

Machine Learning Exercise Sheet 10

Deep Learning II

Anastasia Stamatouli 03710902

05/01/2020

Problem (L)

$$(L) \quad f_w(x_i) = W_{L+1}^T \left(\text{Relu} \left(W_L^T \left(\dots \text{Relu} \left(W_2^T \left(\text{Relu} \left(W_1^T (W^T x) \right) \right) \right) \right) \right) \right)$$

a) In the case of W_{NN}^* non-negative, the Relu function in (L) could be omitted since both W_{NN} and x_i are non-negative.

$$\text{Therefore } (L) \Rightarrow f_w(x_i) = (W_{L+1}^T * W_L^T * \dots * W_2^T * W_1^T * W^T) x$$

$$= W_F^T x$$

which means that $f_w(x_i)$ is a linear combination
Thus, the two loss functions $\xrightarrow{\text{turn to}}$ have the same form. and their minimum value should be the same.

$$L_{NN}(W_{NN}^*) = L_{LS}(W_{LS}^*)$$

b) In the case of W_{LS}^* non-negative, Relu at the hidden layers of the NN make sure to retain non-negativity however the first layer and the last layer do not have Relu which means that we might end up with some negative values. So the loss of neural network might be smaller or equal.

$$L_{NN}(W_{NN}^*) \leq L_{LS}(W_{LS}^*)$$

exercise_10_notebookFinal

January 5, 2020

```
[0]: import copy
import numpy as np
import torch
import torch.nn as nn
import torch.nn.functional as F
import torchvision
import torchvision.transforms as transforms
```

1 PyTorch

In this notebook you will gain some hands-on experience with [PyTorch](#), one of the major frameworks for deep learning. To install PyTorch run `conda install pytorch torchvision cudatoolkit=10.1 -c pytorch`, with cudatoolkit set to whichever CUDA version you have installed. You can check this by running `nvcc --version`. If you do not have an Nvidia GPU you can run `conda install pytorch torchvision cpuonly -c pytorch` instead. However, in this case we recommend using [Google Colab](#).

You will start by re-implementing some common features of deep neural networks (dropout and batch normalization) and then implement a very popular modern architecture for image classification (ResNet) and improve its training loop.

2 1. Dropout

Dropout is a form of regularization for neural networks. It works by randomly setting activations (values) to 0, each one with equal probability p . The values are then scaled by a factor $\frac{1}{1-p}$ to conserve their mean.

Dropout effectively trains a pseudo-ensemble of models with stochastic gradient descent. During evaluation we want to use the full ensemble and therefore have to turn off dropout. Use `self.training` to check if the model is in training or evaluation mode.

Do not use any dropout implementation from PyTorch for this!

```
[0]: class Dropout(nn.Module):
    """
    Dropout, as discussed in the lecture and described here:
    https://pytorch.org/docs/stable/nn.html#torch.nn.Dropout
```

```

Args:
    p: float, dropout probability
    """
def __init__(self, p):
    super().__init__()
    self.p = p

def forward(self, input):
    """
    The module's forward pass.
    This has to be implemented for every PyTorch module.
    PyTorch then automatically generates the backward pass
    by dynamically generating the computational graph during
    execution.

    Args:
        input: PyTorch tensor, arbitrary shape

    Returns:
        PyTorch tensor, same shape as input
    """

    # TODO: Set values randomly to 0.
    U1 = (np.random.rand(*input.shape) < self.p) / (1-self.p)

    U1 = torch.from_numpy(U1)
    output = input* U1

    return output

```

```

[0]: # Test dropout
test = torch.ones(10_000)
dropout = Dropout(0.5)
test_dropped = dropout(test)

# These assertions can in principle fail due to bad luck, but
# if implemented correctly they should almost always succeed.
assert np.isclose(test_dropped.sum().item(), 10_000, atol=400)
assert np.isclose((test_dropped > 0).sum().item(), 5_000, atol=200)

```

3 2. Batch normalization

Batch normalization is a trick use to smoothen the loss landscape and improve training. It is defined as the function

$$y = \frac{x - \mu_x}{\sigma_x + \epsilon} \cdot \gamma + \beta$$

, where γ and β are learnable parameters and ϵ is a some small number to avoid dividing by zero. The Statistics μ_x and σ_x are taken separately for each feature. In a CNN this means averaging over the batch and all pixels.

Do not use any batch normalization implementation from PyTorch for this!

```
[0]: class BatchNorm(nn.Module):
    """
    Batch normalization, as discussed in the lecture and similar to
    https://pytorch.org/docs/stable/nn.html#torch.nn.BatchNorm1d

    Only uses batch statistics (no running mean for evaluation).
    Batch statistics are calculated for a single dimension.
    Gamma is initialized as 1, beta as 0.

    Args:
        num_features: Number of features to calculate batch statistics for.
    """
    def __init__(self, num_features):
        super().__init__()

        # TODO: Initialize the required parameters
        self.gamma = nn.Parameter(torch.ones(1, num_features, 1))
        self.beta = nn.Parameter(torch.zeros(1, num_features, 1))
        self.mu = torch.zeros(1, num_features, 1)
        self.var = torch.zeros(1, num_features, 1)

    def forward(self, input):
        """
        Batch normalization over the dimension C of (N, C, L).

        Args:
            input: PyTorch tensor, shape [N, C, L]

        Return:
            PyTorch tensor, same shape as input
        """
        eps = 1e-5

        # TODO: Implement the required transformation
        mean = input.mean(axis=(0, 2), keepdims=True)
        var = ((input - mean) ** 2).mean(axis=(0, 2), keepdims=True)
        #print(mean.size())

        X_hat = (input - mean) / np.sqrt(var + eps)
        #print(X_hat.size())
```

```
Y = self.gamma * X_hat + self.beta
return Y
```

```
[0]: # Tests the batch normalization implementation
torch.random.manual_seed(42)
test = torch.randn(8, 2, 4)

b1 = BatchNorm(2)
test_b1 = b1(test)

b2 = nn.BatchNorm1d(2, affine=False, track_running_stats=False)
test_b2 = b2(test)

assert torch.allclose(test_b1, test_b2, rtol=0.02)
```

4 3. ResNet

ResNet is the models that first introduced residual connections (a form of skip connections). It is a rather simple, but successful and very popular architecture. In this part of the exercise we will re-implement it step by step.

