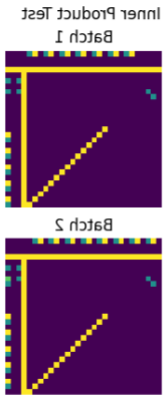
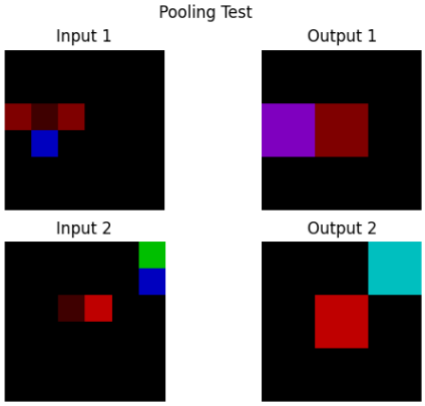
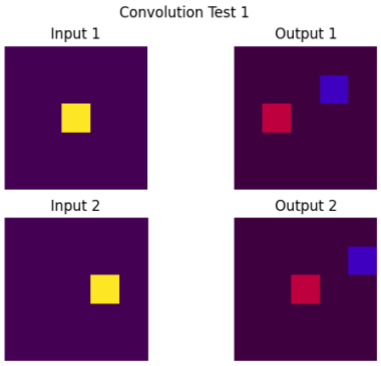
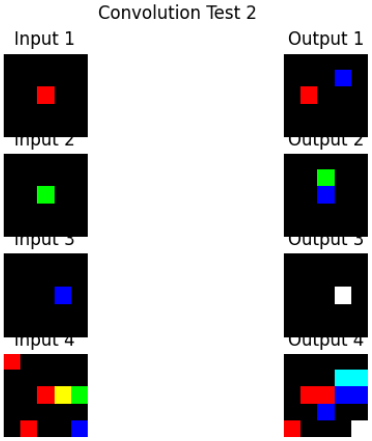
Project1

