Asadullah Hill Galib
CSE 803: Computer Vision
Homework 5
November, 29

1 Fashion-MNIST Classification

1.1 Model Architecture

The best model includes two sequential components followed by three fully connected layers and a dropout layer. Architecture of the best model (sequentially):

- First Sequential component:
  - Conv2d (in_channels=1, out_channels=32, kernel_size=3, padding=1)
  - BatchNorm2d (32)
  - ReLU()
  - MaxPool2d (kernel_size=2, stride=2)
- Second Sequential component:
  - Conv2d (in_channels=32, out_channels=64, kernel_size=3, padding=0)
  - BatchNorm2d (64)
  - ReLU()
  - MaxPool2d (kernel_size=2, stride=2)
- First Fully connected layer: Linear(in_features=64 * 6 * 6, out_features=600)
- Dropout Layer: Dropout2d (0.25) // 25% dropout
- Second Fully connected layer: Linear(in_features=600, out_features=120)
- First Fully connected layer: Linear(in_features=120, out_features=10)

1.2 Hyperparameters

The hyperparameters are as follows:

- Learning Rate: 0.001
- Weight Decay: 0.0001
- Batch Size: 128
- Number of Epochs: 20

1.3 Training and Validation Loss across Iterations

Training and validation loss across iterations is shown in Fig. 7

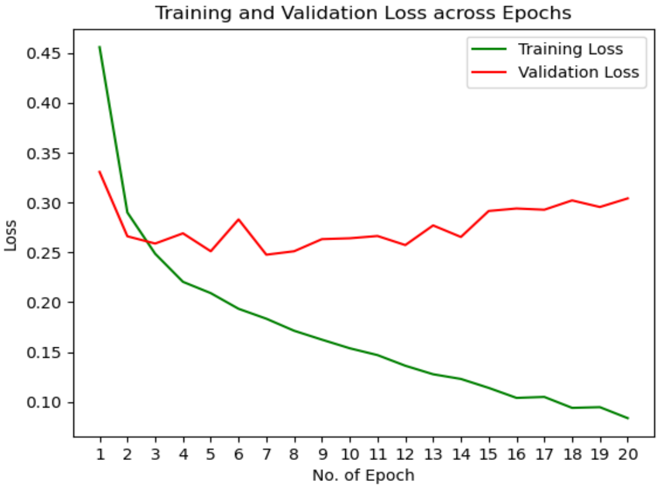
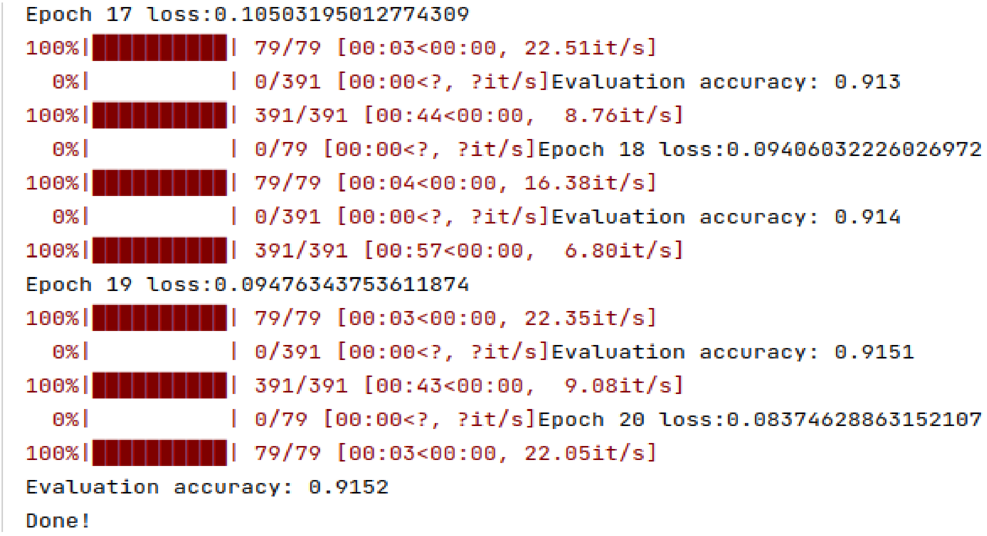