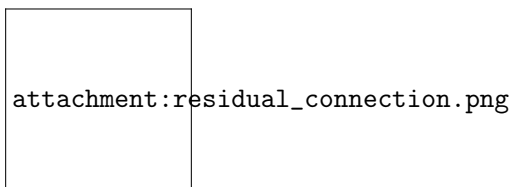
Note that there is also an [improved version of ResNet](#) with optimized residual blocks. Here we will implement the [original version](#) for CIFAR-10. Your dropout and batchnorm implementations won't help you here. Just use PyTorch's own layers.

This is just a convenience function to make e.g. `nn.Sequential` more flexible. It is e.g. useful in combination with `x.squeeze()`.

```
[0]: class Lambda(nn.Module):
    def __init__(self, func):
        super().__init__()
        self.func = func

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

We begin by implementing the residual blocks. The block is illustrated by this sketch:



Note that we use 'SAME' padding, no bias, and batch normalization after each convolution. You

do not need `nn.Sequential` here. The skip connection is already implemented as `self.skip`. It can handle different strides and increases in the number of channels.

```
[0]: class ResidualBlock(nn.Module):
    """
    The residual block used by ResNet.

    Args:
        in_channels: The number of channels (feature maps) of the incoming
        embedding
        out_channels: The number of channels after the first convolution
        stride: Stride size of the first convolution, used for downsampling
    """

    def __init__(self, in_channels, out_channels, stride=1):
        super().__init__()
        if stride > 1 or in_channels != out_channels:
            # Add strides in the skip connection and zeros for the new channels.
            self.skip = Lambda(lambda x: F.pad(x[:, :, ::stride, ::stride],
        (0, 0, 0, 0, 0, 0, out_channels -
        in_channels),
        mode="constant", value=0))
        else:
            self.skip = nn.Sequential()

        # TODO: Initialize the required layers
        self.conv1 = nn.Conv2d(in_channels, out_channels, kernel_size=3,
        stride=stride, padding = 1, bias = False)
        self.bn1 = nn.BatchNorm2d(out_channels)
        self.relu = nn.ReLU()
        self.conv2 = nn.Conv2d(out_channels, out_channels, kernel_size=3,
        stride=1, padding = 1, bias = False)
        self.bn2 = nn.BatchNorm2d(out_channels)
        self.relu2 = nn.ReLU()

    def forward(self, input):
        # TODO: Execute the required layers and functions
        outskip=self.skip(input)
        out = self.conv1(input)
        out = self.bn1 (out)
        out = self.relu (out)
        out = self.conv2 (out)
        out = self.bn2 (out)
        out = out + outskip
        out = self.relu2(out)
        return out
```

Next we implement a stack of residual blocks for convenience. The first layer in the block is the

one changing the number of channels and downsampling. You can use `nn.ModuleList` to use a list of child modules.

```
[0]: class ResidualStack(nn.Module):
    """
    A stack of residual blocks.

    Args:
        in_channels: The number of channels (feature maps) of the incoming
        ↪ embedding
        out_channels: The number of channels after the first layer
        stride: Stride size of the first layer, used for downsampling
        num_blocks: Number of residual blocks
    """

    def __init__(self, in_channels, out_channels, stride, num_blocks):
        super().__init__()

        # TODO: Initialize the required layers (blocks)

        self.stack = nn.ModuleList()
        self.stack.append(ResidualBlock(in_channels, out_channels,
        ↪ stride=stride))
        for i in range(num_blocks-1):
            self.stack.append(ResidualBlock(out_channels, out_channels))

    def forward(self, input):
        # TODO: Execute the layers (blocks)
        x=input
        for i,l in enumerate(self.stack):
            x = self.stack[i](x)
        return x
```

Now we are finally ready to implement the full model! To do this, use the `nn.Sequential` API and carefully read the following paragraph from the paper (Fig. 3 is not important):

attachment:resnet_cifar10_description.png

Note that a convolution layer is always convolution + batch norm + activation (ReLU), that each `ResidualBlock` contains 2 layers, and that you might have to squeeze the embedding before the

dense (fully-connected) layer.

```
[0]: n = 5
      num_classes = 10

      # TODO: Implement ResNet via nn.Sequential

      resnet = nn.Sequential(
          nn.Conv2d(3, 16, kernel_size=3, stride=1, padding = 1, bias = False),
          nn.BatchNorm2d(16),
          nn.ReLU(),
          ResidualStack(16, 16, stride=1, num_blocks = n),
          ResidualStack(16, 32, stride=2, num_blocks = n),
          ResidualStack(32, 64, stride=2, num_blocks = n),
          nn.AvgPool2d(8),
          nn.Flatten(),
          nn.Linear(64, 10)
      )
```

Next we need to initialize the weights of our model.

```
[0]: def initialize_weight(module):
      if isinstance(module, (nn.Linear, nn.Conv2d)):
          nn.init.kaiming_normal_(module.weight, nonlinearity='relu')
      elif isinstance(module, nn.BatchNorm2d):
          nn.init.constant_(module.weight, 1)
          nn.init.constant_(module.bias, 0)

      resnet.apply(initialize_weight);
```

5 4. Training

So now we have a shiny new model, but that doesn't really help when we can't train it. So that's what we do next.

First we need to load the data. Note that we split the official training data into train and validation sets, because you must not look at the test set until you are completely done developing your model and report the final results. Some people don't do this properly, but you should not copy other people's bad habits.

```
[0]: class CIFAR10Subset(torchvision.datasets.CIFAR10):
      """
```

```
Get a subset of the CIFAR10 dataset, according to the passed indices.
"""
```

```
def __init__(self, *args, idx=None, **kwargs):
    super().__init__(*args, **kwargs)

    if idx is None:
        return

    self.data = self.data[idx]
    targets_np = np.array(self.targets)
    self.targets = targets_np[idx].tolist()
```

We next define transformations that change the images into PyTorch tensors, standardize the values according to the precomputed mean and standard deviation, and provide data augmentation for the training set.

