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

ARTICLE INFO

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

ABSTRACT

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/ © 2018 Elsevier Inc. All rights reserved.

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

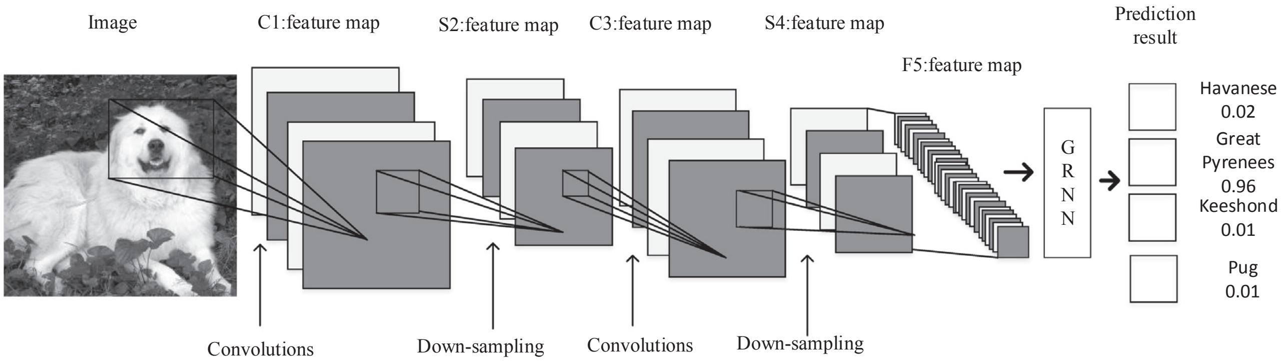

Fig. 1. Schematic illustration of the CNN-GRNN. The CNN extracted the image representation through convolution and down-sampling, and then the GRNN use the extracted feature to make prediction.

is to magnanimity the difference between the input representation and the stored feature. The summation layer has only two neurons, one calculate the total output of all neurons in the pattern layer and another calculate weighted summation of all neurons in the pattern layer. The first one act as numerator in computing while the second neuron act as denominator in computing. The summation layer is full connected with the upper node as the pattern layer. The final prediction result of the output layer is calculated by the two output of the summation.

Given the input representations are $X=(x_1,x_2,...,x_n)^T$, then the output of GRNN can be calculated by Eq. (5).

$$\hat{Y}(X)=\frac{\sum_{i=1}^{n} Y_i e^{-t_i}}{\sum_{i=1}^{n} e^{-t_i}}$$
$$t_i=\frac{(X-X_i)^T(X-X_i)}{2\sigma^2} \qquad (5)$$

where σ is the spread parameter which can control the smooth of functional approximation. The σ can be estimated with large number of samples, but our experiment is devoted to small database and the estimated σ may not work. In the experiment part, we will adjust the σ according to [23] to get better performance.

3.4. Processing strategy

This part is the processing strategy about CNN-GRNN. (Fig. 4) First, the training images are sent to the CNN. After several epochs of back-propagation, the extracted CNN feature accompanied with image labels are sent into GRNN for the training. At last the well-prepared CNN-GRNN can be got through the procedure. The testing images can be sent into the well-prepared CNN-GRNN for image recognition, and the final recognition result can be got.

4. Experiment

Here we test the performance of CNN-GRNN through the accuracy of recognition and the model time-consuming on the Oxford-IIIT Pet Dataset database and the Keck Gesture Dataset. The COIL-100 library is mainly used to test the effection on the image texture recognition of the model. The gesture database is used to test if the model effective for the recognition of shape and the edge information.

4.1. The settings of the experiment

The experiment was running on Matlab2013a platform, and the configuration of our computer is Win7 system (7 Intel(R)Cores(TM) of 2.8 GHz and 8 GB DDR4 RAM).

We will do our experiment on the Oxford-IIIT Pet Dataset and the Keck Gesture Dataset. Before doing the experiment, we will do some preprocessing work to guarantee the accuracy of our experiment [27,28].

