Face feature extraction based on deep learning

Bing Chen, Lin Luo, Xin Jiang

School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644000, China;

Abstract

Face feature extraction is an important part of face recognition, and convolutional neural network, as one of the representative structures of deep learning, has a very important role in feature extraction. Compared with traditional methods, convolutional neural networks have stronger feature learning and feature representation capabilities. Based on the features of convolutional neural networks in image feature extraction, the paper proposes a convolutional neural network model for face feature extraction. The paper firstly uses different convolutional layers in the convolutional neural network and the traditional Local Binary Patterns（LBP） method to extract features from human faces and analyze the extracted features respectively, and then uses the support vector machine （SVM） to predict and classify the extracted features.

Keywords

Face recognition，Feature extraction，Deep learning，Convolutional neural network，Support vector machine.

1. Introduction

When studying problems such as face similarity or local feature differences of faces, it becomes very important to extract stable and reliable face features with high expressiveness. Traditional face recognition techniques, such as Local Binary Pattern (LBP) [1], are based on shallow feature models, which leads to an inability to improve the accuracy of face recognition. There are many traditional feature extraction methods, such as Local Binary Pattern (LBP), which mainly describes the distribution relationship between image pixel points and the gray space between image fields, does not change the visual change of the image due to the light intensity, and has the advantages of constant rotation and constant gray value. Scale-Invariant Feature Transform (SIFT) in literature [2] has scale invariance to detect key points in images, and its basic idea is to introduce a parameter considered as scale in the visual information image information processing model, and obtain visual processing information at different scales by continuously varying the scale parameter, The information is then synthesized to deeply explore the features of the image.

2. Feature extraction based on convolutional neural network

2.1. Convolutional neural network structure

A common CNN network model consists of an input layer (Input), a convolutional layer (Conv), an activation function (ReLU), a normalization (Norm), a pooling layer (Pooling), a fully connected layer (Fc), and so on. Figure 1 shows the classical CNN structure VGG16 [3], which can be used to build new networks by combining different functional layers.

Convolutional neural networks are able to do point-to-point training of each convolutional kernel with its corresponding location pixel. The convolutional kernels act as filters, and each filter extracts some specific feature [4]. The image is transformed into a matrix with $x_{i,j}$

denoting the i-th row and j-th column elements of the image. Each weight of the convolution kernel is numbered, and the mth row and nth column weight of the convolution kernel is denoted by $w_{m,n}$. Then each element of the feature image is numbered, and $x_{i,j}$ denotes the i-th row and j-th column element of the feature image. With f denoting the activation function, the convolution formula is calculated as.

$$a_{i,j} = f(\sum_{m=0}^{2}\sum_{n=0}^{2} w_{m,n}x_{i+m,j+n}) \tag{1}$$

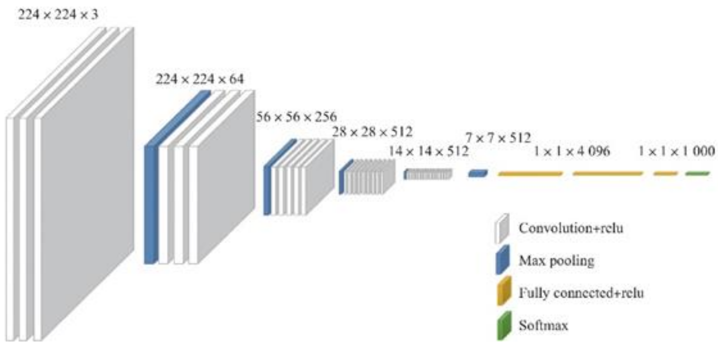

Figure 1: VGG16 network structure

2.2. Fully connected layers

Fully connected layers (FC) [5] are added at the end of the CNN to play the role of classifier. The final fully connected layer is used for feature extraction and classification for end-to-end learning, and the number of feature extraction dimensions is equal to the number of neurons.

3. Experiment

This is because when the size of the convolutional kernel is small, the extraction contains too much information and there are many invalid features which will affect the classification accuracy and make the classification accuracy low. When the size of the convolution kernel is large, a lot of information is filtered out and fewer features are retained so the classification accuracy is reduced.

3.1. Two or more

The experimental data is part of the AR face dataset, 640 frontal unobstructed face images of this dataset are taken, 80 people in total, 8 different images per person, each type is self-labeled. The ratio of 8:2 is divided into training set and test set Some of the face images are shown in Figure 2.

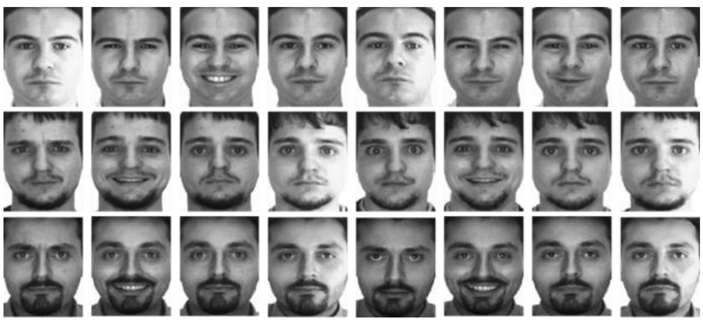

Figure 2: Partial image preview of the dataset

To verify the extraction performance of the model proposed in the paper, it is compared with Alexnet combined with SVM classification separately. alexnet contains 4 convolutional layers with a learning rate of 0.01 iterations of 500. RGB face images of the same sample at the same size location are input to Alexnet for training and feature extraction, and later combined with SVM for classification and recording the accuracy.

3.2. Results and Analysis

According to the above results, the deeper the convolution layer is, the more obvious the features are. The fewer the number of features, the fewer the fine patches caused by noise in the feature map. the LBP feature extracted image can reflect the texture of each region more clearly compared with the original image, while at the same time fading the smoothed region which is not useful for the study of features, and at the same time reducing the number of feature dimensions. The results are shown in Table 1.

Table 1: Accuracy time comparison of different models

Models	Accuracy		Feature extraction time /ms	Classification time /ms
	Conv2	Conv3		
DNNet+SVM	73.7%	55%	243	82
AlexNet+SVM	70.0%	58.7%	604	91

4. Conclusion

Aiming at the traditional methods to extract face features with low accuracy and susceptible to interference, the paper proposes to combine the powerful feature learning and representation capability of CNN for face image feature extraction. Different convolutional layers are used for feature extraction of faces, and then SVM is used for feature classification experiments. The experiments are compared and analyzed for feature classification accuracy, and the results are compared and analyzed with the traditional feature extraction method LBP classification.

References

[1] Park Y K , Kim J K : Fast adaptive smoothing based on LBP for robust face recognition, Electronics Letters, vol. 43 (2007) No. 24,p. 1350-1351.

[2] YONG Shi, ZHAO Lv, NING Bi, et al: An improved SIFT algorithm for robust emotion recognition under various face poses and illuminations, Neural Computing and Applications, vol. 32 (2020) No. 12, p. 9267-9281.

[3] ZHENG yuanpan , LI guangyang , LI ye: Survey of Application of Deep Learning in Image Recognition, Computer Engineering and Applications, vol. 55 (2019) No. 12, p. 20-36.

[4] Hinton G E , Salakhutdinov R R : Reducing the Dimensionality of Data with Neural Networks, Science, vol. 5786 (2006) No. 313, p. 504-511.

[5] LOU Guangxin, SHI Hongzhen: Face Image Recognition Based on Convolutional Neural Network, China Communications, vol. 17 (2020) No. 2, p. 125-132.