```
[0]: normalize = transforms.Normalize(mean=[0.485, 0.456, 0.406],
                                     std=[0.229, 0.224, 0.225])

transform_train = transforms.Compose([
    transforms.RandomHorizontalFlip(),
    transforms.RandomCrop(32, 4),
    transforms.ToTensor(),
    normalize,
])

transform_eval = transforms.Compose([
    transforms.ToTensor(),
    normalize
])
```

```
[13]: ntrain = 45_000
train_set = CIFAR10Subset(root='./data', train=True, idx=range(ntrain),
                          download=True, transform=transform_train)
val_set = CIFAR10Subset(root='./data', train=True, idx=range(ntrain, 50_000),
                        download=True, transform=transform_eval)
test_set = torchvision.datasets.CIFAR10(root='./data', train=False,
                                         download=True, transform=transform_eval)
```

```
Files already downloaded and verified
Files already downloaded and verified
Files already downloaded and verified
```

```
[0]: dataloaders = {}
dataloaders['train'] = torch.utils.data.DataLoader(train_set, batch_size=128,
                                                  shuffle=True, num_workers=2,
                                                  pin_memory=True)
dataloaders['val'] = torch.utils.data.DataLoader(val_set, batch_size=128,
                                                  shuffle=False, num_workers=2,
```

```

                                pin_memory=True)
dataloaders['test'] = torch.utils.data.DataLoader(test_set, batch_size=128,
                                                shuffle=False, num_workers=2,
                                                pin_memory=True)

```

Next we push the model to our GPU (if there is one).

```

[0]: device = torch.device('cuda') if torch.cuda.is_available() else torch.
    ↪device('cpu')
    resnet.to(device);

```

Next we define a helper method that does one epoch of training or evaluation. We have only defined training here, so you need to implement the necessary changes for evaluation!

```

[0]: def run_epoch(model, optimizer, dataloader, train):
    """
    Run one epoch of training or evaluation.

    Args:
        model: The model used for prediction
        optimizer: Optimization algorithm for the model
        dataloader: Dataloader providing the data to run our model on
        train: Whether this epoch is used for training or evaluation

    Returns:
        Loss and accuracy in this epoch.
    """
    # TODO: Change the necessary parts to work correctly during evaluation
    ↪(train=False)
    device = next(model.parameters()).device

    # Set model to training mode (for e.g. batch normalization, dropout)

    if train==True:
        model.train()
    else:
        model.eval()

    epoch_loss = 0.0

    epoch_acc = 0.0

    # Iterate over data
    for xb, yb in dataloader:
        xb, yb = xb.to(device), yb.to(device)

```

```

    # zero the parameter gradients
    if train==True:
        optimizer.zero_grad()

    # forward
    with torch.set_grad_enabled(train):
        pred = model(xb)
        loss = F.cross_entropy(pred, yb)
        top1 = torch.argmax(pred, dim=1)
        ncorrect = torch.sum(top1 == yb)

        if train==True:
            loss.backward()
            optimizer.step()

    # statistics
    epoch_loss += loss.item()
    epoch_acc += ncorrect.item()

epoch_loss /= len(dataloader.dataset)
epoch_acc /= len(dataloader.dataset)
return epoch_loss, epoch_acc

```

Next we implement a method for fitting (training) our model. For many models early stopping can save a lot of training time. Your task is to add early stopping to the loop (based on validation accuracy). Early stopping usually means exiting the training loop if the validation accuracy hasn't improved for patience number of steps. Don't forget to save the best model parameters according to validation accuracy. You will need `copy.deepcopy` and the `state_dict` for this.

```

[0]: def fit(model, optimizer, lr_scheduler, dataloaders, max_epochs, patience):
    """
    Fit the given model on the dataset.

    Args:
        model: The model used for prediction
        optimizer: Optimization algorithm for the model
        lr_scheduler: Learning rate scheduler that improves training
            in late epochs with learning rate decay
        dataloaders: Dataloaders for training and validation
        max_epochs: Maximum number of epochs for training
        patience: Number of epochs to wait with early stopping the
            training if validation loss has decreased

    Returns:
        Loss and accuracy in this epoch.
    """

```

```

best_acc = 0
curr_patience = 0

for epoch in range(max_epochs):
    train_loss, train_acc = run_epoch(model, optimizer,
    ↪dataloaders['train'], train=True)
    lr_scheduler.step()
    print(f"Epoch {epoch + 1: >3}/{max_epochs}, train loss: {train_loss:.
    ↪2e}, accuracy: {train_acc * 100:.2f}%")

    val_loss, val_acc = run_epoch(model, None, dataloaders['val'],
    ↪train=False)
    print(f"Epoch {epoch + 1: >3}/{max_epochs}, val loss: {val_loss:.2e},
    ↪accuracy: {val_acc * 100:.2f}%")

    # TODO: Add early stopping and save the best weights (in
    ↪best_model_weights)

    if (val_acc > best_acc):
        curr_patience = 0
        best_acc = val_acc
        best_model_weights = copy.deepcopy(model.state_dict())
    else:
        curr_patience += 1

    if curr_patience > patience:

        print("early_stopping")
        break

model.load_state_dict(best_model_weights)

```

In most cases you should just use the Adam optimizer for training, because it works well out of the box. However, a well-tuned SGD (with momentum) will in most cases outperform Adam. And since the original paper gives us a well-tuned SGD we will just use that.

```

[18]: optimizer = torch.optim.SGD(resnet.parameters(), lr=0.1, momentum=0.9,
    ↪weight_decay=1e-4)
lr_scheduler = torch.optim.lr_scheduler.MultiStepLR(optimizer, milestones=[100,
    ↪150], gamma=0.1)

