Preprocessing method for training dataset in character recognition using Convolutional Neural Network

Taku Akase, Hitoshi Nakao, Lifeng Zhang

Kyushu Institute of Technology, Kitakyushu, 804-8550, Japan

*Corresponding Author: akase.taku574@mail.kyutech.jp

Abstract

Nowadays, image processing technology is used in the logistic industry widely to eliminate a labor shortage and prevent human errors. However, the existing image processing technology cannot be applied on some work processes such as to identify the product name written in characters that printed on an unstable package surface, because the color and font are not uniform and the shapes are always deformed. Therefore, we propose a system previously, which enabled this task using character recognition technology by Convolutional Neural Network (CNN) with a certain accuracy. In this research, we focus on how to make an optimal training set which can obtain high accuracy. As a result, the accuracy is improved to 99.8% by adding the diversity of the image’s background for the learning process.

Keywords: Convolutional Neural Network, Deep Learning, Image Recognition.

1. Introduction

In recent years, in the Japanese logistics industry, a system using image processing technology has been introduced to improve work efficiency. These are expected to be effective against the problems of labor shortage and human error. In Japan, where there is a concern that the workforce will decline, a system has been introduced to streamline work, and a new system will still be needed.

However, the existing image processing technology cannot be applied on some work processes such as to identify the product name written in characters that printed on an unstable package surface, because the color and font are not uniform and the shapes are always deformed. The reason is that Optical Character Recognition (OCR) technology for identifying characters cannot recognize characters extracted from natural scenes while recognizing characters accurately for scanned documents. It is because the accuracy is low. In the research of natural character recognition conducted so far, high recognition accuracy is realized by using CNN. However, when learning color characters images by using CNN, not only the shape of the characters but also the color of the background is learned, and this is because the recognition accuracy becomes low. So, in this research, when learning the character of the color image extracted from natural scenes by CNN, we propose the processing method of the data set for learning only the form of the character. As a result of the experiment, in the data set used in this research, recognition accuracy improved from 98.0% to 99.8%, and it is thought that only the shape of the character was learned.

In this paper, Chapter 2 explains the CNN principle and the problems of natural character recognition. Chapter 3 shows how to process the data set to solve the problem. Chapter 4 shows the contents and results of experiments to measure recognition accuracy. Chapter 5 presents the conclusions of this study based on experiments.

2. Related works

2.1 Convolutional Neural Network

CNN is a neural network that consists of convolution layer, pooling layer, and fully connected layer. These layers are arranged in a feed-forward structure as shown in Fig. 1.

In the convolutional layer, feature quantities of an image are extracted by applying kernels to an image. Unlike Multi Layer Perceptron (MLP), the number of optimized parame-
ters doesn’t increase as an image size increases because the number of optimized parameters depends on a kernel size rather than an image size. Fig. 2 shows how a convolution is performed using an image and a kernel.

In convolution, the kernel is slid within the image and it is feature quantities that the sum of the image data values which overlap the kernel multiplied by the kernel value. When an image size is \(M \times N\) and a kernel size is \(M' \times N'\), the convolution corresponding to the \(k\)th kernel is shown as Equation (1). In Equation (1), \(w\) is a kernel weight. However, this equation is grayscale convolution, and RGB three-channel convolution requires adding one parameter to the kernel.

\[
z_{i,j}^{(k)} = \sum_{s=0}^{m-1} \sum_{t=0}^{n-1} W_{st}^{(k)} x_{(i+s)(j+t)}
\]

After completion of the convolution, all output values are activated by the activation function. In this paper, we describe the Rectified Linear Unit (ReLU) function used in this research. The activation using ReLU is shown as Equation (2). In Equation (2), \(b\) is a bias and a one-dimensional array.

\[
a_{i,j}^{(k)} = h(z_{i,j}^{(k)} + b^{(k)}) = max(0, z_{i,j}^{(k)} + b^{(k)})
\]

In the pooling layer, reducing an image size makes feature quantities robust against small positional changes. There are several pooling methods, but in this research we use the most used max pooling. Fig. 3 shows how max pooling is applied to the feature map. In the max pooling layer, input data is divided into sub data sets so that there is no overlap, and the maximum value in each sub data is output. This not only can maintain the movement invariance of the network, but also can reduce the amount of calculation.

The calculation of the max pooling layer is shown as Equation (3). In Equation (3), \(l_1 \) and \(l_2\) are the size of the max pooling filter, and \(s \in [0, l_1], t \in [0, l_2]\).

\[
y_{i,j}^{(k)} = max(a_{(i+s)(j+t)}^{(k)})
\]

2.2 Character Recognition using CNN

In the machine learning community, there is MNIST dataset \(^{1}\) to measure the accuracy of character recognition. This dataset has become a standard benchmark for learning, and consists of 70,000 images of numbers 0 to 9. Multiple research groups have published accuracies above 99.7% \(^{2-7}\).

When 10 types of product numbers as shown in Fig. 4 are prepared and learned by CNN, the accuracy is 98.0% \(^{8}\). It was found that when the image of a product number was mistakenly recognized, the background color of the product number of the correct answer is the same as the background color of the incorrectly determined product number. Fig. 5 shows that images of product numbers which are likely to be misidentified each other. The cause of this problem is considered to be learning not only the character shape but also the background color when learning with CNN.

3. Proposed Method

We propose a new processing method for training dataset to learn only the character shape when training a character image dataset of color image with CNN. The processing method is shown in Fig. 6.

First, a foreground image is created by photographing a character with a single color such as green as background and transmitting only the character. Next, images of various textures are prepared, and these background images and the created foreground image are synthesized. Finally, the synthesized images are learned by CNN as a training dataset.
4. Experiments

In this research, we prepare ten kinds of product number shown as Fig.7. The characters of product number are written on transparent sheets wrapping products with numbers and indexed names of products. In the two experiments we conduct this time, we prepare training dataset of 20,000 images and validation dataset of 1,000 images. In Experiment 1, the images as shown in Fig.6 is used to learn the network as it is. In Experiment 2, we create from the images shown as Fig.8 the images shown as Fig.7 with proposed method, and they are used to learn the network.

4.1 Experiment1

In Experiment 1, we use the original images as shown in Fig.7 as a training dataset without processing. When we measure the accuracy of the model learned using 1,000 evaluation images, the accuracy is 98.0%.

4.2 Experiment2

In Experiment 2, we use the images created by the proposed method as shown in Fig. 8 as a training data set. When the accuracy of the learned model is measured using 1,000 evaluation images of Experiment 1, the accuracy is 99.8%.

5. Conclusions

In this paper, we verify the preprocessing method for a training dataset to improve character recognition accuracy. In recognition of the character extracted from the natural scene, when learning an image of a character by CNN, not only the shape of the character but also color information is learned. Therefore, we aim to improve the accuracy by applying the proposed preprocessing to the training dataset.

In Experiment 1, when the character image is learned without processing, the accuracy is 98.0%, but in many cases, it is erroneously recognized as a character having the same color. In Experiment 2, when the preprocessing of the proposed method is added to the image used in Experiment 1, the accuracy is 99.8%, and there is no misrecognition as in Experiment 1. From the two experimental results, it is considered that only the character shape can be learned by preprocessing of the proposed method.

We were able to make the product identification system, which is not influenced by color and the distortion of the character, but improvement for the practical use is necessary because a photography condition is severe. It is necessary to examine the most suitable learning method to raise versatility as a future problem.

References


(8) T. Akase, H. Nakao, L. Zhang, “Image classification system not affected by background and text color”, ICIAE, 2018