Face Image Feature Extraction based on Deep Learning Algorithm

Qing Kuang

Institute of Technology, East China Jiaotong University, 330052

*Corresponding author email: kuangqing@ecjtu.edu.cn

Abstract. In recent years, due to the rapid development of computer technology, artificial intelligence technology in the computer field has begun to integrate into people's life, and facial recognition, as a unique biometric recognition method, is the core of artificial intelligence technology. Based on this, this paper discusses the local feature extraction and global feature extraction based on the deep learning algorithm, and proposes a training classification method based on the deep learning model combined with local pattern and GLQP representation feature extraction algorithm. In this paper, the local quantization method is used to input the data set preprocessed by the filter into the network. The depth of CNN network is selected as 4 layers, and the network is trained to produce high-resolution features. Experiments show that the accuracy of the trained deep network model is 92.2% in the test set. Therefore, compared with the traditional methods, deep learning has the advantages of powerful visualization and automatic face feature extraction, overcomes the shortcomings of deep learning model in the process of shallow feature learning, and shows higher recognition efficiency and generalization.

Keywords: Biometric Recognition, Face Recognition, Deep Learning Algorithm, Feature Extraction

1. Introduction

With the continuous development of society, the demand for fast and effective identification technology [1-2] is growing. Biometric technology [3-4] has been rapidly improved in diversity, reliability and convenience, which is the subversion of traditional identification technology. Biometric technology uses computer programs to process human biometrics and to verify a person's identity based on fingerprints, iris, face, voice and handwriting. Because the field of computer vision is the most convenient and accepted by users, it has always been the focus of research in the field of computer vision.

Traditional face recognition technology includes three parts: face recognition [5], feature extraction [6-7] and classifier selection. The accuracy of face recognition and the ability of facial expression recognition are affected by the accuracy of facial expression training. After deep learning, face recognition systems based on deep learning technology [8-9] emerge in endlessly. In the structure of multilayer neural network [10], deep learning can provide representation information from low level to
high level in each hidden network layer, and the original data in the result layer can provide more abstract high-level features. For complex nonlinear models, deep learning can use the powerful modeling function of multiple neural networks to simulate the distribution of complex models.

Therefore, this paper discusses facial expression research based on deep learning algorithm. In this paper, we use the depth learning model to generate facial features, and then use the first pixel of the face image as the depth network input. Finally, we integrate the captured local features and global facial features into a common feature. Research shows that deep learning can extract features that are more conducive to classification from high-dimensional data by multi-level expression of original information. In face recognition, the recognition rate on face database is improved to more than 99% by using deep learning technology, which is close to or even better than that of human.

2. Face Image Feature Extraction Method based on Deep Learning

2.1. Face Image Feature Extraction Method

Deep learning model can make facial expression automatic, and the extraction process does not involve human resources, so as to save manual labor and improve the performance of behavior recognition. In this paper, the GLQP environment processor and CNN deep learning model are used to study the unconstrained facial expression. The general process of feature extraction is as follows:

1. Modify the face image by eye location algorithm, reduce the test sample and training sample to 90 × 150 pixels by bilinear interpolation method, and perform the preprocessing process of histogram equation and other data.

2. Use 40 Gabor filters (using 40 directions and different scales) to refine the preview image to obtain all samples.

3. Using LQP operator, Gabor image and test sample are divided into 9 × 15 = 135 small rectangular blocks, and each LQP operator is used to separate the texture of each sub block and connect each sub function. Concatenate the features of each sub block to establish a sample GLQP attribute.

4. GLQP was used to characterize the training samples according to CNN. It is used as input and iterates 500 times.

5. CNN is used to optimize the contour features of experimental samples, and the fusion features of local features and global features are used as partition input to classify and recognize facial images, and the results are counted to establish a complete face recognition system.