# Fit model
fit(resnet, optimizer, lr_scheduler, dataloaders, max_epochs=200, patience=50)

```

Epoch 1/200, train loss: 1.64e-02, accuracy: 23.81%
 Epoch 1/200, val loss: 1.31e-02, accuracy: 38.98%
 Epoch 2/200, train loss: 1.15e-02, accuracy: 45.72%
 Epoch 2/200, val loss: 1.18e-02, accuracy: 50.06%
 Epoch 3/200, train loss: 9.07e-03, accuracy: 58.27%
 Epoch 3/200, val loss: 8.50e-03, accuracy: 60.98%
 Epoch 4/200, train loss: 7.60e-03, accuracy: 65.17%
 Epoch 4/200, val loss: 7.13e-03, accuracy: 69.42%
 Epoch 5/200, train loss: 6.58e-03, accuracy: 70.28%
 Epoch 5/200, val loss: 7.91e-03, accuracy: 67.98%
 Epoch 6/200, train loss: 5.77e-03, accuracy: 74.08%
 Epoch 6/200, val loss: 6.54e-03, accuracy: 72.84%
 Epoch 7/200, train loss: 5.17e-03, accuracy: 76.96%
 Epoch 7/200, val loss: 6.04e-03, accuracy: 75.96%
 Epoch 8/200, train loss: 4.71e-03, accuracy: 79.26%
 Epoch 8/200, val loss: 7.04e-03, accuracy: 73.92%
 Epoch 9/200, train loss: 4.41e-03, accuracy: 80.38%
 Epoch 9/200, val loss: 5.77e-03, accuracy: 76.54%
 Epoch 10/200, train loss: 4.20e-03, accuracy: 81.43%
 Epoch 10/200, val loss: 4.83e-03, accuracy: 80.40%
 Epoch 11/200, train loss: 3.90e-03, accuracy: 82.81%
 Epoch 11/200, val loss: 7.78e-03, accuracy: 71.28%
 Epoch 12/200, train loss: 3.76e-03, accuracy: 83.36%
 Epoch 12/200, val loss: 4.06e-03, accuracy: 82.54%
 Epoch 13/200, train loss: 3.57e-03, accuracy: 84.33%
 Epoch 13/200, val loss: 5.05e-03, accuracy: 80.32%
 Epoch 14/200, train loss: 3.43e-03, accuracy: 84.79%
 Epoch 14/200, val loss: 4.09e-03, accuracy: 83.90%
 Epoch 15/200, train loss: 3.33e-03, accuracy: 85.34%
 Epoch 15/200, val loss: 4.54e-03, accuracy: 82.10%
 Epoch 16/200, train loss: 3.20e-03, accuracy: 85.73%
 Epoch 16/200, val loss: 4.26e-03, accuracy: 83.64%
 Epoch 17/200, train loss: 3.08e-03, accuracy: 86.37%
 Epoch 17/200, val loss: 5.07e-03, accuracy: 81.68%
 Epoch 18/200, train loss: 3.04e-03, accuracy: 86.65%
 Epoch 18/200, val loss: 4.99e-03, accuracy: 79.64%
 Epoch 19/200, train loss: 2.95e-03, accuracy: 86.76%
 Epoch 19/200, val loss: 4.46e-03, accuracy: 83.12%
 Epoch 20/200, train loss: 2.82e-03, accuracy: 87.53%
 Epoch 20/200, val loss: 3.77e-03, accuracy: 85.86%
 Epoch 21/200, train loss: 2.75e-03, accuracy: 87.79%
 Epoch 21/200, val loss: 4.71e-03, accuracy: 82.12%
 Epoch 22/200, train loss: 2.72e-03, accuracy: 87.96%
 Epoch 22/200, val loss: 3.87e-03, accuracy: 84.48%
 Epoch 23/200, train loss: 2.61e-03, accuracy: 88.26%
 Epoch 23/200, val loss: 4.08e-03, accuracy: 84.86%
 Epoch 24/200, train loss: 2.57e-03, accuracy: 88.55%
 Epoch 24/200, val loss: 4.23e-03, accuracy: 84.26%

Epoch 25/200, train loss: 2.51e-03, accuracy: 88.76%
 Epoch 25/200, val loss: 3.93e-03, accuracy: 84.84%
 Epoch 26/200, train loss: 2.54e-03, accuracy: 88.67%
 Epoch 26/200, val loss: 4.54e-03, accuracy: 83.04%
 Epoch 27/200, train loss: 2.47e-03, accuracy: 89.09%
 Epoch 27/200, val loss: 4.87e-03, accuracy: 81.48%
 Epoch 28/200, train loss: 2.37e-03, accuracy: 89.32%
 Epoch 28/200, val loss: 4.42e-03, accuracy: 82.78%
 Epoch 29/200, train loss: 2.36e-03, accuracy: 89.55%
 Epoch 29/200, val loss: 3.83e-03, accuracy: 85.10%
 Epoch 30/200, train loss: 2.31e-03, accuracy: 89.62%
 Epoch 30/200, val loss: 4.31e-03, accuracy: 83.24%
 Epoch 31/200, train loss: 2.28e-03, accuracy: 89.99%
 Epoch 31/200, val loss: 3.82e-03, accuracy: 86.22%
 Epoch 32/200, train loss: 2.22e-03, accuracy: 90.01%
 Epoch 32/200, val loss: 4.45e-03, accuracy: 84.36%
 Epoch 33/200, train loss: 2.24e-03, accuracy: 90.05%
 Epoch 33/200, val loss: 5.47e-03, accuracy: 81.30%
 Epoch 34/200, train loss: 2.17e-03, accuracy: 90.24%
 Epoch 34/200, val loss: 3.67e-03, accuracy: 85.42%
 Epoch 35/200, train loss: 2.15e-03, accuracy: 90.45%
 Epoch 35/200, val loss: 3.82e-03, accuracy: 85.88%
 Epoch 36/200, train loss: 2.07e-03, accuracy: 90.71%
 Epoch 36/200, val loss: 3.72e-03, accuracy: 86.62%
 Epoch 37/200, train loss: 2.08e-03, accuracy: 90.65%
 Epoch 37/200, val loss: 3.72e-03, accuracy: 85.70%
 Epoch 38/200, train loss: 2.04e-03, accuracy: 90.86%
 Epoch 38/200, val loss: 3.84e-03, accuracy: 86.68%
 Epoch 39/200, train loss: 2.06e-03, accuracy: 90.75%
 Epoch 39/200, val loss: 3.68e-03, accuracy: 86.10%
 Epoch 40/200, train loss: 2.04e-03, accuracy: 90.93%
 Epoch 40/200, val loss: 3.64e-03, accuracy: 86.82%
 Epoch 41/200, train loss: 1.97e-03, accuracy: 91.24%
 Epoch 41/200, val loss: 4.36e-03, accuracy: 84.12%
 Epoch 42/200, train loss: 1.98e-03, accuracy: 91.08%
 Epoch 42/200, val loss: 3.50e-03, accuracy: 87.14%
 Epoch 43/200, train loss: 1.99e-03, accuracy: 91.00%
 Epoch 43/200, val loss: 4.21e-03, accuracy: 85.12%
 Epoch 44/200, train loss: 1.94e-03, accuracy: 91.42%
 Epoch 44/200, val loss: 4.33e-03, accuracy: 83.64%
 Epoch 45/200, train loss: 1.92e-03, accuracy: 91.34%
 Epoch 45/200, val loss: 4.23e-03, accuracy: 84.56%
 Epoch 46/200, train loss: 1.88e-03, accuracy: 91.64%
 Epoch 46/200, val loss: 5.53e-03, accuracy: 80.46%
 Epoch 47/200, train loss: 1.91e-03, accuracy: 91.48%
 Epoch 47/200, val loss: 4.50e-03, accuracy: 83.36%
 Epoch 48/200, train loss: 1.82e-03, accuracy: 91.80%
