc: 0.7313
Epoch 294/1000, Train Loss: 0.9986, Train Acc: 0.7571, Validation Loss: 1.2868, Validation Acc: 0.7313
Epoch 295/1000, Train Loss: 0.9815, Train Acc: 0.7571, Validation Loss: 1.2856, Validation Acc: 0.7305
Epoch 296/1000, Train Loss: 1.0024, Train Acc: 0.7286, Validation Loss: 1.2844, Validation Acc: 0.7309
Epoch 297/1000, Train Loss: 1.0613, Train Acc: 0.6857, Validation Loss: 1.2831, Validation Acc: 0.7305
Epoch 298/1000, Train Loss: 1.0335, Train Acc: 0.7429, Validation Loss: 1.2819, Validation Acc: 0.7305
Epoch 299/1000, Train Loss: 0.9707, Train Acc: 0.7929, Validation Loss: 1.2807, Validation Acc: 0.7301
Epoch 300/1000, Train Loss: 1.0094, Train Acc: 0.7571, Validation Loss: 1.2794, Validation Acc: 0.7301

Epoch 301/1000, Train Loss: 1.0082, Train Acc: 0.7571, Validation Loss: 1.2782, Validation Acc: 0.7313
Epoch 302/1000, Train Loss: 1.0464, Train Acc: 0.7071, Validation Loss: 1.2771, Validation Acc: 0.7317
Epoch 303/1000, Train Loss: 0.9892, Train Acc: 0.7429, Validation Loss: 1.2762, Validation Acc: 0.7313
Epoch 304/1000, Train Loss: 1.0563, Train Acc: 0.6857, Validation Loss: 1.2752, Validation Acc: 0.7317
Epoch 305/1000, Train Loss: 1.0449, Train Acc: 0.7429, Validation Loss: 1.2743, Validation Acc: 0.7317
Epoch 306/1000, Train Loss: 0.9741, Train Acc: 0.7786, Validation Loss: 1.2734, Validation Acc: 0.7317
Epoch 307/1000, Train Loss: 0.9961, Train Acc: 0.7643, Validation Loss: 1.2724, Validation Acc: 0.7301
Epoch 308/1000, Train Loss: 1.0360, Train Acc: 0.7714, Validation Loss: 1.2715, Validation Acc: 0.7301
Epoch 309/1000, Train Loss: 1.0175, Train Acc: 0.7286, Validation Loss: 1.2705, Validation Acc: 0.7309
Epoch 310/1000, Train Loss: 1.0353, Train Acc: 0.7214, Validation Loss: 1.2696, Validation Acc: 0.7309
Epoch 311/1000, Train Loss: 1.0030, Train Acc: 0.7786, Validation Loss: 1.2687, Validation Acc: 0.7313
Epoch 312/1000, Train Loss: 0.9943, Train Acc: 0.7714, Validation Loss: 1.2677, Validation Acc: 0.7317
Epoch 313/1000, Train Loss: 0.9371, Train Acc: 0.7500, Validation Loss: 1.2666, Validation Acc: 0.7313
Epoch 314/1000, Train Loss: 1.0013, Train Acc: 0.7286, Validation Loss: 1.2654, Validation Acc: 0.7317
Epoch 315/1000, Train Loss: 1.0211, Train Acc: 0.7714, Validation Loss: 1.2643, Validation Acc: 0.7321
Epoch 316/1000, Train Loss: 0.9813, Train Acc: 0.7214, Validation Loss: 1.2632, Validation Acc: 0.7325
Epoch 317/1000, Train Loss: 1.0107, Train Acc: 0.7357, Validation Loss: 1.2621, Validation Acc: 0.7329
Epoch 318/1000, Train Loss: 1.0873, Train Acc: 0.7286, Validation Loss: 1.2610, Validation Acc: 0.7325
Epoch 319/1000, Train Loss: 0.9626, Train Acc: 0.7429, Validation Loss: 1.2599, Validation Acc: 0.7321
Epoch 320/1000, Train Loss: 1.1081, Train Acc: 0.7071, Validation Loss: 1.2589, Validation Acc: 0.7317
Epoch 321/1000, Train Loss: 1.0309, Train Acc: 0.7000, Validation Loss: 1.2579, Validation Acc: 0.7313
Epoch 322/1000, Train Loss: 1.0502, Train Acc: 0.7143, Validation Loss: 1.2568, Validation Acc: 0.7313
Epoch 323/1000, Train Loss: 0.9062, Train Acc: 0.7357, Validation Loss: 1.2558, Validation Acc: 0.7317
Epoch 324/1000, Train Loss: 0.9967, Train Acc: 0.8071, Validation Loss: 1.2548, Validation Acc: 0.7313
Epoch 325/1000, Train Loss: 0.9285, Train Acc: 0.7643, Validation Loss: 1.2538, Validation Acc: 0.7313
Epoch 326/1000, Train Loss: 0.9577, Train Acc: 0.7571, Validation Loss: 1.2528, Validation Acc: 0.7313
Epoch 327/1000, Train Loss: 0.9657, Train Acc: 0.6929, Validation Loss: 1.2518, Validation Acc: 0.7317
Epoch 328/1000, Train Loss: 0.9456, Train Acc: 0.7500, Validation Loss: 1.2508, Validation Acc: 0.7321
Epoch 329/1000, Train Loss: 1.0443, Train Acc: 0.6857, Validation Loss: 1.2498, Validation Acc: 0.7325
Epoch 330/1000, Train Loss: 0.9433, Train Acc: 0.7714, Validation Loss: 1.2487, Validation Acc: 0.7325

Epoch 331/1000, Train Loss: 1.0701, Train Acc: 0.7214, Validation Loss: 1.2478, Validation Acc: 0.7325
Epoch 332/1000, Train Loss: 1.0138, Train Acc: 0.8143, Validation Loss: 1.2467, Validation Acc: 0.7325
Epoch 333/1000, Train Loss: 1.0219, Train Acc: 0.7143, Validation Loss: 1.2456, Validation Acc: 0.7329
Epoch 334/1000, Train Loss: 0.9265, Train Acc: 0.7714, Validation Loss: 1.2445, Validation Acc: 0.7329
Epoch 335/1000, Train Loss: 0.9624, Train Acc: 0.7786, Validation Loss: 1.2434, Validation Acc: 0.7329
Epoch 336/1000, Train Loss: 0.9817, Train Acc: 0.7071, Validation Loss: 1.2423, Validation Acc: 0.7333
Epoch 337/1000, Train Loss: 0.9898, Train Acc: 0.7643, Validation Loss: 1.2411, Validation Acc: 0.7344
Epoch 338/1000, Train Loss: 0.9194, Train Acc: 0.8214, Validation Loss: 1.2400, Validation Acc: 0.7352
Epoch 339/1000, Train Loss: 0.9684, Train Acc: 0.7143, Validation Loss: 1.2389, Validation Acc: 0.7352
Epoch 340/1000, Train Loss: 0.9322, Train Acc: 0.7571, Validation Loss: 1.2377, Validation Acc: 0.7375
Epoch 341/1000, Train Loss: 1.0997, Train Acc: 0.7214, Validation Loss: 1.2365, Validation Acc: 0.7383
Epoch 342/1000, Train Loss: 1.0062, Train Acc: 0.7643, Validation Loss: 1.2355, Validation Acc: 0.7403
Epoch 343/1000, Train Loss: 0.9549, Train Acc: 0.7429, Validation Loss: 1.2345, Validation Acc: 0.7407
Epoch 344/1000, Train Loss: 0.9445, Train Acc: 0.7929, Validation Loss: 1.2335, Validation Acc: 0.7410
Epoch 345/1000, Train Loss: 0.8963, Train Acc: 0.7786, Validation Loss: 1.2324, Validation Acc: 0.7426
Epoch 346/1000, Train Loss: 0.9603, Train Acc: 0.7786, Validation Loss: 1.2314, Validation Acc: 0.7438
Epoch 347/1000, Train Loss: 0.9504, Train Acc: 0.7786, Validation Loss: 1.2303, Validation Acc: 0.7449
Epoch 348/1000, Train Loss: 1.0552, Train Acc: 0.7000, Validation Loss: 1.2293, Validation Acc: 0.7457
Epoch 349/1000, Train Loss: 0.9884, Train Acc: 0.7500, Validation Loss: 1.2284, Validation Acc: 0.7469
Epoch 350/1000, Train Loss: 1.0045, Train Acc: 0.7571, Validation Loss: 1.2274, Validation Acc: 0.7469
Epoch 351/1000, Train Loss: 0.9600, Train Acc: 0.7357, Validation Loss: 1.2264, Validation Acc: 0.7481
Epoch 352/1000, Train Loss: 1.0753, Train Acc: 0.7429, Validation Loss: 1.2254, Validation Acc: 0.7492
Epoch 353/1000, Train Loss: 0.9501, Train Acc: 0.7643, Validation Loss: 1.2246, Validation Acc: 0.7504
Epoch 354/1000, Train Loss: 0.9455, Train Acc: 0.7786, Validation Loss: 1.2238, Validation Acc: 0.7508
Epoch 355/1000, Train Loss: 1.0190, Train Acc: 0.7571, Validation Loss: 1.2231, Validation Acc: 0.7512
Epoch 356/1000, Train Loss: 0.9677, Train Acc: 0.7643, Validation Loss: 1.2225, Validation Acc: 0.7516
Epoch 357/1000, Train Loss: 0.9424, Train Acc: 0.7929, Validation Loss: 1.2218, Validation Acc: 0.7527
Epoch 358/1000, Train Loss: 0.9949, Train Acc: 0.7214, Validation Loss: 1.2210, Validation Acc: 0.7531
Epoch 359/1000, Train Loss: 1.0563, Train Acc: 0.6786, Validation Loss: 1.2205, Validation Acc: 0.7543
Epoch 360/1000, Train Loss: 1.0022, Train Acc: 0.7071, Validation Loss: 1.2199, Validation Acc: 0.7551

Epoch 361/1000, Train Loss: 0.9612, Train Acc: 0.7929, Validation Loss: 1.2194, Validation Acc: 0.7570
Epoch 362/1000, Train Loss: 0.9876, Train Acc: 0.8143, Validation Loss: 1.2188, Validation Acc: 0.7578
Epoch 363/1000, Train Loss: 0.9788, Train Acc: 0.7786, Validation Loss: 1.2181, Validation Acc: 0.7593
Epoch 364/1000, Train Loss: 0.8855, Train Acc: 0.7786, Validation Loss: 1.2172, Validation Acc: 0.7593
Epoch 365/1000, Train Loss: 0.8961, Train Acc: 0.7643, Validation Loss: 1.2164, Validation Acc: 0.7613
Epoch 366/1000, Train Loss: 1.0178, Train Acc: 0.7143, Validation Loss: 1.2155, Validation Acc: 0.7625
Epoch 367/1000, Train Loss: 0.9813, Train Acc: 0.7643, Validation Loss: 1.2146, Validation Acc: 0.7629
Epoch 368/1000, Train Loss: 0.9295, Train Acc: 0.8357, Validation Loss: 1.2137, Validation Acc: 0.7625
Epoch 369/1000, Train Loss: 0.9151, Train Acc: 0.8000, Validation Loss: 1.2126, Validation Acc: 0.7617
Epoch 370/1000, Train Loss: 0.9379, Train Acc: 0.7929, Validation Loss: 1.2116, Validation Acc: 0.7617
Epoch 371/1000, Train Loss: 0.9344, Train Acc: 0.7786, Validation Loss: 1.2105, Validation Acc: 0.7625
Epoch 372/1000, Train Loss: 1.0083, Train Acc: 0.8286, Validation Loss: 1.2094, Validation Acc: 0.7629
Epoch 373/1000, Train Loss: 0.9491, Train Acc: 0.8286, Validation Loss: 1.2084, Validation Acc: 0.7629
Epoch 374/1000, Train Loss: 0.9821, Train Acc: 0.7643, Validation Loss: 1.2076, Validation Acc: 0.7629
Epoch 375/1000, Train Loss: 0.9394, Train Acc: 0.7214, Validation Loss: 1.2068, Validation Acc: 0.7629
Epoch 376/1000, Train Loss: 0.9476, Train Acc: 0.7071, Validation Loss: 1.2060, Validation Acc: 0.7644
Epoch 377/1000, Train Loss: 0.9390, Train Acc: 0.7929, Validation Loss: 1.2052, Validation Acc: 0.7652
Epoch 378/1000, Train Loss: 0.8885, Train Acc: 0.7786, Validation Loss: 1.2044, Validation Acc: 0.7660
Epoch 379/1000, Train Loss: 0.9615, Train Acc: 0.7857, Validation Loss: 1.2037, Validation Acc: 0.7664
Epoch 380/1000, Train Loss: 0.9670, Train Acc: 0.7214, Validation Loss: 1.2029, Validation Acc: 0.7667
Epoch 381/1000, Train Loss: 0.9704, Train Acc: 0.8143, Validation Loss: 1.2021, Validation Acc: 0.7671
Epoch 382/1000, Train Loss: 0.9995, Train Acc: 0.7571, Validation Loss: 1.2014, Validation Acc: 0.7667
Epoch 383/1000, Train Loss: 0.9257, Train Acc: 0.7929, Validation Loss: 1.2006, Validation Acc: 0.7675
Epoch 384/1000, Train Loss: 1.0046, Train Acc: 0.7214, Validation Loss: 1.1998, Validation Acc: 0.7679
Epoch 385/1000, Train Loss: 0.9703, Train Acc: 0.7286, Validation Loss: 1.1991, Validation Acc: 0.7683
Epoch 386/1000, Train Loss: 0.9135, Train Acc: 0.7929, Validation Loss: 1.1985, Validation Acc: 0.7691
Epoch 387/1000, Train Loss: 0.9270, Train Acc: 0.7714, Validation Loss: 1.1977, Validation Acc: 0.7691
Epoch 388/1000, Train Loss: 0.8991, Train Acc: 0.7714, Validation Loss: 1.1968, Validation Acc: 0.7691
Epoch 389/1000, Train Loss: 0.9229, Train Acc: 0.7500, Validation Loss: 1.1960, Validation Acc: 0.7699
Epoch 390/1000, Train Loss: 0.9969, Train Acc: 0.7786, Validation Loss: 1.1951, Validation Acc: 0.7699

Epoch 391/1000, Train Loss: 0.9260, Train Acc: 0.8000, Validation Loss: 1.1941, Validation Acc: 0.7706
Epoch 392/1000, Train Loss: 0.8796, Train Acc: 0.7857, Validation Loss: 1.1931, Validation Acc: 0.7706
Epoch 393/1000, Train Loss: 0.9434, Train Acc: 0.7643, Validation Loss: 1.1922, Validation Acc: 0.7702
Epoch 394/1000, Train Loss: 0.8606, Train Acc: 0.7429, Validation Loss: 1.1912, Validation Acc: 0.7702
Epoch 395/1000, Train Loss: 1.0229, Train Acc: 0.7929, Validation Loss: 1.1904, Validation Acc: 0.7702
Epoch 396/1000, Train Loss: 0.9610, Train Acc: 0.7929, Validation Loss: 1.1896, Validation Acc: 0.7702
Epoch 397/1000, Train Loss: 0.9875, Train Acc: 0.7143, Validation Loss: 1.1889, Validation Acc: 0.7706
Epoch 398/1000, Train Loss: 0.9446, Train Acc: 0.7786, Validation Loss: 1.1883, Validation Acc: 0.7718
Epoch 399/1000, Train Loss: 0.9841, Train Acc: 0.7643, Validation Loss: 1.1876, Validation Acc: 0.7722
Epoch 400/1000, Train Loss: 0.9059, Train Acc: 0.8357, Validation Loss: 1.1868, Validation Acc: 0.7722
Epoch 401/1000, Train Loss: 1.0101, Train Acc: 0.7571, Validation Loss: 1.1863, Validation Acc: 0.7726
Epoch 402/1000, Train Loss: 0.8856, Train Acc: 0.7857, Validation Loss: 1.1857, Validation Acc: 0.7722
Epoch 403/1000, Train Loss: 0.9274, Train Acc: 0.8071, Validation Loss: 1.1852, Validation Acc: 0.7734
Epoch 404/1000, Train Loss: 0.9160, Train Acc: 0.7429, Validation Loss: 1.1848, Validation Acc: 0.7738
Epoch 405/1000, Train Loss: 0.8659, Train Acc: 0.8214, Validation Loss: 1.1844, Validation Acc: 0.7745
Epoch 406/1000, Train Loss: 0.9078, Train Acc: 0.7929, Validation Loss: 1.1839, Validation Acc: 0.7749
Epoch 407/1000, Train Loss: 0.9291, Train Acc: 0.7500, Validation Loss: 1.1833, Validation Acc: 0.7753
Epoch 408/1000, Train Loss: 0.9726, Train Acc: 0.7286, Validation Loss: 1.1827, Validation Acc: 0.7753
Epoch 409/1000, Train Loss: 0.8885, Train Acc: 0.8071, Validation Loss: 1.1822, Validation Acc: 0.7749
Epoch 410/1000, Train Loss: 0.9332, Train Acc: 0.7929, Validation Loss: 1.1816, Validation Acc: 0.7753
Epoch 411/1000, Train Loss: 1.0533, Train Acc: 0.7357, Validation Loss: 1.1813, Validation Acc: 0.7749
Epoch 412/1000, Train Loss: 0.9922, Train Acc: 0.6786, Validation Loss: 1.1810, Validation Acc: 0.7745
Epoch 413/1000, Train Loss: 1.0061, Train Acc: 0.7786, Validation Loss: 1.1808, Validation Acc: 0.7745
Epoch 414/1000, Train Loss: 0.8881, Train Acc: 0.7786, Validation Loss: 1.1804, Validation Acc: 0.7753
Epoch 415/1000, Train Loss: 0.9510, Train Acc: 0.7714, Validation Loss: 1.1801, Validation Acc: 0.7769
Epoch 416/1000, Train Loss: 0.9184, Train Acc: 0.7571, Validation Loss: 1.1797, Validation Acc: 0.7769
Epoch 417/1000, Train Loss: 0.9243, Train Acc: 0.7714, Validation Loss: 1.1793, Validation Acc: 0.7769
Epoch 418/1000, Train Loss: 0.9338, Train Acc: 0.7500, Validation Loss: 1.1789, Validation Acc: 0.7769
Epoch 419/1000, Train Loss: 0.8836, Train Acc: 0.7357, Validation Loss: 1.1784, Validation Acc: 0.7765
Epoch 420/1000, Train Loss: 0.8852, Train Acc: 0.8000, Validation Loss: 1.1779, Validation Acc: 0.7761

