FRMN - A Face Recognition Model Based on Convolutional Neural Network

FRMN- A Face Recognition Model Based on Convolutional Neural Network

Shifeng Shang, Haiyan Liu, Qiang Qu, Guannan Li, Jie Cao

Department of information and communication, Armored Forces Academy of PLA Army, Beijing, China

Abstract. A face recognition model based on Convolutional Neural Network was proposed in this paper. This model mixed local and global image features at the largest extent. By acquiring the good local feature from mouth, nose and mouth, it is great improvement to enhance the accuracy for face recognition. From the experiment results, this model can better abstract the global feature and show a great improvement in face recognition.

1. Introduction

The research on face recognition by means of computer has been carried out since very early, including pattern recognition, image processing, and physiology and so on. At present, with the continuous improvement of deep learning algorithm and its good robustness for feature extraction, many research teams begin to apply deep learning to face recognition, and great progress has been made. For face recognition, it involves face detection, face key location, face alignment, face feature value extraction and face comparison. For face alignment, some face recognition algorithms do not adopt local feature extraction for face comparison, and still achieve good results [1].

Deep learning at present more popular approach is the convolution neural network (CNN, convolutional neural network). The network is composed of forward propagation network and reverse propagation network. In the forward propagation, the input training set is continuously learned and the weight is updated. In the back propagation, the input test set is continuously tested and the weight is updated according to the reverse error.

The deep learning system needs a large number of samples to learn better features. By means of machine learning, the input samples are constantly featured and the objective function constraints are used to achieve the optimal learning process. Among them, the learning rate, momentum and other parameters will have a certain impact on the network. In addition, for the deep learning, rich sample is also essential, how well the samples in a certain extent, also affect the effect of network training, if not enough sample set definition and difference between sample classification is too small will be unstable or because of the network learning process continuously, even a picture exceptions can cause the class accuracy of small drop sharply, so for the sample selection need to preferential selection [2] [3] [4].

Among them, face recognition in the unconstrained environment is an important part to test the robustness of the algorithm. The unconstrained environment involves the change of light, the change of face Angle, the change of expression, etc., which is better than the flexible features of the face. The feature extraction of the face in different environments will also be different. All these changes will bring huge challenges to face recognition by computer. Currently, for the flexible face processing, it is often used for alignment, involving face completion, 3d transformation and other graphic calculations, and then face mapping to the front image, and then feature extraction, face comparison and other operations, so as to achieve the purpose of unconstrained face processing [4][5]. For the key point detection of human face is a very important link, through the positioning of human eyes, nose, mouth, etc., can be very good to judge the Angle of the face, the degree of deformation, and then the system will determine the face 3d mapping Angle, part of the completion of parameters, and finally the actual three-dimensional transformation.

2. The Network Structure of Model

Face recognition model structure based on neural network not only has good extraction ability for coarse granularity, but also for fine granularity classification, which is also an important factor for it to be competent in detection, key point extraction and feature extraction at the same time. The model is built based on different layers, there are input image layer, face detection layer, which are connected by network to form a more complex CNN training model. In face detection, it can better identify the eyes, nose and mouth of people by strengthening the facial detail features, so as to better identify the facial areas formed by these areas.

Convolutional neural network is a multi-layer neural network, each layer is composed of multiple two-dimensional planes, and each plane is composed of multiple independent neurons. The input image layer is convolved with three trainable filters and a bias. Three feature maps are generated in C1 layer after convolution. These maps are then filtered to the C3 layer. This hierarchy then generates S4, just like S2. Finally, these pixel values are rasterized and connected into a vector input to the traditional neural network, which is calculated by the s-curve function and produces the output.

The most prominent feature of convolutional neural network is sparse connection (local sensing) and weight sharing, which are mainly composed of basic components such as input layer, convolution layer, sub-sampling layer, full connection layer, classification layer and output layer. The input layer is the original image, which is generally three-channel (R,G,B) sampling to get the grayscale photos of each channel, and then the final output can be obtained through convolution, subsampling, full connection and other calculations, which can be used for classification, retrieval or recognition. Some research teams have found through experiments that the eigenvalues obtained through the full connection layer can be well used for retrieval.