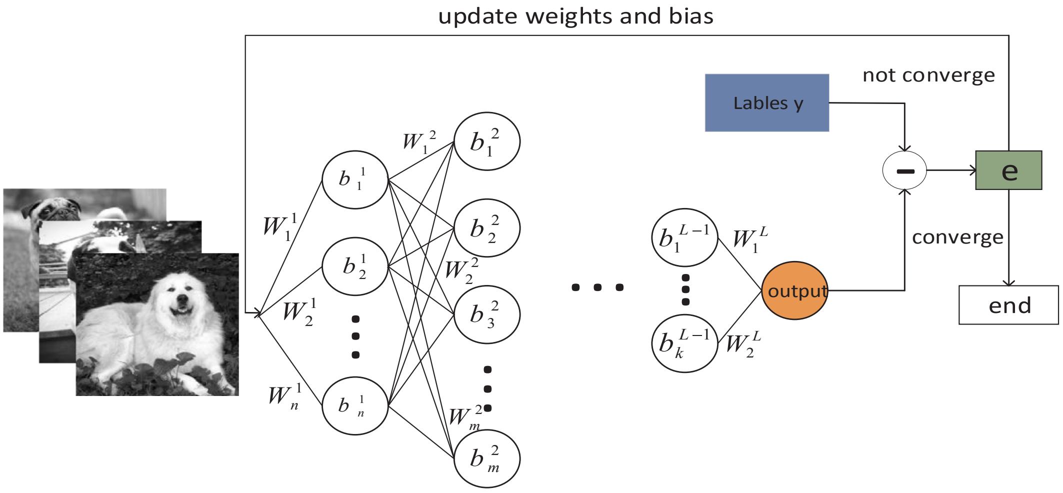

Fig. 2. The training procedure of the CNN. First initialize weights and biases in all convolution kernels. Through the forward propagation with the training set, the output O can be obtained. Then the CNN can learn the error E through comparing the output O with the labels y. the CNN judge the model 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. Update the weights and bias in each layer with the learning rate α.

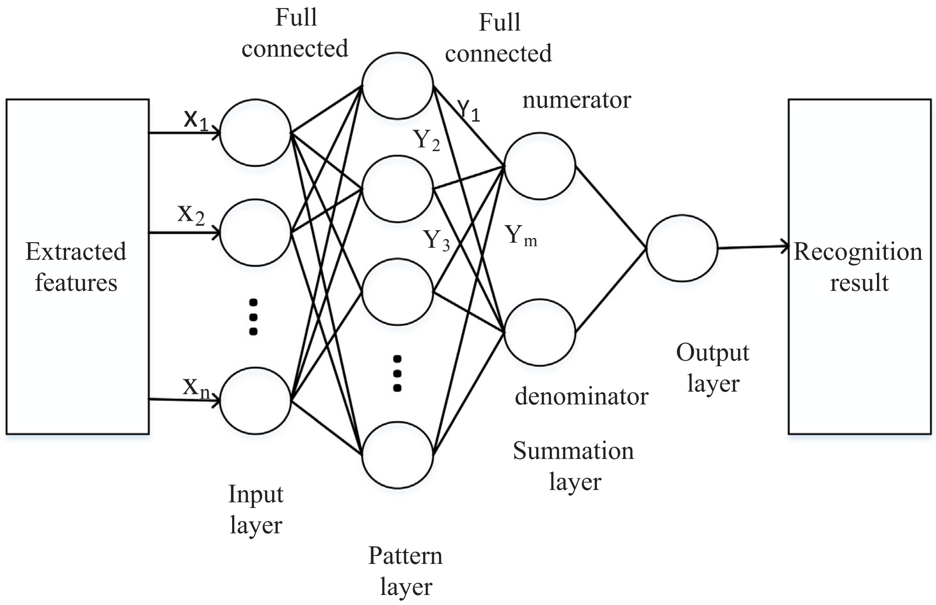

Fig. 3. The structure of GRNN for image classification.

We randomly select 600 samples from the Oxford-IIIT Pet Dataset, it contains 12 different dogs and each type contains 50 samples. Among them, 150 samples are used for training and 450 samples are used for testing.

Considering the limited processing capacity of the model, in order to get the excellent recognition result, we should do preprocessing on the image. We change the original RGB image into the gray image and resize the image into 80 × 80. There are several processed samples in Fig. 5.

We randomly select 700 samples from the Keck Gesture Dataset, it contains 14 different gestures and each gesture contains 50 images. Among them, 400 samples are used for training and 300 samples are used for testing.

In order to get better performance on image recognition, we not only change these images into gray image, but we employ binarization on these processed gray image. There are several processed samples in Fig. 6.