Epoch 421/1000, Train Loss: 0.9405, Train Acc: 0.7714, Validation Loss: 1.1777, Validation Acc: 0.7757
Epoch 422/1000, Train Loss: 0.9362, Train Acc: 0.7714, Validation Loss: 1.1775, Validation Acc: 0.7757
Epoch 423/1000, Train Loss: 0.8826, Train Acc: 0.8286, Validation Loss: 1.1773, Validation Acc: 0.7757
Epoch 424/1000, Train Loss: 0.9260, Train Acc: 0.7786, Validation Loss: 1.1770, Validation Acc: 0.7749
Epoch 425/1000, Train Loss: 0.9288, Train Acc: 0.8000, Validation Loss: 1.1767, Validation Acc: 0.7753
Epoch 426/1000, Train Loss: 0.9643, Train Acc: 0.7500, Validation Loss: 1.1765, Validation Acc: 0.7761
Epoch 427/1000, Train Loss: 0.9829, Train Acc: 0.7786, Validation Loss: 1.1762, Validation Acc: 0.7753
Epoch 428/1000, Train Loss: 0.8998, Train Acc: 0.7429, Validation Loss: 1.1759, Validation Acc: 0.7757
Epoch 429/1000, Train Loss: 0.8890, Train Acc: 0.7857, Validation Loss: 1.1755, Validation Acc: 0.7757
Epoch 430/1000, Train Loss: 0.9278, Train Acc: 0.7714, Validation Loss: 1.1751, Validation Acc: 0.7761
Epoch 431/1000, Train Loss: 0.9686, Train Acc: 0.7786, Validation Loss: 1.1746, Validation Acc: 0.7757
Epoch 432/1000, Train Loss: 0.8178, Train Acc: 0.8214, Validation Loss: 1.1739, Validation Acc: 0.7749
Epoch 433/1000, Train Loss: 0.8516, Train Acc: 0.7714, Validation Loss: 1.1732, Validation Acc: 0.7757
Epoch 434/1000, Train Loss: 0.9425, Train Acc: 0.7786, Validation Loss: 1.1725, Validation Acc: 0.7753
Epoch 435/1000, Train Loss: 0.8869, Train Acc: 0.8143, Validation Loss: 1.1717, Validation Acc: 0.7753
Epoch 436/1000, Train Loss: 0.9334, Train Acc: 0.8071, Validation Loss: 1.1710, Validation Acc: 0.7753
Epoch 437/1000, Train Loss: 0.8929, Train Acc: 0.7500, Validation Loss: 1.1702, Validation Acc: 0.7749
Epoch 438/1000, Train Loss: 0.8977, Train Acc: 0.7857, Validation Loss: 1.1694, Validation Acc: 0.7753
Epoch 439/1000, Train Loss: 0.8238, Train Acc: 0.8143, Validation Loss: 1.1686, Validation Acc: 0.7757
Epoch 440/1000, Train Loss: 0.9620, Train Acc: 0.7929, Validation Loss: 1.1677, Validation Acc: 0.7753
Epoch 441/1000, Train Loss: 0.8354, Train Acc: 0.8429, Validation Loss: 1.1669, Validation Acc: 0.7749
Epoch 442/1000, Train Loss: 0.9131, Train Acc: 0.8143, Validation Loss: 1.1661, Validation Acc: 0.7753
Epoch 443/1000, Train Loss: 0.9347, Train Acc: 0.7357, Validation Loss: 1.1652, Validation Acc: 0.7753
Epoch 444/1000, Train Loss: 0.9689, Train Acc: 0.7643, Validation Loss: 1.1645, Validation Acc: 0.7753
Epoch 445/1000, Train Loss: 0.8632, Train Acc: 0.7857, Validation Loss: 1.1639, Validation Acc: 0.7749
Epoch 446/1000, Train Loss: 0.8756, Train Acc: 0.7929, Validation Loss: 1.1632, Validation Acc: 0.7753
Epoch 447/1000, Train Loss: 0.9244, Train Acc: 0.7571, Validation Loss: 1.1627, Validation Acc: 0.7753
Epoch 448/1000, Train Loss: 0.8386, Train Acc: 0.7643, Validation Loss: 1.1622, Validation Acc: 0.7753
Epoch 449/1000, Train Loss: 0.8664, Train Acc: 0.8143, Validation Loss: 1.1617, Validation Acc: 0.7749
Epoch 450/1000, Train Loss: 0.9371, Train Acc: 0.7857, Validation Loss: 1.1613, Validation Acc: 0.7757

Epoch 451/1000, Train Loss: 0.8037, Train Acc: 0.7929, Validation Loss: 1.1609, Validation Acc: 0.7757
Epoch 452/1000, Train Loss: 0.9442, Train Acc: 0.7714, Validation Loss: 1.1604, Validation Acc: 0.7757
Epoch 453/1000, Train Loss: 0.8964, Train Acc: 0.8071, Validation Loss: 1.1599, Validation Acc: 0.7761
Epoch 454/1000, Train Loss: 0.8877, Train Acc: 0.7786, Validation Loss: 1.1593, Validation Acc: 0.7761
Epoch 455/1000, Train Loss: 0.8390, Train Acc: 0.7786, Validation Loss: 1.1589, Validation Acc: 0.7761
Epoch 456/1000, Train Loss: 0.9066, Train Acc: 0.8214, Validation Loss: 1.1584, Validation Acc: 0.7761
Epoch 457/1000, Train Loss: 0.9352, Train Acc: 0.8071, Validation Loss: 1.1578, Validation Acc: 0.7765
Epoch 458/1000, Train Loss: 0.9237, Train Acc: 0.7643, Validation Loss: 1.1574, Validation Acc: 0.7769
Epoch 459/1000, Train Loss: 0.8978, Train Acc: 0.7857, Validation Loss: 1.1570, Validation Acc: 0.7769
Epoch 460/1000, Train Loss: 0.8965, Train Acc: 0.8071, Validation Loss: 1.1564, Validation Acc: 0.7769
Epoch 461/1000, Train Loss: 0.8936, Train Acc: 0.8000, Validation Loss: 1.1557, Validation Acc: 0.7773
Epoch 462/1000, Train Loss: 0.8716, Train Acc: 0.7286, Validation Loss: 1.1552, Validation Acc: 0.7776
Epoch 463/1000, Train Loss: 0.9998, Train Acc: 0.7714, Validation Loss: 1.1549, Validation Acc: 0.7780
Epoch 464/1000, Train Loss: 0.8672, Train Acc: 0.8000, Validation Loss: 1.1544, Validation Acc: 0.7780
Epoch 465/1000, Train Loss: 0.8917, Train Acc: 0.7786, Validation Loss: 1.1538, Validation Acc: 0.7780
Epoch 466/1000, Train Loss: 0.8529, Train Acc: 0.8286, Validation Loss: 1.1532, Validation Acc: 0.7780
Epoch 467/1000, Train Loss: 0.8643, Train Acc: 0.8071, Validation Loss: 1.1525, Validation Acc: 0.7784
Epoch 468/1000, Train Loss: 0.8874, Train Acc: 0.7714, Validation Loss: 1.1519, Validation Acc: 0.7780
Epoch 469/1000, Train Loss: 0.8650, Train Acc: 0.7714, Validation Loss: 1.1512, Validation Acc: 0.7773
Epoch 470/1000, Train Loss: 0.8434, Train Acc: 0.7929, Validation Loss: 1.1506, Validation Acc: 0.7773
Epoch 471/1000, Train Loss: 0.9308, Train Acc: 0.7786, Validation Loss: 1.1500, Validation Acc: 0.7773
Epoch 472/1000, Train Loss: 0.8937, Train Acc: 0.8000, Validation Loss: 1.1496, Validation Acc: 0.7773
Epoch 473/1000, Train Loss: 0.8965, Train Acc: 0.7714, Validation Loss: 1.1493, Validation Acc: 0.7776
Epoch 474/1000, Train Loss: 0.8380, Train Acc: 0.8000, Validation Loss: 1.1489, Validation Acc: 0.7765
Epoch 475/1000, Train Loss: 0.9785, Train Acc: 0.8143, Validation Loss: 1.1487, Validation Acc: 0.7765
Epoch 476/1000, Train Loss: 0.7888, Train Acc: 0.8714, Validation Loss: 1.1484, Validation Acc: 0.7769
Epoch 477/1000, Train Loss: 0.8277, Train Acc: 0.8214, Validation Loss: 1.1480, Validation Acc: 0.7769
Epoch 478/1000, Train Loss: 0.9043, Train Acc: 0.7929, Validation Loss: 1.1476, Validation Acc: 0.7769
Epoch 479/1000, Train Loss: 0.8570, Train Acc: 0.8071, Validation Loss: 1.1472, Validation Acc: 0.7765
Epoch 480/1000, Train Loss: 0.8148, Train Acc: 0.8143, Validation Loss: 1.1468, Validation Acc: 0.7765

Epoch 481/1000, Train Loss: 0.9042, Train Acc: 0.8286, Validation Loss: 1.1466, Validation Acc: 0.7765
Epoch 482/1000, Train Loss: 0.9798, Train Acc: 0.7786, Validation Loss: 1.1462, Validation Acc: 0.7765
Epoch 483/1000, Train Loss: 0.8494, Train Acc: 0.8500, Validation Loss: 1.1459, Validation Acc: 0.7769
Epoch 484/1000, Train Loss: 0.8743, Train Acc: 0.8143, Validation Loss: 1.1456, Validation Acc: 0.7776
Epoch 485/1000, Train Loss: 0.7844, Train Acc: 0.8286, Validation Loss: 1.1453, Validation Acc: 0.7776
Epoch 486/1000, Train Loss: 0.8453, Train Acc: 0.8500, Validation Loss: 1.1451, Validation Acc: 0.7788
Epoch 487/1000, Train Loss: 0.8791, Train Acc: 0.8214, Validation Loss: 1.1448, Validation Acc: 0.7792
Epoch 488/1000, Train Loss: 0.9142, Train Acc: 0.8143, Validation Loss: 1.1445, Validation Acc: 0.7792
Epoch 489/1000, Train Loss: 0.8037, Train Acc: 0.8000, Validation Loss: 1.1443, Validation Acc: 0.7792
Epoch 490/1000, Train Loss: 0.8288, Train Acc: 0.8714, Validation Loss: 1.1439, Validation Acc: 0.7788
Epoch 491/1000, Train Loss: 0.8579, Train Acc: 0.7786, Validation Loss: 1.1435, Validation Acc: 0.7788
Epoch 492/1000, Train Loss: 0.9518, Train Acc: 0.7571, Validation Loss: 1.1433, Validation Acc: 0.7788
Epoch 493/1000, Train Loss: 0.8543, Train Acc: 0.7643, Validation Loss: 1.1430, Validation Acc: 0.7784
Epoch 494/1000, Train Loss: 0.8994, Train Acc: 0.8143, Validation Loss: 1.1427, Validation Acc: 0.7780
Epoch 495/1000, Train Loss: 0.8368, Train Acc: 0.8071, Validation Loss: 1.1425, Validation Acc: 0.7773
Epoch 496/1000, Train Loss: 0.9523, Train Acc: 0.7429, Validation Loss: 1.1422, Validation Acc: 0.7765
Epoch 497/1000, Train Loss: 0.7886, Train Acc: 0.7929, Validation Loss: 1.1420, Validation Acc: 0.7757
Epoch 498/1000, Train Loss: 0.8492, Train Acc: 0.8429, Validation Loss: 1.1417, Validation Acc: 0.7757
Epoch 499/1000, Train Loss: 0.9279, Train Acc: 0.7857, Validation Loss: 1.1414, Validation Acc: 0.7757
Epoch 500/1000, Train Loss: 0.9008, Train Acc: 0.7857, Validation Loss: 1.1413, Validation Acc: 0.7753
Epoch 501/1000, Train Loss: 0.8418, Train Acc: 0.8143, Validation Loss: 1.1413, Validation Acc: 0.7761
Epoch 502/1000, Train Loss: 0.7720, Train Acc: 0.8286, Validation Loss: 1.1412, Validation Acc: 0.7776
Epoch 503/1000, Train Loss: 0.7779, Train Acc: 0.8429, Validation Loss: 1.1410, Validation Acc: 0.7776
Epoch 504/1000, Train Loss: 0.8417, Train Acc: 0.8214, Validation Loss: 1.1408, Validation Acc: 0.7776
Epoch 505/1000, Train Loss: 0.8709, Train Acc: 0.8286, Validation Loss: 1.1407, Validation Acc: 0.7769
Epoch 506/1000, Train Loss: 0.8766, Train Acc: 0.7643, Validation Loss: 1.1406, Validation Acc: 0.7769
Epoch 507/1000, Train Loss: 0.8614, Train Acc: 0.8000, Validation Loss: 1.1405, Validation Acc: 0.7769
Epoch 508/1000, Train Loss: 0.8968, Train Acc: 0.8214, Validation Loss: 1.1404, Validation Acc: 0.7765
Epoch 509/1000, Train Loss: 0.8542, Train Acc: 0.7929, Validation Loss: 1.1402, Validation Acc: 0.7765
Epoch 510/1000, Train Loss: 0.9421, Train Acc: 0.7857, Validation Loss: 1.1401, Validation Acc: 0.7761

Epoch 511/1000, Train Loss: 0.8407, Train Acc: 0.8000, Validation Loss: 1.1399, Validation Acc: 0.7769
Epoch 512/1000, Train Loss: 0.8025, Train Acc: 0.8143, Validation Loss: 1.1396, Validation Acc: 0.7761
Epoch 513/1000, Train Loss: 0.8589, Train Acc: 0.7857, Validation Loss: 1.1392, Validation Acc: 0.7761
Epoch 514/1000, Train Loss: 0.8661, Train Acc: 0.8000, Validation Loss: 1.1387, Validation Acc: 0.7757
Epoch 515/1000, Train Loss: 0.8832, Train Acc: 0.8071, Validation Loss: 1.1383, Validation Acc: 0.7757
Epoch 516/1000, Train Loss: 0.8175, Train Acc: 0.8357, Validation Loss: 1.1380, Validation Acc: 0.7757
Epoch 517/1000, Train Loss: 0.8294, Train Acc: 0.8143, Validation Loss: 1.1377, Validation Acc: 0.7765
Epoch 518/1000, Train Loss: 0.8183, Train Acc: 0.8500, Validation Loss: 1.1372, Validation Acc: 0.7773
Epoch 519/1000, Train Loss: 0.8634, Train Acc: 0.7929, Validation Loss: 1.1367, Validation Acc: 0.7776
Epoch 520/1000, Train Loss: 0.9340, Train Acc: 0.7643, Validation Loss: 1.1362, Validation Acc: 0.7780
Epoch 521/1000, Train Loss: 0.8900, Train Acc: 0.8286, Validation Loss: 1.1358, Validation Acc: 0.7784
Epoch 522/1000, Train Loss: 0.8273, Train Acc: 0.8214, Validation Loss: 1.1355, Validation Acc: 0.7780
Epoch 523/1000, Train Loss: 0.8074, Train Acc: 0.8143, Validation Loss: 1.1350, Validation Acc: 0.7780
Epoch 524/1000, Train Loss: 0.9395, Train Acc: 0.7714, Validation Loss: 1.1346, Validation Acc: 0.7773
Epoch 525/1000, Train Loss: 0.8788, Train Acc: 0.7714, Validation Loss: 1.1340, Validation Acc: 0.7773
Epoch 526/1000, Train Loss: 0.7966, Train Acc: 0.8357, Validation Loss: 1.1335, Validation Acc: 0.7765
Epoch 527/1000, Train Loss: 0.9450, Train Acc: 0.7500, Validation Loss: 1.1332, Validation Acc: 0.7769
Epoch 528/1000, Train Loss: 0.8674, Train Acc: 0.8429, Validation Loss: 1.1327, Validation Acc: 0.7769
Epoch 529/1000, Train Loss: 0.8050, Train Acc: 0.8429, Validation Loss: 1.1321, Validation Acc: 0.7769
Epoch 530/1000, Train Loss: 0.8971, Train Acc: 0.8071, Validation Loss: 1.1317, Validation Acc: 0.7773
Epoch 531/1000, Train Loss: 0.8547, Train Acc: 0.7929, Validation Loss: 1.1311, Validation Acc: 0.7769
Epoch 532/1000, Train Loss: 0.7874, Train Acc: 0.8286, Validation Loss: 1.1306, Validation Acc: 0.7769
Epoch 533/1000, Train Loss: 0.8976, Train Acc: 0.7571, Validation Loss: 1.1301, Validation Acc: 0.7776
Epoch 534/1000, Train Loss: 0.9156, Train Acc: 0.7714, Validation Loss: 1.1297, Validation Acc: 0.7776
Epoch 535/1000, Train Loss: 0.8731, Train Acc: 0.7786, Validation Loss: 1.1293, Validation Acc: 0.7780
Epoch 536/1000, Train Loss: 0.8897, Train Acc: 0.8214, Validation Loss: 1.1289, Validation Acc: 0.7788
Epoch 537/1000, Train Loss: 0.9042, Train Acc: 0.7929, Validation Loss: 1.1285, Validation Acc: 0.7788
Epoch 538/1000, Train Loss: 0.8573, Train Acc: 0.8071, Validation Loss: 1.1282, Validation Acc: 0.7792
Epoch 539/1000, Train Loss: 0.9024, Train Acc: 0.8143, Validation Loss: 1.1279, Validation Acc: 0.7792
Epoch 540/1000, Train Loss: 0.8215, Train Acc: 0.7786, Validation Loss: 1.1277, Validation Acc: 0.7796