Figure 1: Training and validation loss across iterations

1.4 Accuracy

Best model accuracy is : **91.52%**



```
Epoch 17 loss:0.10503195012774309
100%|██████████| 79/79 [00:03<00:00, 22.51it/s]
  0%|          | 0/391 [00:00<?, ?it/s]Evaluation accuracy: 0.913
100%|██████████| 391/391 [00:44<00:00,  8.76it/s]
  0%|          | 0/79 [00:00<?, ?it/s]Epoch 18 loss:0.09406032226026972
100%|██████████| 79/79 [00:04<00:00, 16.38it/s]
  0%|          | 0/391 [00:00<?, ?it/s]Evaluation accuracy: 0.914
100%|██████████| 391/391 [00:57<00:00,  6.80it/s]
Epoch 19 loss:0.09476343753611874
100%|██████████| 79/79 [00:03<00:00, 22.35it/s]
  0%|          | 0/391 [00:00<?, ?it/s]Evaluation accuracy: 0.9151
100%|██████████| 391/391 [00:43<00:00,  9.08it/s]
  0%|          | 0/79 [00:00<?, ?it/s]Epoch 20 loss:0.08374628863152107
100%|██████████| 79/79 [00:03<00:00, 22.05it/s]
Evaluation accuracy: 0.9152
Done!
```

Figure 2: Best model accuracy

2 Activation Visualization

2.1 Model Architecture

The self.base module includes the following (sequentially):

- First Convolutional Layer: Conv2d (in_channels=1, out_channels=16, kernel_size=5)
- ReLU
- MaxPool2d (kernel_size=2, stride=2)
- Second Convolutional Layer: Conv2d (in_channels=16, out_channels=32, kernel_size=5)
- ReLU
- MaxPool2d (kernel_size=2, stride=2)
- Linear Layer: Linear(in_features=11, out_features=11)
- out_channel: 32

2.2 Hyperparameters

The hyperparameters are as follows:

- Learning Rate: 0.001
- Weight Decay: 0.0001
- Batch Size: 128
- Number of Epochs: 25

2.3 Accuracy

Model accuracy on test sett is : **81.03%**

2.4 Correctly Classifier Image and the Activation Maps

The correctly classifier image's index is: 2. It is shown in Fig. 3

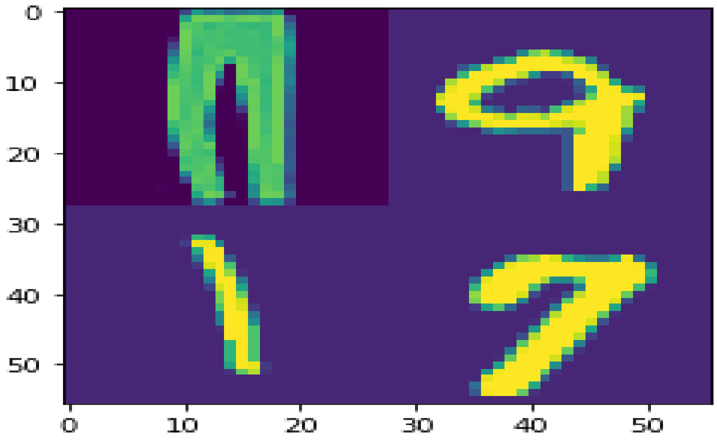


Figure 3: Correctly Classified Image: Trouser

Its corresponding activation maps is shown in Fig. 4. The second block is showing most activation as the trouser's class is 2. We see that at ground truth class for trouser (class -2), activation is higher at the position of the Fashion-MNIST image in the input image, implying that our model has learned to "look at" only the Fashion-MNIST images for classification

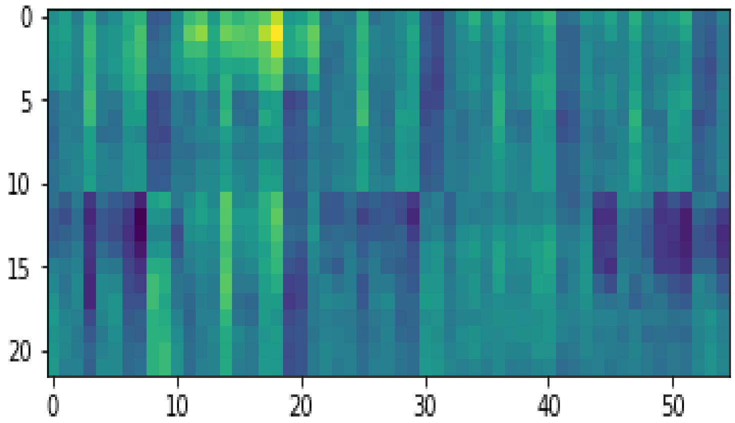


Figure 4: Heatmap of the image (Trouser - class 2)