4.2. The experiment result

The model adopts the GRNN to classify the image and the GRNN has only one spread parameter σ, which will undoubtedly has a great

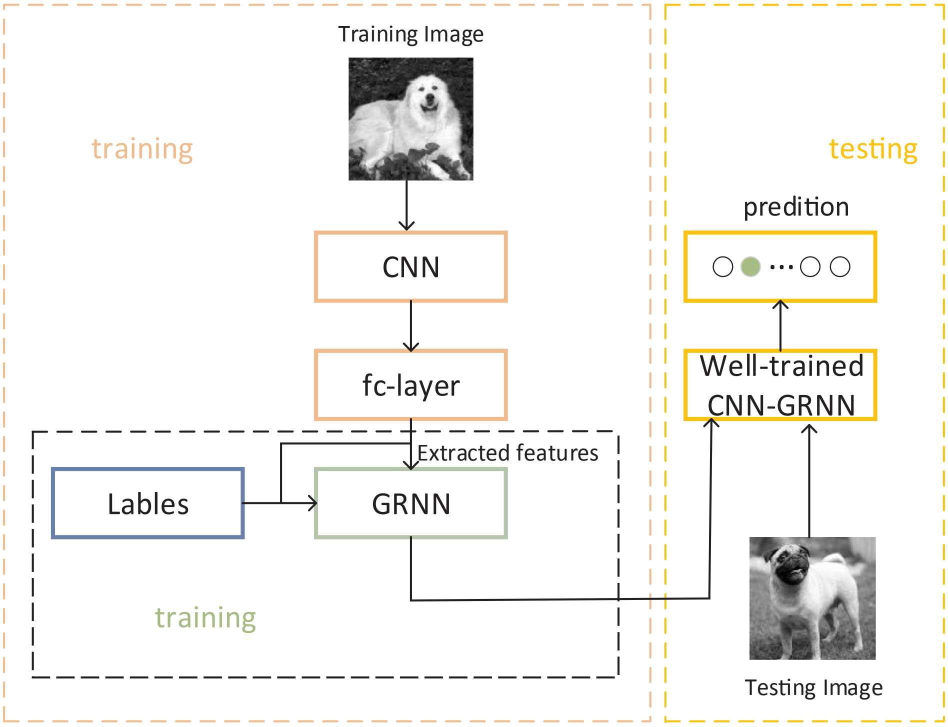

Fig. 4. The processing strategy about CNN-GRNN.