Epoch 541/1000, Train Loss: 0.8946, Train Acc: 0.7500, Validation Loss: 1.1275, Validation Acc: 0.7808
Epoch 542/1000, Train Loss: 0.8369, Train Acc: 0.8143, Validation Loss: 1.1274, Validation Acc: 0.7812
Epoch 543/1000, Train Loss: 0.8588, Train Acc: 0.7929, Validation Loss: 1.1272, Validation Acc: 0.7812
Epoch 544/1000, Train Loss: 0.8403, Train Acc: 0.8214, Validation Loss: 1.1270, Validation Acc: 0.7815
Epoch 545/1000, Train Loss: 0.8284, Train Acc: 0.8143, Validation Loss: 1.1267, Validation Acc: 0.7815
Epoch 546/1000, Train Loss: 0.8547, Train Acc: 0.7571, Validation Loss: 1.1264, Validation Acc: 0.7815
Epoch 547/1000, Train Loss: 0.9072, Train Acc: 0.8000, Validation Loss: 1.1260, Validation Acc: 0.7812
Epoch 548/1000, Train Loss: 0.7975, Train Acc: 0.8571, Validation Loss: 1.1256, Validation Acc: 0.7808
Epoch 549/1000, Train Loss: 0.9781, Train Acc: 0.7286, Validation Loss: 1.1252, Validation Acc: 0.7808
Epoch 550/1000, Train Loss: 0.8940, Train Acc: 0.8000, Validation Loss: 1.1248, Validation Acc: 0.7808
Epoch 551/1000, Train Loss: 0.9125, Train Acc: 0.7643, Validation Loss: 1.1244, Validation Acc: 0.7804
Epoch 552/1000, Train Loss: 0.8945, Train Acc: 0.7500, Validation Loss: 1.1241, Validation Acc: 0.7804
Epoch 553/1000, Train Loss: 0.8135, Train Acc: 0.8214, Validation Loss: 1.1238, Validation Acc: 0.7792
Epoch 554/1000, Train Loss: 0.8335, Train Acc: 0.8429, Validation Loss: 1.1235, Validation Acc: 0.7792
Epoch 555/1000, Train Loss: 0.8750, Train Acc: 0.8000, Validation Loss: 1.1233, Validation Acc: 0.7796
Epoch 556/1000, Train Loss: 0.8236, Train Acc: 0.8286, Validation Loss: 1.1231, Validation Acc: 0.7796
Epoch 557/1000, Train Loss: 0.8717, Train Acc: 0.8000, Validation Loss: 1.1229, Validation Acc: 0.7796
Epoch 558/1000, Train Loss: 0.8836, Train Acc: 0.8286, Validation Loss: 1.1227, Validation Acc: 0.7796
Epoch 559/1000, Train Loss: 0.8449, Train Acc: 0.7857, Validation Loss: 1.1225, Validation Acc: 0.7796
Epoch 560/1000, Train Loss: 0.7919, Train Acc: 0.7929, Validation Loss: 1.1222, Validation Acc: 0.7800
Epoch 561/1000, Train Loss: 0.8494, Train Acc: 0.8071, Validation Loss: 1.1220, Validation Acc: 0.7800
Epoch 562/1000, Train Loss: 0.7506, Train Acc: 0.8000, Validation Loss: 1.1217, Validation Acc: 0.7788
Epoch 563/1000, Train Loss: 0.7793, Train Acc: 0.8286, Validation Loss: 1.1214, Validation Acc: 0.7784
Epoch 564/1000, Train Loss: 0.8245, Train Acc: 0.8000, Validation Loss: 1.1210, Validation Acc: 0.7780
Epoch 565/1000, Train Loss: 0.8544, Train Acc: 0.7929, Validation Loss: 1.1207, Validation Acc: 0.7776
Epoch 566/1000, Train Loss: 0.8216, Train Acc: 0.8071, Validation Loss: 1.1206, Validation Acc: 0.7776
Epoch 567/1000, Train Loss: 0.8731, Train Acc: 0.8214, Validation Loss: 1.1203, Validation Acc: 0.7784
Epoch 568/1000, Train Loss: 0.8400, Train Acc: 0.8214, Validation Loss: 1.1200, Validation Acc: 0.7784
Epoch 569/1000, Train Loss: 0.8092, Train Acc: 0.8000, Validation Loss: 1.1198, Validation Acc: 0.7780
Epoch 570/1000, Train Loss: 0.8610, Train Acc: 0.8500, Validation Loss: 1.1196, Validation Acc: 0.7780

Epoch 571/1000, Train Loss: 0.7743, Train Acc: 0.8214, Validation Loss: 1.1194, Validation Acc: 0.7780
Epoch 572/1000, Train Loss: 0.8864, Train Acc: 0.7357, Validation Loss: 1.1192, Validation Acc: 0.7780
Epoch 573/1000, Train Loss: 0.7978, Train Acc: 0.8500, Validation Loss: 1.1192, Validation Acc: 0.7792
Epoch 574/1000, Train Loss: 0.8450, Train Acc: 0.7786, Validation Loss: 1.1191, Validation Acc: 0.7800
Epoch 575/1000, Train Loss: 0.7875, Train Acc: 0.8357, Validation Loss: 1.1191, Validation Acc: 0.7804
Epoch 576/1000, Train Loss: 0.8718, Train Acc: 0.7786, Validation Loss: 1.1191, Validation Acc: 0.7796
Epoch 577/1000, Train Loss: 0.7818, Train Acc: 0.8000, Validation Loss: 1.1191, Validation Acc: 0.7792
Epoch 578/1000, Train Loss: 0.8473, Train Acc: 0.8071, Validation Loss: 1.1191, Validation Acc: 0.7800
Epoch 579/1000, Train Loss: 0.8688, Train Acc: 0.8143, Validation Loss: 1.1190, Validation Acc: 0.7796
Epoch 580/1000, Train Loss: 0.7703, Train Acc: 0.8714, Validation Loss: 1.1189, Validation Acc: 0.7788
Epoch 581/1000, Train Loss: 0.8336, Train Acc: 0.8500, Validation Loss: 1.1186, Validation Acc: 0.7792
Epoch 582/1000, Train Loss: 0.8225, Train Acc: 0.8286, Validation Loss: 1.1182, Validation Acc: 0.7792
Epoch 583/1000, Train Loss: 0.7880, Train Acc: 0.8571, Validation Loss: 1.1178, Validation Acc: 0.7788
Epoch 584/1000, Train Loss: 0.8414, Train Acc: 0.8071, Validation Loss: 1.1174, Validation Acc: 0.7792
Epoch 585/1000, Train Loss: 0.8804, Train Acc: 0.7857, Validation Loss: 1.1169, Validation Acc: 0.7792
Epoch 586/1000, Train Loss: 0.8796, Train Acc: 0.7929, Validation Loss: 1.1164, Validation Acc: 0.7796
Epoch 587/1000, Train Loss: 0.8381, Train Acc: 0.8071, Validation Loss: 1.1159, Validation Acc: 0.7796
Epoch 588/1000, Train Loss: 0.7936, Train Acc: 0.8500, Validation Loss: 1.1153, Validation Acc: 0.7800
Epoch 589/1000, Train Loss: 0.7674, Train Acc: 0.8429, Validation Loss: 1.1146, Validation Acc: 0.7812
Epoch 590/1000, Train Loss: 0.8095, Train Acc: 0.8429, Validation Loss: 1.1140, Validation Acc: 0.7812
Epoch 591/1000, Train Loss: 0.8148, Train Acc: 0.8000, Validation Loss: 1.1133, Validation Acc: 0.7804
Epoch 592/1000, Train Loss: 0.9063, Train Acc: 0.7571, Validation Loss: 1.1128, Validation Acc: 0.7812
Epoch 593/1000, Train Loss: 0.8972, Train Acc: 0.8000, Validation Loss: 1.1123, Validation Acc: 0.7808
Epoch 594/1000, Train Loss: 0.9158, Train Acc: 0.7857, Validation Loss: 1.1120, Validation Acc: 0.7815
Epoch 595/1000, Train Loss: 0.7470, Train Acc: 0.8429, Validation Loss: 1.1115, Validation Acc: 0.7812
Epoch 596/1000, Train Loss: 0.8228, Train Acc: 0.8071, Validation Loss: 1.1111, Validation Acc: 0.7812
Epoch 597/1000, Train Loss: 0.7551, Train Acc: 0.8571, Validation Loss: 1.1106, Validation Acc: 0.7819
Epoch 598/1000, Train Loss: 0.7334, Train Acc: 0.8857, Validation Loss: 1.1102, Validation Acc: 0.7823
Epoch 599/1000, Train Loss: 0.8184, Train Acc: 0.8643, Validation Loss: 1.1099, Validation Acc: 0.7827
Epoch 600/1000, Train Loss: 0.8049, Train Acc: 0.8214, Validation Loss: 1.1096, Validation Acc: 0.7819

Epoch 601/1000, Train Loss: 0.7643, Train Acc: 0.8429, Validation Loss: 1.1094, Validation Acc: 0.7815
Epoch 602/1000, Train Loss: 0.8010, Train Acc: 0.8429, Validation Loss: 1.1091, Validation Acc: 0.7808
Epoch 603/1000, Train Loss: 0.8131, Train Acc: 0.7714, Validation Loss: 1.1090, Validation Acc: 0.7808
Epoch 604/1000, Train Loss: 0.7906, Train Acc: 0.8357, Validation Loss: 1.1088, Validation Acc: 0.7804
Epoch 605/1000, Train Loss: 0.7845, Train Acc: 0.8571, Validation Loss: 1.1086, Validation Acc: 0.7804
Epoch 606/1000, Train Loss: 0.8690, Train Acc: 0.8143, Validation Loss: 1.1084, Validation Acc: 0.7804
Epoch 607/1000, Train Loss: 0.8261, Train Acc: 0.8214, Validation Loss: 1.1084, Validation Acc: 0.7796
Epoch 608/1000, Train Loss: 0.8353, Train Acc: 0.8357, Validation Loss: 1.1084, Validation Acc: 0.7796
Epoch 609/1000, Train Loss: 0.8005, Train Acc: 0.8000, Validation Loss: 1.1083, Validation Acc: 0.7796
Epoch 610/1000, Train Loss: 0.8132, Train Acc: 0.8071, Validation Loss: 1.1083, Validation Acc: 0.7796
Epoch 611/1000, Train Loss: 0.7960, Train Acc: 0.8429, Validation Loss: 1.1083, Validation Acc: 0.7792
Epoch 612/1000, Train Loss: 0.8361, Train Acc: 0.8143, Validation Loss: 1.1086, Validation Acc: 0.7788
Epoch 613/1000, Train Loss: 0.7813, Train Acc: 0.8214, Validation Loss: 1.1089, Validation Acc: 0.7788
Epoch 614/1000, Train Loss: 0.7473, Train Acc: 0.8643, Validation Loss: 1.1090, Validation Acc: 0.7792
Epoch 615/1000, Train Loss: 0.7906, Train Acc: 0.8714, Validation Loss: 1.1091, Validation Acc: 0.7796
Epoch 616/1000, Train Loss: 0.7876, Train Acc: 0.8000, Validation Loss: 1.1091, Validation Acc: 0.7792
Epoch 617/1000, Train Loss: 0.8328, Train Acc: 0.8143, Validation Loss: 1.1091, Validation Acc: 0.7800
Epoch 618/1000, Train Loss: 0.8841, Train Acc: 0.7929, Validation Loss: 1.1094, Validation Acc: 0.7800
Epoch 619/1000, Train Loss: 0.7727, Train Acc: 0.8286, Validation Loss: 1.1096, Validation Acc: 0.7800
Epoch 620/1000, Train Loss: 0.8515, Train Acc: 0.8429, Validation Loss: 1.1097, Validation Acc: 0.7808
Epoch 621/1000, Train Loss: 0.8107, Train Acc: 0.8000, Validation Loss: 1.1098, Validation Acc: 0.7804
Epoch 622/1000, Train Loss: 0.8727, Train Acc: 0.8000, Validation Loss: 1.1098, Validation Acc: 0.7796
Epoch 623/1000, Train Loss: 0.7859, Train Acc: 0.8500, Validation Loss: 1.1096, Validation Acc: 0.7796
Epoch 624/1000, Train Loss: 0.8489, Train Acc: 0.8357, Validation Loss: 1.1094, Validation Acc: 0.7804
Epoch 625/1000, Train Loss: 0.7768, Train Acc: 0.7929, Validation Loss: 1.1092, Validation Acc: 0.7792
Epoch 626/1000, Train Loss: 0.7966, Train Acc: 0.8357, Validation Loss: 1.1089, Validation Acc: 0.7800
Epoch 627/1000, Train Loss: 0.8329, Train Acc: 0.8214, Validation Loss: 1.1084, Validation Acc: 0.7800
Epoch 628/1000, Train Loss: 0.9109, Train Acc: 0.8071, Validation Loss: 1.1080, Validation Acc: 0.7808
Epoch 629/1000, Train Loss: 0.8687, Train Acc: 0.8214, Validation Loss: 1.1075, Validation Acc: 0.7812
Epoch 630/1000, Train Loss: 0.8707, Train Acc: 0.7857, Validation Loss: 1.1070, Validation Acc: 0.7812

Epoch 631/1000, Train Loss: 0.7628, Train Acc: 0.8357, Validation Loss: 1.1068, Validation Acc: 0.7800
Epoch 632/1000, Train Loss: 0.8312, Train Acc: 0.8071, Validation Loss: 1.1065, Validation Acc: 0.7800
Epoch 633/1000, Train Loss: 0.7979, Train Acc: 0.8429, Validation Loss: 1.1063, Validation Acc: 0.7800
Epoch 634/1000, Train Loss: 0.7003, Train Acc: 0.8429, Validation Loss: 1.1060, Validation Acc: 0.7800
Epoch 635/1000, Train Loss: 0.8169, Train Acc: 0.8214, Validation Loss: 1.1057, Validation Acc: 0.7800
Epoch 636/1000, Train Loss: 0.7835, Train Acc: 0.7929, Validation Loss: 1.1054, Validation Acc: 0.7796
Epoch 637/1000, Train Loss: 0.8434, Train Acc: 0.8286, Validation Loss: 1.1052, Validation Acc: 0.7796
Epoch 638/1000, Train Loss: 0.8119, Train Acc: 0.8571, Validation Loss: 1.1049, Validation Acc: 0.7796
Epoch 639/1000, Train Loss: 0.8193, Train Acc: 0.7857, Validation Loss: 1.1048, Validation Acc: 0.7792
Epoch 640/1000, Train Loss: 0.7951, Train Acc: 0.8214, Validation Loss: 1.1046, Validation Acc: 0.7792
Epoch 641/1000, Train Loss: 0.7817, Train Acc: 0.8214, Validation Loss: 1.1046, Validation Acc: 0.7792
Epoch 642/1000, Train Loss: 0.7382, Train Acc: 0.8714, Validation Loss: 1.1046, Validation Acc: 0.7792
Epoch 643/1000, Train Loss: 0.8717, Train Acc: 0.7857, Validation Loss: 1.1046, Validation Acc: 0.7792
Epoch 644/1000, Train Loss: 0.8348, Train Acc: 0.7929, Validation Loss: 1.1045, Validation Acc: 0.7792
Epoch 645/1000, Train Loss: 0.7349, Train Acc: 0.8357, Validation Loss: 1.1044, Validation Acc: 0.7804
Epoch 646/1000, Train Loss: 0.7862, Train Acc: 0.8429, Validation Loss: 1.1042, Validation Acc: 0.7804
Epoch 647/1000, Train Loss: 0.8126, Train Acc: 0.8286, Validation Loss: 1.1040, Validation Acc: 0.7792
Epoch 648/1000, Train Loss: 0.8059, Train Acc: 0.8643, Validation Loss: 1.1037, Validation Acc: 0.7792
Epoch 649/1000, Train Loss: 0.8058, Train Acc: 0.8143, Validation Loss: 1.1032, Validation Acc: 0.7792
Epoch 650/1000, Train Loss: 0.7728, Train Acc: 0.8357, Validation Loss: 1.1028, Validation Acc: 0.7788
Epoch 651/1000, Train Loss: 0.8174, Train Acc: 0.8357, Validation Loss: 1.1023, Validation Acc: 0.7784
Epoch 652/1000, Train Loss: 0.7794, Train Acc: 0.8214, Validation Loss: 1.1019, Validation Acc: 0.7780
Epoch 653/1000, Train Loss: 0.9088, Train Acc: 0.8071, Validation Loss: 1.1016, Validation Acc: 0.7784
Epoch 654/1000, Train Loss: 0.7777, Train Acc: 0.8643, Validation Loss: 1.1014, Validation Acc: 0.7784
Epoch 655/1000, Train Loss: 0.7568, Train Acc: 0.8143, Validation Loss: 1.1012, Validation Acc: 0.7780
Epoch 656/1000, Train Loss: 0.7947, Train Acc: 0.8500, Validation Loss: 1.1011, Validation Acc: 0.7784
Epoch 657/1000, Train Loss: 0.8709, Train Acc: 0.7929, Validation Loss: 1.1009, Validation Acc: 0.7784
Epoch 658/1000, Train Loss: 0.8317, Train Acc: 0.8500, Validation Loss: 1.1006, Validation Acc: 0.7788
Epoch 659/1000, Train Loss: 0.7607, Train Acc: 0.8286, Validation Loss: 1.1005, Validation Acc: 0.7792
Epoch 660/1000, Train Loss: 0.7411, Train Acc: 0.8786, Validation Loss: 1.1002, Validation Acc: 0.7792