3 Semantic Segmentation

3.1 Model Architecture

The recommended U-net-based architecture [1, 2] is used as the underlying architecture. The architecture is shown in Fig. 5

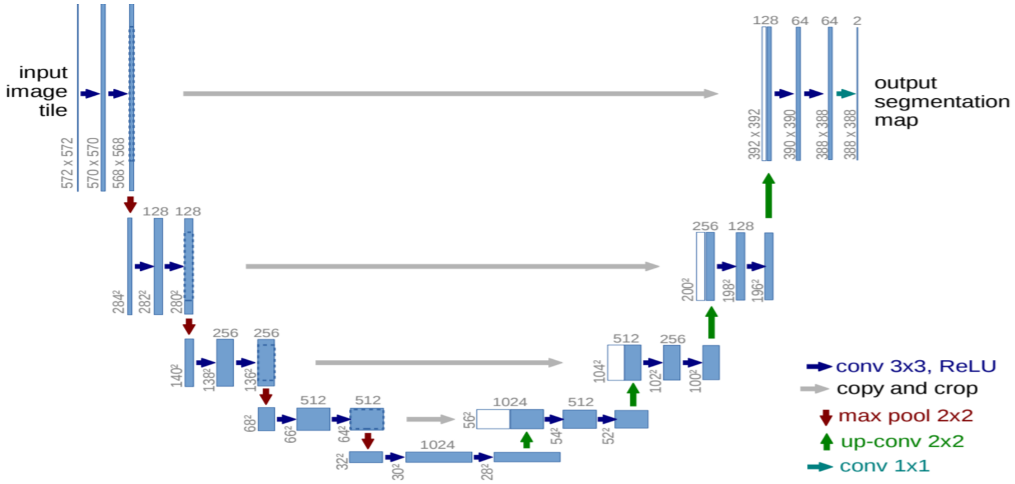


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Figure 5: Heatmap of the image (Trouser - class 2)

3.2 Hyperparameters

The hyperparameters are as follows:

- Learning Rate: 0.001
- Weight Decay: 0.0001
- Batch Size: 1
- n_channels = 3
- n_classes = 5
- Number of Epochs: 10
- Train Data Range: 0 to 250
- Validation Data Range: 250 to 350
- Test Data Range: 0 to 114

3.3 Training and Validation Loss across Iterations

Training and validation loss across iterations is shown in Fig. 7

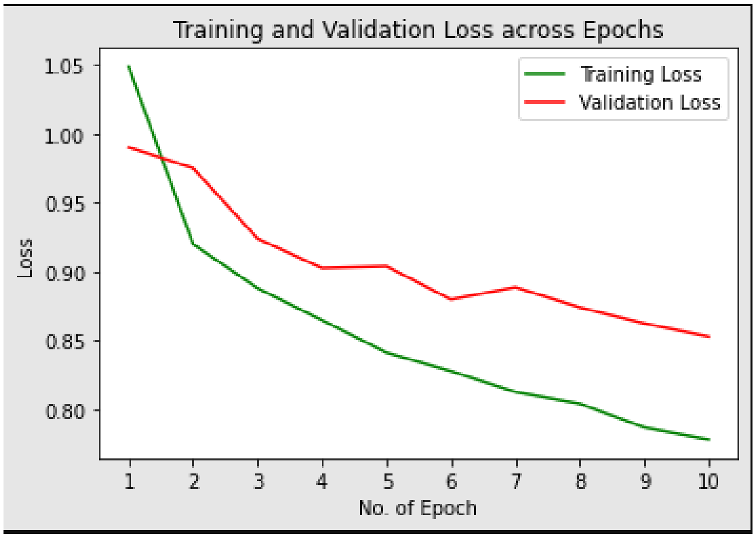
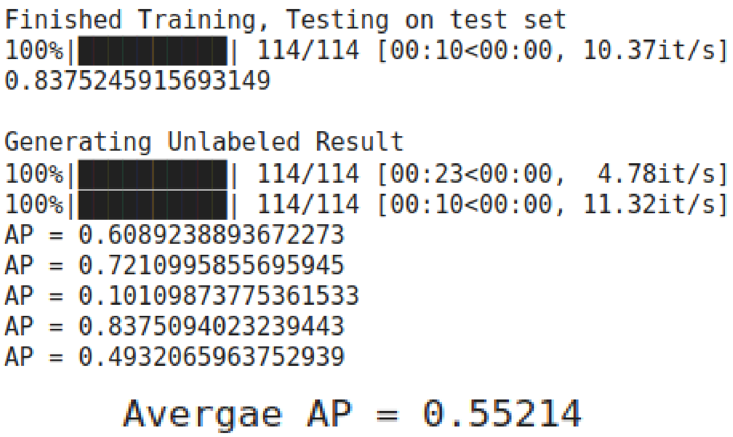


