In [2]:
```
#upload tiny imagenet folder into jupyter project

#import zipfile as zf
#files = zf.ZipFile("tiny-imagenet-200.zip", 'r')
#files.extractall()
#files.close
```

In [3]:
```
%matplotlib inline
import matplotlib.pyplot as plt
import numpy as np
import torch
import torchvision
import torchvision.datasets as datasets
import torch.utils.data as data
from torchvision.utils import make_grid
import torchvision.transforms as transforms
import torch.nn as nn
import torch.nn.functional as F
import torch.optim as optim

import os

import vgg
import resnet
import googlenet
import alexnet
```

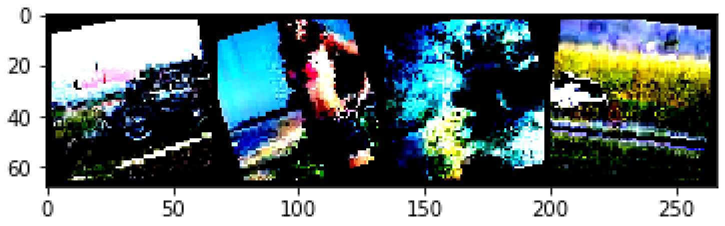
In [4]:
```
# If there are GPUs, choose the first one for computing. Otherwise use CPU.
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
print(device)
# If 'cuda:0' is printed, it means GPU is available.
```

```
cuda:0
```

In [5]:
```
data_transforms = {
    'train': transforms.Compose([
        transforms.RandomRotation(20),
        transforms.RandomHorizontalFlip(0.5),
        transforms.ToTensor(),
        transforms.Normalize([0.4802, 0.4481, 0.3975], [0.2302, 0.2265, 0.2262])
    ]),
    'val': transforms.Compose([
        transforms.ToTensor(),
        transforms.Normalize([0.4802, 0.4481, 0.3975], [0.2302, 0.2265, 0.2262])
    ]),
    'test': transforms.Compose([
        transforms.ToTensor(),
        transforms.Normalize([0.4802, 0.4481, 0.3975], [0.2302, 0.2265, 0.2262])
    ])
}

data_dir = 'tiny-imagenet-200/'

num_workers = {
    'train' : 100,
    'val'   : 0,
    'test'  : 0
}

image_datasets = {x: datasets.ImageFolder(os.path.join(data_dir, x),
                                          data_transforms[x])
                  for x in ['train', 'val','test']}
dataloaders = {x: torch.utils.data.DataLoader(image_datasets[x], batch_size=100,
                                             shuffle=True,num_workers=num_worker
              for x in ['train', 'val', 'test']}
dataset_sizes = {x: len(image_datasets[x]) for x in ['train', 'val', 'test']}
```

In [26]:
```python
def imshow(img):
    img = img.numpy().transpose((1, 2, 0))
    img = np.clip(img, 0, 1)
    plt.imshow(img)

images, labels = next(iter(dataloaders['train']))
print(labels)
grid = make_grid(images[:4], nrow=4)
imshow(grid)
```

```
tensor([164, 120, 196,  20, 146, 131,  38,  12, 139,   0,  72,   9, 188,  11,
        101,   6,  16, 157,  36, 189, 106,  66,  25, 194, 140, 187, 151,  68,
         62, 195, 159, 131,  85,  47,  11, 156,  99, 141,  48, 160,  65, 112,
         65,  68,  14,  35,   0, 101, 190, 167, 156, 168,   2, 149, 186,   7,
        107,  60,  37,  76, 169,   9, 157,  83,  62, 181,  97, 164,  48, 149,
         85,  96, 107, 144, 149,  24, 113,  65,  89,  16,  19,  31, 153,   8,
         59,  98, 135, 116, 110,  75,  44, 175,  37, 187, 180,  47, 148, 167,
        107, 176])
```



In [40]:
```
net = googlenet.GoogLeNet()   # Create the network instance.
net.to(device)
```