Epoch 661/1000, Train Loss: 0.8049, Train Acc: 0.7857, Validation Loss: 1.000, Validation Acc: 0.7792
Epoch 662/1000, Train Loss: 0.7522, Train Acc: 0.8214, Validation Loss: 1.000, Validation Acc: 0.7788
Epoch 663/1000, Train Loss: 0.8913, Train Acc: 0.7857, Validation Loss: 1.0998, Validation Acc: 0.7792
Epoch 664/1000, Train Loss: 0.8216, Train Acc: 0.8214, Validation Loss: 1.0996, Validation Acc: 0.7792
Epoch 665/1000, Train Loss: 0.9563, Train Acc: 0.7571, Validation Loss: 1.0994, Validation Acc: 0.7792
Epoch 666/1000, Train Loss: 0.8313, Train Acc: 0.8286, Validation Loss: 1.0992, Validation Acc: 0.7792
Epoch 667/1000, Train Loss: 0.8735, Train Acc: 0.8143, Validation Loss: 1.0991, Validation Acc: 0.7792
Epoch 668/1000, Train Loss: 0.7827, Train Acc: 0.8071, Validation Loss: 1.0989, Validation Acc: 0.7792
Epoch 669/1000, Train Loss: 0.8350, Train Acc: 0.8071, Validation Loss: 1.0987, Validation Acc: 0.7796
Epoch 670/1000, Train Loss: 0.8306, Train Acc: 0.8071, Validation Loss: 1.0984, Validation Acc: 0.7792
Epoch 671/1000, Train Loss: 0.7359, Train Acc: 0.8571, Validation Loss: 1.0980, Validation Acc: 0.7788
Epoch 672/1000, Train Loss: 0.7621, Train Acc: 0.8500, Validation Loss: 1.0977, Validation Acc: 0.7784
Epoch 673/1000, Train Loss: 0.8347, Train Acc: 0.8143, Validation Loss: 1.0973, Validation Acc: 0.7792
Epoch 674/1000, Train Loss: 0.8240, Train Acc: 0.8500, Validation Loss: 1.0970, Validation Acc: 0.7788
Epoch 675/1000, Train Loss: 0.8091, Train Acc: 0.8000, Validation Loss: 1.0967, Validation Acc: 0.7784
Epoch 676/1000, Train Loss: 0.7813, Train Acc: 0.8000, Validation Loss: 1.0965, Validation Acc: 0.7784
Epoch 677/1000, Train Loss: 0.8338, Train Acc: 0.8000, Validation Loss: 1.0963, Validation Acc: 0.7788
Epoch 678/1000, Train Loss: 0.8848, Train Acc: 0.7929, Validation Loss: 1.0964, Validation Acc: 0.7796
Epoch 679/1000, Train Loss: 0.8500, Train Acc: 0.8429, Validation Loss: 1.0965, Validation Acc: 0.7796
Epoch 680/1000, Train Loss: 0.8296, Train Acc: 0.8143, Validation Loss: 1.0965, Validation Acc: 0.7784
Epoch 681/1000, Train Loss: 0.7587, Train Acc: 0.8071, Validation Loss: 1.0964, Validation Acc: 0.7780
Epoch 682/1000, Train Loss: 0.8197, Train Acc: 0.8643, Validation Loss: 1.0963, Validation Acc: 0.7784
Epoch 683/1000, Train Loss: 0.8606, Train Acc: 0.8429, Validation Loss: 1.0961, Validation Acc: 0.7784
Epoch 684/1000, Train Loss: 0.7792, Train Acc: 0.8071, Validation Loss: 1.0960, Validation Acc: 0.7784
Epoch 685/1000, Train Loss: 0.7741, Train Acc: 0.8357, Validation Loss: 1.0959, Validation Acc: 0.7784
Epoch 686/1000, Train Loss: 0.8077, Train Acc: 0.8000, Validation Loss: 1.0960, Validation Acc: 0.7788
Epoch 687/1000, Train Loss: 0.7848, Train Acc: 0.7929, Validation Loss: 1.0960, Validation Acc: 0.7792
Epoch 688/1000, Train Loss: 0.8074, Train Acc: 0.8000, Validation Loss: 1.0962, Validation Acc: 0.7792
Epoch 689/1000, Train Loss: 0.7902, Train Acc: 0.8286, Validation Loss: 1.0963, Validation Acc: 0.7792
Epoch 690/1000, Train Loss: 0.7402, Train Acc: 0.8143, Validation Loss: 1.0962, Validation Acc: 0.7788

Epoch 691/1000, Train Loss: 0.8025, Train Acc: 0.8214, Validation Loss: 1.0962, Validation Acc: 0.7792
Epoch 692/1000, Train Loss: 0.8629, Train Acc: 0.8000, Validation Loss: 1.0962, Validation Acc: 0.7796
Epoch 693/1000, Train Loss: 0.8002, Train Acc: 0.8286, Validation Loss: 1.0962, Validation Acc: 0.7800
Epoch 694/1000, Train Loss: 0.8496, Train Acc: 0.7786, Validation Loss: 1.0961, Validation Acc: 0.7796
Epoch 695/1000, Train Loss: 0.8044, Train Acc: 0.8571, Validation Loss: 1.0962, Validation Acc: 0.7796
Epoch 696/1000, Train Loss: 0.7920, Train Acc: 0.8714, Validation Loss: 1.0961, Validation Acc: 0.7796
Epoch 697/1000, Train Loss: 0.7072, Train Acc: 0.8214, Validation Loss: 1.0961, Validation Acc: 0.7796
Epoch 698/1000, Train Loss: 0.8040, Train Acc: 0.8429, Validation Loss: 1.0962, Validation Acc: 0.7800
Epoch 699/1000, Train Loss: 0.8561, Train Acc: 0.8429, Validation Loss: 1.0962, Validation Acc: 0.7800
Epoch 700/1000, Train Loss: 0.8347, Train Acc: 0.8143, Validation Loss: 1.0963, Validation Acc: 0.7800
Epoch 701/1000, Train Loss: 0.8768, Train Acc: 0.8000, Validation Loss: 1.0964, Validation Acc: 0.7800
Epoch 702/1000, Train Loss: 0.7647, Train Acc: 0.8357, Validation Loss: 1.0965, Validation Acc: 0.7796
Epoch 703/1000, Train Loss: 0.8188, Train Acc: 0.7929, Validation Loss: 1.0966, Validation Acc: 0.7792
Epoch 704/1000, Train Loss: 0.8055, Train Acc: 0.8000, Validation Loss: 1.0966, Validation Acc: 0.7792
Epoch 705/1000, Train Loss: 0.7684, Train Acc: 0.8786, Validation Loss: 1.0965, Validation Acc: 0.7792
Epoch 706/1000, Train Loss: 0.8047, Train Acc: 0.8143, Validation Loss: 1.0962, Validation Acc: 0.7800
Epoch 707/1000, Train Loss: 0.7978, Train Acc: 0.8143, Validation Loss: 1.0958, Validation Acc: 0.7815
Epoch 708/1000, Train Loss: 0.8542, Train Acc: 0.8143, Validation Loss: 1.0955, Validation Acc: 0.7815
Epoch 709/1000, Train Loss: 0.7760, Train Acc: 0.8143, Validation Loss: 1.0952, Validation Acc: 0.7823
Epoch 710/1000, Train Loss: 0.8351, Train Acc: 0.8143, Validation Loss: 1.0948, Validation Acc: 0.7827
Epoch 711/1000, Train Loss: 0.7719, Train Acc: 0.8429, Validation Loss: 1.0942, Validation Acc: 0.7823
Epoch 712/1000, Train Loss: 0.8162, Train Acc: 0.8286, Validation Loss: 1.0936, Validation Acc: 0.7823
Epoch 713/1000, Train Loss: 0.8043, Train Acc: 0.8571, Validation Loss: 1.0930, Validation Acc: 0.7815
Epoch 714/1000, Train Loss: 0.8249, Train Acc: 0.8286, Validation Loss: 1.0924, Validation Acc: 0.7815
Epoch 715/1000, Train Loss: 0.7734, Train Acc: 0.8571, Validation Loss: 1.0921, Validation Acc: 0.7815
Epoch 716/1000, Train Loss: 0.9215, Train Acc: 0.7500, Validation Loss: 1.0919, Validation Acc: 0.7812
Epoch 717/1000, Train Loss: 0.7700, Train Acc: 0.8500, Validation Loss: 1.0916, Validation Acc: 0.7815
Epoch 718/1000, Train Loss: 0.7841, Train Acc: 0.9000, Validation Loss: 1.0911, Validation Acc: 0.7815
Epoch 719/1000, Train Loss: 0.7213, Train Acc: 0.8429, Validation Loss: 1.0907, Validation Acc: 0.7815
Epoch 720/1000, Train Loss: 0.8270, Train Acc: 0.8143, Validation Loss: 1.0904, Validation Acc: 0.7815

Epoch 721/1000, Train Loss: 0.8245, Train Acc: 0.7786, Validation Loss: 1.0901, Validation Acc: 0.7815
Epoch 722/1000, Train Loss: 0.7839, Train Acc: 0.8500, Validation Loss: 1.0898, Validation Acc: 0.7812
Epoch 723/1000, Train Loss: 0.7344, Train Acc: 0.8643, Validation Loss: 1.0895, Validation Acc: 0.7812
Epoch 724/1000, Train Loss: 0.8249, Train Acc: 0.8214, Validation Loss: 1.0891, Validation Acc: 0.7808
Epoch 725/1000, Train Loss: 0.7719, Train Acc: 0.7929, Validation Loss: 1.0888, Validation Acc: 0.7812
Epoch 726/1000, Train Loss: 0.8180, Train Acc: 0.7929, Validation Loss: 1.0886, Validation Acc: 0.7812
Epoch 727/1000, Train Loss: 0.8292, Train Acc: 0.7929, Validation Loss: 1.0885, Validation Acc: 0.7823
Epoch 728/1000, Train Loss: 0.7736, Train Acc: 0.8143, Validation Loss: 1.0885, Validation Acc: 0.7823
Epoch 729/1000, Train Loss: 0.7619, Train Acc: 0.7929, Validation Loss: 1.0884, Validation Acc: 0.7831
Epoch 730/1000, Train Loss: 0.7623, Train Acc: 0.8500, Validation Loss: 1.0883, Validation Acc: 0.7831
Epoch 731/1000, Train Loss: 0.7519, Train Acc: 0.8357, Validation Loss: 1.0883, Validation Acc: 0.7827
Epoch 732/1000, Train Loss: 0.7951, Train Acc: 0.8500, Validation Loss: 1.0881, Validation Acc: 0.7827
Epoch 733/1000, Train Loss: 0.8240, Train Acc: 0.7929, Validation Loss: 1.0880, Validation Acc: 0.7819
Epoch 734/1000, Train Loss: 0.8244, Train Acc: 0.8071, Validation Loss: 1.0878, Validation Acc: 0.7819
Epoch 735/1000, Train Loss: 0.7979, Train Acc: 0.8357, Validation Loss: 1.0876, Validation Acc: 0.7819
Epoch 736/1000, Train Loss: 0.8254, Train Acc: 0.8000, Validation Loss: 1.0874, Validation Acc: 0.7815
Epoch 737/1000, Train Loss: 0.8047, Train Acc: 0.8571, Validation Loss: 1.0875, Validation Acc: 0.7819
Epoch 738/1000, Train Loss: 0.8185, Train Acc: 0.8857, Validation Loss: 1.0874, Validation Acc: 0.7819
Epoch 739/1000, Train Loss: 0.7376, Train Acc: 0.8357, Validation Loss: 1.0874, Validation Acc: 0.7819
Epoch 740/1000, Train Loss: 0.7311, Train Acc: 0.8500, Validation Loss: 1.0872, Validation Acc: 0.7823
Epoch 741/1000, Train Loss: 0.7366, Train Acc: 0.8571, Validation Loss: 1.0870, Validation Acc: 0.7823
Epoch 742/1000, Train Loss: 0.8147, Train Acc: 0.8214, Validation Loss: 1.0868, Validation Acc: 0.7823
Epoch 743/1000, Train Loss: 0.9259, Train Acc: 0.8214, Validation Loss: 1.0871, Validation Acc: 0.7819
Epoch 744/1000, Train Loss: 0.8151, Train Acc: 0.8214, Validation Loss: 1.0874, Validation Acc: 0.7819
Epoch 745/1000, Train Loss: 0.8867, Train Acc: 0.8143, Validation Loss: 1.0876, Validation Acc: 0.7815
Epoch 746/1000, Train Loss: 0.7940, Train Acc: 0.8286, Validation Loss: 1.0879, Validation Acc: 0.7808
Epoch 747/1000, Train Loss: 0.8574, Train Acc: 0.8000, Validation Loss: 1.0882, Validation Acc: 0.7804
Epoch 748/1000, Train Loss: 0.8539, Train Acc: 0.8214, Validation Loss: 1.0883, Validation Acc: 0.7812
Epoch 749/1000, Train Loss: 0.8302, Train Acc: 0.8286, Validation Loss: 1.0882, Validation Acc: 0.7812
Epoch 750/1000, Train Loss: 0.7499, Train Acc: 0.8357, Validation Loss: 1.0881, Validation Acc: 0.7812

Epoch 751/1000, Train Loss: 0.7921, Train Acc: 0.8500, Validation Loss: 1.0879, Validation Acc: 0.7812
Epoch 752/1000, Train Loss: 0.7861, Train Acc: 0.8429, Validation Loss: 1.0877, Validation Acc: 0.7815
Epoch 753/1000, Train Loss: 0.8025, Train Acc: 0.7857, Validation Loss: 1.0874, Validation Acc: 0.7819
Epoch 754/1000, Train Loss: 0.7486, Train Acc: 0.8571, Validation Loss: 1.0870, Validation Acc: 0.7815
Epoch 755/1000, Train Loss: 0.9048, Train Acc: 0.8071, Validation Loss: 1.0867, Validation Acc: 0.7815
Epoch 756/1000, Train Loss: 0.7359, Train Acc: 0.8429, Validation Loss: 1.0862, Validation Acc: 0.7819
Epoch 757/1000, Train Loss: 0.7245, Train Acc: 0.8429, Validation Loss: 1.0858, Validation Acc: 0.7819
Epoch 758/1000, Train Loss: 0.8150, Train Acc: 0.8143, Validation Loss: 1.0852, Validation Acc: 0.7819
Epoch 759/1000, Train Loss: 0.7523, Train Acc: 0.8571, Validation Loss: 1.0847, Validation Acc: 0.7815
Epoch 760/1000, Train Loss: 0.8865, Train Acc: 0.8143, Validation Loss: 1.0841, Validation Acc: 0.7808
Epoch 761/1000, Train Loss: 0.7976, Train Acc: 0.8786, Validation Loss: 1.0834, Validation Acc: 0.7815
Epoch 762/1000, Train Loss: 0.7830, Train Acc: 0.8500, Validation Loss: 1.0828, Validation Acc: 0.7815
Epoch 763/1000, Train Loss: 0.7866, Train Acc: 0.7857, Validation Loss: 1.0821, Validation Acc: 0.7819
Epoch 764/1000, Train Loss: 0.8274, Train Acc: 0.8357, Validation Loss: 1.0814, Validation Acc: 0.7819
Epoch 765/1000, Train Loss: 0.7921, Train Acc: 0.8714, Validation Loss: 1.0806, Validation Acc: 0.7812
Epoch 766/1000, Train Loss: 0.6363, Train Acc: 0.8929, Validation Loss: 1.0799, Validation Acc: 0.7812
Epoch 767/1000, Train Loss: 0.7530, Train Acc: 0.8286, Validation Loss: 1.0792, Validation Acc: 0.7812
Epoch 768/1000, Train Loss: 0.7608, Train Acc: 0.8429, Validation Loss: 1.0785, Validation Acc: 0.7808
Epoch 769/1000, Train Loss: 0.7609, Train Acc: 0.8143, Validation Loss: 1.0778, Validation Acc: 0.7804
Epoch 770/1000, Train Loss: 0.7971, Train Acc: 0.8071, Validation Loss: 1.0770, Validation Acc: 0.7808
Epoch 771/1000, Train Loss: 0.7597, Train Acc: 0.8000, Validation Loss: 1.0763, Validation Acc: 0.7815
Epoch 772/1000, Train Loss: 0.7845, Train Acc: 0.8500, Validation Loss: 1.0756, Validation Acc: 0.7823
Epoch 773/1000, Train Loss: 0.7913, Train Acc: 0.8357, Validation Loss: 1.0750, Validation Acc: 0.7819
Epoch 774/1000, Train Loss: 0.7524, Train Acc: 0.8357, Validation Loss: 1.0745, Validation Acc: 0.7823
Epoch 775/1000, Train Loss: 0.8159, Train Acc: 0.7929, Validation Loss: 1.0742, Validation Acc: 0.7819
Epoch 776/1000, Train Loss: 0.7151, Train Acc: 0.8214, Validation Loss: 1.0739, Validation Acc: 0.7819
Epoch 777/1000, Train Loss: 0.7726, Train Acc: 0.8214, Validation Loss: 1.0736, Validation Acc: 0.7819
Epoch 778/1000, Train Loss: 0.7976, Train Acc: 0.7714, Validation Loss: 1.0735, Validation Acc: 0.7808
Epoch 779/1000, Train Loss: 0.7672, Train Acc: 0.8143, Validation Loss: 1.0734, Validation Acc: 0.7808
Epoch 780/1000, Train Loss: 0.8029, Train Acc: 0.8143, Validation Loss: 1.0735, Validation Acc: 0.7815