Figure 6: Training and validation loss across iterations

3.4 Average Precision on Test Set

Average Precision on Test Set is : **0.552**

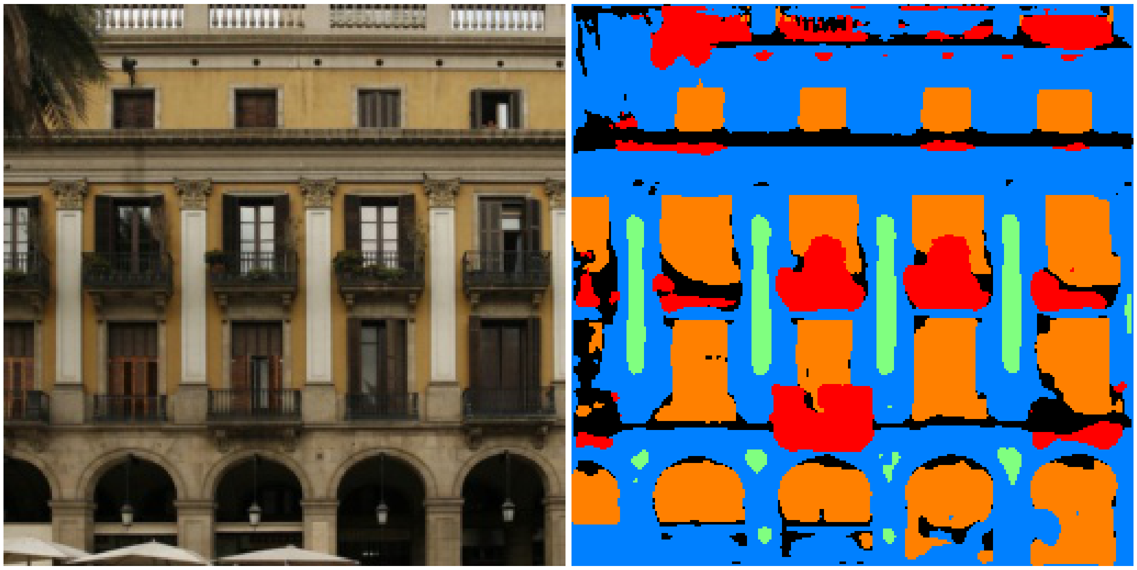


```
Finished Training, Testing on test set
100%|██████████| 114/114 [00:10<00:00, 10.37it/s]
0.8375245915693149

Generating Unlabeled Result
100%|██████████| 114/114 [00:23<00:00,  4.78it/s]
100%|██████████| 114/114 [00:10<00:00, 11.32it/s]
AP = 0.6089238893672273
AP = 0.7210995855695945
AP = 0.10109873775361533
AP = 0.8375094023239443
AP = 0.4932065963752939

        Avergae AP = 0.55214
```

Figure 7: Average Precision on Test Set

3.5 Evaluation on a selected photo of a building

According to Fig. 8, the output image is quite well classified. Pillars (green), windows (orange), facade (blue), balcony (red), and others (black) are almost segmented in the output image.



(a) Input

(b) Output

Figure 8: Semantic Segmentation on a selected photo

References

[1] Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015

[2] https://github.com/milesial/Pytorch-UNet