Ying Zhenyu 23043355g

**8.1**

1.1

1.

Conv layer 1:

Input shape: (16, 1, 28, 28)

Output shape: (16, 16, 24, 24)

Max pooling 1:

Input shape: (16, 16, 24, 24)

Output shape: (16, 16, 12, 12)

Conv layer 2:

Input shape: (16, 16, 12, 12)

Output shape: (16, 32, 10, 10)

Max pooling 2:

Input shape: (16, 32, 10, 10)

Output shape: (16, 32, 5, 5)

Conv layer 3:

Input shape: (16, 32, 5, 5)

Output shape: (16, 16, 5, 5)

FC layer 1:

Input shape: (16, 1655) = (16, 400)

Output shape: (16, 64)

FC layer 2:

Input shape: (16, 64)

Output shape: (16, 10)

2.

Conv layer 1:

Kernel size: 5x5, out\_channels: 16

Trainable parameters: (5 \* 5 \* 1 + 1) \* 16 = 416

Max pooling 1:

No trainable parameters

Conv layer 2:

Kernel size: 3x3, out\_channels: 32

Trainable parameters: (3 \* 3 \* 16 + 1) \* 32 = 4640

Max pooling 2:

No trainable parameters

Conv layer 3:

Kernel size: 1x1, out\_channels: 16

Trainable parameters: (1 \* 1 \* 32 + 1) \* 16 = 528

FC layer 1:

in\_features: 1655 = 400, out\_features: 64

Trainable parameters: (400 + 1) \* 64 = 25664

FC layer 2:

in\_features: 64, out\_features: 10

Trainable parameters: (64 + 1) \* 10 = 650

1.2

1.

文本

描述已自动生成

2.

a.

文本

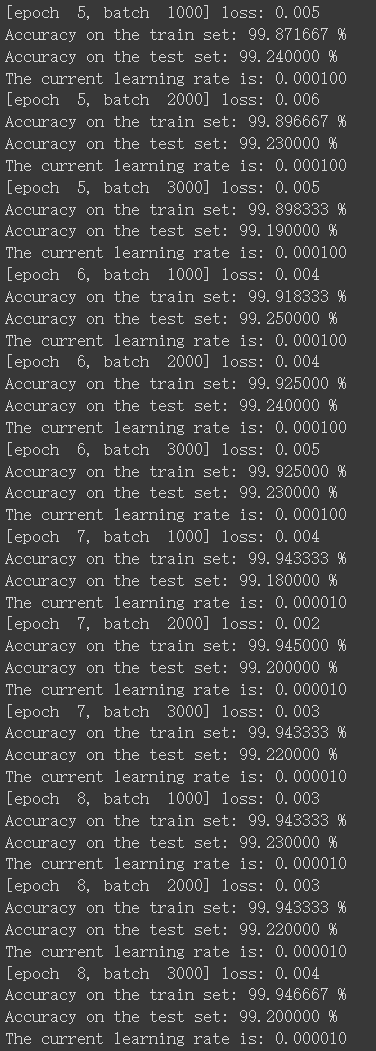
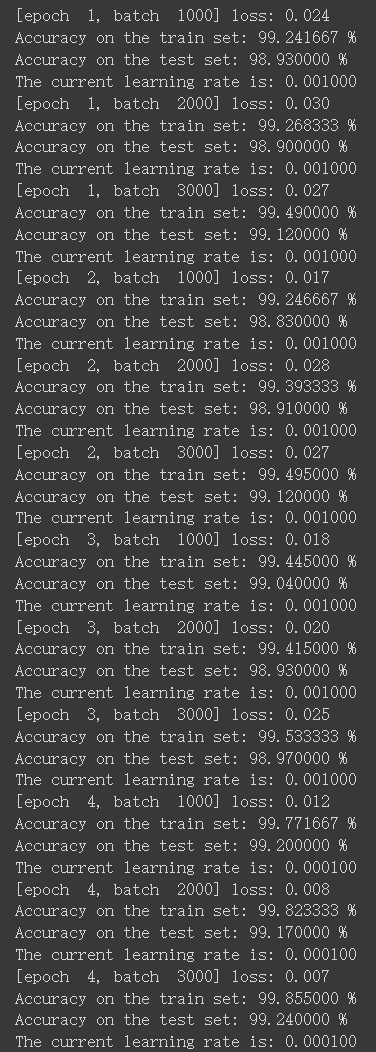
描述已自动生成

