Research on Face Feature Recognition Based on Deep Learning

Shengchao Yang

Shanghai Foreign Language School Affiliated to SISU, China.

Abstract

Convolutional neural network has a good effect in face recognition research, but the extracted face features ignore the local structural features of the face. To solve this problem, a face recognition method based on deep learning and feature fusion is proposed. This algorithm combines the local binary pattern information with the original image information as the input of SDFVGG network, which makes the extracted facial features richer and more representational. Among them, SDFVGG network is a network that combines the deep and shallow features of VGG network. Experiments on CAS-PEAL-R1 face database show that it is very effective to fuse the deep and shallow features of network with the feature information of LBP image and original image in convolution neural network to improve the accuracy of face recognition, and the highest face recognition rate of 98.58% can be obtained, which is better than traditional algorithm and general convolution neural network.

Keywords

Neural Network; Face Recognition; SDFVGG; Feature Extraction; Deep Learning.

1. Introduction

Face recognition is a kind of biometric recognition technology which is based on human face feature information. It has the advantages of good anti-counterfeiting performance and non-invasive. In recent years, face recognition has become a research hotspot in pattern recognition, image processing, machine vision and neural network. It has important research value in national defense security, identity authentication, video surveillance, Internet interaction and other fields. Traditional face recognition process includes four stages: face detection, face alignment, face feature extraction and face classification. Face feature extraction is the key to face recognition, and the quality of feature extraction directly affects the accuracy of classification. In traditional feature extraction methods, Local Binary Pattem (LBP) is an operator used to describe local texture features of images. Because of its simple calculation and strong ability of feature classification, LBP is widely used in face recognition research [1-3]. However, in the unrestricted environment, because of the complexity of face images, traditional feature extraction methods cannot achieve the desired results, and the expression of features relies too much on manual selection.

In recent years, in-depth learning has attracted more and more researchers' attention. Compared with shallow model, it has obvious advantages in feature extraction. Deep learning is a hierarchical machine learning method including multi-level non-linear transformation. It combines low-level features to form more abstract and effective high-level representations, which have good generalization ability [4]. Among them, convolutional neural network is a classical and widely used method of in-depth learning, and its connections between neurons are inspired by the visual cortex of animals. The characteristics of convolutional neural network, such as local perception, weight sharing and pooling operation, make it closer to biological neural network, which can effectively reduce the complexity of network and model learning parameters. At the same time, the model has a certain degree of invariance to displacement, scaling, rotation or other forms of deformation, and has strong robustness and fault tolerance [5-6]. In face recognition tasks, compared with the features extracted by traditional methods, convolution neural network has more advantages in strong representation ability automatically learned through a series of operations such as convolution, activation function and pooling, and the recognition rate of authentication on LFW data sets has exceeded the recognition rate of human eyes [7-8]. However, the image features extracted by convolution neural network

ignore the local structure features of the image, and the network will learn unfavorable feature representation because of illumination and other factors. The traditional feature extraction method, LBP, is an operator used to describe the local texture features of an image. It has the characteristics of light insensitivity, translation invariance and rotation invariance. The complementarity between the traditional feature extraction method LBP and convolution neural network can improve the discriminability of feature extraction.

VGGl [9] is a classical convolution neural network. The combination of convolution layers of several small filters in its structure can enhance the expression of features while using fewer parameters. In this paper, the deep and shallow features of VGG network are fused and called SDFVGG network. A new method based on LBP and SDFVGG network is proposed. This method combines the LBP face feature map with the original image as input of SDFVGG network, so that SDFVGG network can not only automatically learn the information of the original face image, but also learn the LBP texture information.