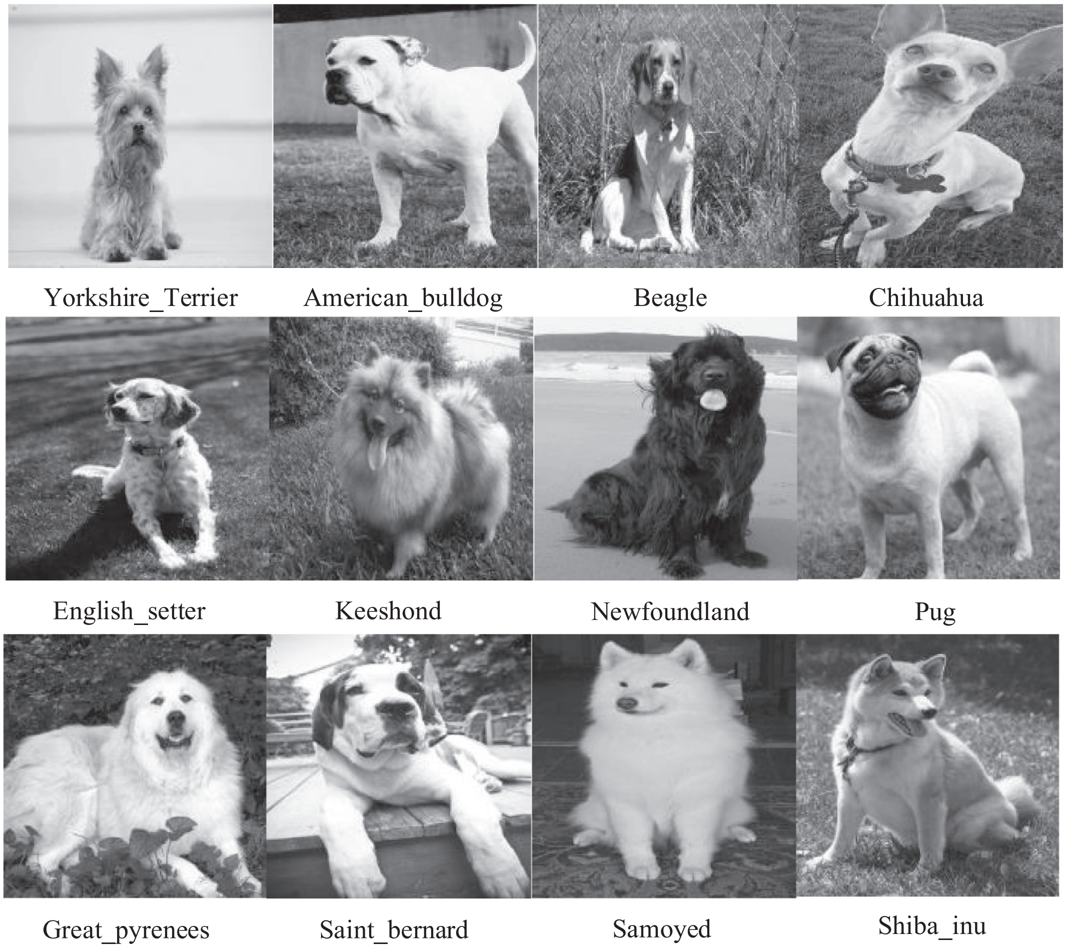

Fig. 5. Some processed samples in the Oxford-IIIT Pet Dataset. They were transformed from RGB image into gray image.

impact on the recognition accuracy. We first fix the iteration times into 100 and fix the learning rate into 0.5, and then do experiment on Oxford-IIIT Pet Dataset and the Keck Gesture Dataset with different spread parameter σ [29].The recognition accuracy can be seen in Fig. 7.As we can see in Fig. 7, the CNN-GRNN model can get the most excellent performance when the spread parameter σ is set to be 0.01.

The model uses the CNN as the representation extractor, so the iteration times of the CNN will affect the recognition accuracy as the spread parameter σ. In consideration of the above experiment result, so we set the spread parameter σ to 0.01, and then do experiment on the Oxford-IIIT Pet Dataset and the Keck Gesture Dataset with different iteration times. The recognition accuracy can be seen in Fig. 8.As we can see in Fig. 8(a), for the Oxford-IIIT Pet Dataset, the recognition accuracy tend to be stable when the iteration times come to 90. As we can see in Fig. 8(b), for the Keck Gesture Dataset, the recognition rate becomes stable when the iteration times reach 100.

The model uses the CNN as the representation extractor, so the iteration times of the CNN will affect the recognition accuracy as the spread parameter σ. In consideration of the above experiment result, so we set the spread parameter σ to 0.01, and then do experiment on the Oxford-IIIT Pet Dataset and the Keck Gesture Dataset with different iteration times. The recognition accuracy can be seen in Fig. 8. As we can see in Fig. 8(a), for the Oxford-IIIT Pet Dataset, the recognition accuracy tend to be stable when the iteration times come to 90. As we can see in Fig. 8(b), for the Keck Gesture Dataset, the recognition rate becomes stable when the iteration times reach 100.

Besides upside experiment, we also did experiment using Gray Level Co-occurrency (GLCM) and HU invariant moments combined with GRNN to do recognition on the Oxford-IIIT Pet Dataset and the Keck Gesture Dataset. GLCM and HU invariant moments extract image feature manually. GLCM extract contrast, energy, entropy and correlation for recognition. HU invariant moments construct 7 invariant moments using the first normalization central moment and the second normalization central moment and it uses the invariant moment combined with image labels to recognize image. We set the spread parameter σ to 0.01 and we set the iterate times to 100.

The recognition accuracy using above methods on the Oxford-IIIT Pet Dataset and the Keck Gesture Dataset is shown in Table 1. As we can see in Table 1, our method is superior to other method in image recognition. And the method using CNN extractor all achieved better performance than the method using other feature extractor.

The time consumed on training using different methods is shown in Table 2. Considering both recognition accuracy and time consumed in training, our method is superior to other method.

The time consumed in real time recognition using above methods on the Oxford-IIIT Pet Dataset and the Keck Gesture Dataset is shown in Table 3. As we can see in Table 3, our model costs less time than other methods.

As we can see in above tables, CNN-GRNN outperforms other methods on recognition accuracy and the time consumed on training and testing.

