Investigation of Convolutional Neural Network Structure for Low Resolution Character Recognition

Kyoko Maeda, Tomo Miyazaki, Yoshihiro Sugaya, and Shinichiro Omachi
Department of Communications Engineering, Graduate School of Engineering
Tohoku University
Sendai, Japan

Abstract— **In this work we present a Convolutional Neural Network (CNN) model for recognition of low resolution characters. We examined CNN model structures suitable for this problem. We aim to recognize images smaller than 18×18 pixels with high accuracy.**

Keywords- Character Recognition; Deep Learning; Convolutional Neural Networks

I. Introduction

A lot of object recognition methods using convolutional neural networks (CNNs) have been proposed and they have produced astonishing results so far since ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012. CNNs made great achievements especially in character recognition. For example, LeNet [1] designed by LeCun recognized the MNIST handwritten character dataset with the accuracy of more than 99%. However, low resolution character recognition is a challenging problem and there is still room for improvement. In this paper, we propose a low resolution character recognition method using CNNs considering the characteristics of low resolution images.

II. Existing Method

We used a model contained in MatConvNet [2] toolbox as an existing method in experiments. The network structure was based on LeNet and optimized for MNIST dataset. Fig. 1 is the structure of the existing method. We used it as a base of our model. We changed the structure and added some preprocessing. The existing method is described as follows:

A. Preprocessing

Mean subtraction is applied to all images. The mean over all training images are subtracted from training and testing images.

B. Training and recognition

Network structure is almost the same as Fig. 1. Rectified Linear Unit (ReLU) is used as an activation function. Max pooling is employed in all pooling layers.

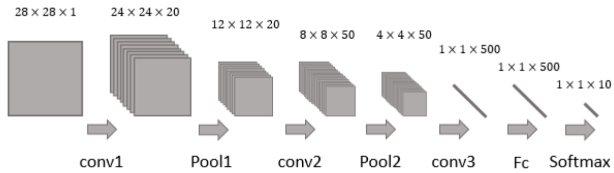


Figure 1. Structure of the existing network

III. Proposed Method

We propose a model suitable for low resolution character recognition based on an existing one contained in MatConvNet toolbox of which network structure is based on LeNet. The proposed method is described as follows:

A. Preprocessing

All images are rescaled to 28×28 pixels and then converted to grayscale images. We apply mean subtraction and ZCA whitening [3] to both training and testing samples. ZCA whitening is a processing to make the input less redundant. We added it to reduce inequality in contrast of low quality character images.

B. Training and recognition

The structure of our CNN is shown in Fig. 2. The network structure and hyperparameters are determined through preliminary experiments. This network is shallower than the network of LeNet because our images are small and low quality. Our model has only two convolutional layers. The size of each kernel is larger than that of LeNet (e.g., Conv1 is 11×11, Conv2 is 8×8) to deal with blurred images.

We employed several kinds of activation function such as Rectified Linear Unit (ReLU) or maxout [4], which are said to be in common use recently, between Pool2 layer and Fully-connected layer. ReLU sets input values less than zero to zero. The output value $h_i(x)$ of ReLU is described as

$$h_i(x) = \max(0,x) \tag{1}$$

where x is input value. On the other hand, Maxout takes maximum value from k models from the given input. The output value $h_i(x)$ of Maxout is described as

$$h_i(x) = \max_{j \in [1,k]} z_{ij} \qquad (2)$$

where z_{ij} is output value of previous units. It is described as follows:

$$z_{ij} = x^T W_{\ldots ij} + b_{ij} \qquad (3)$$

where weight W and bias b are learned parameters. To avoid overfitting problem, we also used dropout procedure [5]. We will evaluate the effectiveness of those functions in next section.

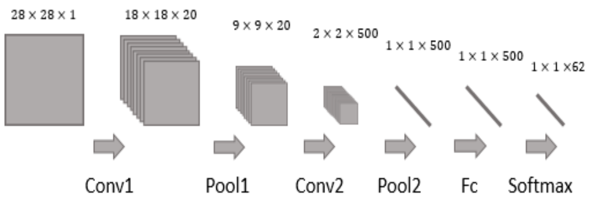


Figure 2. Structure of our CNN

IV. EXPERIMENTS

To show the effectiveness of the proposed method, some experiments were carried out. We evaluated the existing method with our own data at first, then conducted experiments to evaluate the proposed CNN model with four kinds of activation functions. Our models were implemented with MatConvNet Toolbox.

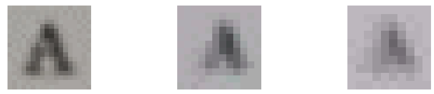
We used character images captured from videos at a distance from 50 to 100 centimeters at ten-centimeter intervals as training and testing samples. Three cameras were used to take videos: a Tablet (Sony SGPT1211), a Web Camera (Logicool Qcam Orbit AF) and a Digital Camera (Fujifilm FinePix F770EXR). In this paper, we show the result using images from the Web Camera as testing samples and others as training samples. Instances of images from the Web Camera taken from 50cm, 80cm and 100cm are shown in Fig.3.

The images contain 62 categories: capital letters (A-Z), small letters (a-z) and digits (0-9). Each category has 10 images per distance. Thus, the total number of testing images is 620 and that of training images is 1240.

We used mini-batch training with batch size 10. 20 epochs were enough to train the CNN. These were determined through experiments. We also conducted an experiment about learning rates and we got the best result with 0.001.

The result is shown in Table 1. This is the average error rates from all distances. All of the proposed methods outperformed the existing method. ReLU-Dropout-Maxout structure have achieved the best performance.

Considering these results, shallower network and larger kernels seem to work well and a combination of activation functions may be effective.



(a) 50cm (18×18 pixels) (b) 80cm (12×12 pixels) (c) 100cm (10×10 pixels)

Figure 3. Training and testing samples

TABLE 1. THE CLASSIFICATION ERROR RATES (%)

Existing Method[2]	***ReLU***	***ReLU+ Dropout***	***Maxout+ Dropout***	***ReLU+ Dropout+ Maxout***
49.81	20.89	20.99	21.69	19.68

V. CONCLUSION

In this paper, we proposed a method of low resolution character recognition using CNNs. Considering characteristics of low resolution images, we added some processing such as ZCA whitening and changed network structure from the existing model.

The accuracy of low resolution character recognition has improved by our method. The ReLU-Dropout-Maxout structure was revealed to be effective as well.

REFERENCES

[1] Y. LeCun, et al., "Gradient-based learning applied to document recognition," Proc. of the IEEE vol.86, no.11, pp.2278-2324 (1998).

[2] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," Proc. of the 23rd Annual ACM Conference on Multimedia Conference. ACM, pp.689-692 (2015).

[3] A.J. Bell and T.J. Sejnowski, "The "independent components" of natural scenes are edge filters," Vision research, vol.37, no.23, pp.3327-3338 (1997).

[4] I.J. Goodfellow, et al., "Maxout Networks," In Proc. of The 30th International Conference on Machine Learning, pp.1319-1327 (2013).

[5] G.E. Hinton, et al., "Improving neural networks by preventing co-adaptation of feature detectors," arXiv preprint arXiv:1207.0580 (2012).