3. The Algorithm of Face Recognition

Compared with the traditional face recognition algorithm, face recognition can be performed directly by classification method because face features have been extracted relatively well. Traditional face recognition algorithms include LGBP, Gabor feature selection and discriminant analysis method based on AdaBoost, Kernel discriminant analysis method based on SV sv-kfd, and face recognition method based on specific human face subspace, etc. [5][6][7].

The LGBP method first convolutes it with several Gabor filters of different scales and directions (the convolution result is called Gabor feature map) to obtain multi-resolution transform images. Then each Gabor feature maps fan. If mutually disjoint of local space area, the brightness of the local neighborhood pixels on each region extraction change model, and the extraction of the variation pattern of each local space area histogram space area, all the Gabor feature maps, histogram concatenated into a high dimensional feature in all areas of the histogram to encode human face image. And through the similarity matching technology between histograms (such as histogram intersection operation) to achieve the final face recognition. LGBP method has the advantages of fast calculation speed, no need for large sample learning and strong generalization ability. Typical methods of face recognition using Gabor features include elastic graph matching (EGM) and Gabor feature discrimination classification (GFC). In practical application, EGM needs to solve the positioning problem of key feature points, and it is difficult to improve its speed. However, GFC directly reduced the dimensions of Gabor features in the down-sampling with PCA and carried out discriminant analysis. Although this avoids the problem of accurately locating key feature points, the feature dimension of down-sampling is still on the high side, and the simple down-sampling strategy is likely to miss a lot of useful features. In addition, Eigenface is one of the most famous algorithms in the field of face recognition. In essence, it is adopted by PCA to obtain the linear subspace of face image distribution, which reflects the common features of face image distribution from the perspective of optimal reconstruction [8], [9], [10].

Softmax Regression method was used to realize the classification and detection of different human faces. This algorithm is based on the existing face training. If there are 1000 categories of faces in the library, softmax outputs 1000 categories. In the process of face recognition, according to different faces, the output is a vector with a one-dimensional length of 1000, and its maximum value represents the most likely category. Because the model can accurately extract face features, so in the algorithm implementation, Softmax can be used to get a good recognition effect. In logistic regression, the training set is$\{(x^{(1)},y^{(1)}), \ldots,(x^{(m)},y^{(m)})\}$ }, where, m marks the number of samples, $y^{(i)} \in \{0, 1\}$. Suppose the model is,

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1+\exp(-\theta^T x)}$$

Logistic is essentially a linear regression model, but it adds a layer of function mapping on the continuous value result of regression, sums the characteristic linear, and then USES g(z) as the mapping to map the continuous value to the discrete value 0/1. If it is sigmoid function, it is classified as 0/1. If it's a hyperbolic sine function, it's classified as 1 over minus 1.

For Softmax regression, the target result is multiple discrete values, which is the extension of logistic model in multiple classification problems, , $y^{(i)} \in \{1,2,3\ldots k\}$, for a given test x, assume that the model estimates the probability value p(y=j|x) for each category, then assume that the function form is,

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 \mid x^{(i)};\theta) \\ p(y^{(i)} = 2 \mid x^{(i)};\theta) \\ \cdots \\ p(y^{(i)} = k \mid x^{(i)};\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \cdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix}$$

The corresponding cost function is,

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right]$$

The probability that Softmax divides x into categories j is,

$$p(y^{(i)} = j \mid x^{(i)};\theta) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}$$

For this cost function, calculate its gradient to minimize J(), and the gradient formula is as follows:

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[x^{(i)}(1\{y^{(i)} = j\} - p(y^{(i)} = j \mid x^{(i)};\theta))]$$

For each iteration,

$$\theta_j = \theta_j - \alpha\nabla_{\theta_j} J(\theta)(j = 1,2,\ldots,k)$$

Through the model introduced above, face recognition can be carried out after the completion of training. In addition, for an image to be recognized no matter it is rotated, translated or mirrored, it is always the image itself. Based on the above preprocessing, the recognition accuracy can be improved better.

The general process of face recognition algorithm is as follows:

Input: face image to be recognized

Output: print the most similar face category and similarity size

1. Input face picture p to be recognized
2. Rotate, translate and mirror p to generate batch face image pp

3. Input pp into face recognition model for batch classification and detection, and generate batch output qq corresponding to pp

4. Select the type output with the largest similarity in qq, which can best represent the category of input face

4. Experimental Results and Analysis

Through the mixed training of local facial features (eyes, nose and mouth) and global features (face features), through the comparison experiment, it can be seen that the system can better extract important local features and conduct face recognition more accurately by integrating local features into global features for training.

Table 1. Face recognition results

Number	training method	face recognition accuracy
1	without adding any local features	72.30%
2	add eye category for mixed training	76.40%
3	Add nose type for mixed training	78.90%
4	Add mouth type for mixed training	79.40%
5	Add eyes, nose and mouth for mixed training	81.80%

As can be seen from table 1, after adding local facial features, the overall recognition accuracy of the system is higher than that without adding any local features. Further analysis shows that since eyes are the most important feature for the whole face, the training of eye local features is more conducive to improving the accuracy of face recognition than nose and mouth. At the same time, after adding the three categories of eyes, nose and mouth into the training, it can be seen that the accuracy is significantly improved compared with the single local feature.

5. Summary

Based on convolutional neural network theory, this paper puts forward a sample set for a global features and local features fusion of training methods, through the model training, the method can extract maximum face key parts such as features, in areas such as the mouth, nose, eyes and at different face feature extraction, can better from these parts of the face more explicit feature vector. From the experimental results, since eyes play a very important role in face recognition, it is more important for feature extraction than nose and mouth. In conclusion, by integrating local features with global features for training, the model can better extract global features, so as to better carry out face recognition.

In future work, will further more local features convergence to the global hybrid training, such as eye categories can be divided into monocular and binocular, eyes, nose, eyes and nose can be divided into different categories, the purpose is to extract more from the local characteristics of more sensitive rich characteristics, to facilitate more accurate classification.

6. Reference

[1] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensional ity: High-dimensional feature and its efficient compression for face verification(C). Computer Vision and Pattern Recognition 2013: 3025–3032

[2] Y. Bengio, A. Courville, and P. Vincent. Representation learn ing: a review and new perspectives [J]. IEEE transactions on pattern analysis and machine intelligence, Aug. 2013, vol. 35, no. 8, pp. 1798–828.

[3] http://www.jdl.ac.cn/project/faceId/res-identify.htm

[4] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extrac tion [J]. Artificial Neural Networks and Machine Learning. 2011, vol. 6791, pp.52-59

[5] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, vol. 1, pp. 886–893.

[6] J. Yang, J. Yang, D. Zhang, and J. Lu. Feature fusion: parallel strategy vs. serial strategy [J]. Pattern Recognition, 2003, vol. 36, no. 6, pp. 1369–1381.

[7] Wolf, L., Hassner, T., Taigman, Y. Descriptor based methods in the wild [C]. In: Workshop on Faces in Real-Life Images: Detection, Alignment, and Recognition. (2008)

[8] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications [J]. Neural Networks, 2000, vol. 13, no. 4–5, pp. 411–430.

[9] J. Zou, Q. Ji, and G. Nagy. A comparative study of local matching approach for face recognition [J]. IEEE Trans.Image Processing, 2007, 16 (10): 2617-2628.

[10] P.N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigen faces vs. Fisherfaces: recognition using class specific linear projection [J]. IEEE transactions on pattern analysis and ma chine intelligence, Jul. 1997, vol. 19, no. 7, pp. 711–720.