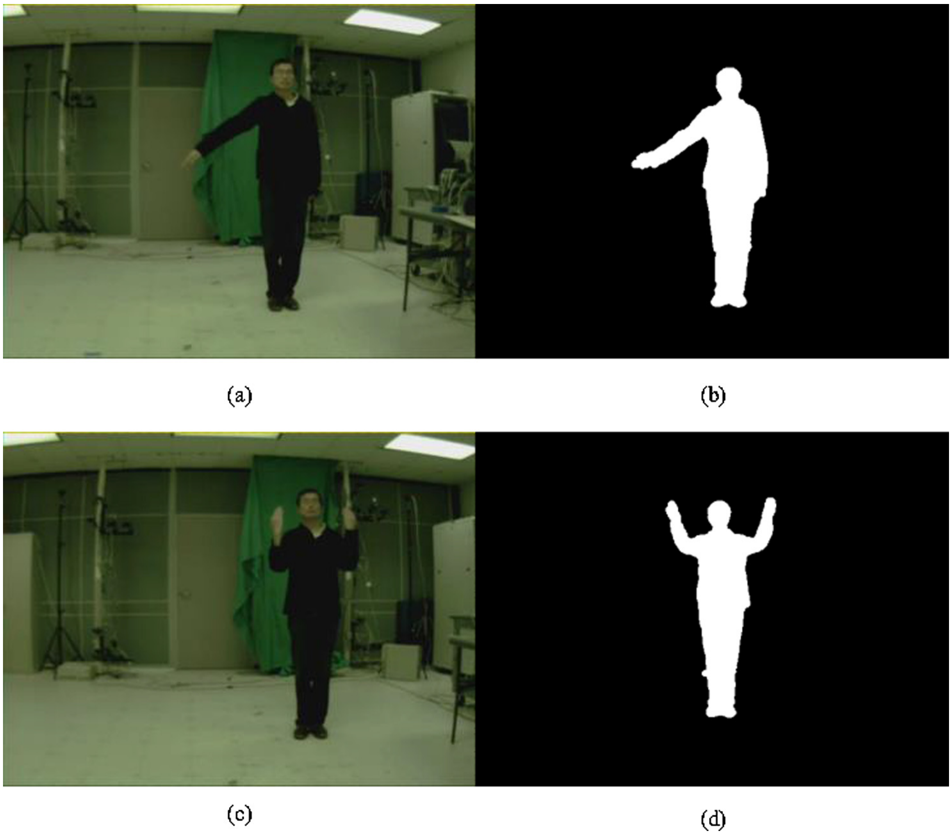

Fig. 6. Some processed samples in the Keck Gesture Dataset. First, we transform these RGB images into gray image. And then we employ binarization on these gray images. (a) and (c) is the original image, image (b) and (d) is binary image.

4.3. Result analysis

In this part, we will analyze above experiment result in two aspects, model recognition accuracy and model time consumed on training and testing respectively.

CNN can learn image multilayer high-dimensional feature itself, but the representations GLCM and HU invariant moments extracted is single-layer low dimensional feature, which cannot fully represent image information necessary for image recognition. Therefore, CNN exactor is superior to GLCM and HU invariant moments in model recognition accuracy. Due to performance deficiency of CNN, the recognition accuracy of it is inferior to CNN-SVM model and our model. GRNN and SVM are superior to the BP neural network in the classification and generalization task. And the SVM and GRNN are suitable for small sample dataset. But the SVM is a typical purpose two class classifier and its essence is to find an optimal hyper plane to maximize

