Handwriting Recognition Performance of Different Neural Network Structure

Author Ping-Hsuan-Hsi, Pei-Fan-Chen

Abstract

AlexNet was the pioneer to usher in a new era of CNN, convolution neural network, in the image recognition field. Two crucial new techniques in CNN are the convolution layer and the pooling layer. Thus, we want to compare the handwriting recognition performance of the fully-connected neural network and the AlexNet with different pooling methods. We will evaluate the performance based on the MNIST dataset, added Gaussian noise MNIST dataset, and the seventy numbers written by ourselves with normal and added noise version.

1. Introduction

In this project, we want to compare the learning performance between the fully connected neural network and the AlexNet, one of the most important CNN structures. We think that this is critical for beginners to learn the importance of the convolution layer and the difference between max pooling and average pooling in the CNN structure. The accuracy of the MNIST testing data is the most common method to evaluate the learning performance, but we think that the model's accuracy on other handwriting testing data is also a crucial metric to evaluate the generalization. In addition, we are curious about the stability of a model when learning and detecting the added noise dataset. Thus, we will implement three neural network structures, fully connected, AlexNet with max pooling, and AlexNet with average pooling, and evaluate their performance under two situations and two kinds of testing data respectively. In the first situation, we use the original MNIST dataset to train and test the model, as well as the original handwriting numbers written by ourselves to test the models. In the second situation, all the datasets will be added Gaussian noise in the training and testing stage.

2. Related Work

The first version of a fully connected neural network, Perceptron, created by Frank Rosenblatt in the 1950s. [2] AlexNet proposed by Alex Krizhevsky et al. in 2012. [1] Compared to fully connected neural network, AlexNet introduced new techniques including convolution layer, max pooling layer, ReLU Nonlinearity, and the Dropout process which contributed greatly to the improvement of the image recognization accuracy.

3. Methodology

3.1. Datasets

All images have 28 x 28 size and 10 class labels from (0-9).

3.1.1 Normal Dataset

In the first experiment, we use the MNIST dataset with 60000 training images and 10000 testing images. Apart from that, we use 70 handwritten numbers by ourselves as the additional testing images. The dataset is fed into 3 classifiers, namely: fully connected neural network(5-layer), AlexNet with max pooling(8-layer), and AlexNet with average pooling(8-layer).

3.1.2 Added Noise Dataset

In the second experiment, we have all the datasets in the first experiment(MNIST 60000 training + 10000 testing and 70 handwritten testing numbers by ourselves) to be added Gaussian noise with mean = 0 and std = 0.3, and also be fed into the same 3 classifiers.

3.2. Classifiers

In all 3 classifiers, the number of learning epochs is 10, the learning rate is 0.01, and the batch size is 64.

3.2.1 Fully Connected Neural Network

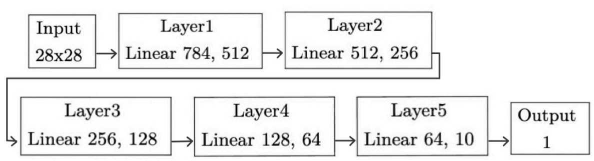

图 1. Fully connected neural network structure

3.2.2 AlexNet - Max Pooling

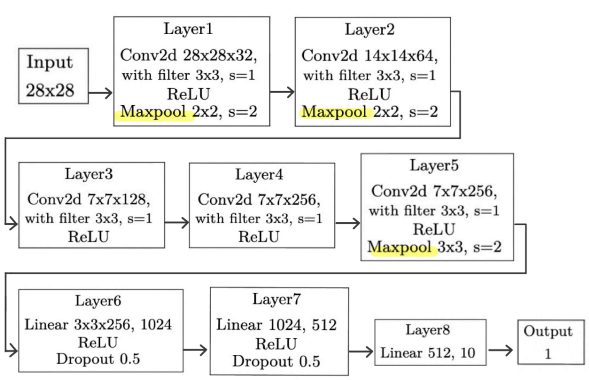

图 2. Structure of AlexNet with max pooling for MNIST dataset

3.2.3 AlexNet - Average Pooling

The overall structure is the same as the AlexNet-max pooling structure shown in 3.2.2, but the pooling method in layer1, 2, and 5 is changed to average pooling.

4. Experiments

4.1. Evaluation Method

For MNIST testing data, we use the last 3 epochs' accuracies' average as the result accuracy. For 70 testing numbers by ourselves, we simply count images number $\frac{correct}{70=total}$ as the result accuracy. All the result accuracy is rounded off to the 2nd decimal place.

4.2. Result

	original data		data with gaussian noise	
	MNIST	Ours	MNIST	Ours
fully-connected	97.46%	48.57% (34/70)	97.07%	35.71% (25/70)
AlexNet & max	99.06%	92.86% (65/70)	99.14%	84.26% (59/70)
AlexNet & avg	98.67%	81.43% (57/70)	98.97%	82.86% (58/70)

*Ours: test data written by ourselves

图 3. Result

4.3. Discussion

Observation1 - AlexNet with max pooling has the best performance in all 4 tests, and has the highest generalization ability. The fully connected neural network uses the similarity between two images to detect the numbers. On the other hand, the convolution layers in AlexNet extract the important local features from an image. In the number recognition field, classifying a number by its important features is much more decisive than by the similarity. In addition, the max pooling method can help extract the features in a more precise way. Thus, AlexNet with max pooling has the best performance and the highest generalization ability.

Observation2 - AlexNet performs better when the dataset is added Gaussian noise. The max pooling layer can help eliminate noise in an image, but still may be influenced since there is a certain probability that we choose the value that greatly is affected by the noise in the max pooling process. As for the AlexNet with average pooling, it is the least affected model in the noise test. We suppose that is because we use the Gaussian noise which is normally distributed with mean=0, so the average noise of a block of pixels can be expected to be 0, which is non-affected.

Observation3 - We find that AlexNet has high error rate on one specific handwritten number by ourselves. However, the fully connected neural network succeed in classifying the number correctly. The reason is

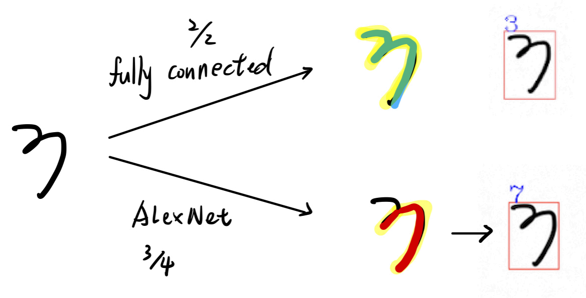

图 4. Misclassified number of AlexNet

that this handwritten number 3 has the local feature fit number 7. In this case, we can see that there is still drawback when classifying mainly based on the local features, whose range would be relatively small compared to the whole picture. To solve this problem from our AlexNet structure, we can try the bigger filter in the first two convolution layers, as the filter size suggested by the AlexNet paper, or train with augmented dataset including rotated data.

参考文献

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc. 1

[2] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65 6:386–408, 1958. 1