Out[40]:
```
GoogLeNet(
  (conv1): BasicConv2d(
    (conv): Conv2d(3, 64, kernel_size=(7, 7), stride=(2, 2), padding=(3, 3),
bias=False)
    (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_running
_stats=True)
  )
  (maxpool1): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_
mode=True)
  (conv2): BasicConv2d(
    (conv): Conv2d(64, 64, kernel_size=(1, 1), stride=(1, 1), bias=False)
    (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_running
_stats=True)
  )
  (conv3): BasicConv2d(
    (conv): Conv2d(64, 192, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
    (bn): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True, track_runnin
g_stats=True)
  )
  (maxpool2): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_
mode=True)
  (inception3a): Inception(
    (branch1): BasicConv2d(
      (conv): Conv2d(192, 64, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_runni
ng_stats=True)
    )
    (branch2): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(192, 96, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(96, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(96, 128, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch3): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(192, 16, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(16, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(16, 32, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
        (bn): BatchNorm2d(32, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
    (branch4): Sequential(
      (0): MaxPool2d(kernel_size=3, stride=1, padding=1, dilation=1, ceil_mod
e=True)
      (1): BasicConv2d(
        (conv): Conv2d(192, 32, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(32, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
  )
  (inception3b): Inception(
    (branch1): BasicConv2d(
      (conv): Conv2d(256, 128, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_runn
ing_stats=True)
    )
    (branch2): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(256, 128, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(128, 192, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch3): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(256, 32, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(32, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(32, 96, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
        (bn): BatchNorm2d(96, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
    (branch4): Sequential(
      (0): MaxPool2d(kernel_size=3, stride=1, padding=1, dilation=1, ceil_mod
e=True)
      (1): BasicConv2d(
        (conv): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
  )
  (maxpool3): MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_
mode=True)
  (inception4a): Inception(
    (branch1): BasicConv2d(
      (conv): Conv2d(480, 192, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True, track_runn
ing_stats=True)
    )
    (branch2): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(480, 96, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(96, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(96, 208, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(208, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch3): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(480, 16, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(16, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(16, 48, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
        (bn): BatchNorm2d(48, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
    (branch4): Sequential(
      (0): MaxPool2d(kernel_size=3, stride=1, padding=1, dilation=1, ceil_mod
e=True)
      (1): BasicConv2d(
        (conv): Conv2d(480, 64, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
  )
  (inception4b): Inception(
    (branch1): BasicConv2d(
      (conv): Conv2d(512, 160, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True, track_runn
ing_stats=True)
    )
    (branch2): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(512, 112, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(112, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(112, 224, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(224, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch3): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(512, 24, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(24, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(24, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
        (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
    (branch4): Sequential(
      (0): MaxPool2d(kernel_size=3, stride=1, padding=1, dilation=1, ceil_mod
e=True)
      (1): BasicConv2d(
        (conv): Conv2d(512, 64, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
  )
  (inception4c): Inception(
    (branch1): BasicConv2d(
      (conv): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_runn
ing_stats=True)
    )
    (branch2): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(128, 256, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch3): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(512, 24, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(24, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(24, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
        (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