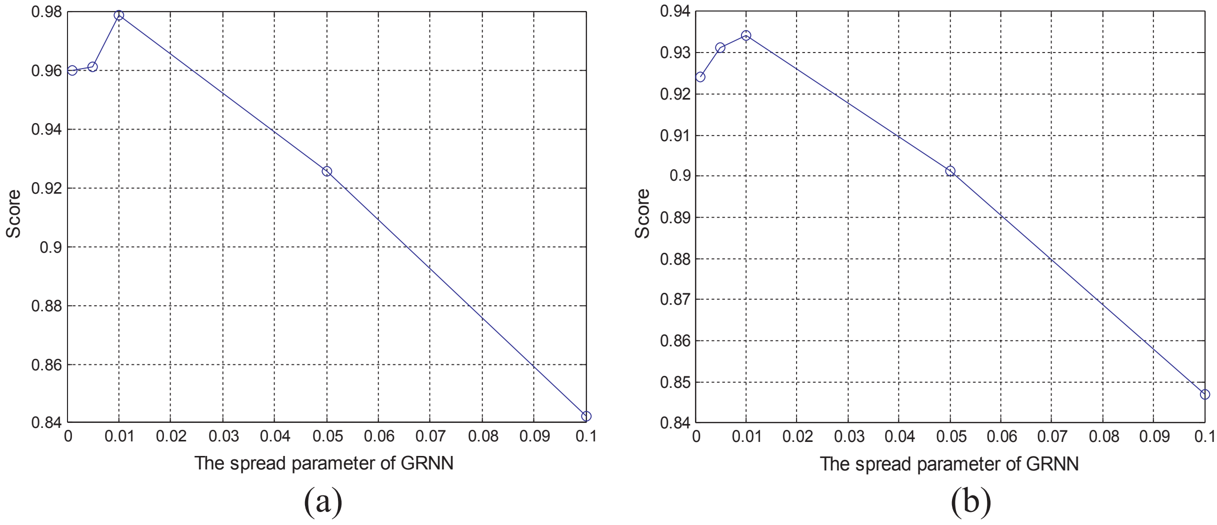

Fig. 7. The recognition accuracy of our model under different spread parameter when testing on (a) the Oxford-IIIT Pet Dataset and (b) the Keck Gesture Dataset.

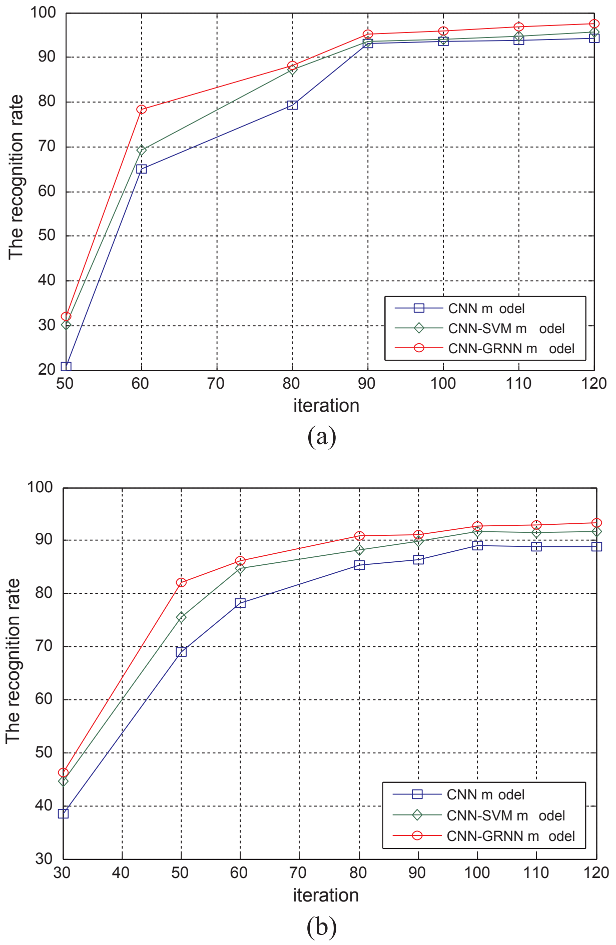

Fig. 8. The recognition accuracy using our model versus the CNN and CNN-SVM model under different iterations times on (a) the Oxford-IIIT Pet Dataset and (b) the Keck Gesture Dataset.

Table 1
The recognition accuracy comparison on two datasets.

Method	The Oxford-IIIT Pet Dataset	The Keck Gesture Dataset
GLCM + GRNN	0.5610	–
HU + GRNN	–	0.8000
CNN	0.9450	0.8887
CNN-SVM	0.9570	0.9172
CNN-GRNN	0.9720	0.9315

Table 2
The time consumed on training on two datasets using different methods.

Method	The Oxford-IIIT Pet Dataset	The Keck Gesture Dataset
GLCM + GRNN	0.042 s	–
HU + GRNN	–	0.048 s
CNN-SVM	1.885 s	1.589 s
CNN-GRNN	1.301 s	0.761 s

Table 3
The time consumed on real-time recognition comparison on two datasets.