Part 1: Forward Pass



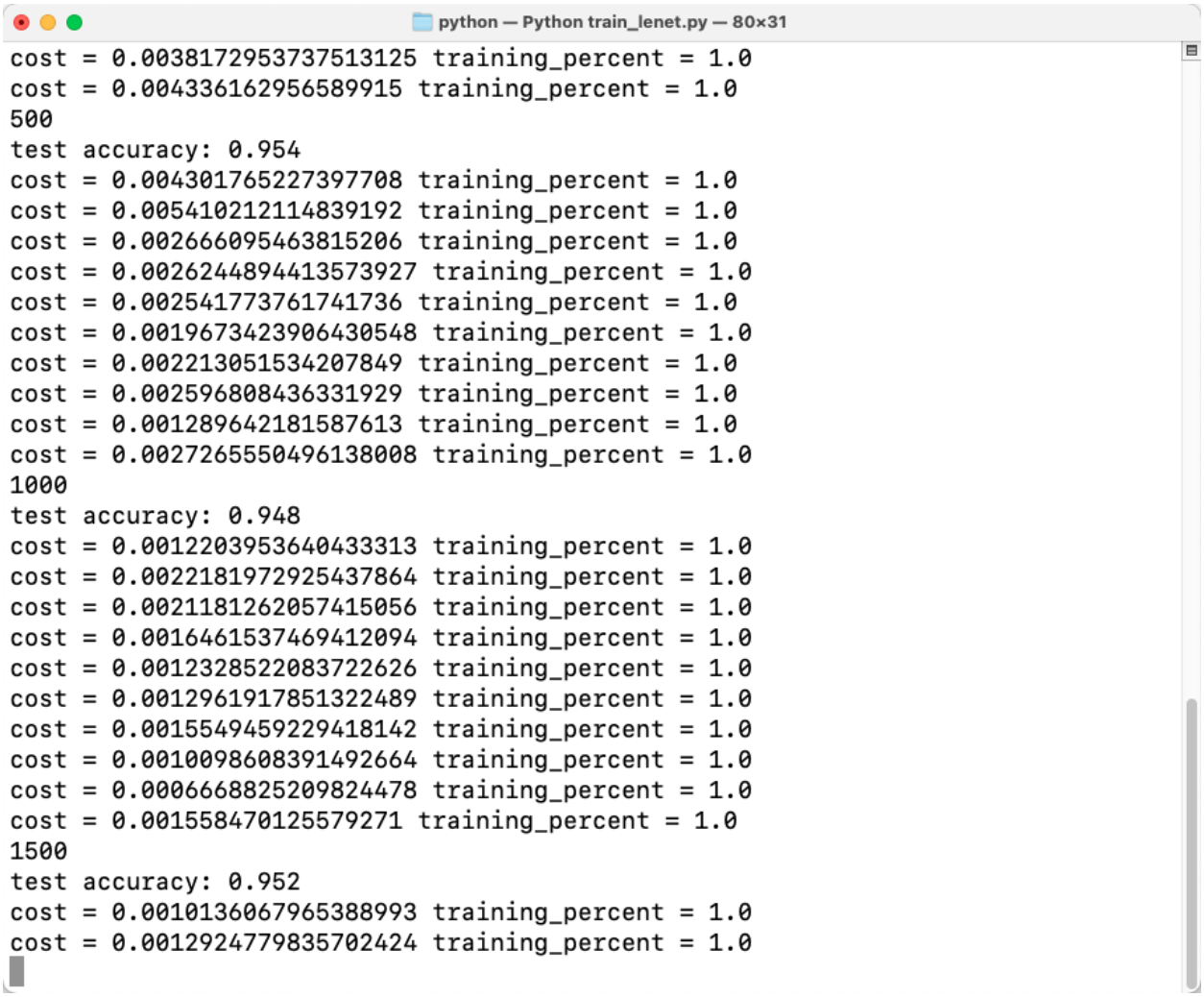






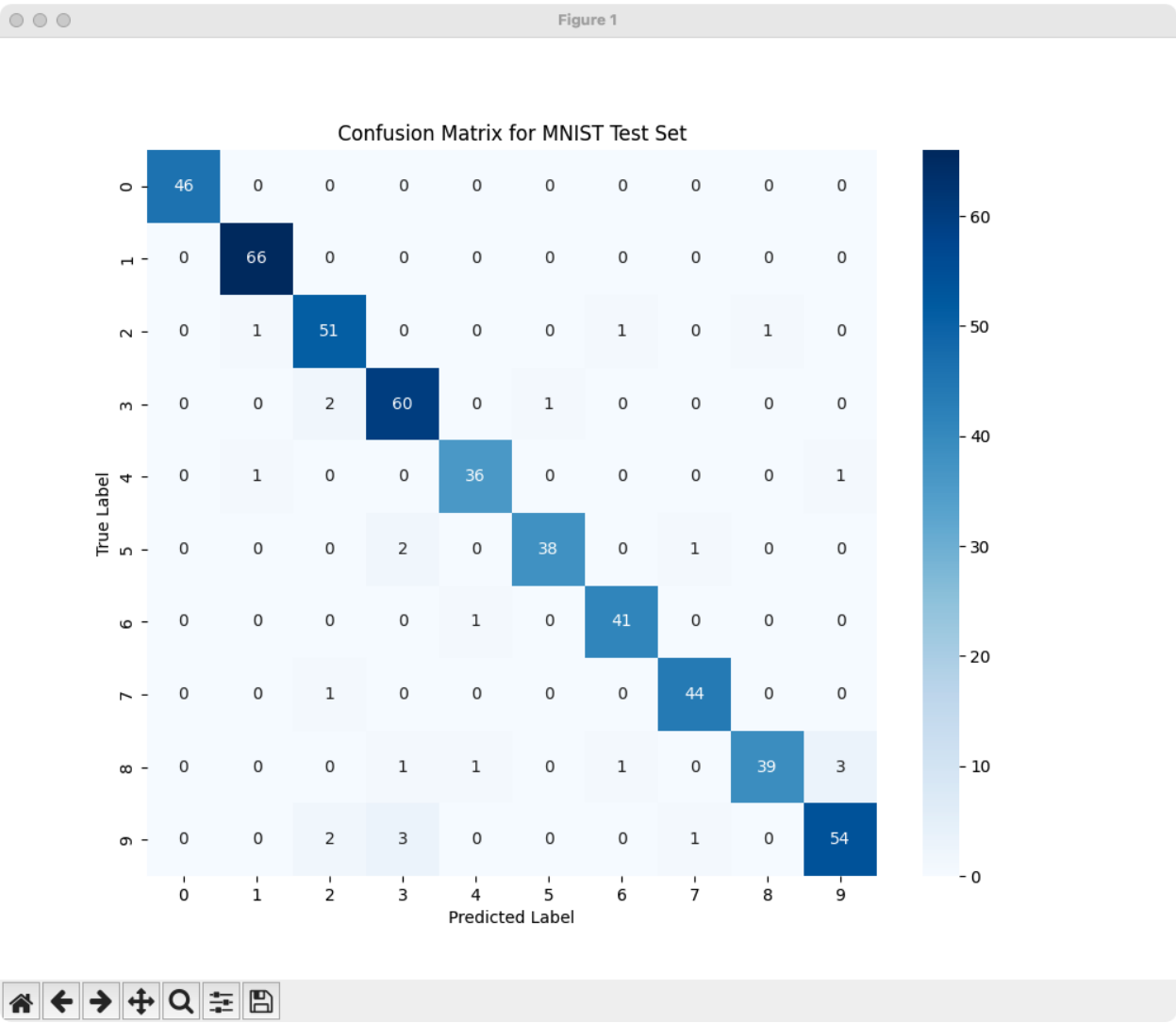
Part 3 Training

Q 3.1 Training



The test accuracy reached 95.2%.

Q 3.2 Test the network

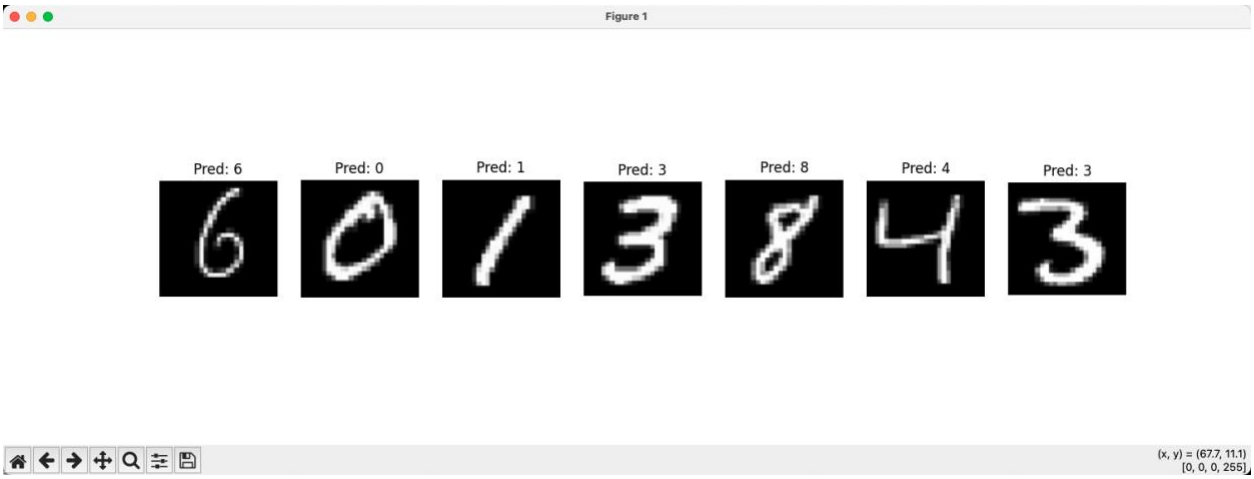



```
The two most confused class pairs:
Class 8 was misclassified as 9 3 times
Class 9 was misclassified as 3 3 times
```

1. **8 and 9**: because their shapes are quite similar. Both numbers have circular shapes, and if the bottom part of 9 is not clearly separated, it can look like 8. When handwriting is unclear or when the model is not trained with many examples, it becomes hard for the model to tell the difference between these two numbers.
2. **9 and 3**: because the top part of both numbers looks similar. In certain handwriting or fonts, the curve at the top of 9 can look very much like the top of 3. This makes it

difficult for the machine learning model to correctly identify them, especially if the quality of the image is poor or if the training data lacks variety.

Q 3.3 Real-world testing



Here are the test digits I found on the internet. All of the 7 digits are correctly predicted.