    (branch4): Sequential(
      (0): MaxPool2d(kernel_size=3, stride=1, padding=1, dilation=1, ceil_mod
e=True)
      (1): BasicConv2d(
        (conv): Conv2d(512, 64, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
  )
  (inception4d): Inception(
    (branch1): BasicConv2d(
      (conv): Conv2d(512, 112, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(112, eps=0.001, momentum=0.1, affine=True, track_runn
ing_stats=True)
    )
    (branch2): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(512, 144, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(144, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(144, 288, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(288, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch3): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(512, 32, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(32, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(32, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
        (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
```

```
      )
    )
    (branch4): Sequential(
      (0): MaxPool2d(kernel_size=3, stride=1, padding=1, dilation=1, ceil_mod
e=True)
      (1): BasicConv2d(
        (conv): Conv2d(512, 64, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(64, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
    )
  )
  (inception4e): Inception(
    (branch1): BasicConv2d(
      (conv): Conv2d(528, 256, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True, track_runn
ing_stats=True)
    )
    (branch2): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(528, 160, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(160, 320, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(320, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch3): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(528, 32, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(32, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(32, 128, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch4): Sequential(
      (0): MaxPool2d(kernel_size=3, stride=1, padding=1, dilation=1, ceil_mod
e=True)
      (1): BasicConv2d(
        (conv): Conv2d(528, 128, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
  )
  (maxpool4): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_
mode=True)
  (inception5a): Inception(
    (branch1): BasicConv2d(
      (conv): Conv2d(832, 256, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(256, eps=0.001, momentum=0.1, affine=True, track_runn
ing_stats=True)
    )
    (branch2): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(832, 160, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(160, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(160, 320, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(320, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch3): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(832, 32, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(32, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(32, 128, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch4): Sequential(
      (0): MaxPool2d(kernel_size=3, stride=1, padding=1, dilation=1, ceil_mod
e=True)
      (1): BasicConv2d(
        (conv): Conv2d(832, 128, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
  )
  (inception5b): Inception(
    (branch1): BasicConv2d(
      (conv): Conv2d(832, 384, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(384, eps=0.001, momentum=0.1, affine=True, track_runn
ing_stats=True)
    )
    (branch2): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(832, 192, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(192, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(192, 384, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(384, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch3): Sequential(
      (0): BasicConv2d(
        (conv): Conv2d(832, 48, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(48, eps=0.001, momentum=0.1, affine=True, track_run
ning_stats=True)
      )
      (1): BasicConv2d(
        (conv): Conv2d(48, 128, kernel_size=(3, 3), stride=(1, 1), padding=
(1, 1), bias=False)
        (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
    (branch4): Sequential(
      (0): MaxPool2d(kernel_size=3, stride=1, padding=1, dilation=1, ceil_mod
e=True)
      (1): BasicConv2d(
        (conv): Conv2d(832, 128, kernel_size=(1, 1), stride=(1, 1), bias=Fals
e)
        (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_ru
nning_stats=True)
      )
    )
  )
  (aux1): InceptionAux(
    (conv): BasicConv2d(
      (conv): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_runn
ing_stats=True)
    )
    (fc1): Linear(in_features=2048, out_features=1024, bias=True)
    (fc2): Linear(in_features=1024, out_features=200, bias=True)
  )
  (aux2): InceptionAux(
    (conv): BasicConv2d(
      (conv): Conv2d(528, 128, kernel_size=(1, 1), stride=(1, 1), bias=False)
      (bn): BatchNorm2d(128, eps=0.001, momentum=0.1, affine=True, track_runn
ing_stats=True)
    )
    (fc1): Linear(in_features=2048, out_features=1024, bias=True)
    (fc2): Linear(in_features=1024, out_features=200, bias=True)
  )
  (avgpool): AdaptiveAvgPool2d(output_size=(1, 1))
  (dropout): Dropout(p=0.2, inplace=False)
  (fc): Linear(in_features=1024, out_features=200, bias=True)
)
```