Method	The Oxford-IIIT Pet Dataset	The Keck Gesture Dataset
GLCM + GRNN	1.54 s	–
HU + GRNN	–	1.16 s
CNN	1.15 s	0.75 s
CNN-SVM	1.20 s	0.78 s
CNN-GRNN	1.13 s	0.76 s

margin between two classes. Even if we can solve the problem through using multiple SVM classifiers, but the recognition accuracy will not be permitted when two classes has same scores. GRNN based on artificial neural network has advantages in multi-classification and it has better generalization and robustness.

As shown in Tables 2 and 3, the time consumed on training of CNN-SVM and CNN-GRNN with CNN feature extractors is a bit more than that of traditional method, because the extracted feature is more complex. But the time consumed on online recognition of CNN, CNN-SVM and CNN-GRNN with CNN feature extractors is less than that of traditional image recognition method, which guarantees the real-time performance of the model. The result benefits from the unique weight sharing mechanism and the unique local sensing field of CNN. Among these methods, GRNN achieved best performance. Because GRNN calculate regression values according to the extracted feature directly from the perspective of probability density function without iteration.

5. Summary

In this paper, we had introduced a hybrid model based on CNN for image classification. As CNN can extract features from images, we then use General Regression Neural Network with powerful function approximation to recognize image according to extracted representations. CNN can learn image multilayer high-dimensional feature itself which is superior to other traditional methods. And General Regression Neural Network we used can strengthen the classification and processing capacity of CNN, and it can make the model converging more quickly. Experiments show that our model is effective on image recognition. However, CNN model contains a large number of weights which leads to a large number of iterations, so it is time-consuming. Therefore, we will pay attention to optimize the training process of our model in the future.

Conflict of interest

There is no conflict of interest.

Acknowledgement

This work is partially supported by National Natural Science Foundation of China (61672201).

References

[1] Richang Hong, Luming Zhang, Dacheng Tao, Unified photo enhancement by discovering aesthetic communities from Flickr, IEEE Trans. Image Process. 25 (3) (2016) 1124–1135.

[2] Richang Hong, Yang Yang, Meng Wang, Xian-Sheng Hua, Learning visual semantic relationships for efficient visual retrieval, IEEE Trans. Big Data 1 (4) (2015) 152–161.

[3] Richang Hong, Zhenzhen Hu, Ruxin Wang, Meng Wang, Dacheng Tao, Multi-view object retrieval via multi-scale topic models, IEEE Trans. Image Process. 25 (12) (2016) 5814–5827.

[4] Luming Zhang, Yahong Han, Yi Yang, Mingli Song, Shuicheng Yan, Qi Tian, Discovering discriminative graphlets for aerial image categories recognition, IEEE Trans. Image Process. (T-IP) 22 (12) (2013) 5071–5084 (IF:3.199, CCF A, JCR 2).

[5] Luming Zhang, Mingli Song, Qi Zhao, Xiao Liu, Jiajun Bu, Chun Chen, Probabilistic Graphlet Transfer for Photo Cropping, IEEE Trans. Image Process. (T-IP) 21 (5) (2013) 803–815 (IF:3.199, CCF A, JCR 2).

[6] Luming Zhang, Yue Gao, Rongrong Ji, Qionghai Dai, Xuelong Li, Actively learning human gaze shifting paths for photo cropping, IEEE Trans. Image Process. (T-IP), 23 (5) (2014) 2235–2245 (IF:3.199, CCF A, JCR 2).

[7] Luming Zhang, Yue Gao, Roger Zimmermann, Qi Tian, Xuelong Li, Fusion of Multi-Channel Local and Global Structural Cues for Photo Aesthetics Evaluation, IEEE Trans. Image Process. (T-IP) 23 (3) (2014) 1419–1429 (IF:3.199, CCF A, JCR 2).

[8] Luming Zhang, Yi Yang, Yue Gao, Changbo Wang, Yi Yu, Xuelong Li, A probabilistic associative model for segmenting weakly-supervised images, IEEE Trans. Image Process (T-IP) 23 (9) (2014) 4150–4159 (IF:3.199, CCF A, JCR 2).

[9] Luming Zhang, Yingjie Xia, Rongrong Ji, Xuelong Li, Spatial-aware object-level saliency prediction by learning graphlet hierarchies, IEEE Trans. Ind. Electron. (T-

IE) 62 (2) (2015) 1301–1308 (IF: 5.165, JCR 1).