Epoch 781/1000, Train Loss: 0.7222, Train Acc: 0.8357, Validation Loss: 1.0736, Validation Acc: 0.7808
Epoch 782/1000, Train Loss: 0.8041, Train Acc: 0.8071, Validation Loss: 1.0738, Validation Acc: 0.7812
Epoch 783/1000, Train Loss: 0.8937, Train Acc: 0.7571, Validation Loss: 1.0741, Validation Acc: 0.7812
Epoch 784/1000, Train Loss: 0.8126, Train Acc: 0.8500, Validation Loss: 1.0745, Validation Acc: 0.7815
Epoch 785/1000, Train Loss: 0.7730, Train Acc: 0.8357, Validation Loss: 1.0749, Validation Acc: 0.7823
Epoch 786/1000, Train Loss: 0.7483, Train Acc: 0.8500, Validation Loss: 1.0753, Validation Acc: 0.7823
Epoch 787/1000, Train Loss: 0.7126, Train Acc: 0.8571, Validation Loss: 1.0755, Validation Acc: 0.7819
Epoch 788/1000, Train Loss: 0.8255, Train Acc: 0.8429, Validation Loss: 1.0758, Validation Acc: 0.7823
Epoch 789/1000, Train Loss: 0.7527, Train Acc: 0.8429, Validation Loss: 1.0760, Validation Acc: 0.7827
Epoch 790/1000, Train Loss: 0.7900, Train Acc: 0.8214, Validation Loss: 1.0761, Validation Acc: 0.7827
Epoch 791/1000, Train Loss: 0.8032, Train Acc: 0.8429, Validation Loss: 1.0760, Validation Acc: 0.7839
Epoch 792/1000, Train Loss: 0.8062, Train Acc: 0.7643, Validation Loss: 1.0759, Validation Acc: 0.7839
Epoch 793/1000, Train Loss: 0.7101, Train Acc: 0.8786, Validation Loss: 1.0759, Validation Acc: 0.7835
Epoch 794/1000, Train Loss: 0.8484, Train Acc: 0.7786, Validation Loss: 1.0760, Validation Acc: 0.7831
Epoch 795/1000, Train Loss: 0.7151, Train Acc: 0.8929, Validation Loss: 1.0761, Validation Acc: 0.7835
Epoch 796/1000, Train Loss: 0.7792, Train Acc: 0.8143, Validation Loss: 1.0763, Validation Acc: 0.7831
Epoch 797/1000, Train Loss: 0.7685, Train Acc: 0.8500, Validation Loss: 1.0767, Validation Acc: 0.7823
Epoch 798/1000, Train Loss: 0.7843, Train Acc: 0.8214, Validation Loss: 1.0769, Validation Acc: 0.7819
Epoch 799/1000, Train Loss: 0.7653, Train Acc: 0.8429, Validation Loss: 1.0770, Validation Acc: 0.7815
Epoch 800/1000, Train Loss: 0.7962, Train Acc: 0.8429, Validation Loss: 1.0768, Validation Acc: 0.7815
Epoch 801/1000, Train Loss: 0.7788, Train Acc: 0.8286, Validation Loss: 1.0766, Validation Acc: 0.7819
Epoch 802/1000, Train Loss: 0.8125, Train Acc: 0.8643, Validation Loss: 1.0763, Validation Acc: 0.7815
Epoch 803/1000, Train Loss: 0.7807, Train Acc: 0.8071, Validation Loss: 1.0760, Validation Acc: 0.7812
Epoch 804/1000, Train Loss: 0.7835, Train Acc: 0.8429, Validation Loss: 1.0756, Validation Acc: 0.7812
Epoch 805/1000, Train Loss: 0.8169, Train Acc: 0.7929, Validation Loss: 1.0750, Validation Acc: 0.7819
Epoch 806/1000, Train Loss: 0.7331, Train Acc: 0.8429, Validation Loss: 1.0746, Validation Acc: 0.7819
Epoch 807/1000, Train Loss: 0.8592, Train Acc: 0.8071, Validation Loss: 1.0739, Validation Acc: 0.7819
Epoch 808/1000, Train Loss: 0.7814, Train Acc: 0.8214, Validation Loss: 1.0733, Validation Acc: 0.7815
Epoch 809/1000, Train Loss: 0.7979, Train Acc: 0.8357, Validation Loss: 1.0728, Validation Acc: 0.7815
Epoch 810/1000, Train Loss: 0.8137, Train Acc: 0.8357, Validation Loss: 1.0723, Validation Acc: 0.7819

Epoch 811/1000, Train Loss: 0.8329, Train Acc: 0.8000, Validation Loss: 1.0721, Validation Acc: 0.7815
Epoch 812/1000, Train Loss: 0.7783, Train Acc: 0.8071, Validation Loss: 1.0719, Validation Acc: 0.7815
Epoch 813/1000, Train Loss: 0.7241, Train Acc: 0.8643, Validation Loss: 1.0717, Validation Acc: 0.7819
Epoch 814/1000, Train Loss: 0.7418, Train Acc: 0.8214, Validation Loss: 1.0715, Validation Acc: 0.7823
Epoch 815/1000, Train Loss: 0.7640, Train Acc: 0.7857, Validation Loss: 1.0712, Validation Acc: 0.7823
Epoch 816/1000, Train Loss: 0.7623, Train Acc: 0.8500, Validation Loss: 1.0710, Validation Acc: 0.7823
Epoch 817/1000, Train Loss: 0.7692, Train Acc: 0.8429, Validation Loss: 1.0707, Validation Acc: 0.7827
Epoch 818/1000, Train Loss: 0.8241, Train Acc: 0.8214, Validation Loss: 1.0704, Validation Acc: 0.7827
Epoch 819/1000, Train Loss: 0.7992, Train Acc: 0.8071, Validation Loss: 1.0702, Validation Acc: 0.7823
Epoch 820/1000, Train Loss: 0.7867, Train Acc: 0.8357, Validation Loss: 1.0701, Validation Acc: 0.7823
Epoch 821/1000, Train Loss: 0.6854, Train Acc: 0.8786, Validation Loss: 1.0699, Validation Acc: 0.7831
Epoch 822/1000, Train Loss: 0.6646, Train Acc: 0.8643, Validation Loss: 1.0697, Validation Acc: 0.7835
Epoch 823/1000, Train Loss: 0.7591, Train Acc: 0.8000, Validation Loss: 1.0697, Validation Acc: 0.7831
Epoch 824/1000, Train Loss: 0.7852, Train Acc: 0.8214, Validation Loss: 1.0695, Validation Acc: 0.7831
Epoch 825/1000, Train Loss: 0.7014, Train Acc: 0.8214, Validation Loss: 1.0693, Validation Acc: 0.7827
Epoch 826/1000, Train Loss: 0.7480, Train Acc: 0.8286, Validation Loss: 1.0691, Validation Acc: 0.7823
Epoch 827/1000, Train Loss: 0.7275, Train Acc: 0.8786, Validation Loss: 1.0687, Validation Acc: 0.7827
Epoch 828/1000, Train Loss: 0.8039, Train Acc: 0.8571, Validation Loss: 1.0686, Validation Acc: 0.7823
Epoch 829/1000, Train Loss: 0.7897, Train Acc: 0.8286, Validation Loss: 1.0684, Validation Acc: 0.7827
Epoch 830/1000, Train Loss: 0.7768, Train Acc: 0.7571, Validation Loss: 1.0685, Validation Acc: 0.7819
Epoch 831/1000, Train Loss: 0.8644, Train Acc: 0.7500, Validation Loss: 1.0689, Validation Acc: 0.7831
Epoch 832/1000, Train Loss: 0.7305, Train Acc: 0.8357, Validation Loss: 1.0694, Validation Acc: 0.7827
Epoch 833/1000, Train Loss: 0.7929, Train Acc: 0.8500, Validation Loss: 1.0698, Validation Acc: 0.7827
Epoch 834/1000, Train Loss: 0.7537, Train Acc: 0.8286, Validation Loss: 1.0701, Validation Acc: 0.7823
Epoch 835/1000, Train Loss: 0.8162, Train Acc: 0.8571, Validation Loss: 1.0702, Validation Acc: 0.7823
Epoch 836/1000, Train Loss: 0.7881, Train Acc: 0.8143, Validation Loss: 1.0703, Validation Acc: 0.7819
Epoch 837/1000, Train Loss: 0.7542, Train Acc: 0.8571, Validation Loss: 1.0703, Validation Acc: 0.7819
Epoch 838/1000, Train Loss: 0.7489, Train Acc: 0.8500, Validation Loss: 1.0703, Validation Acc: 0.7823
Epoch 839/1000, Train Loss: 0.8666, Train Acc: 0.7643, Validation Loss: 1.0704, Validation Acc: 0.7823
Epoch 840/1000, Train Loss: 0.7782, Train Acc: 0.8071, Validation Loss: 1.0706, Validation Acc: 0.7823

Epoch 841/1000, Train Loss: 0.7495, Train Acc: 0.8429, Validation Loss: 1.0707, Validation Acc: 0.7823
Epoch 842/1000, Train Loss: 0.7494, Train Acc: 0.8571, Validation Loss: 1.0706, Validation Acc: 0.7819
Epoch 843/1000, Train Loss: 0.6643, Train Acc: 0.8643, Validation Loss: 1.0706, Validation Acc: 0.7819
Epoch 844/1000, Train Loss: 0.8082, Train Acc: 0.7929, Validation Loss: 1.0707, Validation Acc: 0.7819
Epoch 845/1000, Train Loss: 0.7950, Train Acc: 0.8143, Validation Loss: 1.0708, Validation Acc: 0.7812
Epoch 846/1000, Train Loss: 0.7642, Train Acc: 0.8143, Validation Loss: 1.0709, Validation Acc: 0.7804
Epoch 847/1000, Train Loss: 0.7980, Train Acc: 0.7929, Validation Loss: 1.0710, Validation Acc: 0.7808
Epoch 848/1000, Train Loss: 0.7603, Train Acc: 0.8571, Validation Loss: 1.0712, Validation Acc: 0.7812
Epoch 849/1000, Train Loss: 0.7333, Train Acc: 0.8571, Validation Loss: 1.0713, Validation Acc: 0.7812
Epoch 850/1000, Train Loss: 0.8280, Train Acc: 0.8571, Validation Loss: 1.0712, Validation Acc: 0.7812
Epoch 851/1000, Train Loss: 0.7587, Train Acc: 0.8357, Validation Loss: 1.0712, Validation Acc: 0.7812
Epoch 852/1000, Train Loss: 0.7101, Train Acc: 0.8500, Validation Loss: 1.0710, Validation Acc: 0.7812
Epoch 853/1000, Train Loss: 0.7634, Train Acc: 0.8714, Validation Loss: 1.0709, Validation Acc: 0.7812
Epoch 854/1000, Train Loss: 0.8094, Train Acc: 0.8143, Validation Loss: 1.0708, Validation Acc: 0.7812
Epoch 855/1000, Train Loss: 0.6676, Train Acc: 0.8929, Validation Loss: 1.0707, Validation Acc: 0.7812
Epoch 856/1000, Train Loss: 0.7355, Train Acc: 0.8714, Validation Loss: 1.0705, Validation Acc: 0.7812
Epoch 857/1000, Train Loss: 0.7718, Train Acc: 0.8571, Validation Loss: 1.0702, Validation Acc: 0.7812
Epoch 858/1000, Train Loss: 0.8537, Train Acc: 0.7929, Validation Loss: 1.0699, Validation Acc: 0.7815
Epoch 859/1000, Train Loss: 0.8432, Train Acc: 0.8071, Validation Loss: 1.0698, Validation Acc: 0.7815
Epoch 860/1000, Train Loss: 0.7691, Train Acc: 0.8357, Validation Loss: 1.0695, Validation Acc: 0.7812
Epoch 861/1000, Train Loss: 0.7562, Train Acc: 0.8571, Validation Loss: 1.0692, Validation Acc: 0.7815
Epoch 862/1000, Train Loss: 0.7451, Train Acc: 0.7929, Validation Loss: 1.0689, Validation Acc: 0.7808
Epoch 863/1000, Train Loss: 0.7583, Train Acc: 0.8357, Validation Loss: 1.0686, Validation Acc: 0.7808
Epoch 864/1000, Train Loss: 0.7988, Train Acc: 0.7786, Validation Loss: 1.0683, Validation Acc: 0.7812
Epoch 865/1000, Train Loss: 0.7484, Train Acc: 0.8357, Validation Loss: 1.0682, Validation Acc: 0.7812
Epoch 866/1000, Train Loss: 0.7953, Train Acc: 0.8357, Validation Loss: 1.0680, Validation Acc: 0.7808
Epoch 867/1000, Train Loss: 0.7413, Train Acc: 0.8643, Validation Loss: 1.0679, Validation Acc: 0.7804
Epoch 868/1000, Train Loss: 0.7604, Train Acc: 0.8500, Validation Loss: 1.0676, Validation Acc: 0.7804
Epoch 869/1000, Train Loss: 0.8328, Train Acc: 0.8357, Validation Loss: 1.0672, Validation Acc: 0.7812
Epoch 870/1000, Train Loss: 0.6911, Train Acc: 0.8214, Validation Loss: 1.0668, Validation Acc: 0.7815

Epoch 871/1000, Train Loss: 0.7744, Train Acc: 0.8214, Validation Loss: 1.0665, Validation Acc: 0.7812
Epoch 872/1000, Train Loss: 0.7938, Train Acc: 0.8500, Validation Loss: 1.0663, Validation Acc: 0.7812
Epoch 873/1000, Train Loss: 0.7066, Train Acc: 0.8286, Validation Loss: 1.0661, Validation Acc: 0.7812
Epoch 874/1000, Train Loss: 0.8562, Train Acc: 0.8000, Validation Loss: 1.0658, Validation Acc: 0.7815
Epoch 875/1000, Train Loss: 0.7402, Train Acc: 0.8000, Validation Loss: 1.0656, Validation Acc: 0.7819
Epoch 876/1000, Train Loss: 0.8063, Train Acc: 0.8286, Validation Loss: 1.0655, Validation Acc: 0.7819
Epoch 877/1000, Train Loss: 0.7620, Train Acc: 0.8071, Validation Loss: 1.0654, Validation Acc: 0.7823
Epoch 878/1000, Train Loss: 0.7469, Train Acc: 0.8571, Validation Loss: 1.0654, Validation Acc: 0.7823
Epoch 879/1000, Train Loss: 0.7761, Train Acc: 0.8286, Validation Loss: 1.0653, Validation Acc: 0.7827
Epoch 880/1000, Train Loss: 0.8066, Train Acc: 0.8143, Validation Loss: 1.0651, Validation Acc: 0.7827
Epoch 881/1000, Train Loss: 0.7411, Train Acc: 0.8143, Validation Loss: 1.0648, Validation Acc: 0.7827
Epoch 882/1000, Train Loss: 0.7663, Train Acc: 0.8500, Validation Loss: 1.0645, Validation Acc: 0.7823
Epoch 883/1000, Train Loss: 0.7375, Train Acc: 0.8643, Validation Loss: 1.0642, Validation Acc: 0.7819
Epoch 884/1000, Train Loss: 0.7471, Train Acc: 0.8429, Validation Loss: 1.0638, Validation Acc: 0.7823
Epoch 885/1000, Train Loss: 0.7412, Train Acc: 0.8357, Validation Loss: 1.0636, Validation Acc: 0.7827
Epoch 886/1000, Train Loss: 0.7535, Train Acc: 0.8143, Validation Loss: 1.0633, Validation Acc: 0.7827
Epoch 887/1000, Train Loss: 0.7529, Train Acc: 0.8071, Validation Loss: 1.0633, Validation Acc: 0.7827
Epoch 888/1000, Train Loss: 0.7821, Train Acc: 0.8500, Validation Loss: 1.0632, Validation Acc: 0.7827
Epoch 889/1000, Train Loss: 0.8325, Train Acc: 0.8071, Validation Loss: 1.0631, Validation Acc: 0.7823
Epoch 890/1000, Train Loss: 0.7683, Train Acc: 0.8214, Validation Loss: 1.0631, Validation Acc: 0.7823
Epoch 891/1000, Train Loss: 0.7410, Train Acc: 0.8429, Validation Loss: 1.0632, Validation Acc: 0.7823
Epoch 892/1000, Train Loss: 0.7468, Train Acc: 0.8643, Validation Loss: 1.0633, Validation Acc: 0.7819
Epoch 893/1000, Train Loss: 0.7084, Train Acc: 0.8286, Validation Loss: 1.0633, Validation Acc: 0.7819
Epoch 894/1000, Train Loss: 0.7259, Train Acc: 0.8500, Validation Loss: 1.0633, Validation Acc: 0.7815
Epoch 895/1000, Train Loss: 0.7650, Train Acc: 0.8786, Validation Loss: 1.0633, Validation Acc: 0.7815
Epoch 896/1000, Train Loss: 0.7900, Train Acc: 0.8143, Validation Loss: 1.0632, Validation Acc: 0.7815
Epoch 897/1000, Train Loss: 0.6957, Train Acc: 0.8786, Validation Loss: 1.0632, Validation Acc: 0.7827
Epoch 898/1000, Train Loss: 0.7364, Train Acc: 0.8500, Validation Loss: 1.0631, Validation Acc: 0.7827
Epoch 899/1000, Train Loss: 0.8363, Train Acc: 0.8143, Validation Loss: 1.0632, Validation Acc: 0.7827
Epoch 900/1000, Train Loss: 0.8005, Train Acc: 0.8000, Validation Loss: 1.0632, Validation Acc: 0.7819

