Journal of Visual Communication and Image Representation 55 (2018) 640–647

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

journal homepage: www.elsevier.com/locate/jvci

Small sample image recognition using improved Convolutional Neural Network☆

Jiajia Zhang*, Kun Shao, Xing Luo

School of Computer and Information, Hefei University of Technology, Hefei, China

A R T I C L E I N F O

Keywords:
Image recognition
Convolutional Neural Network (CNN)
General Regression Neural Network (GRNN)
Small sample
Real-time

A B S T R A C T

In recent years, with the raise of the neural network and deep learning, significant progress has been achieved in the field of image recognition. Convolutional Neural Network (CNN) has been widely used in multiple image recognition tasks, but the recognition accuracy still has a lot of room for improvement. In this paper, we proposed a hybrid model CNN-GRNN to improve recognition accuracy. The model uses CNN to extract multilayer image representation and it uses General Regression Neural Network (GRNN) to classify image using the extracted feature. The CNN-GRNN model replace Back propagation (BP) neural network inside CNN with GRNN to improve generalization and robustness of CNN. Furthermore, we validate our model on the Oxford-IIIT Pet Dataset database and the Keck Gesture Dataset, the experiment result indicate that our model is superior to Gray Level Co-occurrency (GLCM),HU invariant moments, CNN and CNN_SVM on small sample dataset. Our model has favorable real-time characteristic at the same time.

1. Introduction

With the development of science and technology, image processing technology has improved a lot [1–6]. A lot of works has been done about image processing [7–9]. Image recognition is an important field of artificial intelligence. As the development of image processing, image recognition technology has been gradually applied in many fields [10–14].At the same time, the accuracy, reliability and real time requirements of image recognition are becoming stricter. Recent years, as the development of deep learning, CNN based on it has been used in many field about image processing. Owing to the close connection between the layers of CNN and the sufficient space information of CNN, CNN can work well on image processing and understanding task. CNN can even extract rich correlated features automatically from images. On account of above features of CNN, it has achieved excellent results in all kinds of image recognition tasks such as face recognition, eye detection and pedestrian detection [15–17].

Though CNN has achieved big success in image recognition, it still has its own limitations. The BP neural network inside CNN model is too simple, so it needs multiple iterations with a large number of training samples. In other words, it can learn the image representation well only when it has enough training data and iterate sufficient times. The BP neural network adopts the descending gradient training method, which make the model converge slowly and it easily come to the local optimization, affecting the final recognition accuracy.

So we propose a new hybrid model CNN-GRNN, it can get excellent performance even with small sample. GRNN can get ideal recognition result even it do not has enough feature and it do not need iteration. Our method on image recognition consists of two parts: 1) In the training time, the CNN-GRNN model use CNN to extract the image representation, and then it let the full connected layer to do the prediction work. 2) In the testing time, CNN is responsible for the representation extraction task as in the training time, and then GRNN will classify the image using the extracted feature, which is different from the training. This model aims to establish relevance between the image representations and objective prediction result.

2. Related work

In this part, we will introduce some previous methods that related to our current work.

Fu et al. put forward a traffic signs detection method that uses Hu invariant moments to do eigenvalues extraction and traffic signs detection [18].The method is rapid and reliable with a high recognition rate. But the representation it extracted is low dimension feature and without hierarchy information. So the method is restricted to simple detection and recognition work which do not need more image information.

☆ This paper has been recommended for acceptance by Luming Zhang.

* Corresponding author.
E-mail address: 1293742589@qq.com (J. Zhang).

https://doi.org/10.1016/j.jvcir.2018.07.011
Received 9 May 2018; Received in revised form 20 June 2018; Accepted 19 July 2018
Available online 21 July 2018
1047-3203/

Shen et al. proposed a method that uses Gray Level Co-occurrency (GLCM) to do splice image forgery detection [19]. It uses Gray Level Co-occurrency (GLCM) to extract the texture information of image and use the learned representations to make recognition. The method is superior to many methods which only extracts low dimension feature vector from images. Owing to the method can only extract low dimension feature manually as the Hu invariant moments, the method is to be improved by combining with other methods.

Kunihiko Fukushima proposed a method that uses a neural network model for a mechanism of visual pattern recognition [20]. The network is named "neocognitron". It is nearly the first time that people put forward Convolution Neural Network in its neurocognitive mechanism. The experiment result shows the network works well without any instructions about categories.

Le Cun et al. proposed a method that uses the error gradient to train CNN and it achieved good results [21]. The method uses the CNN model to extract the image presentation and utilized the BP neural network classifier inside the CNN to classify the image. The method is almost the first time that uses the network itself learn the image features only with image labels. As it turns out, the learning ability of the CNN is powerful.