[10] Luming Zhang, Yue Gao, Yingjie Xia, Qionghai Dai, Xuelong Li, A fine-grained image categorization system by cellet-encoded spatial pyramid modeling, IEEE Trans. Ind. Electron. (T-IE) 62 (1) (2015) 564–571 (IF: 5.165, JCR 1).

[11] Luming Zhang, Yingjie Xia, Kuang Mao, Zhengyu Shan, An effective video summarization framework toward handheld devices, IEEE Trans. Ind. Electron. (T-IE) 62 (2) (2015) 1309–1316 (IF: 5.165, JCR 1).

[12] Luming Zhang, Yue Gao, Chaoqun Hong, Yinfu Feng, Jianke Zhu, Deng Cai, Feature correlation hypergraph: exploiting high-order potentials for multimodal recognition, IEEE Trans. Cybern. (T-CYB) 44 (8) (2014) 1408–1419 (IF:3.236, CCF B, JCR 1).

[13] Luming Zhang, Yue Gao, Rongrong Ji, Lv Ke, Jiale Shen, Representative discovery of structure cues for weakly-supervised image segmentation, IEEE Trans. Multimedia (T-MM) 16 (2) (2014) 470–479 (IF:1.754, CCF B, JCR 2).

[14] Luming Zhang, Mingli Song, Yi Yang, Qi Zhao, Zhao Chen, Nicu Sebe, Weakly supervised photo cropping, IEEE Trans. Multimedia (T-MM) 16 (1) (2014) 94–107 (IF:1.754, CCF B, JCR 2).

[15] Y. Chen, L. Zhang, X. Liu, et al., Pedestrian detection by learning a mixture mask model and its implementation, Inf. Sci. Int. J. 372 (C) (2016) 148–161.

[16] F.H.C. Tivive, A. Bouzerdoum, A new class of convolutional neural networks (SICoNNets) and their application of face detection, International Joint Conference on Neural Networks, vol. 3., IEEE, 2003, pp. 2157–2162.

[17] F.H.C. Tivive, A. Bouzerdown, An eye feature detector based on convolutional neural network, Eighth International Symposium on Signal Processing and ITS Applications, IEEE, 2006, pp. 90–93.

[18] M.Y. Fu, F.Y. Liu, Y. Yang, et al., Background pixels mutation detection and Hu invariant moments based traffic signs detection on autonomous vehicles Control Conference. IEEE, 2014:670-674.

[19] X. Shen, Z. Shi, H. Chen, Splicing image forgery detection using textural features based on the grey level co-occurrence matrices, IET Image Proc. 11 (1) (2017) 44–53.

[20] K. Fukushima, Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol. Cybern. 36 (4) (1980) 193–202.

[21] Y. Le Cun, B. Boser, J.S. Denker, et al., Backpropagation to Handwritten Code Applied Zip recognition, Neural Comput. 1 (4) (1989) 541–551.

[22] Niu Xiaoxiao, C.Y. Suen, A novel hybrid CNN-SVM classifier for recognizing handwritten digits, Pattern Recogn. 45 (4) (2012) 1318–1325.

[23] H. Zheng, J. Fu, T. Mei, J. Luo, Learning multi-attentionconvolutional neural network for fine-grained image recognition, in: The IEEE International Conference on Computer Vision (ICCV), Oct 2017. 2, 3.

[24] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2) (2004) 91–110.

[25] Shaode Yu, Fan Jiang1, Leida Li, Yaoqin Xie. CNN-GRNN for Image Sharpness Assessment. ACCV 2016 Workshops, Part I, LNCS 10116, pp. 50–61, 2017.

[26] C. Li, A.C. Bovik, X. Wu, Blind image quality assessment using a general regression neural network, IEEE Trans. Neural Netw. 22 (5) (2011) 793–799.

[27] Y. Chen, D. Pan, Y. Pan, et al., Indoor scene understanding via monocular RGB-D images, Inf. Sci. 320 (C) (2015) 361–371.

[28] Y. Chen, T.V. Nguyen, M. Kankanhalli, et al., Audio matters in visual attention, IEEE Trans. Circuits Syst. Video Technol. 24 (11) (2014) 1992–2003.

[29] D.F. Speeht, A general regression neural network, IEEE Trans. Neural Networks 2 (6) (1991) 568–576.