Epoch 901/1000, Train Loss: 0.8149, Train Acc: 0.8571, Validation Loss: 1.0634, Validation Acc: 0.7819
Epoch 902/1000, Train Loss: 0.6516, Train Acc: 0.8357, Validation Loss: 1.0636, Validation Acc: 0.7812
Epoch 903/1000, Train Loss: 0.8205, Train Acc: 0.7857, Validation Loss: 1.0637, Validation Acc: 0.7812
Epoch 904/1000, Train Loss: 0.7530, Train Acc: 0.8286, Validation Loss: 1.0637, Validation Acc: 0.7812
Epoch 905/1000, Train Loss: 0.7664, Train Acc: 0.8429, Validation Loss: 1.0638, Validation Acc: 0.7819
Epoch 906/1000, Train Loss: 0.7766, Train Acc: 0.8643, Validation Loss: 1.0639, Validation Acc: 0.7815
Epoch 907/1000, Train Loss: 0.7451, Train Acc: 0.8786, Validation Loss: 1.0639, Validation Acc: 0.7819
Epoch 908/1000, Train Loss: 0.8384, Train Acc: 0.8071, Validation Loss: 1.0640, Validation Acc: 0.7823
Epoch 909/1000, Train Loss: 0.6712, Train Acc: 0.8429, Validation Loss: 1.0643, Validation Acc: 0.7819
Epoch 910/1000, Train Loss: 0.7739, Train Acc: 0.8500, Validation Loss: 1.0644, Validation Acc: 0.7812
Epoch 911/1000, Train Loss: 0.8074, Train Acc: 0.8143, Validation Loss: 1.0644, Validation Acc: 0.7808
Epoch 912/1000, Train Loss: 0.7541, Train Acc: 0.8071, Validation Loss: 1.0645, Validation Acc: 0.7808
Epoch 913/1000, Train Loss: 0.6507, Train Acc: 0.8643, Validation Loss: 1.0646, Validation Acc: 0.7808
Epoch 914/1000, Train Loss: 0.8386, Train Acc: 0.8286, Validation Loss: 1.0645, Validation Acc: 0.7808
Epoch 915/1000, Train Loss: 0.7952, Train Acc: 0.7714, Validation Loss: 1.0644, Validation Acc: 0.7808
Epoch 916/1000, Train Loss: 0.7625, Train Acc: 0.8071, Validation Loss: 1.0643, Validation Acc: 0.7812
Epoch 917/1000, Train Loss: 0.7450, Train Acc: 0.8571, Validation Loss: 1.0641, Validation Acc: 0.7812
Epoch 918/1000, Train Loss: 0.7306, Train Acc: 0.8429, Validation Loss: 1.0640, Validation Acc: 0.7812
Epoch 919/1000, Train Loss: 0.7231, Train Acc: 0.9000, Validation Loss: 1.0637, Validation Acc: 0.7815
Epoch 920/1000, Train Loss: 0.7357, Train Acc: 0.8429, Validation Loss: 1.0635, Validation Acc: 0.7819
Epoch 921/1000, Train Loss: 0.8074, Train Acc: 0.8214, Validation Loss: 1.0633, Validation Acc: 0.7823
Epoch 922/1000, Train Loss: 0.8179, Train Acc: 0.7929, Validation Loss: 1.0631, Validation Acc: 0.7823
Epoch 923/1000, Train Loss: 0.7759, Train Acc: 0.8571, Validation Loss: 1.0628, Validation Acc: 0.7819
Epoch 924/1000, Train Loss: 0.7124, Train Acc: 0.8286, Validation Loss: 1.0624, Validation Acc: 0.7819
Epoch 925/1000, Train Loss: 0.6265, Train Acc: 0.8786, Validation Loss: 1.0621, Validation Acc: 0.7819
Epoch 926/1000, Train Loss: 0.7825, Train Acc: 0.8000, Validation Loss: 1.0618, Validation Acc: 0.7819
Epoch 927/1000, Train Loss: 0.7452, Train Acc: 0.8286, Validation Loss: 1.0616, Validation Acc: 0.7819
Epoch 928/1000, Train Loss: 0.7683, Train Acc: 0.8357, Validation Loss: 1.0613, Validation Acc: 0.7819
Epoch 929/1000, Train Loss: 0.7565, Train Acc: 0.8429, Validation Loss: 1.0609, Validation Acc: 0.7815
Epoch 930/1000, Train Loss: 0.8446, Train Acc: 0.8000, Validation Loss: 1.0607, Validation Acc: 0.7819

Epoch 931/1000, Train Loss: 0.6840, Train Acc: 0.8571, Validation Loss: 1.0606, Validation Acc: 0.7819
Epoch 932/1000, Train Loss: 0.6928, Train Acc: 0.8357, Validation Loss: 1.0606, Validation Acc: 0.7823
Epoch 933/1000, Train Loss: 0.8183, Train Acc: 0.8071, Validation Loss: 1.0607, Validation Acc: 0.7823
Epoch 934/1000, Train Loss: 0.7297, Train Acc: 0.8643, Validation Loss: 1.0607, Validation Acc: 0.7823
Epoch 935/1000, Train Loss: 0.7488, Train Acc: 0.8429, Validation Loss: 1.0605, Validation Acc: 0.7815
Epoch 936/1000, Train Loss: 0.7138, Train Acc: 0.8571, Validation Loss: 1.0603, Validation Acc: 0.7812
Epoch 937/1000, Train Loss: 0.7463, Train Acc: 0.8571, Validation Loss: 1.0601, Validation Acc: 0.7819
Epoch 938/1000, Train Loss: 0.8177, Train Acc: 0.8214, Validation Loss: 1.0598, Validation Acc: 0.7819
Epoch 939/1000, Train Loss: 0.7743, Train Acc: 0.8429, Validation Loss: 1.0595, Validation Acc: 0.7819
Epoch 940/1000, Train Loss: 0.7686, Train Acc: 0.8000, Validation Loss: 1.0592, Validation Acc: 0.7819
Epoch 941/1000, Train Loss: 0.7705, Train Acc: 0.8071, Validation Loss: 1.0589, Validation Acc: 0.7819
Epoch 942/1000, Train Loss: 0.8219, Train Acc: 0.7929, Validation Loss: 1.0586, Validation Acc: 0.7815
Epoch 943/1000, Train Loss: 0.7106, Train Acc: 0.8929, Validation Loss: 1.0583, Validation Acc: 0.7815
Epoch 944/1000, Train Loss: 0.7988, Train Acc: 0.8143, Validation Loss: 1.0580, Validation Acc: 0.7812
Epoch 945/1000, Train Loss: 0.7174, Train Acc: 0.8000, Validation Loss: 1.0579, Validation Acc: 0.7812
Epoch 946/1000, Train Loss: 0.7131, Train Acc: 0.8786, Validation Loss: 1.0577, Validation Acc: 0.7812
Epoch 947/1000, Train Loss: 0.7090, Train Acc: 0.8429, Validation Loss: 1.0575, Validation Acc: 0.7808
Epoch 948/1000, Train Loss: 0.7510, Train Acc: 0.7929, Validation Loss: 1.0572, Validation Acc: 0.7808
Epoch 949/1000, Train Loss: 0.7547, Train Acc: 0.8071, Validation Loss: 1.0570, Validation Acc: 0.7808
Epoch 950/1000, Train Loss: 0.7958, Train Acc: 0.8214, Validation Loss: 1.0568, Validation Acc: 0.7808
Epoch 951/1000, Train Loss: 0.8297, Train Acc: 0.8357, Validation Loss: 1.0567, Validation Acc: 0.7800
Epoch 952/1000, Train Loss: 0.7355, Train Acc: 0.8357, Validation Loss: 1.0565, Validation Acc: 0.7800
Epoch 953/1000, Train Loss: 0.7115, Train Acc: 0.8357, Validation Loss: 1.0565, Validation Acc: 0.7804
Epoch 954/1000, Train Loss: 0.7310, Train Acc: 0.8500, Validation Loss: 1.0563, Validation Acc: 0.7804
Epoch 955/1000, Train Loss: 0.7407, Train Acc: 0.8714, Validation Loss: 1.0561, Validation Acc: 0.7804
Epoch 956/1000, Train Loss: 0.7518, Train Acc: 0.8429, Validation Loss: 1.0560, Validation Acc: 0.7808
Epoch 957/1000, Train Loss: 0.7030, Train Acc: 0.8286, Validation Loss: 1.0562, Validation Acc: 0.7804
Epoch 958/1000, Train Loss: 0.8048, Train Acc: 0.8286, Validation Loss: 1.0561, Validation Acc: 0.7796
Epoch 959/1000, Train Loss: 0.7941, Train Acc: 0.7857, Validation Loss: 1.0560, Validation Acc: 0.7800
Epoch 960/1000, Train Loss: 0.8006, Train Acc: 0.8357, Validation Loss: 1.0560, Validation Acc: 0.7800

Epoch 961/1000, Train Loss: 0.7446, Train Acc: 0.8714, Validation Loss: 1.0560, Validation Acc: 0.7796
Epoch 962/1000, Train Loss: 0.8084, Train Acc: 0.8214, Validation Loss: 1.0560, Validation Acc: 0.7796
Epoch 963/1000, Train Loss: 0.7939, Train Acc: 0.8286, Validation Loss: 1.0559, Validation Acc: 0.7796
Epoch 964/1000, Train Loss: 0.8108, Train Acc: 0.8214, Validation Loss: 1.0559, Validation Acc: 0.7796
Epoch 965/1000, Train Loss: 0.8095, Train Acc: 0.7857, Validation Loss: 1.0558, Validation Acc: 0.7792
Epoch 966/1000, Train Loss: 0.7067, Train Acc: 0.8786, Validation Loss: 1.0557, Validation Acc: 0.7796
Epoch 967/1000, Train Loss: 0.7024, Train Acc: 0.8429, Validation Loss: 1.0559, Validation Acc: 0.7796
Epoch 968/1000, Train Loss: 0.7696, Train Acc: 0.7857, Validation Loss: 1.0558, Validation Acc: 0.7800
Epoch 969/1000, Train Loss: 0.7304, Train Acc: 0.8857, Validation Loss: 1.0556, Validation Acc: 0.7804
Epoch 970/1000, Train Loss: 0.7468, Train Acc: 0.8429, Validation Loss: 1.0553, Validation Acc: 0.7804
Epoch 971/1000, Train Loss: 0.7515, Train Acc: 0.8429, Validation Loss: 1.0551, Validation Acc: 0.7804
Epoch 972/1000, Train Loss: 0.7303, Train Acc: 0.8857, Validation Loss: 1.0549, Validation Acc: 0.7800
Epoch 973/1000, Train Loss: 0.7761, Train Acc: 0.8000, Validation Loss: 1.0547, Validation Acc: 0.7804
Epoch 974/1000, Train Loss: 0.8202, Train Acc: 0.8000, Validation Loss: 1.0543, Validation Acc: 0.7804
Epoch 975/1000, Train Loss: 0.7324, Train Acc: 0.8214, Validation Loss: 1.0540, Validation Acc: 0.7812
Epoch 976/1000, Train Loss: 0.7845, Train Acc: 0.8643, Validation Loss: 1.0536, Validation Acc: 0.7812
Epoch 977/1000, Train Loss: 0.7232, Train Acc: 0.8357, Validation Loss: 1.0531, Validation Acc: 0.7815
Epoch 978/1000, Train Loss: 0.7623, Train Acc: 0.8357, Validation Loss: 1.0526, Validation Acc: 0.7815
Epoch 979/1000, Train Loss: 0.7305, Train Acc: 0.8643, Validation Loss: 1.0521, Validation Acc: 0.7815
Epoch 980/1000, Train Loss: 0.7613, Train Acc: 0.8786, Validation Loss: 1.0515, Validation Acc: 0.7808
Epoch 981/1000, Train Loss: 0.7495, Train Acc: 0.8500, Validation Loss: 1.0509, Validation Acc: 0.7812
Epoch 982/1000, Train Loss: 0.8566, Train Acc: 0.7500, Validation Loss: 1.0505, Validation Acc: 0.7812
Epoch 983/1000, Train Loss: 0.7836, Train Acc: 0.8357, Validation Loss: 1.0501, Validation Acc: 0.7812
Epoch 984/1000, Train Loss: 0.7189, Train Acc: 0.9000, Validation Loss: 1.0497, Validation Acc: 0.7812
Epoch 985/1000, Train Loss: 0.7334, Train Acc: 0.8643, Validation Loss: 1.0494, Validation Acc: 0.7808
Epoch 986/1000, Train Loss: 0.7803, Train Acc: 0.8286, Validation Loss: 1.0491, Validation Acc: 0.7808
Epoch 987/1000, Train Loss: 0.7021, Train Acc: 0.8929, Validation Loss: 1.0487, Validation Acc: 0.7800
Epoch 988/1000, Train Loss: 0.8173, Train Acc: 0.7786, Validation Loss: 1.0485, Validation Acc: 0.7800
Epoch 989/1000, Train Loss: 0.8462, Train Acc: 0.7643, Validation Loss: 1.0483, Validation Acc: 0.7800
Epoch 990/1000, Train Loss: 0.7328, Train Acc: 0.8143, Validation Loss: 1.0481, Validation Acc: 0.7800

```
Epoch 991/1000, Train Loss: 0.6917, Train Acc: 0.8286, Validation Loss: 1.0480, Validation Acc: 0.7804
Epoch 992/1000, Train Loss: 0.7212, Train Acc: 0.8429, Validation Loss: 1.0479, Validation Acc: 0.7800
Epoch 993/1000, Train Loss: 0.8003, Train Acc: 0.8429, Validation Loss: 1.0478, Validation Acc: 0.7800
Epoch 994/1000, Train Loss: 0.7350, Train Acc: 0.8214, Validation Loss: 1.0477, Validation Acc: 0.7800
Epoch 995/1000, Train Loss: 0.7142, Train Acc: 0.8786, Validation Loss: 1.0477, Validation Acc: 0.7800
Epoch 996/1000, Train Loss: 0.8023, Train Acc: 0.8429, Validation Loss: 1.0476, Validation Acc: 0.7800
Epoch 997/1000, Train Loss: 0.7228, Train Acc: 0.8214, Validation Loss: 1.0476, Validation Acc: 0.7800
Epoch 998/1000, Train Loss: 0.7295, Train Acc: 0.8357, Validation Loss: 1.0476, Validation Acc: 0.7788
Epoch 999/1000, Train Loss: 0.7356, Train Acc: 0.8714, Validation Loss: 1.0476, Validation Acc: 0.7788
Epoch 1000/1000, Train Loss: 0.6793, Train Acc: 0.8714, Validation Loss: 1.0475, Validation Acc: 0.7784
Optimization Finished!
Total time elapsed: 495.8697s
Test Loss: 1.0475, Test Acc: 0.7784

Out[ ]: (1.0474703311920166,
          0.7784267912772586,
          tensor([0, 4, 1, ..., 6, 4, 2], device='cuda:0'))
```

Question: (Your task)

Compare the evaluation results for Vanilla GCN (from the tutorial session) and GAT. Comment on the discrepancy in their performance (if any) and briefly explain why you think it's the case.

Your answer here:

Comparison of Accuracies:

- Vanilla GCN: Achieved 70.33% accuracy on test set.
- GAT: Achieved **77.84%** accuracy on test set.

Comments:

GAT clearly outperforms vanilla GCN by 7.5%. The reason for this is the attention mechanism that GAT uses. GAT assigns different values to different neighbours of a node and uses the important neighbours and ignores the others. On the other hand, vanilla GCN treats all the neighbours same because it uses fixed weights based on the graph structure. This attention mechanism helps GAT to capture more complex patterns.

Part 2: Transformers

In the tutorial session, you learned how to use a transformer model for language modeling. This is a type of decoder-only model, where we directly use the input embeddings to generate text. Here, we will focus on a different task instead: translation. To do this, we will need to use an encoder-decoder architecture.

Part 2.1: Transformer Layers

Here, you will need to implement a few different layers.

1. A multi-head attention block.
2. A feedforward network.
3. A positional encoding (in the tutorial we used a learned positional encoding. In this assignment, you should implement a sinusoidal positional encoding).

Note that the template code provided for you below is only a template, and you may modify the parameters for the functions as needed.

Once you have these, you will need to combine them together into an encoder and a decoder. Each of these should have N sets of multi-head attention + feedforward. However, the encoder and decoder have to be implemented slightly differently, as the decoder requires masking.

Additionally, since this is a translation task, we will need two sets of word embeddings (the input embeddings corresponding to the starting language, and the output embeddings corresponding to the translated language).

Multi-Head Attention

Multi-head attention should be implemented slightly differently here, compared to the tutorial. Specifically, your multi-head attention implementation should allow for masking. This means that if an input mask is provided, everything that is masked out (a value of 0 in the mask) should not contribute to the attention calculations.

It is crucial that the decoder is properly masked. This is because we do not want to allow the decoder to apply attention to output positions it has not seen yet.

```
In [ ]: class MultiHeadAttention(nn.Module):
    def __init__(self, input_dimension, hidden_dim, num_heads):
        super().__init__()
        self.num_heads = num_heads
        self.head_dim = hidden_dim // num_heads
        self.query = nn.Linear(input_dimension, hidden_dim)
        self.key = nn.Linear(input_dimension, hidden_dim)
        self.value = nn.Linear(input_dimension, hidden_dim)
        self.out = nn.Linear(hidden_dim, input_dimension)

    def forward(self, x, context=None, mask=None, return_attn=False):
        B, T, C = x.shape
        q = self.query(x)
        k = self.key(context)
```

```

    v = self.value(context)

    q = q.view(B, T, self.num_heads, self.head_dim).transpose(1, 2)
    k = k.view(B, -1, self.num_heads, self.head_dim).transpose(1, 2)
    v = v.view(B, -1, self.num_heads, self.head_dim).transpose(1, 2)

    attention_scores = torch.matmul(q, k.transpose(-2, -1)) / math.sqrt(self.head_dim)

    if mask is not None:
        attention_scores = attention_scores.masked_fill(mask == 0, float("-inf"))

    attention_weights = F.softmax(attention_scores, dim=-1)
    self.last_attention_weights = attention_weights.detach()

    out = torch.matmul(attention_weights, v)
    out = out.transpose(1, 2).contiguous().view(B, T, -1)
    out = self.out(out)

    if return_attn:
        return out, attention_weights
    return out

class FeedForward(nn.Module):
    def __init__(self, input_dimension, hidden_dim, output_dim, n_layers, dropout=0.1):
        super().__init__()

        self.layers = nn.ModuleList()
        if n_layers == 1:
            self.layers.append(nn.Linear(input_dimension, output_dim))
        else:
            self.layers.append(nn.Linear(input_dimension, hidden_dim))
            for _ in range(n_layers - 2):
                self.layers.append(nn.Linear(hidden_dim, hidden_dim))
            self.layers.append(nn.Linear(hidden_dim, output_dim))
        self.activation = nn.ReLU()
        self.dropout = nn.Dropout(dropout)

    def forward(self, x):
        for i, layer in enumerate(self.layers):
            x = layer(x)
            if i < len(self.layers) - 1:
                x = self.activation(x)
                x = self.dropout(x)
        return x

class PositionalEncoding(nn.Module):
    def __init__(self, hidden_dim, max_len=5000):
        super().__init__()
        position = torch.arange(0, max_len).unsqueeze(1)
        div_term = torch.exp(torch.arange(0, hidden_dim, 2) * (-math.log(10000.0) / hidden_dim))
        positional_encoding = torch.zeros(max_len, hidden_dim)
        positional_encoding[:, 0::2] = torch.sin(position * div_term)
        positional_encoding[:, 1::2] = torch.cos(position * div_term)
        self.register_buffer('positional_encoding', positional_encoding)

    def forward(self, x):
        return x + self.positional_encoding[:x.size(1)]