2. Basic Principles

2.1 LBP algorithm

LBP refers to the local binary pattern, which is an operator used to describe the local texture features of an image. The basic principle is that the original LBP operator is defined in a 3×3 domain of a pixel, and the gray value of eight adjacent pixels is compared with the threshold value with the adjustment value of the neighborhood central pixel. If the adjacent pixel value is greater than the reading value, the position of the pixel is marked as 1: otherwise, 0. Therefore, 8-bit binary numbers are generated by comparing 8 points in a 3×3 neighborhood. Then 8-bit binary numbers are arranged in sequence to form a series of binary codes, which are then converted into decimal numbers, which are the LBP mode of the central pixel.

The LBP operator is derived from the following formula

$$LBP(x_c, y_c) = \sum_{p=0}^{7} s(i_p - i_c)\, 2^p \tag{1}$$

Among them, (x_c, y_c) is the coordinate of the central pixel; i_c is the gray value of the central pixel; $i_p (p = 0, 1,\ldots,7)$ represents eight pixel values on the central neighborhood; *s(x)* is defined as a symbolic function

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2}$$

The LBP pattern is obtained by the LBP operator as shown in Figure 1.

32	20	34
53	37	39
53	17	31

(a) Grayscale map of local ar

0	0	0
1		1
1	0	0

(b) Binarization processing

Fig. 1. LBP patterns

The LBP mode of the central pixel is $(00010011)_2$= 19. Because local binary pattern is the local information feature of face and has the characteristics of light insensitivity, gray translation invariance and rotation invariance, the combination of original image and LBP image as the input of convolution neural network makes the face feature extracted by convolution neural network richer and more representational.

2.2 VGG

VGG is a deep convolution neural network developed by researchers from the Visual Geometry Group of Oxford University and Google DeepMind. Based on AlexNet, VGG replaces single layer network with stacked 3×3 convolution layer and 2×2 maximum pooling layer, reduces convolution layer parameters and deepens network structure, improves performance, and successfully constructs 16-19 layer deep convolution neural network. Compared with the previous state-of-the-art network structure, the VGG error rate decreased significantly, and achieved the second place in ILSVRC [10] 2014 classification events and the first place in positioning events. In addition, VGG has strong expansibility and good generalization when migrating to other image data.

The convolution core size of VGG16 network is 3×3, the convolution step size is 1, the maximum pooling size is 2×2, and the step size is 2. The network has five convolutions, two convolution layers in the first two segments and three convolution layers in the last three segments. The number of convolution cores in each segment is 64, 128, 256, 512 and 512, respectively. The two 3×3 convolution layers stack with 5×5 field sizes and the three 3×3 convolution layers stack with 7×7 field sizes. Compared with a 7×7 convolution layer, using three 3×3 convolution layer stacks has the following advantages:(1) The former has more non-linear transformations than the latter, i.e. the former can use three ReLU [11] activation functions, while the latter only has one, which makes the convolution neural network more capable of learning features; (2) Three convolution layers connected in series have fewer parameters than one convolution layer of 7×7. At the same time, after each convolution, a maximum pooling layer is connected to reduce the size of the picture, thus reducing the parameters in the final full connection layer. The structure of VGG16 is shown in Figure 2.

Conv denotes the convolution layer of the network, Maxpool denotes the maximum pooling layer, and FC denotes the full connection layer of the network.

In this paper, the migration learning method is used to train the VGGl6 model pre-trained on the ImageNet data set by fine-tuning. The so-called transfer learning is to apply the model trained on a problem to a new problem through simple adjustment. Transfer learning solves the problem of insufficient training data and training time.

3. SDFVGG algorithm

3.1 Feature fusion method

This paper presents a method to fuse the deep and shallow features of VGG network. Its basic process and principle are shown in Figure 3.

(1) Different shallow features of the network are extracted by parallel multi-layer convolution layer of different scales, which enhances the ability of feature expression.

(2) Integrating different shallow features and network deep features through Concat layer to generate fusion features;

(3) Different fusion features are generated by the parallel multi-layer convolution blocks. These different fusion features are fused with the deeper features of the network to generate the final fusion features.

