Image classification system not affected by background and text color

Taku Akase*, Hitoshi Nakao, Lifeng Zhang

Kyusyu Institute of Technology, Kitakyusyu, 804-8550, Japan

*Corresponding Author: zhang@ele.kyutech.ac.jp

Abstract

In recent years, along with the growth of international transactions and EC transactions, the inspection work becomes more complicated due to the diversification of parcels and the increase in small-lot cargo. Therefore, the efficiency of shipping inspection process and prevention of shipping error are required. Currently, a classification system using image processing is introduced for inspection work. However, in existing systems, black characters with a white background can be read correctly, but for non-uniformed background and colored character, the accuracy of reading characters is low. In this research, we develop a system that classifies products from the package images automatically using machine learning, without being influenced by background and text color. The final goal of this research is to improve this accuracy for practical use.

Keywords: Deep learning, classification system.

1. Introduction

In the logistics field, product inspection is a critical procedure before put in or take out a product from the warehouse. As the human error cannot be avoided completely, an automated inspection system using the computer image processing is introduced in this work. The system which distinguishes a product by reading the package on where the product name is printed. However, it may not read a letter depending on the combination of background colors and text colors. Therefore, the system which could distinguish a product regardless of the combination of background colors and text colors is necessary.

In this study, what is the suitable input data for machine learning is examined for making the product inspection system using the machine learning. The actual product

packages are used in the experiment, and several combinations of background colors and text colors are prepared as a simulation. Furthermore, how does the precision of identification changed by applying preprocessing to the images learned were examined.

The rest of paper is organized as follows. Section 2 shows the principle of machine learning. Section 3 depicts and explains a flow of the processing to make an identification system. Section 4 shows the experiments, and their results discuss them. Finally, Section 5 concludes this paper and indicates the future work.

2. Principle

2.1 Convolutional Neural Network

Deep Learning is the machine learning algorithm using the deep neural network (DNN), which is used here as a machine identify the product form the image data automatically. The algorithm that assumed a human being and the cranial nerve circuit of the animal which were designed to do the pattern recognition called the neural network with DNN a model multi-layer structure.

The algorithm of machine learning and the deep learning suggested that input data were one-dimensional indirectly at the start. However, target data may not be one-dimensional all the time when they look at the real world. The most typical case is image data.

Seeing from the implementation, it could convert two-dimensional or more high-dimensional data into a one-dimensional array. However, it is better to construct a model which can use two-dimensional data without converting. A devised method is Convolutional Neural Network to solve this problem. This method is called CNN. Figure 1 shows the structure of basic CNN. In CNN, feature quantities are extracted from two-dimensional input data through a

convolutional layer and a pooling layer and are conveyed to a general multi-layer perceptron. This algorithm was suggested in reference to the structure of the human visual cortex.

Fig. 1. Structure of CNN

From Figure 1, the first step is the convolution. The layer extracts a characteristic by applying some filters for an image. The filter which this characteristic extraction is used for a kernel and call for plural characteristics to be a kernel with feature map. Figure 2 illustrates the convolution procedure of two Dimension data and kernel. A kernel slides in the image and returns the sum of the product which crossed a value and a kernel level of the image data to overlap with a kernel. It means that it can extract various characteristics by changing the value of the kernel here. CNN learns the weight of the kernel which is an appropriate value namely a parameter in own through learning algorithm and comes to be able to classify image data with very high precision finally.

Fig. 2. Convolution

Next step is the pooling. After downsampling of the image data with the pooling layer and was propagated from the layer; conversion to reduce the quantity of data. A downsampling is necessary because a network keeps movement unchangeableness. The technique used in the method of the downsampling most is max pooling. Figure 3 shows the summary.

Fig. 3. Max Pooling

With the max pooling layer, the input data are divided into the sub-dataset that there is no heap, and the maximum in each sub-data is output. Not only the movement unchangeableness of the network is continued holding, but also can in this way connect it with reduction of the computational complexity in a layer spreading out after the pooling layer. With the max pooling layer, the input data are divided into the sub-dataset that there is no heap, and the maximum in each sub-data is output. Not only the movement unchangeableness of the network is continued holding, but also can in this way connect it with reduction of the computational complexity in a layer spreading out after the pooling layer.