In [36]:
```
# We use cross-entropy as loss function.
loss_func = nn.CrossEntropyLoss()
# We use stochastic gradient descent (SGD) as optimizer.
opt = optim.SGD(net.parameters(), lr=0.001, momentum=0.9)
```

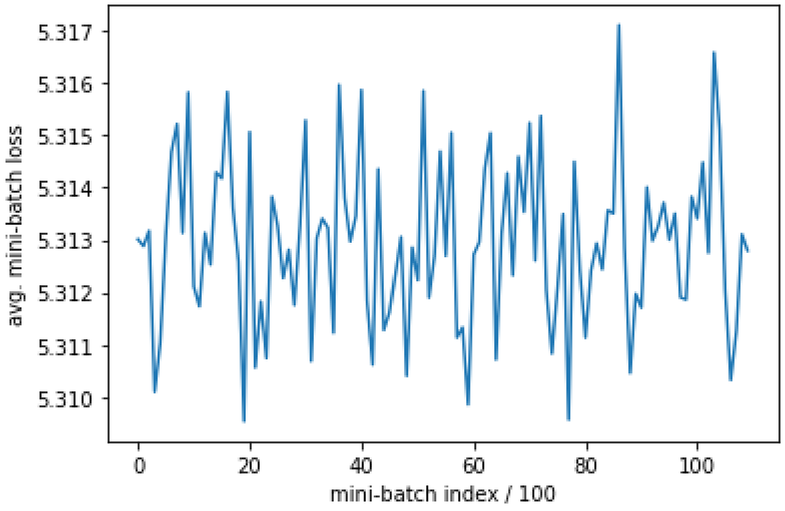
In [41]:
```
avg_losses = []   # Avg. losses.
epochs = 11       # Total epochs.
print_freq = 100  # Print frequency.

for epoch in range(epochs):  # Loop over the dataset multiple times.
    running_loss = 0.0       # Initialize running loss.
    for i, data in enumerate(dataloaders['train'], 0):

        net.train()

        # Get the inputs
        inputs, labels = data

        # Move the inputs to the specified device.
        inputs, labels = inputs.to(device), labels.to(device)

        # Zero the parameter gradients.
        opt.zero_grad()

        # Forward step.
        outputs = net(inputs)
        loss = loss_func(outputs, labels)

        # Backward step.
        loss.backward()

        # Optimization step (update the parameters).
        opt.step()

        # Print statistics.
        running_loss += loss.item()
        if i % print_freq == print_freq - 1: # Print every several mini-batches.
            avg_loss = running_loss / print_freq
            print('[epoch: {}, i: {:5d}] avg mini-batch loss: {:.3f}'.format(
                epoch, i, avg_loss))
            avg_losses.append(avg_loss)
            running_loss = 0.0

print('Finished Training.')
```

```
[epoch: 0, i:    99] avg mini-batch loss: 5.313
[epoch: 0, i:   199] avg mini-batch loss: 5.313
[epoch: 0, i:   299] avg mini-batch loss: 5.313
[epoch: 0, i:   399] avg mini-batch loss: 5.310
[epoch: 0, i:   499] avg mini-batch loss: 5.311
[epoch: 0, i:   599] avg mini-batch loss: 5.313
[epoch: 0, i:   699] avg mini-batch loss: 5.315
[epoch: 0, i:   799] avg mini-batch loss: 5.315
[epoch: 0, i:   899] avg mini-batch loss: 5.313
[epoch: 0, i:   999] avg mini-batch loss: 5.316
[epoch: 1, i:    99] avg mini-batch loss: 5.312
[epoch: 1, i:   199] avg mini-batch loss: 5.312
[epoch: 1, i:   299] avg mini-batch loss: 5.313
[epoch: 1, i:   399] avg mini-batch loss: 5.313
[epoch: 1, i:   499] avg mini-batch loss: 5.314
[epoch: 1, i:   599] avg mini-batch loss: 5.314
```

```
[epoch: 1, i:   699] avg mini-batch loss: 5.316
[epoch: 1, i:   799] avg mini-batch loss: 5.314
[epoch: 1, i:   899] avg mini-batch loss: 5.313
[epoch: 1, i:   999] avg mini-batch loss: 5.310
[epoch: 2, i:    99] avg mini-batch loss: 5.315
[epoch: 2, i:   199] avg mini-batch loss: 5.311
[epoch: 2, i:   299] avg mini-batch loss: 5.312
[epoch: 2, i:   399] avg mini-batch loss: 5.311
[epoch: 2, i:   499] avg mini-batch loss: 5.314
[epoch: 2, i:   599] avg mini-batch loss: 5.313
[epoch: 2, i:   699] avg mini-batch loss: 5.312
[epoch: 2, i:   799] avg mini-batch loss: 5.313
[epoch: 2, i:   899] avg mini-batch loss: 5.312
[epoch: 2, i:   999] avg mini-batch loss: 5.313
[epoch: 3, i:    99] avg mini-batch loss: 5.315
[epoch: 3, i:   199] avg mini-batch loss: 5.311
[epoch: 3, i:   299] avg mini-batch loss: 5.313
[epoch: 3, i:   399] avg mini-batch loss: 5.313
[epoch: 3, i:   499] avg mini-batch loss: 5.313
[epoch: 3, i:   599] avg mini-batch loss: 5.311
[epoch: 3, i:   699] avg mini-batch loss: 5.316
[epoch: 3, i:   799] avg mini-batch loss: 5.314
[epoch: 3, i:   899] avg mini-batch loss: 5.313
[epoch: 3, i:   999] avg mini-batch loss: 5.313
[epoch: 4, i:    99] avg mini-batch loss: 5.316
[epoch: 4, i:   199] avg mini-batch loss: 5.312
[epoch: 4, i:   299] avg mini-batch loss: 5.311
[epoch: 4, i:   399] avg mini-batch loss: 5.314
[epoch: 4, i:   499] avg mini-batch loss: 5.311
[epoch: 4, i:   599] avg mini-batch loss: 5.312
[epoch: 4, i:   699] avg mini-batch loss: 5.312
[epoch: 4, i:   799] avg mini-batch loss: 5.313
