MASC 520 Project 3 Report
Handwritten Digit Recognition using Convolutional Neural Network (CNN) on MNIST Dataset

Ruoxi Li
5817808450

Abstract

A convolutional neural network (CNN) is a class of deep neural networks, most commonly applied to analyzing visual imagery. In this project, I trained a CNN to classify handwritten digit images in MNIST dataset with 10 different labels from 0 to 9. The CNN has two convolution layers and two fully connected layer, each followed by a ReLU nonlinearity which has non-saturing property, and a final 10-way softmax. To reduce the complexity of the network, max pooling is used after the last convolution layer. Dropout is used for fully connected layer to overcome overfitting problem. On the test data, I achieved 99% accuracy after 14 epochs.

1 Method

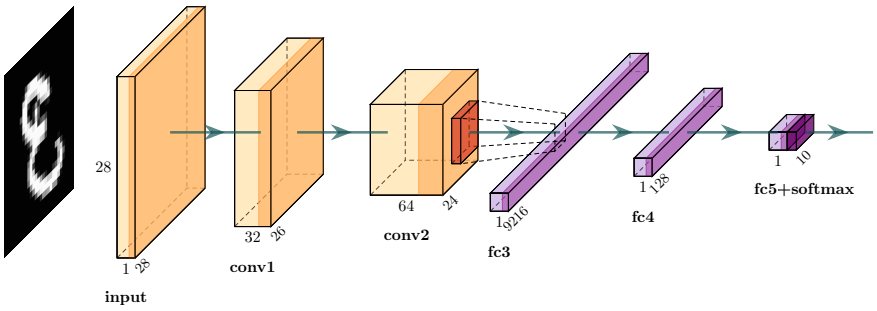

Figure 1: Schematic of the neural network architecture to recognize handwritten digit

The CNN architecture to recognize MNIST handwritten digit is presented in Figure 1. It consists of two convolutional layers (conv2 and conv3), one max pooling layers (after conv3) and two fully-connected layers (fc5 and fc6).The output of fc6 is passed to a 10-way softmax, which produces a probability distribution over ten labels (i.e. '0'-'9').

The first layer (conv1) takes the original 28×28 image as the input. The first convolutional layer (conv2) filters the grayscale image with 32 kernels of size 3×3, while the second convolutional layer (conv3) filters the $26 \times 26 \times 32$ feature maps with 64 kernels of size $3 \times 3 \times 32$. Both convolutional layers use a unit stride. Down-sampling occurs at layer conv3 by applying 2×2 non-overlapping max pooling. Then the $12 \times 12 \times 64$ feature maps are flatten to a 1×9216 vector (fc4). Finally, the two fully-connected layers, fc5 and fc6, have 128 and 10 neurons, respectively.

Preprint. Under review.

To speedup the training process and to reduce overfitting, dropout layer is inserted after max pooling and fc4 with dropout rate 0.25 and 0.5 respectively. ReLU nonlinear function is used at the output of the convolutional and fully-connected layers, which is much efficient to compute than the sigmoid or the hyperbolic tangent functions and also proved to be trained faster for the same accuracy.

2 Results

Same as what we have done in the Project 1, the training data are normalized by the mean and standard deviation computed on the training set in order to make the training faster and robust. The model is trained using Adam optimizer with 0.01 learning rate and a minibatch of 64 images. We evaluate the model after 14 epochs.

The CNN architecture shown in Figure 1 is built using Pytorch. The training and testing loss is shown in Figure 2, and the testing accuracy is shown in Figure 3. The loss decreases from 2.7 to 0.13 after 14 epochs. The training loss is evaluated every iteration and the testing loss is calculated every epoch. After only 14 training epochs, the accuracy on training set is 99%, and the accuracy on test set is 99%.

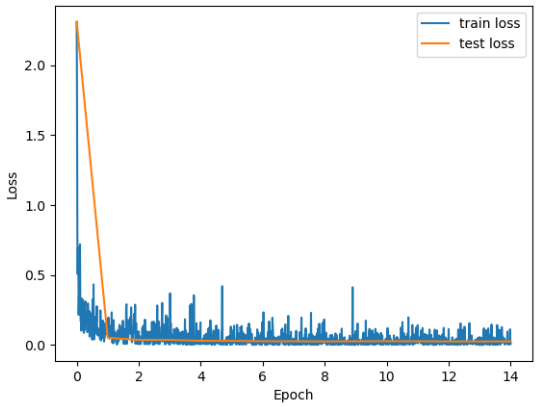

Figure 2: Training and Testing loss

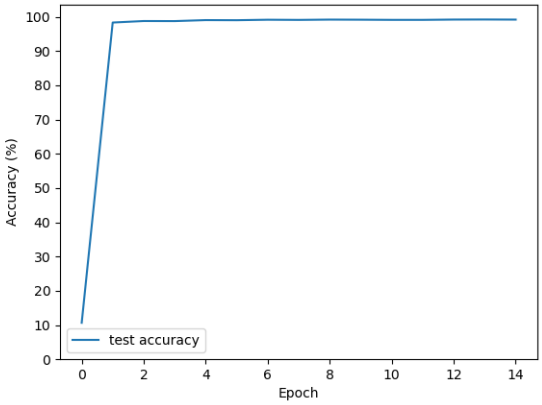

Figure 3: Testing accuracy for recognizing handwritten digit

In theory, the CNN should have much better performance than the MLP. However, since this task is relative easy (even support vector machine (SVM) can achieve 90% accuracy), the CNN cannot show

its advantage. If we train CNN on a more complicated dataset, it can achieve much higher accuracy than MLP.

To illustrate the result, we sample 10 classified images from test set. The result is shown in Figure 4. All the samples are classified correctly.

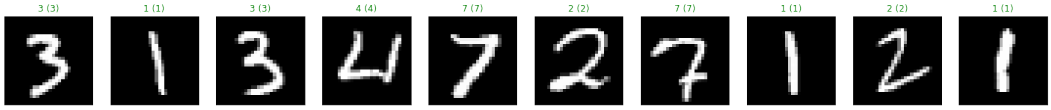

Figure 4: Classification result

3 Conclusion

In this project, a simple CNN is trained to classify handwritten digit. The CNN is a very powerful neural network and can handle this problem easily. ReLU are used as the activation function and Adams are used as the optimizer. To reduce the complexity of the CNN, maxpooling is used after the last convolution layer. After tuning some hyperparameters, the model achieves 99% test accuracy after only 14 epochs. Although we only get 1% accuracy higher than the result made by MLP, the CNN can achieve much higher accuracy than MLP on more difficult tasks.