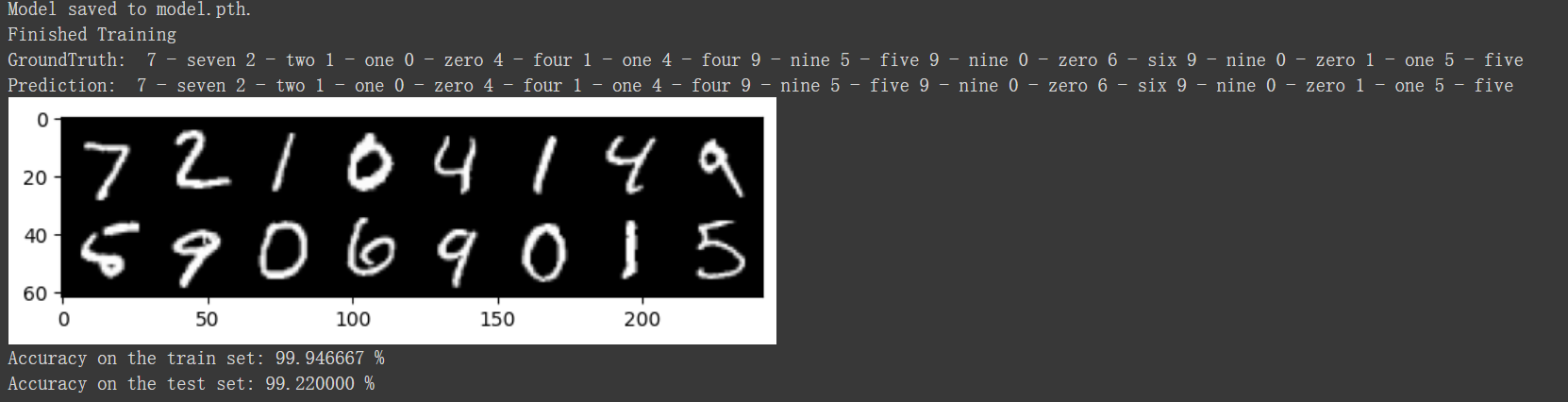
文本

描述已自动生成

b.

Result of Adam:





Result of SGD:

文本

描述已自动生成 文本

描述已自动生成

图形用户界面

描述已自动生成

Discussion:

Adam converges faster than SGD, has lower requirements for super parameters, and it has higher accuracy in this experiment. However, Adam may lead to over fitting, which should also be noted.

c.



d.

The training accuracy, testing accuracy and learning rate every 1000 mini-batches are listed in section b.

The file model.pth is attached. This model is gained with Adam optimizer.

**8.2**

2.1

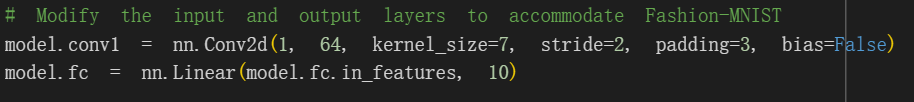
手机屏幕的截图

描述已自动生成

2.2

1.





2.

The result with torchvision.transforms.RandomHorizontalFlip and torchvision.transforms.RandomRotation:

文本

描述已自动生成

torchvision.transforms.RandomHorizontalFlip and torchvision.transforms.RandomRotation did data augmentation, which avoids overfitting, increasing the final classification accuracy on test set.

2.3

The file fashion\_mnist.pth is attached.