```

In [63]: `class TransformerEncoder(nn.Module):`

`def __init__(self, input_dim, hidden_dim, num_heads, num_layers, output_dim):`

```

super().__init__()
# TODO
self.layers = nn.ModuleList([
    nn.ModuleDict({
        "attention": MultiHeadAttention(input_dim, hidden_dim, num_heads),
        "feedforward": FeedForward(input_dim, hidden_dim * 4, input_dim),
        "norm1": nn.LayerNorm(input_dim),
        "norm2": nn.LayerNorm(input_dim),
        "dropout": nn.Dropout(dropout)
    })
    for _ in range(num_layers)
])
self.positional_encoding = PositionalEncoding(input_dim)

def forward(self, x, mask=None, return_attn_maps=False):
    # TODO
    x = self.positional_encoding(x)
    attn_maps = []
    for layer in self.layers:
        if return_attn_maps:
            attention_output, attn = layer["attention"](x, context=None)
            attn_maps.append(attn)
        else:
            attention_output = layer["attention"](x, context=None, mask=mask)
            x = layer["norm1"](x + layer["dropout"](attention_output))

        feedforward_output = layer["feedforward"](x)
        x = layer["norm2"](x + feedforward_output)

    if return_attn_maps:
        return x, attn_maps
    return x

class TransformerDecoder(nn.Module):
    def __init__(self, input_dim, hidden_dim, num_heads, num_layers, output_dim):
        super().__init__()
        self.layers = nn.ModuleList([
            nn.ModuleDict({
                "self_attention": MultiHeadAttention(input_dim, hidden_dim, num_heads),
                "cross_attention": MultiHeadAttention(input_dim, hidden_dim, num_heads),
                "feedforward": FeedForward(input_dim, hidden_dim * 4, input_dim),
                "norm1": nn.LayerNorm(input_dim),
                "norm2": nn.LayerNorm(input_dim),
                "norm3": nn.LayerNorm(input_dim),
                "dropout": nn.Dropout(dropout)
            })
            for _ in range(num_layers)
        ])
        self.positional_encoding = PositionalEncoding(input_dim)

    def forward(self, x, encoder_output, src_mask=None, tgt_mask=None, return_attn_maps=False):
        x = self.positional_encoding(x)
        self_attn_maps = []
        for layer in self.layers:

            if return_attn_maps:
                self_attention_output, self_attn = layer["self_attention"](
                    x, context=None, mask=src_mask)
                self_attn_maps.append(self_attn)
            else:
                self_attention_output = layer["self_attention"](
                    x, context=None, mask=src_mask)
                x = layer["norm1"](x + layer["dropout"](self_attention_output))

            cross_attention_output = layer["cross_attention"](
                encoder_output, context=x, context_mask=tgt_mask)
            x = layer["norm2"](x + layer["dropout"](cross_attention_output))

            feedforward_output = layer["feedforward"](x)
            x = layer["norm3"](x + feedforward_output)

        if return_attn_maps:
            return x, self_attn_maps
        return x

```

```
        self_attention_output = layer["self_attention"](x, context)
        x = layer["norm1"](x + layer["dropout"](self_attention_output))

        cross_attention_output = layer["cross_attention"](x, context)
        x = layer["norm2"](x + layer["dropout"](cross_attention_output))

        feedforward_output = layer["feedforward"](x)
        x = layer["norm3"](x + feedforward_output)

    if return_attn_maps:
        return x, self_attn_maps
    return x
```

```
In [64]: class Transformer(nn.Module):
    def __init__(self, input_dim, hidden_dim, num_heads, num_layers, output_dim):
        super().__init__()
        self.hidden_dim = hidden_dim
        self.src_embedding = nn.Embedding(input_dim, hidden_dim)
        self.tgt_embedding = nn.Embedding(output_dim, hidden_dim)
        self.dropout = nn.Dropout(dropout)

        self.encoder = TransformerEncoder(
            input_dim=hidden_dim,
            hidden_dim=hidden_dim,
            num_heads=num_heads,
            num_layers=num_layers,
            output_dim=hidden_dim,
            dropout=dropout
        )
        self.decoder = TransformerDecoder(
            input_dim=hidden_dim,
            hidden_dim=hidden_dim,
            num_heads=num_heads,
            num_layers=num_layers,
            output_dim=hidden_dim,
            dropout=dropout
        )

        self.output_layer = nn.Linear(hidden_dim, output_dim)

    def forward(self, src, tgt, src_mask=None, tgt_mask=None, return_attn_maps=False):
        src_emb = self.src_embedding(src) * math.sqrt(self.hidden_dim)
        src_emb = self.dropout(src_emb)

        encoder_output = self.encoder(src_emb, src_mask)
        tgt_emb = self.tgt_embedding(tgt) * math.sqrt(self.hidden_dim)
        tgt_emb = self.dropout(tgt_emb)

        if return_attn_maps:
            decoder_output, decoder_self_attn = self.decoder(
                tgt_emb, encoder_output, src_mask, tgt_mask, return_attn_maps
            )
        else:
            decoder_output = self.decoder(tgt_emb, encoder_output, src_mask)

        output = self.output_layer(decoder_output)

        if return_attn_maps:
```

```

return output, {"decoder_self": decoder_self_attn}
return output

```

Part 2.2: Training the Transformer

We will be training our transformer on English -> Turkish translation. In the provided dataset, there are files labeled `{language}.{split}`, for example `en.train`. These are paired data, meaning that each line in the english file has the corresponding translation in the turkish file. There are also 0.5, 0.25, and 0.125 files, which correspond to smaller versions of the training data. Based on training times you encounter, you can use the smaller versions if needed.

As with the language modeling task, the model will output logits for each word, which you will then use for the loss.

Setting up the Data

Here, we will use word tokens, similar to the RNN assignment. However, you will need to have two sets of tokens, one for the starting language and one for the ending language. You are free to copy the data loading code from assignment 3 and make the necessary modifications, or you may implement the data loading in your own way. However, since we are no longer using fixed length sentences, you will need to implement padding (recall assignment 1).

In [65]:

```

import os
import os
from collections import Counter
from torch.utils.data import Dataset, DataLoader
from torch.nn.utils.rnn import pad_sequence

PAD_TOKEN = "<PAD>"
SOS_TOKEN = "<SOS>"
EOS_TOKEN = "<EOS>"
UNK_TOKEN = "<UNK>"

PAD_IDX = 0
SOS_IDX = 1
EOS_IDX = 2
UNK_IDX = 3

```

In [66]:

```

class Vocabulary:
    def __init__(self, min_freq=1):
        self.word2idx = {PAD_TOKEN: PAD_IDX, SOS_TOKEN: SOS_IDX, EOS_TOKEN: EOS_IDX, UNK_TOKEN: UNK_IDX}
        self.idx2word = {PAD_IDX: PAD_TOKEN, SOS_IDX: SOS_TOKEN, EOS_IDX: EOS_TOKEN, UNK_IDX: UNK_TOKEN}
        self.word_count = Counter()
        self.min_freq = min_freq

    def build_vocab(self, sentences):
        for sentence in sentences:
            self.word_count.update(sentence)

        for word, count in self.word_count.items():
            if count >= self.min_freq and word not in self.word2idx:
                idx = len(self.word2idx)
                self.word2idx[word] = idx
                self.idx2word[idx] = word

```

```

        self.word2idx[word] = idx
        self.idx2word[idx] = word

    def __len__(self):
        return len(self.word2idx)

    def encode(self, sentence):
        return [SOS_IDX] + [self.word2idx.get(word, UNK_IDX) for word in sentence]

    def decode(self, indices):
        return [self.idx2word.get(idx, UNK_TOKEN) for idx in indices]

```

In [67]:

```

class TranslationDataset(Dataset):
    def __init__(self, src_sentences, tgt_sentences, src_vocab, tgt_vocab):
        self.src_sentences = src_sentences
        self.tgt_sentences = tgt_sentences
        self.src_vocab = src_vocab
        self.tgt_vocab = tgt_vocab

    def __len__(self):
        return len(self.src_sentences)

    def __getitem__(self, idx):
        src = self.src_vocab.encode(self.src_sentences[idx])
        tgt = self.tgt_vocab.encode(self.tgt_sentences[idx])
        return torch.tensor(src), torch.tensor(tgt)

```

In [68]:

```

def collate_fn(batch):
    src_batch, tgt_batch = zip(*batch)

    src_padded = pad_sequence(src_batch, batch_first=True, padding_value=0)
    tgt_padded = pad_sequence(tgt_batch, batch_first=True, padding_value=0)

    return src_padded, tgt_padded

def load_sentences(filepath):
    sentences = []
    with open(filepath, 'r', encoding='utf-8') as f:
        for line in f:
            tokens = line.strip().lower().split()
            if tokens:
                sentences.append(tokens)
    return sentences

def load_data(base_dir, suffix=''):
    en_train = load_sentences(os.path.join(base_dir, f'en.train{suffix}'))
    tr_train = load_sentences(os.path.join(base_dir, f'tr.train{suffix}'))
    en_dev = load_sentences(os.path.join(base_dir, "en.dev"))
    tr_dev = load_sentences(os.path.join(base_dir, "tr.dev"))
    en_test = load_sentences(os.path.join(base_dir, "en.test"))
    tr_test = load_sentences(os.path.join(base_dir, "tr.test"))

    en_vocab = Vocabulary(min_freq=2)
    tr_vocab = Vocabulary(min_freq=2)
    en_vocab.build_vocab(en_train)
    tr_vocab.build_vocab(tr_train)

    print(f"English vocabulary size: {len(en_vocab)}")
    print(f"Turkish vocabulary size: {len(tr_vocab)}")

```

```

print(f"Training samples: {len(en_train)}")
print(f"Validation samples: {len(en_dev)}")
print(f"Test samples: {len(en_test)}")

train_dataset = TranslationDataset(tr_train, en_train, tr_vocab, en_v
dev_dataset = TranslationDataset(tr_dev, en_dev, tr_vocab, en_vocab)
test_dataset = TranslationDataset(tr_test, en_test, tr_vocab, en_voca

return train_dataset, dev_dataset, test_dataset, en_vocab, tr_vocab

```

In [69]:

```

train_dataset, dev_dataset, test_dataset, en_vocab, tr_vocab = load_data()

BATCH_SIZE = 64

train_loader = DataLoader(train_dataset, batch_size=BATCH_SIZE, shuffle=True)
dev_loader = DataLoader(dev_dataset, batch_size=BATCH_SIZE, shuffle=False)
test_loader = DataLoader(test_dataset, batch_size=BATCH_SIZE, shuffle=False)

```

English vocabulary size: 11541
 Turkish vocabulary size: 20531
 Training samples: 22806
 Validation samples: 4045
 Test samples: 5029

Training the Model

To train the model, you can use a similar approach to the tutorial. However, to calculate the loss, you will need both the source and target language tokens.

You should test with multiple sets of hyperparameters and provide the values and results of the best set on the validation data. For this, you should implement a function to calculate the accuracy.

In [70]:

```

def create_masks(src, tgt):
    src_mask = (src != PAD_IDX).unsqueeze(1).unsqueeze(2)
    tgt_pad_mask = (tgt != PAD_IDX).unsqueeze(1).unsqueeze(2)

    tgt_len = tgt.size(1)
    causal_mask = torch.tril(torch.ones(tgt_len, tgt_len, device=tgt.device))
    causal_mask = causal_mask.unsqueeze(0).unsqueeze(1)

    tgt_mask = torch.logical_and(tgt_pad_mask, causal_mask)

    return src_mask, tgt_mask

```

In [71]:

```

def train(model, train_loader, optimizer, criterion, device):
    model.train()
    total_loss = 0
    correct = 0
    total = 0

    for src, tgt in train_loader:
        src, tgt = src.to(device), tgt.to(device)

        tgt_input = tgt[:, :-1]
        tgt_output = tgt[:, 1:]

        src_mask, tgt_mask = create_masks(src, tgt_input)

```

```
optimizer.zero_grad()

output = model(src, tgt_input, src_mask, tgt_mask)

output = output.reshape(-1, output.size(-1))
tgt_output = tgt_output.reshape(-1)

loss = criterion(output, tgt_output)
loss.backward()
torch.nn.utils.clip_grad_norm_(model.parameters(), max_norm=1.0)
optimizer.step()
total_loss += loss.item()

predictions = output.argmax(dim=-1)
mask = tgt_output != PAD_IDX
correct += ((predictions == tgt_output) & mask).sum().item()
total += mask.sum().item()

accuracy = correct / total if total > 0 else 0
return total_loss / len(train_loader), accuracy

def evaluate(model, data_loader, criterion, device):
    model.eval()
    total_loss = 0
    correct = 0
    total = 0

    with torch.no_grad():
        for src, tgt in data_loader:
            src, tgt = src.to(device), tgt.to(device)

            tgt_input = tgt[:, :-1]
            tgt_output = tgt[:, 1:]

            src_mask, tgt_mask = create_masks(src, tgt_input)

            output = model(src, tgt_input, src_mask, tgt_mask)

            output_dim = output.shape[-1]
            output = output.contiguous().view(-1, output_dim)
            tgt_output = tgt_output.contiguous().view(-1)

            loss = criterion(output, tgt_output)
            total_loss += loss.item()

            predictions = output.argmax(dim=-1)
            mask = tgt_output != PAD_IDX
            correct += ((predictions == tgt_output) & mask).sum().item()
            total += mask.sum().item()

            accuracy = correct / total if total > 0 else 0
    return total_loss / len(data_loader), accuracy

HIDDEN_DIM = 256
NUM_HEADS = 4
NUM_LAYERS = 3
DROPOUT = 0.1
LEARNING_RATE = 0.001
```

```
NUM_EPOCHS = 20

model = Transformer(
    input_dim=len(tr_vocab),
    hidden_dim=HIDDEN_DIM,
    num_heads=NUM_HEADS,
    num_layers=NUM_LAYERS,
    output_dim=len(en_vocab),
    dropout=DROPOUT
).to(device)

criterion = nn.CrossEntropyLoss(ignore_index=PAD_IDX)
optimizer = optim.Adam(model.parameters(), lr=LEARNING_RATE)

train_losses = []
train_accuracies = []
val_losses = []
val_accuracies = []

for epoch in range(NUM_EPOCHS):
    start_time = time.time()

    train_loss, train_acc = train(model, train_loader, optimizer, criterion)
    val_loss, val_acc = evaluate(model, dev_loader, criterion, device)

    train_losses.append(train_loss)
    train_accuracies.append(train_acc)
    val_losses.append(val_loss)
    val_accuracies.append(val_acc)

    elapsed = time.time() - start_time

    print(f"Epoch {epoch+1}/{NUM_EPOCHS} | Time: {elapsed:.1f}s")
    print(f" Train Loss: {train_loss:.4f} | Train Accuracy: {train_acc:.4f}")
    print(f" Val Loss: {val_loss:.4f} | Val Accuracy: {val_acc:.4f}")
```