 Epoch 48/200, val loss: 4.53e-03, accuracy: 84.36%

Epoch 49/200, train loss: 1.85e-03, accuracy: 91.74%
Epoch 49/200, val loss: 3.98e-03, accuracy: 84.38%
Epoch 50/200, train loss: 1.77e-03, accuracy: 92.10%
Epoch 50/200, val loss: 3.54e-03, accuracy: 86.28%
Epoch 51/200, train loss: 1.82e-03, accuracy: 91.93%
Epoch 51/200, val loss: 4.31e-03, accuracy: 84.34%
Epoch 52/200, train loss: 1.85e-03, accuracy: 91.61%
Epoch 52/200, val loss: 4.67e-03, accuracy: 83.76%
Epoch 53/200, train loss: 1.80e-03, accuracy: 91.92%
Epoch 53/200, val loss: 3.98e-03, accuracy: 86.04%
Epoch 54/200, train loss: 1.76e-03, accuracy: 92.08%
Epoch 54/200, val loss: 3.79e-03, accuracy: 86.26%
Epoch 55/200, train loss: 1.74e-03, accuracy: 92.18%
Epoch 55/200, val loss: 4.63e-03, accuracy: 83.96%
Epoch 56/200, train loss: 1.75e-03, accuracy: 92.09%
Epoch 56/200, val loss: 3.99e-03, accuracy: 85.46%
Epoch 57/200, train loss: 1.77e-03, accuracy: 92.06%
Epoch 57/200, val loss: 3.76e-03, accuracy: 86.18%
Epoch 58/200, train loss: 1.68e-03, accuracy: 92.43%
Epoch 58/200, val loss: 4.17e-03, accuracy: 85.10%
Epoch 59/200, train loss: 1.68e-03, accuracy: 92.50%
Epoch 59/200, val loss: 3.70e-03, accuracy: 86.46%
Epoch 60/200, train loss: 1.71e-03, accuracy: 92.19%
Epoch 60/200, val loss: 3.74e-03, accuracy: 86.02%
Epoch 61/200, train loss: 1.69e-03, accuracy: 92.37%
Epoch 61/200, val loss: 3.74e-03, accuracy: 86.96%
Epoch 62/200, train loss: 1.69e-03, accuracy: 92.35%
Epoch 62/200, val loss: 3.50e-03, accuracy: 86.72%
Epoch 63/200, train loss: 1.64e-03, accuracy: 92.69%
Epoch 63/200, val loss: 3.53e-03, accuracy: 87.58%
Epoch 64/200, train loss: 1.61e-03, accuracy: 92.61%
Epoch 64/200, val loss: 3.69e-03, accuracy: 86.22%
Epoch 65/200, train loss: 1.63e-03, accuracy: 92.78%
Epoch 65/200, val loss: 4.64e-03, accuracy: 83.58%
Epoch 66/200, train loss: 1.62e-03, accuracy: 92.84%
Epoch 66/200, val loss: 3.62e-03, accuracy: 86.90%
Epoch 67/200, train loss: 1.62e-03, accuracy: 92.74%
Epoch 67/200, val loss: 3.68e-03, accuracy: 87.64%
Epoch 68/200, train loss: 1.64e-03, accuracy: 92.59%
Epoch 68/200, val loss: 3.50e-03, accuracy: 87.10%
Epoch 69/200, train loss: 1.60e-03, accuracy: 92.88%
Epoch 69/200, val loss: 3.25e-03, accuracy: 88.22%
Epoch 70/200, train loss: 1.57e-03, accuracy: 93.03%
Epoch 70/200, val loss: 4.13e-03, accuracy: 86.36%
Epoch 71/200, train loss: 1.57e-03, accuracy: 92.96%
Epoch 71/200, val loss: 3.82e-03, accuracy: 86.58%
Epoch 72/200, train loss: 1.56e-03, accuracy: 92.95%
Epoch 72/200, val loss: 3.48e-03, accuracy: 87.40%

Epoch 73/200, train loss: 1.55e-03, accuracy: 93.00%
 Epoch 73/200, val loss: 3.83e-03, accuracy: 86.12%
 Epoch 74/200, train loss: 1.57e-03, accuracy: 92.96%
 Epoch 74/200, val loss: 3.65e-03, accuracy: 87.34%
 Epoch 75/200, train loss: 1.59e-03, accuracy: 92.95%
 Epoch 75/200, val loss: 3.36e-03, accuracy: 87.02%
 Epoch 76/200, train loss: 1.59e-03, accuracy: 92.89%
 Epoch 76/200, val loss: 3.93e-03, accuracy: 86.02%
 Epoch 77/200, train loss: 1.57e-03, accuracy: 93.04%
 Epoch 77/200, val loss: 3.02e-03, accuracy: 88.10%
 Epoch 78/200, train loss: 1.54e-03, accuracy: 93.04%
 Epoch 78/200, val loss: 3.37e-03, accuracy: 88.10%
 Epoch 79/200, train loss: 1.53e-03, accuracy: 93.13%
 Epoch 79/200, val loss: 4.64e-03, accuracy: 83.98%
 Epoch 80/200, train loss: 1.54e-03, accuracy: 93.08%
 Epoch 80/200, val loss: 3.40e-03, accuracy: 87.68%
 Epoch 81/200, train loss: 1.55e-03, accuracy: 93.13%
 Epoch 81/200, val loss: 4.84e-03, accuracy: 84.26%
 Epoch 82/200, train loss: 1.53e-03, accuracy: 92.97%
 Epoch 82/200, val loss: 3.49e-03, accuracy: 87.72%
 Epoch 83/200, train loss: 1.48e-03, accuracy: 93.35%
 Epoch 83/200, val loss: 3.18e-03, accuracy: 87.90%
 Epoch 84/200, train loss: 1.50e-03, accuracy: 93.16%
 Epoch 84/200, val loss: 4.33e-03, accuracy: 84.80%
 Epoch 85/200, train loss: 1.50e-03, accuracy: 93.30%
 Epoch 85/200, val loss: 2.95e-03, accuracy: 88.94%
 Epoch 86/200, train loss: 1.48e-03, accuracy: 93.42%
 Epoch 86/200, val loss: 3.93e-03, accuracy: 85.64%
 Epoch 87/200, train loss: 1.47e-03, accuracy: 93.29%
 Epoch 87/200, val loss: 5.08e-03, accuracy: 82.56%
 Epoch 88/200, train loss: 1.49e-03, accuracy: 93.19%
 Epoch 88/200, val loss: 3.70e-03, accuracy: 86.72%
 Epoch 89/200, train loss: 1.47e-03, accuracy: 93.36%
 Epoch 89/200, val loss: 5.36e-03, accuracy: 82.80%
 Epoch 90/200, train loss: 1.51e-03, accuracy: 93.19%
 Epoch 90/200, val loss: 3.63e-03, accuracy: 86.60%
 Epoch 91/200, train loss: 1.45e-03, accuracy: 93.38%
 Epoch 91/200, val loss: 3.81e-03, accuracy: 86.54%
 Epoch 92/200, train loss: 1.47e-03, accuracy: 93.54%
 Epoch 92/200, val loss: 3.74e-03, accuracy: 87.32%
 Epoch 93/200, train loss: 1.48e-03, accuracy: 93.37%
 Epoch 93/200, val loss: 3.98e-03, accuracy: 86.22%
 Epoch 94/200, train loss: 1.44e-03, accuracy: 93.49%
 Epoch 94/200, val loss: 3.06e-03, accuracy: 88.40%
 Epoch 95/200, train loss: 1.46e-03, accuracy: 93.48%
 Epoch 95/200, val loss: 4.01e-03, accuracy: 85.36%
 Epoch 96/200, train loss: 1.43e-03, accuracy: 93.60%
 Epoch 96/200, val loss: 3.76e-03, accuracy: 86.08%