3. Structure

Figure 4 shows a flow of the handling of machine learning that we used in a study. In the learning phase, we make data set for learning from the set of the raw data that want to let firstly you learn it. Here, we may process data for optimization to learning. We perform depths learning using data set for learning that we prepared for next. We used algorithm called CNN this time. An identification result is output when we input the new data which we want to distinguish into the model that have been learned that we made in the recognition phase. In this study, we inspected what kind of difference produced it in identification precision by changing data set for learning.

Fig. 4. Structure of proposal technique

4. Experiment

The examination conditions assumed in this research are written on transparent sheets wrapping products with

numbers and indexed names of products. It is assumed that the combination of the content and the color of the print is decided in advance. Prepare images and correct labels, machine learning, and create learned models. The test image is input to the learned model created earlier, and the accuracy is calculated by correcting the output.

We experimented on two this time. The product assumes it ten kinds in both experiments. Figure 5 shows the image of the product which we used. We predict which product the image of the product which we input fits in these ten kinds. We compared a correct answer with the output which it predicted to measure precision.

Fig. 5. 10 kinds of product

4.1 Experiment 1

As the first experiment, we prepare about 2,500 pieces of images for learning and 10 pieces of images for evaluations per one kind of product. An angle to photograph didn't have the limit and took a photo from various directions. We resize the images to three sizes like a figure 7 and learn them.

Fig. 6. Image for learning in Ex1

4.2 Experiment 2

As second experiment, we use the same number of images for learning and add a photography condition. The direction to photograph makes only the front and suppresses the reflection of the light. We resize the images to 108*192 and learn them.

5. Conclusions

The correct answer rate became like table 1.

Table 1. Result of experiment 1

Image size	28x28	64x64	96x96
Accuracy	32%	38%	22%

We expected that the accuracy rose as increasing the image size, but the accuracy didn't go up it like table 1. We think that it could not extract the feature of the image as a cause well.

As the result of experiment 2, the accuracy became 98.407% to improve more drastically than experiment 1.

We experimented on two in this study. As a result, it may be said that we were able to improve precision by setting conditions on a reflection of the light and an angle to photograph when we photograph an image for learning. At the experiment 1, too much information was not able to extract feature quantity because there is some reflection of the light in the image. In addition, because there were many things which the combination of background color and text color is same of correct product when it misrecognized, we think that it recognized in the difference in color of the image. It's thought that it cannot extract it to the characteristic of the letter because the precision didn't improve even though we increase image size and do a letter clearly.

At the experiment 2, we photographed an image for learning based on a result of experiment 1 once again. As a result, the recognition was improved to approximately 98%. It is thought that only a product looked that we got rid of a reflection of the light as much as possible to you as an animation with the camera of the smartphone and use what we took out of the animation every frame as an image for learning. By this experiment, we don't the chroma of the image. A lot of function is prepared in TensorFlow which is the machine learning library. It is expected that recognizable environmental width opens exactly by processing it into an image for learning.

From these experiments, the identification system was easy to be influenced by the environment that photographed an image, but we know that we could realize a highly precise identification device by attaching a photography condition.

We were able to make the product identification system which is not influenced by a color and the distortion of the letter, but improvement for the practical use is necessary because a photography condition is severe. Using other depths learning methods, it is necessary to examine the most

suitable learning method to raise versatility as a future problem.

References

- (1) Yusuke Sugomori : “Deep Learning Java programing A theory and implementation of the depths Learning”, Impress Corporation, pp. 121-131, 2016
- (2) “About overlearning and Dropout”,
<<http://sonickun.hatenablog.com/entry/2016/07/18/191656>>
- (3) Ryo Kuratsume : “-CVIM tutorial series- Computer vision the most advanced guide 1”, Advanced Communication Media Corporation, pp. 139-153, 2009
- (4) TensorFlow : “TensorFlow Official Page”,
< <https://www.TensorFlow.org/>>