3.2 SDFVGG Network Architecture

The SDFVGG network structure obtained by using the feature fusion method shown in Figure 3 is shown in Figure 4. The dotted line frame in the figure shows the parallel branches of feature extraction and fusion. SDFVGG network fuses features step by step through connection layer, and the output of Concat-2 layer is the final fusion feature.

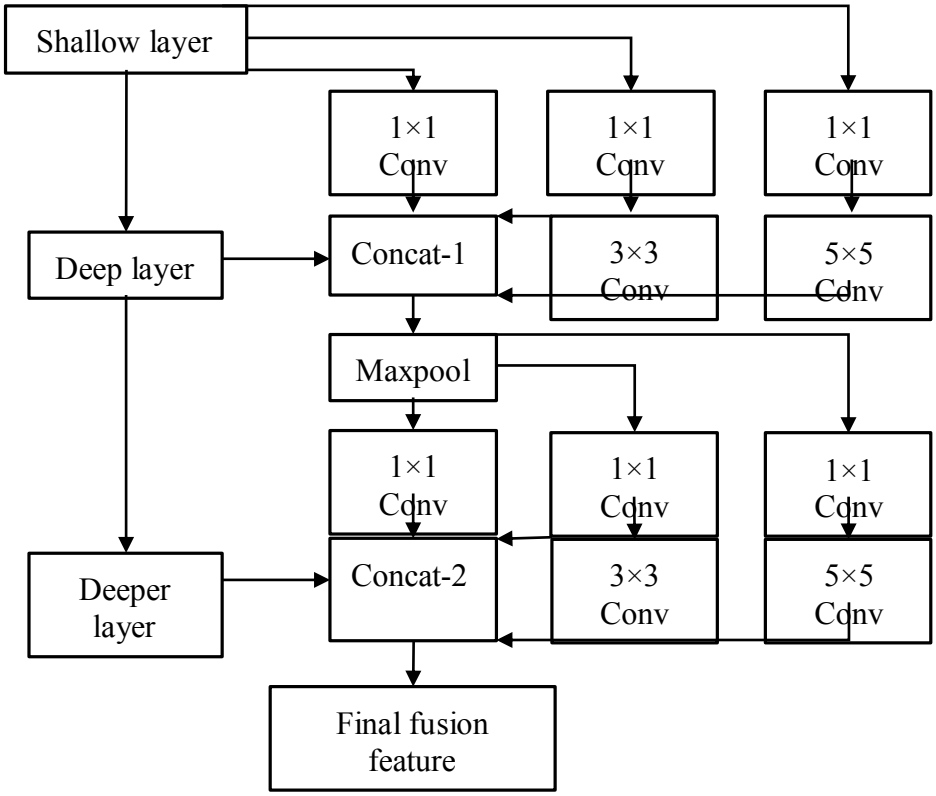

Fig. 2. Network structure of VGG16

Among them, the number of Conv6-1, Conv6-2, Conv6-4 and Conv7-1, Conv7-2, Conv7-4 convolution cores in parallel branches is 64: Conv6-3, Conv6-5 and Conv7-3, Conv7-5 have 128 convolution cores, while Max-pool 6 and VGG network have the same maximum pooling layer Fig. 3. Feature fusion method

parameters. In parallel architecture, although the 1×l convolution layer enhances the non-linear characteristics of activation function, it does not expand the receptive field: the parallel connection of convolution layers of different scales increases the width of the network, improves the performance of the network, and enriches the features extracted by the network. However, when the number of parameters increases with the expansion of the network, it is prone to over-fitting. Therefore, adding Batch Normalization to the first and second full connection layers of the network can control over-fitting and speed up convergence. The last parameter of the full connection layer is set to the number of classes. By calculating the probability of each class, the Soft-max layer obtains the corresponding maximum probability category.

4. Input feature fusion algorithm

This paper combines the original image information with the local binary pattern information as the input of SDFVGG network, so that SDFVGG network can not only learn the global original image information, but also learn the local features of the image, thus making the features extracted by the network more sufficient and more representational. The specific input feature fusion method is shown in Figure 5.