Epoch 97/200, train loss: 1.46e-03, accuracy: 93.50%
Epoch 97/200, val loss: 3.59e-03, accuracy: 87.52%
Epoch 98/200, train loss: 1.42e-03, accuracy: 93.56%
Epoch 98/200, val loss: 3.69e-03, accuracy: 86.70%
Epoch 99/200, train loss: 1.45e-03, accuracy: 93.38%
Epoch 99/200, val loss: 3.56e-03, accuracy: 87.32%
Epoch 100/200, train loss: 1.41e-03, accuracy: 93.64%
Epoch 100/200, val loss: 4.45e-03, accuracy: 85.50%
Epoch 101/200, train loss: 7.76e-04, accuracy: 96.64%
Epoch 101/200, val loss: 2.40e-03, accuracy: 91.58%
Epoch 102/200, train loss: 5.63e-04, accuracy: 97.62%
Epoch 102/200, val loss: 2.39e-03, accuracy: 92.02%
Epoch 103/200, train loss: 4.81e-04, accuracy: 98.02%
Epoch 103/200, val loss: 2.41e-03, accuracy: 91.72%
Epoch 104/200, train loss: 4.39e-04, accuracy: 98.21%
Epoch 104/200, val loss: 2.48e-03, accuracy: 92.02%
Epoch 105/200, train loss: 3.99e-04, accuracy: 98.38%
Epoch 105/200, val loss: 2.49e-03, accuracy: 92.14%
Epoch 106/200, train loss: 3.87e-04, accuracy: 98.39%
Epoch 106/200, val loss: 2.43e-03, accuracy: 92.20%
Epoch 107/200, train loss: 3.39e-04, accuracy: 98.60%
Epoch 107/200, val loss: 2.44e-03, accuracy: 92.24%
Epoch 108/200, train loss: 3.36e-04, accuracy: 98.64%
Epoch 108/200, val loss: 2.53e-03, accuracy: 92.26%
Epoch 109/200, train loss: 3.07e-04, accuracy: 98.82%
Epoch 109/200, val loss: 2.50e-03, accuracy: 92.10%
Epoch 110/200, train loss: 2.85e-04, accuracy: 98.87%
Epoch 110/200, val loss: 2.49e-03, accuracy: 92.24%
Epoch 111/200, train loss: 2.68e-04, accuracy: 98.91%
Epoch 111/200, val loss: 2.57e-03, accuracy: 92.26%
Epoch 112/200, train loss: 2.47e-04, accuracy: 99.08%
Epoch 112/200, val loss: 2.57e-03, accuracy: 92.28%
Epoch 113/200, train loss: 2.42e-04, accuracy: 99.02%
Epoch 113/200, val loss: 2.63e-03, accuracy: 92.36%
Epoch 114/200, train loss: 2.46e-04, accuracy: 99.02%
Epoch 114/200, val loss: 2.67e-03, accuracy: 92.30%
Epoch 115/200, train loss: 2.30e-04, accuracy: 99.09%
Epoch 115/200, val loss: 2.74e-03, accuracy: 92.44%
Epoch 116/200, train loss: 2.24e-04, accuracy: 99.16%
Epoch 116/200, val loss: 2.63e-03, accuracy: 92.22%
Epoch 117/200, train loss: 2.11e-04, accuracy: 99.22%
Epoch 117/200, val loss: 2.67e-03, accuracy: 92.30%
Epoch 118/200, train loss: 1.93e-04, accuracy: 99.29%
Epoch 118/200, val loss: 2.68e-03, accuracy: 92.26%
Epoch 119/200, train loss: 1.91e-04, accuracy: 99.26%
Epoch 119/200, val loss: 2.66e-03, accuracy: 92.22%
Epoch 120/200, train loss: 1.78e-04, accuracy: 99.38%
Epoch 120/200, val loss: 2.75e-03, accuracy: 92.12%

Epoch 121/200, train loss: 1.75e-04, accuracy: 99.34%
Epoch 121/200, val loss: 2.73e-03, accuracy: 92.36%
Epoch 122/200, train loss: 1.66e-04, accuracy: 99.40%
Epoch 122/200, val loss: 2.77e-03, accuracy: 92.16%
Epoch 123/200, train loss: 1.60e-04, accuracy: 99.41%
Epoch 123/200, val loss: 2.75e-03, accuracy: 92.36%
Epoch 124/200, train loss: 1.65e-04, accuracy: 99.40%
Epoch 124/200, val loss: 2.73e-03, accuracy: 92.38%
Epoch 125/200, train loss: 1.57e-04, accuracy: 99.44%
Epoch 125/200, val loss: 2.82e-03, accuracy: 92.28%
Epoch 126/200, train loss: 1.49e-04, accuracy: 99.45%
Epoch 126/200, val loss: 2.74e-03, accuracy: 92.36%
Epoch 127/200, train loss: 1.40e-04, accuracy: 99.46%
Epoch 127/200, val loss: 2.86e-03, accuracy: 92.24%
Epoch 128/200, train loss: 1.39e-04, accuracy: 99.48%
Epoch 128/200, val loss: 2.77e-03, accuracy: 92.22%
Epoch 129/200, train loss: 1.41e-04, accuracy: 99.49%
Epoch 129/200, val loss: 2.91e-03, accuracy: 92.32%
Epoch 130/200, train loss: 1.42e-04, accuracy: 99.50%
Epoch 130/200, val loss: 2.88e-03, accuracy: 92.14%
Epoch 131/200, train loss: 1.34e-04, accuracy: 99.50%
Epoch 131/200, val loss: 2.82e-03, accuracy: 92.24%
Epoch 132/200, train loss: 1.33e-04, accuracy: 99.51%
Epoch 132/200, val loss: 2.75e-03, accuracy: 92.46%
Epoch 133/200, train loss: 1.25e-04, accuracy: 99.53%
Epoch 133/200, val loss: 2.83e-03, accuracy: 92.18%
Epoch 134/200, train loss: 1.31e-04, accuracy: 99.51%
Epoch 134/200, val loss: 2.96e-03, accuracy: 92.02%
Epoch 135/200, train loss: 1.24e-04, accuracy: 99.58%
Epoch 135/200, val loss: 3.00e-03, accuracy: 92.24%
Epoch 136/200, train loss: 1.16e-04, accuracy: 99.60%
Epoch 136/200, val loss: 2.88e-03, accuracy: 92.20%
Epoch 137/200, train loss: 1.19e-04, accuracy: 99.60%
Epoch 137/200, val loss: 2.86e-03, accuracy: 92.46%
Epoch 138/200, train loss: 1.14e-04, accuracy: 99.58%
Epoch 138/200, val loss: 2.87e-03, accuracy: 92.40%
Epoch 139/200, train loss: 1.11e-04, accuracy: 99.64%
Epoch 139/200, val loss: 2.82e-03, accuracy: 92.36%
Epoch 140/200, train loss: 1.12e-04, accuracy: 99.61%
Epoch 140/200, val loss: 3.00e-03, accuracy: 92.10%
Epoch 141/200, train loss: 1.11e-04, accuracy: 99.58%
Epoch 141/200, val loss: 3.02e-03, accuracy: 92.08%
Epoch 142/200, train loss: 1.09e-04, accuracy: 99.56%
Epoch 142/200, val loss: 2.98e-03, accuracy: 92.24%
Epoch 143/200, train loss: 1.03e-04, accuracy: 99.66%
Epoch 143/200, val loss: 3.01e-03, accuracy: 92.20%
Epoch 144/200, train loss: 9.70e-05, accuracy: 99.66%
Epoch 144/200, val loss: 2.91e-03, accuracy: 92.30%