Q 4.1

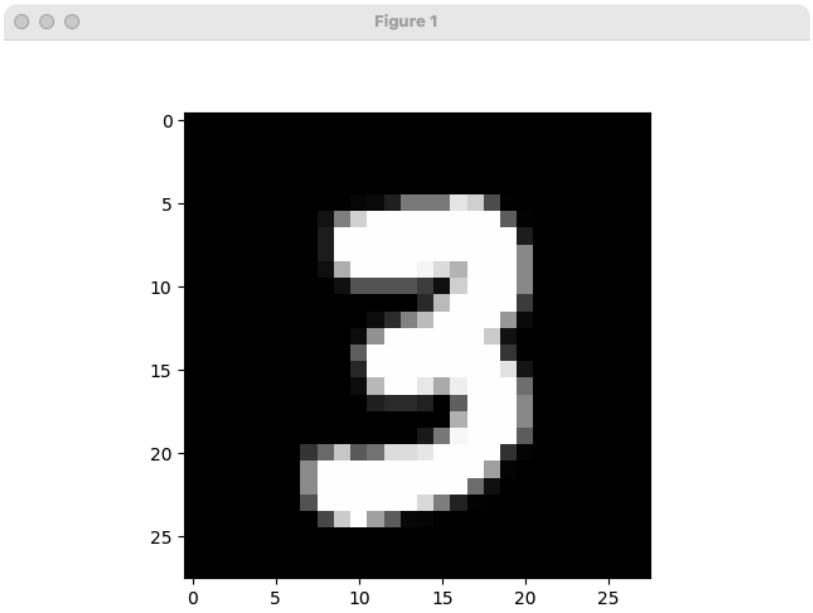


Figure 1
0
5
10
15
20
25
0
5
10
15
20
25

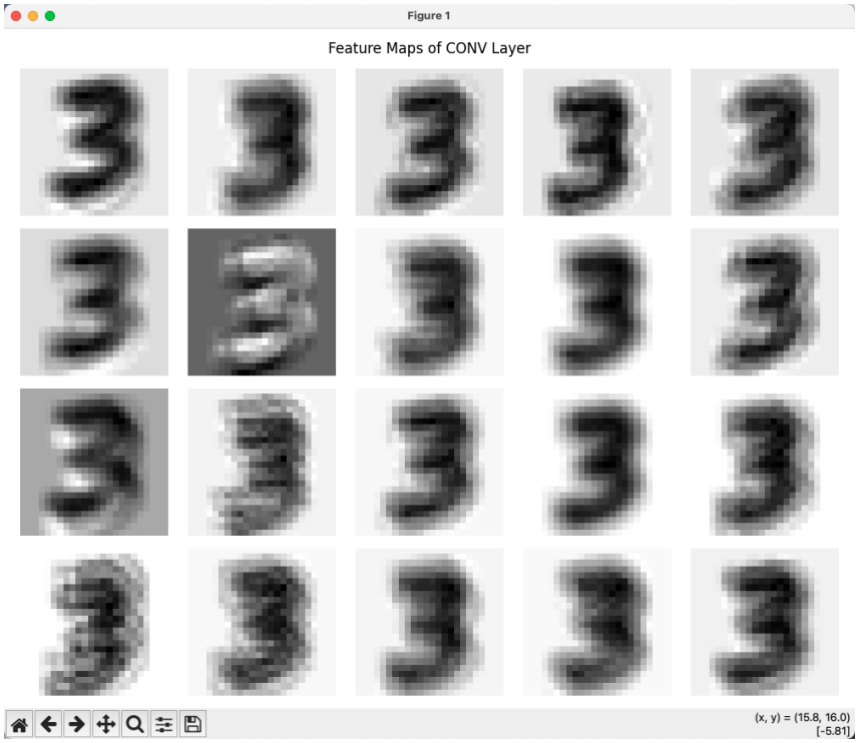
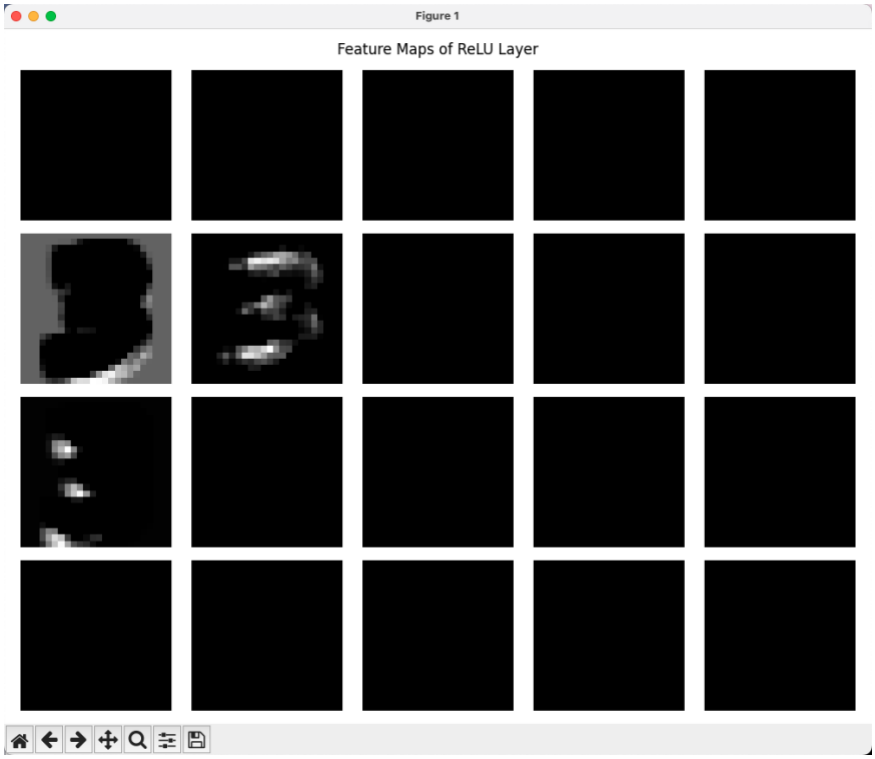