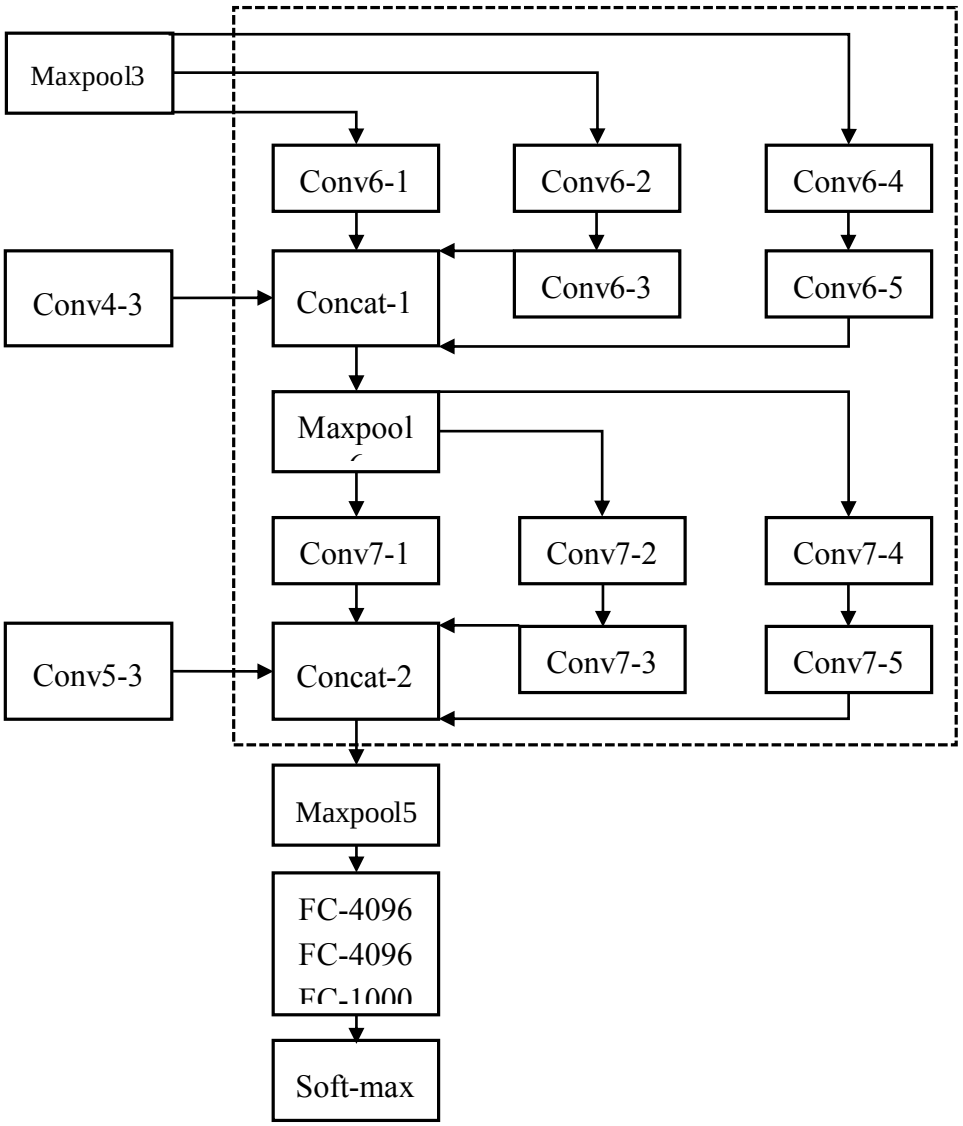

Fig. 4. Network structure of SDFVGG

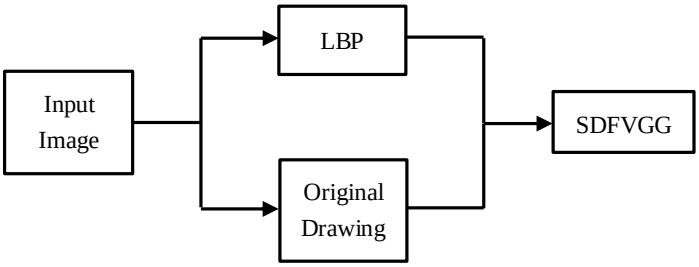

Fig. 5. Input feature fusion

5. Experimental process

5.1 Database

CAS-PEAL [12] contains 99594 photos of 1040 people, including 595 males and 445 females. The images cover various changes in posture, expression, accessories, lighting and background. CAS-PEAL-R1 is a subset of CAS-PEAL, which contains 30863 images of 1040 people. These images

belong to two subsets of front and side. In the front subset, all images are taken by a specific camera, and the subject is facing the camera. Among them, 377 people had 6 images of different expressions; 438 people had images with six different accessories; 233 people had images obtained under at least nine light changes; 297 people took photos under 2 to 4 different backgrounds; 296 people had images with different distances from the camera. In addition, 66 people recorded images in two trials over a six-month interval. The side subset contains 21 different postures of 1040 people.

In the experiment, the most representative three groups of face sets are used, namely, expression set (PE), accessory set (PA) and illumination set (PL).PE contains 1884 facial images of 377 people, PA contains 2616 facial images of 438 people and PL contains 2450 facial images of 233 people. Each set image is divided into training data set and test data set according to 9:1 ratio, and all face images are cut and scaled to 230×200 pixel images according to eye coordinates.

Fig 6. Facial image of CAS-PEAL-R1

5.2 Experimental results and analysis

In order to verify the effectiveness of the proposed method, three groups of face recognition experiments were carried out on CAS-PEAL-R1 face data set.

(1) To compare the effects of network feature fusion on experimental results. The original image was used as the input of VGG and SDFVGG networks respectively. Experiments were carried out on three subsets of PE, PA and PL to compare the effects of network feature fusion on the experimental results. As shown in Table 1, for the two subsets of PA and PL, the recognition accuracy of SDFVGG network is higher than that of VGG network, which shows that the combination of deep and shallow features of network can enhance the expression of features and improve the recognition accuracy.