Epoch 145/200, train loss: 9.39e-05, accuracy: 99.71%
Epoch 145/200, val loss: 3.01e-03, accuracy: 92.32%
Epoch 146/200, train loss: 9.25e-05, accuracy: 99.70%
Epoch 146/200, val loss: 2.93e-03, accuracy: 92.22%
Epoch 147/200, train loss: 1.02e-04, accuracy: 99.64%
Epoch 147/200, val loss: 3.09e-03, accuracy: 91.92%
Epoch 148/200, train loss: 9.56e-05, accuracy: 99.69%
Epoch 148/200, val loss: 3.04e-03, accuracy: 92.28%
Epoch 149/200, train loss: 9.43e-05, accuracy: 99.69%
Epoch 149/200, val loss: 2.98e-03, accuracy: 92.08%
Epoch 150/200, train loss: 9.29e-05, accuracy: 99.71%
Epoch 150/200, val loss: 3.03e-03, accuracy: 92.26%
Epoch 151/200, train loss: 8.36e-05, accuracy: 99.73%
Epoch 151/200, val loss: 2.99e-03, accuracy: 92.32%
Epoch 152/200, train loss: 7.74e-05, accuracy: 99.76%
Epoch 152/200, val loss: 2.95e-03, accuracy: 92.34%
Epoch 153/200, train loss: 6.97e-05, accuracy: 99.76%
Epoch 153/200, val loss: 2.96e-03, accuracy: 92.38%
Epoch 154/200, train loss: 7.06e-05, accuracy: 99.80%
Epoch 154/200, val loss: 2.95e-03, accuracy: 92.50%
Epoch 155/200, train loss: 6.52e-05, accuracy: 99.82%
Epoch 155/200, val loss: 2.98e-03, accuracy: 92.42%
Epoch 156/200, train loss: 7.05e-05, accuracy: 99.80%
Epoch 156/200, val loss: 2.95e-03, accuracy: 92.32%
Epoch 157/200, train loss: 6.61e-05, accuracy: 99.79%
Epoch 157/200, val loss: 2.98e-03, accuracy: 92.40%
Epoch 158/200, train loss: 6.76e-05, accuracy: 99.82%
Epoch 158/200, val loss: 2.93e-03, accuracy: 92.24%
Epoch 159/200, train loss: 6.59e-05, accuracy: 99.84%
Epoch 159/200, val loss: 2.95e-03, accuracy: 92.26%
Epoch 160/200, train loss: 6.35e-05, accuracy: 99.81%
Epoch 160/200, val loss: 3.01e-03, accuracy: 92.44%
Epoch 161/200, train loss: 6.24e-05, accuracy: 99.85%
Epoch 161/200, val loss: 2.95e-03, accuracy: 92.24%
Epoch 162/200, train loss: 6.09e-05, accuracy: 99.82%
Epoch 162/200, val loss: 2.99e-03, accuracy: 92.32%
Epoch 163/200, train loss: 6.07e-05, accuracy: 99.83%
Epoch 163/200, val loss: 2.99e-03, accuracy: 92.44%
Epoch 164/200, train loss: 6.51e-05, accuracy: 99.80%
Epoch 164/200, val loss: 3.01e-03, accuracy: 92.16%
Epoch 165/200, train loss: 6.19e-05, accuracy: 99.83%
Epoch 165/200, val loss: 2.97e-03, accuracy: 92.26%
Epoch 166/200, train loss: 6.46e-05, accuracy: 99.82%
Epoch 166/200, val loss: 2.97e-03, accuracy: 92.40%
Epoch 167/200, train loss: 6.42e-05, accuracy: 99.82%
Epoch 167/200, val loss: 3.00e-03, accuracy: 92.48%
Epoch 168/200, train loss: 6.14e-05, accuracy: 99.84%
Epoch 168/200, val loss: 3.00e-03, accuracy: 92.28%

Epoch 169/200, train loss: 6.30e-05, accuracy: 99.82%
Epoch 169/200, val loss: 2.98e-03, accuracy: 92.48%
Epoch 170/200, train loss: 5.80e-05, accuracy: 99.84%
Epoch 170/200, val loss: 2.99e-03, accuracy: 92.36%
Epoch 171/200, train loss: 6.01e-05, accuracy: 99.84%
Epoch 171/200, val loss: 2.97e-03, accuracy: 92.20%
Epoch 172/200, train loss: 6.57e-05, accuracy: 99.81%
Epoch 172/200, val loss: 3.01e-03, accuracy: 92.40%
Epoch 173/200, train loss: 6.03e-05, accuracy: 99.83%
Epoch 173/200, val loss: 2.98e-03, accuracy: 92.28%
Epoch 174/200, train loss: 6.00e-05, accuracy: 99.82%
Epoch 174/200, val loss: 3.00e-03, accuracy: 92.42%
Epoch 175/200, train loss: 5.87e-05, accuracy: 99.82%
Epoch 175/200, val loss: 2.95e-03, accuracy: 92.44%
Epoch 176/200, train loss: 5.89e-05, accuracy: 99.85%
Epoch 176/200, val loss: 2.97e-03, accuracy: 92.28%
Epoch 177/200, train loss: 6.25e-05, accuracy: 99.83%
Epoch 177/200, val loss: 2.98e-03, accuracy: 92.44%
Epoch 178/200, train loss: 5.42e-05, accuracy: 99.87%
Epoch 178/200, val loss: 3.00e-03, accuracy: 92.36%
Epoch 179/200, train loss: 6.05e-05, accuracy: 99.85%
Epoch 179/200, val loss: 2.96e-03, accuracy: 92.32%
Epoch 180/200, train loss: 6.17e-05, accuracy: 99.84%
Epoch 180/200, val loss: 2.93e-03, accuracy: 92.36%
Epoch 181/200, train loss: 5.81e-05, accuracy: 99.82%
Epoch 181/200, val loss: 2.96e-03, accuracy: 92.38%
Epoch 182/200, train loss: 5.66e-05, accuracy: 99.84%
Epoch 182/200, val loss: 2.99e-03, accuracy: 92.38%
Epoch 183/200, train loss: 6.40e-05, accuracy: 99.82%
Epoch 183/200, val loss: 2.97e-03, accuracy: 92.40%
Epoch 184/200, train loss: 5.32e-05, accuracy: 99.88%
Epoch 184/200, val loss: 2.98e-03, accuracy: 92.56%
Epoch 185/200, train loss: 5.29e-05, accuracy: 99.90%
Epoch 185/200, val loss: 2.99e-03, accuracy: 92.46%
Epoch 186/200, train loss: 5.42e-05, accuracy: 99.85%
Epoch 186/200, val loss: 2.98e-03, accuracy: 92.40%
Epoch 187/200, train loss: 5.51e-05, accuracy: 99.83%
Epoch 187/200, val loss: 2.99e-03, accuracy: 92.30%
Epoch 188/200, train loss: 5.78e-05, accuracy: 99.85%
Epoch 188/200, val loss: 2.95e-03, accuracy: 92.36%
Epoch 189/200, train loss: 5.34e-05, accuracy: 99.87%
Epoch 189/200, val loss: 2.97e-03, accuracy: 92.32%
Epoch 190/200, train loss: 5.42e-05, accuracy: 99.86%
Epoch 190/200, val loss: 2.98e-03, accuracy: 92.46%
Epoch 191/200, train loss: 5.16e-05, accuracy: 99.87%
Epoch 191/200, val loss: 3.01e-03, accuracy: 92.38%
Epoch 192/200, train loss: 5.77e-05, accuracy: 99.83%
Epoch 192/200, val loss: 3.05e-03, accuracy: 92.38%