Figure 1
Feature Maps of CONV Layer
(x, y) = (15.8, 16.0)
[-5.81]



Q 4.2

The feature maps from the CONV and ReLU layers look very different from the original image. The original image is clear and detailed, while the CONV layer shows different aspects of the number "3." These feature maps highlight specific patterns like edges or curves. In the CONV layer, the filters extract key shapes or lines from the image, so each map focuses on different parts of the number.

For the ReLU layer, most of the maps look dark because ReLU removes all negative values, setting them to zero. Only the positive activations are shown, which explains why some parts of the number are barely visible. The ReLU maps show the strongest activated parts of the image, which are usually the most important features of the network.

In summary, the feature maps simplify and highlight certain details of the original image. The CONV maps extract features like edges, while the ReLU maps only show the most significant parts, ignoring the rest.

Part 5

This code processes images, segments digits, and predicts them using a pre-trained neural network.

1. **Image Processing**: It reads grayscale images and applies adaptive thresholding to convert them into binary (black and white) images.
2. **Segmentation**: The connectedComponentsWithStats() function detects connected components (digits or noise), which are filtered based on area size (adjusted via line 58).
3. **Digit Prediction**: Each digit is resized to 28x28 pixels, normalized, and passed to the neural network to get a prediction.
4. **Display**: The results are displayed with up to 10 digits per row for clarity.

Image1

Accuracy: 6/10

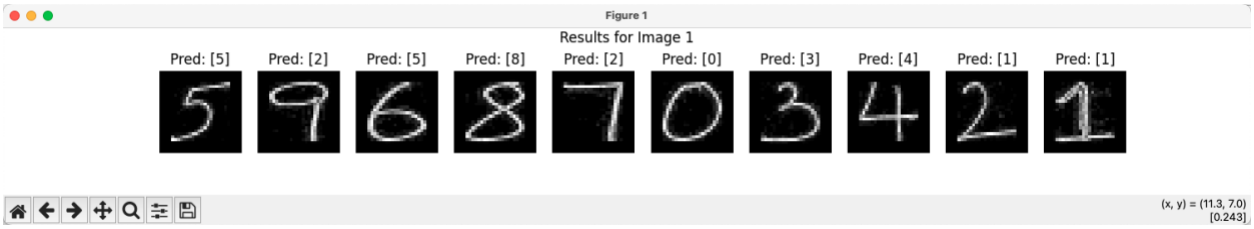


Image2

Accuracy: 5/10

Image3

Accuracy: 4/5

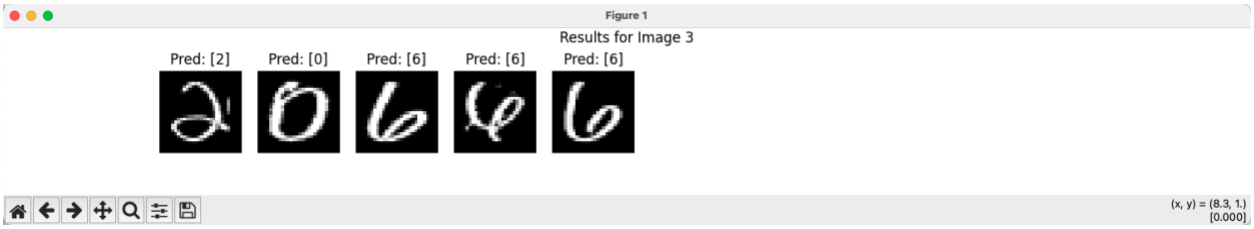


Image4

Accuracy: 24/50