Table 1. The influence of network feature fusion

Network	Recognition Rate /%		
	PE	PA	PL
VGG	99.46	97.66	98.44
SDFVGG	99.69	98.18	99.22

(2) Comparing the effects of different input data types on the experimental results. The input of SDFVGG network is original image, LBP image and image combined with LBP. Under the same experimental conditions, the recognition accuracy of PE, PA and PL is shown in Table 2. It is concluded that the accuracy of face recognition using LBP image as network input is lower than that using original image, because LBP image has information loss compared with original image. However, the accuracy of face recognition by combining the original image with LBP image is higher than that by using LBP image alone, and the generalization ability of the algorithm is stronger.

This is because LBP image better expresses the local features of the image, and the combination of the two not only compensates for the loss of information, but also increases the local feature information of the image, so the recognition rate is improved.

Table 2. Comparison of recognition rates for different input types

Network input	Recognition Rate /%		
	PE	PA	PL
original	99.69	98.18	99.22
LBP	99.06	96.61	97.66
Original+ LBP	98.58	97.12	98.05

(3) The proposed method is compared with other methods on CAS-PEAL-R1 face data set. As shown in Table 3, firstly, we can see that the accuracy of the proposed method is improved by 0.58% and 3.72% on the two subsets of PE and PA, respectively, compared with the existing algorithms, which fully proves the correctness of the algorithm. At the same time, the recognition accuracy of the proposed method on PL subset is 98.05%, which is much higher than that of previous algorithms. It also proves the validity of integrating local feature information of LBP with original image information as input of SDFVGG network.