```
Epoch 193/200, train loss: 5.16e-05, accuracy: 99.88%
Epoch 193/200, val loss: 3.02e-03, accuracy: 92.36%
Epoch 194/200, train loss: 5.61e-05, accuracy: 99.85%
Epoch 194/200, val loss: 3.01e-03, accuracy: 92.42%
Epoch 195/200, train loss: 5.38e-05, accuracy: 99.86%
Epoch 195/200, val loss: 2.99e-03, accuracy: 92.38%
Epoch 196/200, train loss: 5.08e-05, accuracy: 99.88%
Epoch 196/200, val loss: 3.01e-03, accuracy: 92.46%
Epoch 197/200, train loss: 5.30e-05, accuracy: 99.86%
Epoch 197/200, val loss: 2.97e-03, accuracy: 92.44%
Epoch 198/200, train loss: 5.53e-05, accuracy: 99.87%
Epoch 198/200, val loss: 3.00e-03, accuracy: 92.24%
Epoch 199/200, train loss: 5.54e-05, accuracy: 99.86%
Epoch 199/200, val loss: 3.02e-03, accuracy: 92.38%
Epoch 200/200, train loss: 5.45e-05, accuracy: 99.84%
Epoch 200/200, val loss: 2.98e-03, accuracy: 92.38%
```

Once the model is trained we run it on the test set to obtain our final accuracy. Note that we can only look at the test set once, everything else would lead to overfitting. So you *must* ignore the test set while developing your model!

```
[19]: test_loss, test_acc = run_epoch(resnet, None, dataloaders['test'], train=False)
      print(f"Test loss: {test_loss:.1e}, accuracy: {test_acc * 100:.2f}%")
```

```
Test loss: 2.8e-03, accuracy: 92.27%
```

That's almost what was reported in the paper (92.49%) and we didn't even train on the full training set.

6 Optional task: Squeeze out all the juice!

Can you do even better? Have a look at [A Recipe for Training Neural Networks](#) and at the [EfficientNet architecture](#) we discussed in the lecture. Play around with the possibilities PyTorch offers you and see how close you can get to the [state of the art on CIFAR-10](#).

Hint: You can use [Google Colab](#) to access some free GPUs for your experiments.