As the BP neural network inside CNN has some limitations, Niu Xiaoxiao and Suen followed proposed the CNN-SVM model for the recognition of handwritten digitals [22]. It replaces the BP neural network classifier with Support Vecto Machine (SVM) in CNN model, and the recognition rate can reach 99.81%. It uses CNN to extract image feature and then the Support Vector Machine (SVM) will train with the extracted features, the well-trained SVM will classify image at last. The method uses SVM to improve the recognition accuracy of CNN, but it only did experiment on simple handwritten digitals. The author did not do experiment on punctuation mark, let alone other handwritten characters and usual images.

Heliang et al. proposed a novel part learning approach by a multi-attention convolutional neural network (MA-CNN) to do fine-gained image recognition. It focuses on the fact that part localization and fine-grained feature learning. Extensive experiments in the paper demonstrate the superior performance on both multiple-part localization and fine-grained recognition on birds, aircrafts and cars [23].

Lowe D G proposed a method that combine CNN-SVM with Principal component analysis (PCA) [24]. The method is aimed to improve the performance of SVM, and it has achieved good performance. It is not only restricted to handwritten digitals anymore and it works well in texture image classification and image recognition.

The above models or methods have all achieved good results in image recognition. Inspired by above methods, we put forward a hybrid model, CNN-GRNN.

3. The proposed method

We present a novel CNN-GRNN model that uses a simple CNN model for image feature extraction and employ the GRNN model to do classification. The method replaces the BP Neural network with better performanced network GRNN when it still using CNN as feature extractor. The GRNN has a better function approximation capability, and the network has only one variable, which is also superior to BP in network training. At the same time, the network do not need to iterate and it can work with small sample database, which is also superior to the BP neural network in application. With the strategy of this, the identification precision can be improved and the application scope of the convolution neural network can be expanded compared with other method in image recognition.

3.1. 1Structure of the model

Overall, the whole model is composed of CNN feature extractor and GRNN classifier. The procedure of training is as follows (Fig. 1). First, We feed the sample image into the input layer of CNN-GRNN, after multiple convolution and down-sampling, a number of feature images can be obtained. Then, the model stretch the feature image into a column vector, which is the eigenvector extracted from the sample image. At the same time retain the output layer which is full connected with the feature vector for the training of CNN feature extractor. At the last, the classifier outputs the final classification result according to the feature vector.

3.2. CNN

As we can see in Fig. 1, the CNN act as a representation extractor in our proposed method. The CNN is trained end-to-end with gradient descent. The training process is as follows (Fig. 2).

First initialize weights and biases in all convolution kernels. Through the forward propagation with the training set, the output O can be obtained. Then the CNN can learn the error E through comparing the output O with the labels y. Assume the number of the sample set is N, and the number of the sample types are c. We can calculate the error E according to the Eq. (1).

$$E^N = \frac{1}{2}\|y^N - O^N\|^2 = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{c}(y_k^n - O_k^n)^2 \tag{1}$$

Finally, the CNN judge the model converges or not according to the value E. If it converge, then the training is completed. If not, the residual δ of the output is calculated. Given the activation function f, we can get the residual from the Eq. (2). We use the sigmoid function as the activation function in the experiment.

$$\delta^L = \frac{\delta}{\delta z_n^L}\frac{1}{2}\|y^N - O^N\|^2 = -(y^N - O^N)\cdot f'(z^L) \tag{2}$$

The residues are passed from the output layer to the front layer, then we can calculate the residual in every layer (Eq. (3)), δ^l is the residual of the lth layer.

$$\delta^{(l)} = ((W^{(l)})^T\delta^{(l+1)})\cdot f'(z^{(l)}) \tag{3}$$

Update the weights and bias in each layer with the learning rate α (Eq. (4))

$$\begin{aligned} W^{(l)} &= W^{(l)} - \alpha\frac{\partial}{\partial W^{(l)}}E(\omega, b) = W^{(l)} - \alpha\cdot\delta^{(l+1)}(a^{(l)})^T \\ b_i^{(l)} &= b_i^l - \alpha\cdot\frac{\partial}{\partial b_i^{(l)}}E(\omega, b) = b_i^{(l)} - \alpha\cdot\delta^{(l+1)} \end{aligned} \tag{4}$$

The CNN execute the above process until it gets the ideal result we want. And the CNN send the feature vector into the GRNN for validation in the procedure of training. But we should know that the GRNN act as a classifier in the procedure of testing, which is different from what in the training process.

3.3. GRNN

The GRNN we used as the image classifier in the model was proposed by Specht [25], it is transformed from artificial neural network and it is a nonlinear regression neural network based on the parameter estimation[26]. The GRNN not noly has a strong ability of nonlinear mapping and generalization, but it can be used with small sample database. In recent years, the GRNN has achieved better performance than RBF network and BP network on prediction task.

The structure of GRNN is shown as follows (Fig. 3). The classification result can be got without iteration. The GRNN consists of four layers, which are the input layer, the pattern layer, the summation layer and the output layer. The input layer of GRNN is full connected with the pattern layer and the number of neurons in the input layer is equal to the dimension of the learned instances $(X_i;Y_i)_{i=1}^m$. Each neuron in the pattern layer has its own training mode. The output of the pattern layer