Table 3. Comparison with other methods

Different Methods	Recognition Rate /%		
	PE	PA	PL
GPCA+LDA [13]	90.6	82.8	44.8
LGBPHS [14]	95.2	86.8	51.0
PZM-RBFNN [15]	84.8	93.4	63.4
HGPP [14]	96.4	91.9	61.7
LLGP [13]	98.0	92.0	55.0
FHOGC [16]	94.9	90.3	68.7
Ours	98.58	97.12	98.05

6. Conclusion

This paper presents a face recognition method based on LBP and SDFVGG network. The algorithm uses parallel multi-layer convolution layers of different scales to extract the deep and shallow features of VGG network and fuse them to enhance the expression of network features. Face image extracted by LBP operator has the characteristics of light insensitivity, gray translation invariance and rotation invariance. By combining the local structure information of LBP with the original image information as the input of the network, SDFVGG network can extract more discriminant face features. The experimental results on CAS-PEAL-R1 face database show that this method can improve the accuracy of face recognition.

References

[1]Wu Xueqian, Li Feifei, Yan Yan, et al. Face recognition algorithm using LBP and MSF features[J]. Electronic Science and Technology,2018,31(8):18-20,24.

[2] Lihong Z, Fei L, Yongjun W. Face recognition based on LBP and genetic algorithm[C]. Yinchuan:2016 Chinese Control and Decision Conference(CCDC),2016.

[3] Zhang J, Xiao X. Face recognition algorithm based on multi-layer weighted LBP[C]. Hangzhou:8h International Symposium on Computational Intelligence and Design (ISCID), 2015.

[4] Bengio Y, Delalleau O. On the expressive power of deep architectures[C]. Espoo: Discovery Science 14h International Conference,2011.

[5] Wang M, Wang Z, Li J. Deep convolutional neural network applies to face recognition in small and medium databases[C]. Hangzhou:4h Intemational Conference on Systems and Informatics(ICSAI),2017.

[6] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015,521(7553):436-446.

[7] Syaffeza A R, Khalil-Hani M, Liew S S, et al. Convolutional neural network for face recognition with pose andIllumination Variation[J]. International Joumal of Engineering & Technology, 2014,6(1):44-57.
[8] Yan K, lHuang S, Song Yet al. Face recognition based on convolution neural network[C]. Dalian-Chinese Control Conference,2017.
[9] Gu S, Ding L.A complexevalued VGG network based decp learing algorithm for image recognition[C]. Wanzbou: Ninth International Conference on Intelligent Control and Information Processing (ICICIP),2018.
[10] Russakovsky O, Deng J, Su H, et al. lmageNet large scale visual recognition chalenge[J]. Intermational Journal af Computer Vision,2014,115(3)211-252.
[11] Dahl G E, Sainath T N, Hinton G E. Improving deep neural networks for LVCSR using rectified linear units and dropout[C],. Vancouver. IEEE International Conference on Acoustics,2013.
[12] CGao W, Member S, Cao B, ct al. The CASPEAL large-scale chinese face database and baseline evalustions[J]. IEEE Trans on System Man&Cybermeticx,2008,38(1):149.161.
[13] Xie S, Shan S, Chen Xet al. Learned local Gabor paterms for face representation and recognition[J]. Signal Processing,2009,89 12:2333-2344.
[14] Zhang B, Shan S, Chen X, et al. Histogram of Gabor Phase Patterns(HGPP):a novel object representation approech for face recognition[J], IEEE Transactions on Image Processing, 2007,16(1):57-68.
[15] Farajzadeh N, Faez K, Pan G. Study on the performance of moments as invariant deseriptors for practical face recognition systems[J] IET Cocmputer Vision,2010.,4(4): 272-285.
[16] Tan H, Yang B, Ma Z. Face recognition based on the fusion of global and local HOG features of face images[J] IET Computer Vision,2014,8(3):224-234.