[epoch: 4, i:   899] avg mini-batch loss: 5.310
[epoch: 4, i:   999] avg mini-batch loss: 5.313
[epoch: 5, i:    99] avg mini-batch loss: 5.312
[epoch: 5, i:   199] avg mini-batch loss: 5.316
[epoch: 5, i:   299] avg mini-batch loss: 5.312
[epoch: 5, i:   399] avg mini-batch loss: 5.313
[epoch: 5, i:   499] avg mini-batch loss: 5.315
[epoch: 5, i:   599] avg mini-batch loss: 5.313
[epoch: 5, i:   699] avg mini-batch loss: 5.315
[epoch: 5, i:   799] avg mini-batch loss: 5.311
[epoch: 5, i:   899] avg mini-batch loss: 5.311
[epoch: 5, i:   999] avg mini-batch loss: 5.310
[epoch: 6, i:    99] avg mini-batch loss: 5.313
[epoch: 6, i:   199] avg mini-batch loss: 5.313
[epoch: 6, i:   299] avg mini-batch loss: 5.314
[epoch: 6, i:   399] avg mini-batch loss: 5.315
[epoch: 6, i:   499] avg mini-batch loss: 5.311
[epoch: 6, i:   599] avg mini-batch loss: 5.313
[epoch: 6, i:   699] avg mini-batch loss: 5.314
[epoch: 6, i:   799] avg mini-batch loss: 5.312
[epoch: 6, i:   899] avg mini-batch loss: 5.315
[epoch: 6, i:   999] avg mini-batch loss: 5.314
[epoch: 7, i:    99] avg mini-batch loss: 5.315
[epoch: 7, i:   199] avg mini-batch loss: 5.313
[epoch: 7, i:   299] avg mini-batch loss: 5.315
```

```
[epoch: 7, i:   399] avg mini-batch loss: 5.312
[epoch: 7, i:   499] avg mini-batch loss: 5.311
[epoch: 7, i:   599] avg mini-batch loss: 5.312
[epoch: 7, i:   699] avg mini-batch loss: 5.313
[epoch: 7, i:   799] avg mini-batch loss: 5.310
[epoch: 7, i:   899] avg mini-batch loss: 5.314
[epoch: 7, i:   999] avg mini-batch loss: 5.312
[epoch: 8, i:    99] avg mini-batch loss: 5.311
[epoch: 8, i:   199] avg mini-batch loss: 5.312
[epoch: 8, i:   299] avg mini-batch loss: 5.313
[epoch: 8, i:   399] avg mini-batch loss: 5.312
[epoch: 8, i:   499] avg mini-batch loss: 5.314
[epoch: 8, i:   599] avg mini-batch loss: 5.314
[epoch: 8, i:   699] avg mini-batch loss: 5.317
[epoch: 8, i:   799] avg mini-batch loss: 5.313
[epoch: 8, i:   899] avg mini-batch loss: 5.310
[epoch: 8, i:   999] avg mini-batch loss: 5.312
[epoch: 9, i:    99] avg mini-batch loss: 5.312
[epoch: 9, i:   199] avg mini-batch loss: 5.314
[epoch: 9, i:   299] avg mini-batch loss: 5.313
[epoch: 9, i:   399] avg mini-batch loss: 5.313
[epoch: 9, i:   499] avg mini-batch loss: 5.314
[epoch: 9, i:   599] avg mini-batch loss: 5.313
[epoch: 9, i:   699] avg mini-batch loss: 5.314
[epoch: 9, i:   799] avg mini-batch loss: 5.312
[epoch: 9, i:   899] avg mini-batch loss: 5.312
[epoch: 9, i:   999] avg mini-batch loss: 5.314
[epoch: 10, i:    99] avg mini-batch loss: 5.313
[epoch: 10, i:   199] avg mini-batch loss: 5.314
[epoch: 10, i:   299] avg mini-batch loss: 5.313
[epoch: 10, i:   399] avg mini-batch loss: 5.317
[epoch: 10, i:   499] avg mini-batch loss: 5.315
[epoch: 10, i:   599] avg mini-batch loss: 5.312
[epoch: 10, i:   699] avg mini-batch loss: 5.310
[epoch: 10, i:   799] avg mini-batch loss: 5.311
[epoch: 10, i:   899] avg mini-batch loss: 5.313
[epoch: 10, i:   999] avg mini-batch loss: 5.313
Finished Training.
```

In [42]:
```
plt.plot(avg_losses)
plt.xlabel('mini-batch index / {}'.format(print_freq))
plt.ylabel('avg. mini-batch loss')
plt.show()
```



In [43]:
```
# Get test accuracy.
correct = 0
total = 0
with torch.no_grad():
    for i, data in enumerate(dataloaders['test']):
        net.eval()
        images, labels = data
        images, labels = images.to(device), labels.to(device)
        outputs = net(images)
        _, predicted = torch.max(outputs.data, 1)
        total += labels.size(0)
        correct += (predicted == labels).sum().item()

print('Accuracy of the network on the 10000 test images: %d %%' % (
    100 * correct / total))
```

```
Accuracy of the network on the 10000 test images: 17 %
```

In [ ]: