

PAPER • OPEN ACCESS

Face Recognition Algorithm Based on VGG Network Model and SVM

To cite this article: Hongling Chen and Chen Haoyu 2019 *J. Phys.: Conf. Ser.* **1229** 012015

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the [collection](#) - download the first chapter of every title for free.

Face Recognition Algorithm Based on VGG Network Model and SVM

Hongling Chen^{1,*} and Chen Haoyu²

¹ College of computer and information science, Southwest University, China

² College of mathematics and Statistics, Southwest University, China

*Email: chenhongling@email.swu.edu.cn

Abstract. The problem that the dimension of facial features is too large does exist with the Deep learning face recognition. This paper proposes a face recognition algorithm based on SVM combined with VGG network model extracting facial features, which can not only accurately extract face features, but also reduce feature dimensions and avoid irrelevant features to participate in the calculation. Firstly, the VGG-16 model is obtained by training the training data set, which is used for feature extraction, on top of this, principal component analysis method (PCA) is used for feature dimensionality reduction, and last, the face recognition is performed by SVM classifier with linear kernel function. In this paper, we conduct a comparative experiment on CelebA dataset and find that the accuracy reaches its peak when the feature dimension is reduced to 400. The experiment is carried out on LFW dataset using 400-dimensional feature data, and comparing with other algorithms, the results show that the algorithm in this paper has reached the level of state-of-art.

1. Introduction

As a biometric recognition technology, face recognition is one of the hot topics in the fields of pattern recognition, image processing, machine vision, neural network and cognitive science. At the same time, face recognition as a high-stability, high-precision, and easily accepted biometric technology has broad application prospects in the fields of identity authentication, security monitoring, and human-computer interaction.

Face recognition uses correlation recognition algorithms for face recognition or discrimination, which is based on the extracted face image features. Choosing the appropriate face feature extraction method and matching strategy is the key to this process. The construction of the recognition framework goes hand in hand with the way of face extraction feature. At present, the main feature extraction methods of face recognition algorithms can be divided into the following categories:

The extraction method based on geometric features of faces [1] [2]. This method not only extract the normalized inter-point distance of each part of the face and some feature points of the face, such as the two-dimensional topological structure composed of the corners of the eyes, the corners of the mouth, and the tip of the nose, but also is an early face recognition method, which requires high quality face images. If the face is tilted in the picture, it will affect its geometric features and lead to the failure of face recognition.

The extraction method based on statistical characteristics. This method treats the face image as a random vector, also uses statistical methods to distinguish different facial feature patterns, and the



typical methods among which include eigenface [3], independent component analysis[4], and singular value decomposition (SVD) [5].

The extraction method based on neural network features[6]. This method uses a large number of neurons to store and remember face features, and makes full use of the characteristics of neural network self-learning to extract the effective features of faces. The face image is accurately identified according to the probability of different neural unit states. At present, the algorithm based on neural network has achieved good results, but neural network methods are very demanding on data and equipment, which requires numerous data for training and sophisticated equipment to increase the amount of computation.

To solve the above problems, we propose a feature extraction method based on deep learning, that is to extract features of the face using the VGG-16 model and to reduce dimensions of the feature vectors by using principal component analysis (PCA), which avoids unrelated features participating in this operation, then, the SVM algorithm is used to predict the sample. We test the algorithm of this paper on the CelebA and LFW datasets. The results show that the algorithm has reached the state-of-art level.

2. Related work

This section mainly introduces the VGGNet model, Principal Component Analysis (PCA) and Support Vector Machines (SVM) algorithm.

2.1. VGGNet

The VGGNet is a deep convolutional neural network which extracts deeper features in the image and has higher feature extraction capabilities[6]. The VGG-16 used in this paper is a 16-layer deep convolutional network and the specific parameters are shown in column D of Figure 1. This network is improved on the basis of AlexNet[7], replacing the 7×7 and the 5×5 convolution kernels in AlexNet with three 3×3 convolution kernels and two 3×3 convolution kernels, respectively, whose depth is improved under the same receptive fields conditions, thereby the effect of it has been improved.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 1. This is the VGGNet structure configuration, and we use the column D networks architecture as VGG-16 Net.

2.2. Principal Component Analysis

Principal component analysis is used to extract the features of the face[8], which not only retains the main information of the face, but also plays the role of data dimensionality reduction. In this paper, we extract the features of each face through the VGG model, and obtain the face image feature matrix $X = [X_1, X_2, \dots, X_n]$ that is an $m \times n$ dimensional matrix, where n is the total number of samples and m is the dimension of the feature. We can find the mean \bar{X} and covariance matrix S of the sample data:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

$$S = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T \quad (2)$$

The dimension of the matrix S is $m \times m$. Due to the large amount of data, it is of great difficulty to directly obtain the eigenvalues and eigenvectors of the matrix S . Therefore, we can construct a matrix

$$R = \sum_{i=1}^n (X_i - \bar{X})^T (X_i - \bar{X}), \text{ to calculate the eigenvalue of this matrix } \lambda_i (i=1, 2, \dots, n) \text{ and the}$$

eigenvector $v_i (i=1, 2, \dots, n)$ corresponding to the eigenvalue λ_i . After that, the eigenvalues of the matrix S and the corresponding eigenvectors are obtained from formula (3), where u_i is the images feature vector. The eigenvalues λ_i are sorted from large to small, and the eigenvectors corresponding to the first k eigenvalues are selected to form a projection subspace according to formula (4), so that the contribution rate is $e \geq 0.9$. Then $Y = [u_1, u_2, \dots, u_k]$ is the projected coordinate system of the original sample matrix, and the face image in the training set is projected in this coordinate system according to formula (5). The matrix formed by the eigenvector Z_i is the feature subspace obtained by the original sample matrix after being processed by PCA.

$$u_i = \left(\frac{1}{\sqrt{|\lambda_i|}} \right) X v_i \quad (3)$$

$$e = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \geq 0.9 \quad (4)$$

$$Z_i = Y^T (X_i - \bar{X}) \quad (5)$$

2.3. SVM

Support Vector Machine (SVM), a classifier with strong learning and generalization ability, has unique advantages in solving problems such as small samples and nonlinearity [9]. The basic idea of SVM is to map the input vector into a high-dimensional space by nonlinear transformation, and then establish the optimal classification surface in this high-dimensional space. This nonlinear transformation is achieved by selecting the appropriate inner product kernel function. Mapping a vector into a high-dimensional space simply changes the inner product computation, and algorithmic complexity does not increase with increasing number of dimensions.

1) The case of linear separable

For the training set (x_i, y_i) , where $x_i \in R_N$, $y_i \in \{-1, 1\}$, $i=1, 2, \dots, n$, the purpose of SVM is to find a hyperplane that can completely separate the two types of samples, and maximize the interval between the two classes. Suppose the hyperplane equation is:

$$\omega \cdot x + b = 0 \quad (6)$$

if $\omega \cdot x + b > 0$, class it as 1, and if $\omega \cdot x + b < 0$, then class it as -1. Maximizing the interval between the two categories is equivalent to minimizing the following formula.

$$J(\omega) = \frac{1}{2} \|\omega\|^2 \quad (7)$$

The constraint of the formula (7) is:

$$y_i(\omega \cdot x + b) \geq 1, \forall i \in \{1, 2, \dots, n\} \quad (8)$$

By introducing the Lagrange multiplier α_i , we can get the Wolf dual form:

$$L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (9)$$

After solving α , the values of the plane parameters ω and b can be determined, thereout, the SVM optimization classification function can be further obtained:

$$f(x) = \text{sgn} \left(\sum_{i=1}^n \alpha_i y_i (x_i \cdot x + b) \right) \quad (10)$$

2) The case of nonlinear separable

The original sample is mapped into a high-dimensional space using the kernel function, which is transformed into a linear separable problem in high-dimensional space. The corresponding objective function is:

$$J(\omega, b, \epsilon) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \epsilon_i \quad (11)$$

Where C is the penalty factor and ϵ is the slack variable. The constraint of formula is:

$$y_i(\omega \cdot x + b) \geq 1 - \epsilon_i, \epsilon_i \geq 0, \forall i \in \{1, 2, \dots, n\} \quad (12)$$

The optimal classification function is:

$$f(x) = \text{sgn} \left(\sum_{i=1}^n \alpha_i y_i K(x_i \cdot x) + b \right) \quad (13)$$

Where $K(x_i \cdot x)$ is the kernel function. The common kernel functions are as follows: linear kernel functions, polynomial kernel functions, radial basis kernel functions, and Sigmoid kernel functions.

3. Algorithm

This section mainly introduces the principle and the framework of the algorithm as well as the face recognition process.

3.1. The principle of the algorithm

The algorithm of this paper is based on the VGG-16 network model, identifying the face by SVM classifier and PCA which is for dimensionality reduction. According to the article[6], VGG-16, a basic network, whose classification performance is beyond compare, has a neat network structure, hence, it is relatively easy to expand. In [6], the author uses Euclidean distance classification to recognize faces. Euclidean distance classification method, the principle of which is simple and easy to understand, but it only takes into account the mean value of each class of samples, ignoring the variance within the class, and the characteristics extracted from the two samples are 2×4096 , many unrelated features are also involved in the calculation.

In order to better consider the variance between categories and remove irrelevant feature, we use SVM classifier for classification and linear kernel function as the kernel function. Since SVM is a medium-dimensional classifier, using VGG-16 to extract 4096-dimensional features for each face image will lead to dimensional disasters. Therefore, it is necessary to reduce the features to achieve better performance. The specific mathematical principle of PCA is shown in the section 2.2. We reduce the facial feature data to 100, 200, 300, 400, 500 and 600 dimensions respectively, and compare the accuracy of these dimensions. The experimental results will be shown in section 4.

3.2. The process of face recognition

The framework of the algorithm in this paper is shown in Figure 2. According to the flow chart, the following steps are taken to recognize faces:

- (1) The face image databases we used are CelebA and LFW. We also use the Dlib algorithm to preprocess the face image, detect the face, then save it.
- (2) The VGG-16 network model is used to extract the features of face images, and the vector output from 'fc7' layer is taken as the final feature so that we can obtain the matrix of $m \times 4096$ dimensions, where m is the total number of images. The VGG-16 network model used in this paper is trained according to the method of [6].
- (3) Divide the sample data into training set and test set. Each person corresponds to multiple images, and after extracting the features, each image of each person corresponds to its own features. The feature set composed of 90% pictures under each person's name is taken as the training set and marked, and 10% data is taken as the test set and marked as well.
- (4) Principal component analysis method (PCA) is used to reduce the dimension of the training set and test set, with the training set that is a matrix of $m_1 \times n$ dimensions obtained, and the basis vector v of the low-dimensional space can be also gotten, where obtaining the test set a $m_2 \times n$ - dimensional matrix. The sample numbers of training set and test set are m_1 and m_2 respectively, also n is the feature dimension after dimension reduction.
- (5) Normalize the data of the training set and testing set, and train the training set and corresponding tags with SVM algorithm to obtain the face recognition model, which is used for face recognition on the test set.

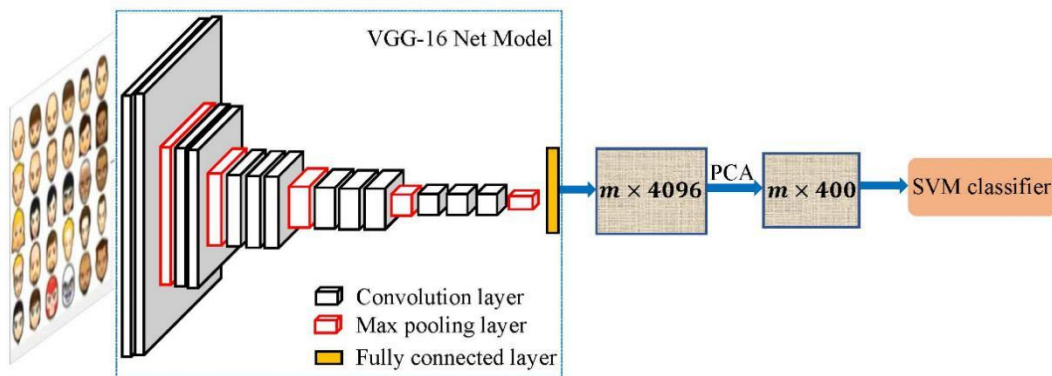


Figure 2. This is the overall framework of our method.

4. Experiment

This section mainly introduces that the algorithm in this paper is evaluated on two data sets.

4.1. Data

The LFW dataset [10] is used to evaluate the algorithm in this experiment as well as the CelebA dataset. LFW (Labeled Face in the Wild) unconstrained natural scene face recognition data set, a data set composed of more than 13,000 pictures concerning worldwide celebrities collected through the Internet, contains more than 5,000 people, where 1680 people correspond to two or more face images. Each face image has its unique ID and serial number to distinguish it.

CelebFaces Attributes Datasets (CelebA) [11] is a large face attribute dataset with more than 200 thousands face images, and each image has 40 attribute annotations. Face images in this dataset not only have large facial expression differences but also cluttered backgrounds. This dataset contains a wealth of attribute annotations, including 10,177 identities and 202,599 facial images, and it is often used for training and testing in computer vision tasks.

4.2. Evaluating Index

The experimental environment of this algorithm is Ubuntu16.04 system, with i7-7700 CPU, 4.20GHz and 32G Memory, video card using GTX1080 as well as video memory for 8G, also the Caffe is used as the deep network framework and the software tool is matlab2017b.

The evaluation indicators of the algorithm in this paper are the ACC and ROC curves, where ACC is the accuracy of algorithm recognition. Under different thresholds, calculate the value of the TPR and the corresponding FPR, and then plot the ROC curve according to the TPR and FPR values. The larger the area under the ROC curve, the better the algorithm is. TPR indicates that the detected positive sample accounts for the proportion of all positive samples, and FPR indicates that the sample tested as a positive sample actually representing a negative sample accounts for how many proportions of all negative samples totality.

The TPR calculation formula is as follows:

$$TPR = TP / (TP + FN) \quad (14)$$

The FPR calculation formula is as follows:

$$FPR = FP / (FP + TN) \quad (15)$$

where TP, FP, TN and FN stand for true positive, false positive, true negative and false negative, respectively.

4.3. Experimental result

Our experiments are conducted on two datasets, respectively. Firstly, we test on the Celeba dataset. We extract different numbers of face features and compare the accuracy of this algorithm with the VGG-16 European distance classification method. The experimental results are shown in Figure 3.

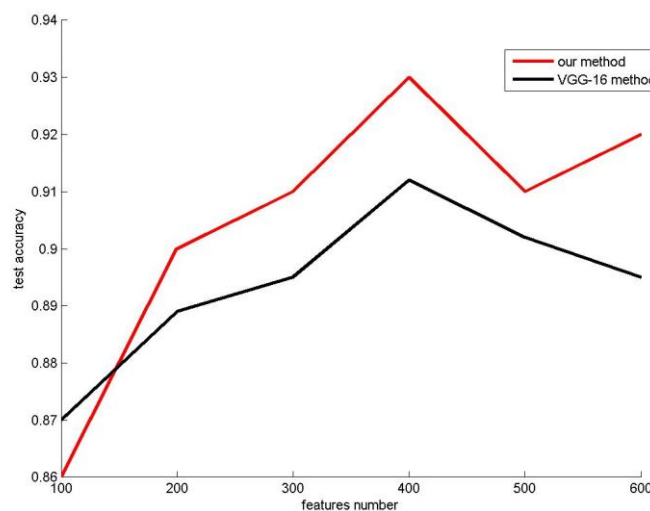


Figure 3. Comparison of two methods with different numbers of dimensions.

Table 1 shows the specific experimental results. We find that the algorithm proposed in this paper has a higher recognition accuracy rate of 93%. At the same time, that the number of features is too high or too low will seriously affect the accuracy. When the feature dimension is 400, the accuracy of face recognition is the highest.

For the sake of the accuracy, we set the feature dimension to 400, perform a second experiment on the LFW dataset and compare the other algorithms to plot the ROC curve, as shown in Figure 4.

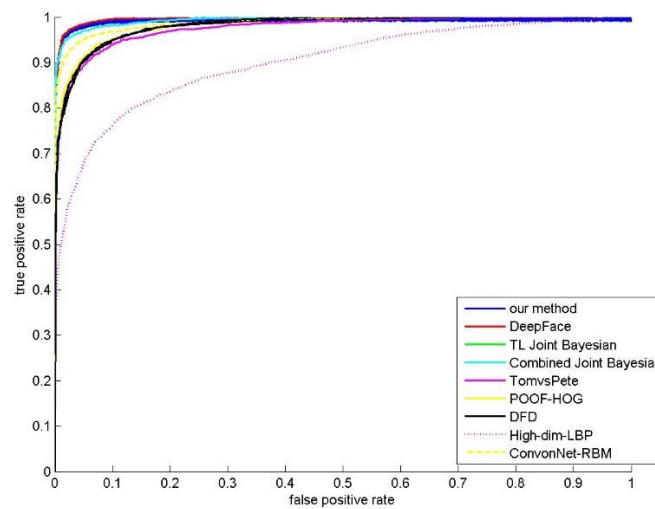


Figure 4. The ROC curves of different algorithms are compared.

Table 1. Comparison of our method and VGG-16 Net, we can get the highest accuracy in 400 dimensions.

Dimensions	Our method	VGG-16
100	0.860	0.870
200	0.900	0.899
300	0.910	0.895
400	0.930	0.912
500	0.910	0.902
600	0.920	0.895

Table 2. Comparison of state-of-the-art face recognition method on LFW dataset.

Methods	Accuracy
Our method	0.9747
DeepFace	0.9735
TL Joint Bayesian	0.9633
Combined Joint Bayesian	0.9242
TomvsPete	0.9310
POOF-HOG	0.9280
DFD	0.8402
High-dim LBP	0.9517
ConvonNet-RBM	0.9252

We compare the ACC values of different algorithms, and the experimental results are shown in Table 2. The ACC value of the algorithm in this paper reaches 97.47%. The DeepFace algorithm also uses VGG-16 model to extract features without the SVM classifier, and its ACC value is 97.35%. It can be seen that the proposed algorithm achieves higher accuracy than the DeepFace method.

5. Conclusion

In this paper, a face recognition algorithm based on VGGNet model combined with SVM is proposed. Firstly, the characteristics of face image are extracted by convolution neural network VGGNet model, then the extracted features' dimensions are reduced by PCA, and finally face recognition is carried out by SVM classification method. The method proposed in this paper is tested on the Celeba dataset. Comparing the effects of different dimensions on the accuracy, it is found that the accuracy is the highest when the feature dimension is reduced to 400 dimensions. We select the feature dimension to be 400 on the LFW dataset for testing, compared with other algorithms, finding that the algorithm in this paper has reached the level of State-of-art.

In addition, the algorithm in this paper needs to be improved in terms of operational efficiency, and not only the VGG-16 network model is too large, but also the number of fully connected layers can be cut appropriately. How to streamline the network structure and design our own feature extraction algorithm is the focus of our future work.

Reference

- [1] Zhang Jun, He Xin, and Li Jiegu. Face recognition based on geometrical feature points extraction. *INFRARED AND LASER ENGINEERING*, 28(4):40–43, 1999.
- [2] Basavaraj Amarapur and Nagaraj Patil. The facial features extraction for face recognition based on geometrical approach. In *Conference on Electrical & Computer Engineering*, 2007.
- [3] B. K. Gunturk, A. U. Batur, Y. Altunbasak, and R. M. Mersereau. Eigenface-domain super-resolution for face recognition. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 12(5):597–606, 2003.
- [4] S. L Fernandes and G. J Bala. A comparative study on ICA and LPP based Face Recognition under varying illuminations and facial expressions. 2013.
- [5] Zhang Qiang and Baoxin Li. Discriminative k-svd for dictionary learning in face recognition. In *Computer Vision & Pattern Recognition*, 2010.
- [6] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.
- [7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems*, pages 1097–1105, 2012.
- [8] Vytautas Perlibakas. Distance measures for pca-based face recognition. *Pattern Recognition Letters*, 25(6):711–724, 2004.
- [9] Sang Ki Kim, Youn Jung Park, Kar Ann Toh, and Sangyoun Lee. Svm-based feature extraction for face recognition. *Pattern Recognition*, 43(8):2871–2881, 2010.
- [10] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.
- [11] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.