Epoch 1/20 | Time: 37.7s
Train Loss: 5.1363 | Train Accuracy: 0.2101
Val Loss: 4.5223 | Val Accuracy: 0.2559
Epoch 2/20 | Time: 39.2s
Train Loss: 4.4970 | Train Accuracy: 0.2587
Val Loss: 4.3134 | Val Accuracy: 0.2780
Epoch 3/20 | Time: 38.1s
Train Loss: 4.1890 | Train Accuracy: 0.2810
Val Loss: 4.1561 | Val Accuracy: 0.2983
Epoch 4/20 | Time: 38.7s
Train Loss: 3.9330 | Train Accuracy: 0.2990
Val Loss: 4.0944 | Val Accuracy: 0.3076
Epoch 5/20 | Time: 38.5s
Train Loss: 3.7022 | Train Accuracy: 0.3149
Val Loss: 4.0748 | Val Accuracy: 0.3123
Epoch 6/20 | Time: 38.4s
Train Loss: 3.4852 | Train Accuracy: 0.3303
Val Loss: 4.0578 | Val Accuracy: 0.3217
Epoch 7/20 | Time: 38.6s
Train Loss: 3.2888 | Train Accuracy: 0.3462
Val Loss: 4.1045 | Val Accuracy: 0.3156
Epoch 8/20 | Time: 38.7s
Train Loss: 3.1053 | Train Accuracy: 0.3622
Val Loss: 4.1423 | Val Accuracy: 0.3217
Epoch 9/20 | Time: 38.5s
Train Loss: 2.9362 | Train Accuracy: 0.3807
Val Loss: 4.1905 | Val Accuracy: 0.3195
Epoch 10/20 | Time: 38.4s
Train Loss: 2.7632 | Train Accuracy: 0.4004
Val Loss: 4.2410 | Val Accuracy: 0.3246
Epoch 11/20 | Time: 38.7s
Train Loss: 2.6208 | Train Accuracy: 0.4193
Val Loss: 4.3046 | Val Accuracy: 0.3231
Epoch 12/20 | Time: 38.6s
Train Loss: 2.4854 | Train Accuracy: 0.4365
Val Loss: 4.3742 | Val Accuracy: 0.3221
Epoch 13/20 | Time: 38.3s
Train Loss: 2.3763 | Train Accuracy: 0.4515
Val Loss: 4.4318 | Val Accuracy: 0.3187
Epoch 14/20 | Time: 38.3s
Train Loss: 2.2657 | Train Accuracy: 0.4681
Val Loss: 4.5140 | Val Accuracy: 0.3183
Epoch 15/20 | Time: 38.2s
Train Loss: 2.1699 | Train Accuracy: 0.4819
Val Loss: 4.5941 | Val Accuracy: 0.3128
Epoch 16/20 | Time: 38.6s
Train Loss: 2.0793 | Train Accuracy: 0.4964
Val Loss: 4.6482 | Val Accuracy: 0.3172
Epoch 17/20 | Time: 38.8s
Train Loss: 2.0014 | Train Accuracy: 0.5094
Val Loss: 4.7254 | Val Accuracy: 0.3171
Epoch 18/20 | Time: 38.5s
Train Loss: 1.9250 | Train Accuracy: 0.5209
Val Loss: 4.7960 | Val Accuracy: 0.3139
Epoch 19/20 | Time: 38.7s
Train Loss: 1.8631 | Train Accuracy: 0.5320
Val Loss: 4.8727 | Val Accuracy: 0.3214
Epoch 20/20 | Time: 38.5s
Train Loss: 1.8104 | Train Accuracy: 0.5407
Val Loss: 4.9605 | Val Accuracy: 0.3107

```
In [74]: fig, axes = plt.subplots(1, 2, figsize=(14, 5))

axes[0].plot(train_losses, label='Train', marker='o', markersize=3)
axes[0].plot(val_losses, label='Validation', marker='s', markersize=3)
axes[0].set_xlabel('Epoch')
axes[0].set_ylabel('Loss')
axes[0].legend()
axes[0].set_title('Training and Validation Loss')
axes[0].grid(True, alpha=0.3)

axes[1].plot(train_accuracies, label='Train', marker='o', markersize=3)
axes[1].plot(val_accuracies, label='Validation', marker='s', markersize=3)
axes[1].set_xlabel('Epoch')
axes[1].set_ylabel('Accuracy')
axes[1].legend()
axes[1].set_title('Training and Validation Accuracy')
axes[1].grid(True, alpha=0.3)

plt.tight_layout()
plt.show()

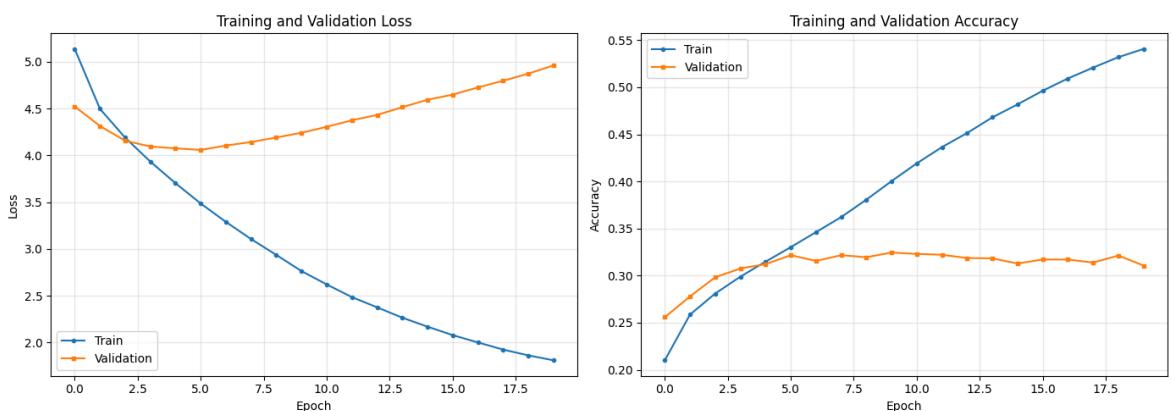
fig, axes = plt.subplots(1, 2, figsize=(10, 4))

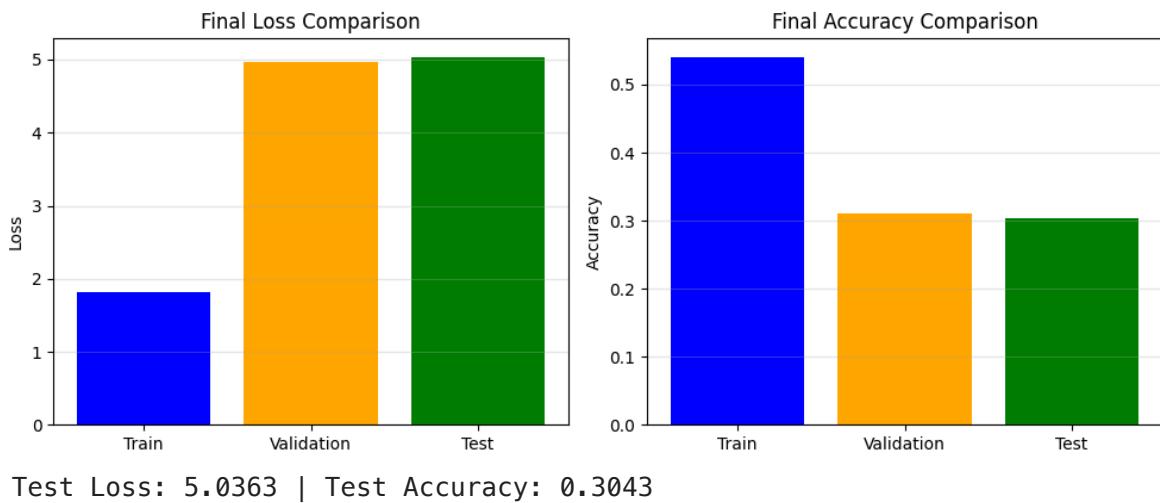
axes[0].bar(['Train', 'Validation', 'Test'],
           [train_losses[-1], val_losses[-1], test_loss],
           color=['blue', 'orange', 'green'])
axes[0].set_ylabel('Loss')
axes[0].set_title('Final Loss Comparison')
axes[0].grid(True, alpha=0.3, axis='y')

axes[1].bar(['Train', 'Validation', 'Test'],
           [train_accuracies[-1], val_accuracies[-1], test_acc],
           color=['blue', 'orange', 'green'])
axes[1].set_ylabel('Accuracy')
axes[1].set_title('Final Accuracy Comparison')
axes[1].grid(True, alpha=0.3, axis='y')

plt.tight_layout()
plt.show()

test_loss, test_acc = evaluate(model, test_loader, criterion, device)
print(f"Test Loss: {test_loss:.4f} | Test Accuracy: {test_acc:.4f}")
```





Experiment 1 (training with small dataset and experimenting with different hyperparameters) :

I first tried to train the model with small dataset (r0.0125). The hyperparameters that I've used to train were like this:

- hidden_dim = 256
- num_layers = 3
- num_heads = 4
- dropout = 0.1
- learning = 0.001
- batch = 64
- epochs = 20

The results I've got is like this:

Dataset	Epochs	Train Loss	Train Acc	Val Loss	Val Acc	Test Loss	Test Acc
r0.125	20	1.77	0.55	5.11	0.31	5.1104	0.3061

- Training loss and accuracy improved over epochs, but validation plateaued around approximately 31-32%.
- Validation loss started to increase after early epochs, this indicates that model overfits on a small dataset.

I also decided to experiment with different hyperparameters:

1. I've trained the model longer (30 epochs). The results are like this: | Dataset | Epochs | Train Loss | Train Acc | Val Loss | Val Acc | Test Loss | Test Acc | |---|---|---|---|---|---|---|---| | r0.125 | 30 | 1.3638 | 0.6264 | 5.5324 | 0.3039 | 5.6142 | 0.2982 |
- Longer training improved training accuracy (from 0.55 to 0.63), but validation accuracy did not improve and test accuracy slightly dropped.

- Validation loss kept increasing after the early epochs, which again confirms overfitting on the small dataset.
- Best validation accuracy was around approximately 32.5% (around epoch 12).
- I decided to not to train the model for 30 epochs, I will be training for 20 epochs for the other experiments.

2. I've trained the model with dropout rate = 0.2. The results are like this:

Dataset	Epochs	Dropout	Train Loss	Train Acc	Val Loss	Val Acc	Test Loss	Test Acc
--- ---: ---: ---: ---: ---: ---: ---: ---	r0.125	20	0.1	1.77	0.55			
5.11	0.31	5.1104	0.3061	r0.125	30	0.1	1.3638	0.6264
							5.5324	0.3039
							5.6142	0.2982
							r0.125	20
							0.2	2.3562
							0.4476	4.5363
							0.3250	4.6137
							0.3169	

- Increasing dropout reduced training accuracy when compared to training the model with 0.1 dropout rate (from 0.55 to 0.45) this indicates a stronger regularization.
- Validation accuracy was approximately 32 % and was slightly more stable, with a small improvement in test accuracy.
- Validation loss still trends upward after early epochs, so overfitting remains but less than th 0.1 dropout training.

Experiment 2 (Training with r0.5)

I trained the model with r0.5 dataset. I used the following hyperparameters:

- hidden_dim = 256
- num_layers = 3
- num_heads = 4
- dropout = 0.1
- learning = 0.001
- batch = 64
- epochs = 20
- For 20 epochs it took approximately 3.4 seconds per epoch so I decided to reduce the number of epochs to 10.

1. Trainin the mode with r0.5 dataset for 10 epochs:

Results are like this:

Dataset	Epochs	Dropout	Train Loss	Train Acc	Val Loss	Val Acc	Test Loss	Test Acc
r0.125	20	0.1	1.7700	0.5500	5.1100	0.3100	5.1104	0.3061
r0.125	30	0.1	1.3638	0.6264	5.5324	0.3039	5.6142	0.2982
r0.125	20	0.2	2.3562	0.4476	4.5363	0.3250	4.6137	0.3169
r0.5	10	0.1	2.7736	0.4179	3.5944	0.3872	3.6628	0.3800

- Validation accuracy steadily improved across epochs without a clear overfitting indicator in 10 epochs.
- r0.5 generalizes better than r0.125 because it achieves higher validation and test accuracy.

2. Training the model with dropout = 0.2:

I've decided to train the model with dropout rate = 0.2. Results are like the following:

Dataset	Epochs	Dropout	Train Loss	Train Acc	Val Loss	Val Acc	Test Loss	Test Acc
r0.125	20	0.1	1.7700	0.5500	5.1100	0.3100	5.1104	0.3061
r0.125	30	0.1	1.3638	0.6264	5.5324	0.3039	5.6142	0.2982
r0.125	20	0.2	2.3562	0.4476	4.5363	0.3250	4.6137	0.3169
r0.5	10	0.1	2.7736	0.4179	3.5944	0.3872	3.6628	0.3800
r0.5	10	0.2	3.1835	0.3723	3.6677	0.3744	3.7323	0.3678

- Higher dropout reduces training accuracy, suggesting stronger regularization.
- Validation and test accuracy are slightly lower than training the model with 0.1 dropout rate, so 0.1 is better on r0.5 in this setup.

Experiment 3 (Training with full dataset):

I've tried to train the model with full dataset but training the model for one epoch took too long so, I was not able to train the model on full dataset on google colab. One epoch was approximately 8.3 minutes long.

Summary of Experiments:

- On the small subset (r0.125), training accuracy improved but validation accuracy plateaued around 31%–33% and validation loss rose after early epochs, indicating overfitting.
- Extending training to 30 epochs increased training accuracy but made validation and test performance slightly worse, so longer training is not beneficial for r0.125.

- Increasing dropout to 0.2 reduced overfitting and slightly improved test accuracy on r0.125, but the gains were not big.
- On the larger subset (r0.5), validation/test accuracy improved substantially (approximately 38%), showing better generalization with more data.
- For r0.5, dropout 0.1 performed better than 0.2 (higher validation and test set accuracy), so 0.1 is the preferred setting in this setup.
- Full-dataset training was not feasible in colab due to runtime (approximately 8.3 minutes per epoch), so results are reported only for r0.125 and r0.5.

Overall, dataset size had the biggest impact on performance; dropout helped modestly, while longer training on small data led to overfitting. Also I want to add that model would achieve higher accuracies for validation and test sets if I could have trained the model on the full dataset.

Part 2.3: Attention Visualization

We will now set up some attention visualization, in order to understand what the model has learned. Given an input of length N , an attention map is an $N \times N$ matrix, where the (i, j) value represents the strength of the attention between the i -th input and the j -th input.

In order to get the attention maps, you should take the output of each head of the multi-head attention at each layer of the model. The provided function below will help visualize the attention maps given some input data. Run this and include some results on different heads/different layers of the model. Given an explanation of what the results mean.

```
In [73]: # Attention visualization helper
def plot_attention_maps(tokens_x, attn_maps, idx=0, tokens_y=None):
    if tokens_y is None:
        tokens_y = tokens_x
    attn_maps = [m[idx].detach().cpu().numpy() for m in attn_maps]
    num_heads = attn_maps[0].shape[0]
    num_layers = len(attn_maps)
    fig_size = 4 if num_heads == 1 else 3
    fig, ax = plt.subplots(num_layers, num_heads, figsize=(num_heads * fig_size, num_layers * 4))
    if num_layers == 1:
        ax = [ax]
    if num_heads == 1:
        ax = [[a] for a in ax]
    for row in range(num_layers):
        for column in range(num_heads):
            ax[row][column].imshow(attn_maps[row][column], origin="lower")
            ax[row][column].set_xticks(list(range(len(tokens_x))))
            ax[row][column].set_xticklabels(tokens_x, rotation=90)
            ax[row][column].set_yticks(list(range(len(tokens_y))))
            ax[row][column].set_yticklabels(tokens_y)
            ax[row][column].set_title(f"Layer {row+1}, Head {column+1}")
    fig.subplots_adjust(hspace=0.5)
    plt.show()
```

```

VIS_MODEL = model
VIS_DATASET = dev_dataset
VIS_SRC_VOCAB = tr_vocab
VIS_TGT_VOCAB = en_vocab
SAMPLE_IDX = 0
MAX_TOKENS = 12

src_tokens = VIS_DATASET.src_sentences[SAMPLE_IDX][:MAX_TOKENS]
tgt_tokens = VIS_DATASET.tgt_sentences[SAMPLE_IDX][:MAX_TOKENS]
src_ids = torch.tensor(VIS_SRC_VOCAB.encode(src_tokens)).unsqueeze(0).to(
tgt_ids = torch.tensor(VIS_TGT_VOCAB.encode(tgt_tokens)).unsqueeze(0).to(
tgt_input = tgt_ids[:, :-1]

src_mask, tgt_mask = create_masks(src_ids, tgt_input)

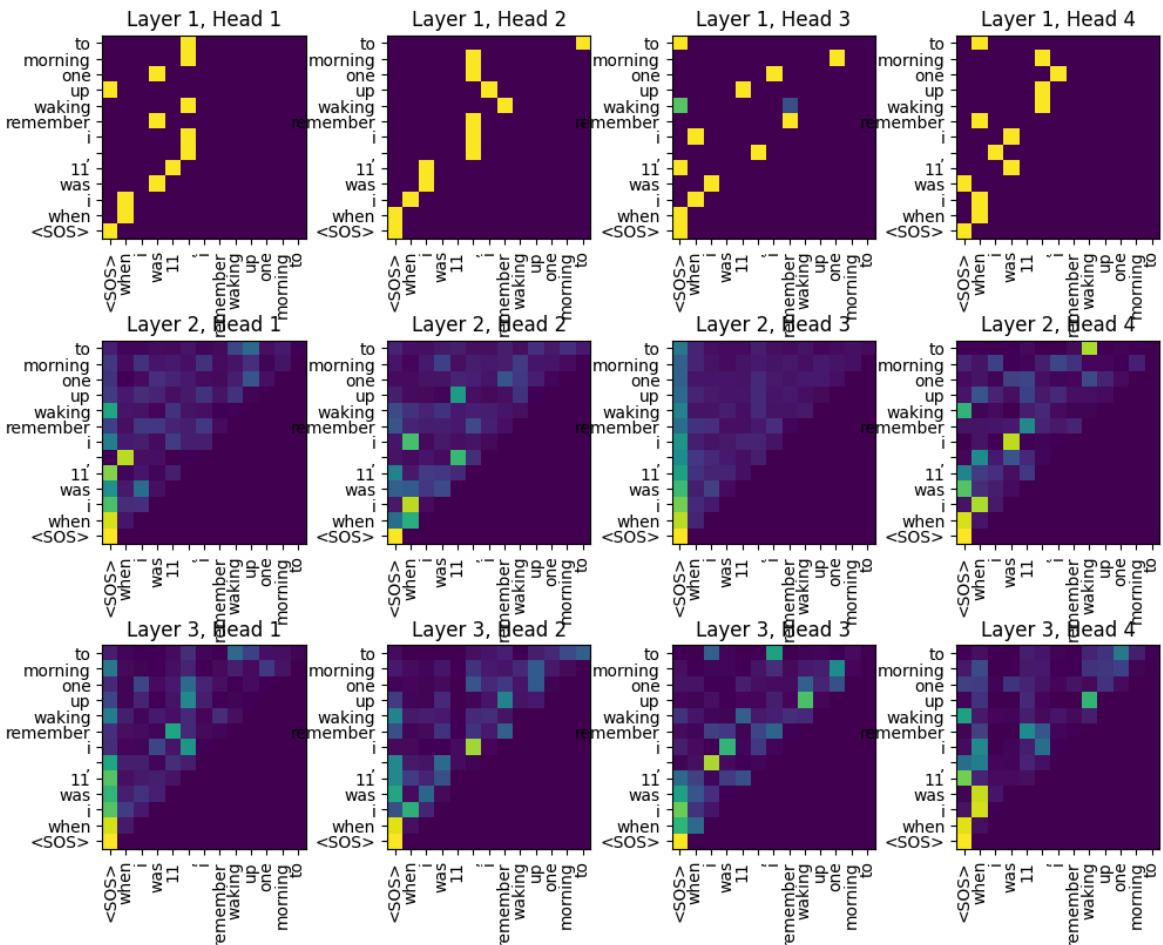
VIS_MODEL.eval()
with torch.no_grad():
    _, attn_maps = VIS_MODEL(src_ids, tgt_input, src_mask, tgt_mask, retu

tgt_labels = VIS_TGT_VOCAB.decode(tgt_input.squeeze(0).tolist())

print("Decoder self-attention")
plot_attention_maps(tgt_labels, attn_maps["decoder_self"], idx=0)

```

Decoder self-attention



The decoder self-attention maps show how each output token attends to previous tokens when generating the translation.

Layer 1: Each token focuses on itself or nearby previous tokens, the model captures short-range context.

Layer 2: Attention becomes more distributed. Model is combining information from a wider window of earlier tokens.

Layer 3: Attention is more diffuse with distinct hotspots, which suggests that different heads specialize in different tokens while others attend more globally. This reflects higher-level sequence dependencies in later layers.

What to submit

As in previous assignments, you have several options on how you can make the submission. The important part is, we need to see your code and your results in order to be able to grade you properly.

- You can write the answers and experiment results to your notebook and turn it into a pdf and submit it.
- You can put your experiment results and explanations to a separate file as a report, and submit this report and the notebook pdf together.
- As long as we can see your code and also the experiment results and answers, you will be fine.

In []:

Late Policy

You may use up to 7 grace days over the course of the semester for the assignments you will take. You can only use up to 3 grace days per assignment.

Academic Integrity

All work on assignments must be done individually unless stated otherwise. Turning in someone else's work, in whole or in part, as your own will be considered as a violation of academic integrity. Please note that the former condition also holds for the material found on the web as everything on the web has been written by someone else.

Acknowledgements

Adapted from University of Toronto, Neural Networks and Deep Learning course (CSC413/2516).