2.2. Research on Deep Learning Algorithm

The deep learning model used in this paper is convolutional neural network (CNN), which is a multilayer feeder network composed of input layer, hidden layer and result layer. Before the data enters the input layer, it needs to go through the preprocessing process, and then it is transferred into the hidden layer, and finally output through the output layer. From beginning to end, the deep learning network automatically extracts useful features from the original image and aggregates them. In this paper, the number of network layers is set to 4 layers, including 1 input layer, 2 hidden layers and 1 output layer. By taking the face image as CNN input, the joint distribution of output layer and hidden layer of CNN meets the following requirements:

\[
P(V,h^1, h^2, \cdots, h^m) = P(V | h^1)P(h^1 | h^2) \cdots P(h^{l-2} | h^{l-1})P(h^{l-1}, h^l)
\]

The relationship between the hidden layer elements of the \(m\) layer and the \(m+1\) layer satisfies the following formula:

\[
P(h^m | h^{m+1}) = \prod_j P(h^m | h^{m+1})
\]

When the \(m\) layer is a binary unit, the output model is as follows:

\[
P(h_1^m = 1 | h^{m+1}) = \sigma(b_1^m + \sum_j W_{ij}^m h_j^{m+1})
\]
In the above formula, $V$ is the GLQP texture feature of the input visual layer, $h^1, h^2, \ldots, h^l$ is the advanced feature of different levels learned by CNN from the input data $V$; $b^m$ represents the offset of the $m$ layer, and $W^m_{ij}$ is the connection weight between the $m$ layer and the $m+1$ layer.

### 3. Experimental Thinking and Design

#### 3.1. Experimental Ideas

This paper discusses facial expression research based on deep learning algorithm. In this paper, the first pixel of the face image is used as the depth network input in the existing method of generating facial features using a deep learning model. The network will learn the expression of facial features, and then integrate the captured local features and global features into a common feature.

#### 3.2. Experimental Design

Personal identity plays a very important role in many fields, such as security, public security, justice and e-commerce. Identity authentication is currently based on methods such as certificate and password input, but these methods are easy to lose, easy to damage and easy to crack password. Therefore, the use of traditional methods may not meet the needs of social development.

Therefore, people are looking for more convenient and safer identification methods, and biometric technology provides an opportunity to achieve this requirement. Biometrics are physiological features that can be measured or verified by everyone, such as fingerprints, facial expressions, iris, voice, etc. Compared with traditional methods, biological characteristics have obvious unique advantages. Therefore, biometric based recognition technology has the advantages of security, reliability and convenience. The definition of deep learning is related to the depth of network learning. Its content is to establish a machine learning framework model with multiple hidden layers, use a large number of data for training, obtain a large number of feature data from classification skills, and then identify the experimental set samples, the ultimate purpose is to improve the accuracy.

### 4. Discussion

#### 4.1. Nonlinear Dynamic Model of Gear Transmission System under Vibration and Noise Factors

The biological characteristics that can be used to identify must be global, unique, stable and collective. Many relatively mature and promising biometric technologies include fingerprint, face, iris, retina and DNA. As shown in Table 1, fingerprints, iris, DNA and other biological features have some shortcomings in concealment and collectability, so they are difficult to be used in crime control. However, face recognition is global, unique, stable and collective, and can be used for escape arrest and other applications.

| biosignature | Uniqueness | Concealment | Acceptability | No deformation | Collectability |
|--------------|------------|-------------|---------------|----------------|---------------|
| Face         | low        | high        | high          | in             | high          |
| fingerprint  | high       | low         | in            | high           | in            |
| retina       | high       | low         | low           | high           | in            |
| iris         | high       | low         | low           | in             | low           |
| voice        | low        | in          | high          | low            | in            |

Basically, deep learning uses neural networks to simulate the structure of neurons in different layers of the human brain, and gradually creates more subtle high-level features by combining high-level features with low-level features. The definition of deep learning is related to the depth of network learning. Its content is to establish a machine learning framework model with multiple hidden layers,
use a large number of data for training, obtain a large number of feature data from classification skills, and then identify the experimental set samples, the ultimate purpose is to improve the accuracy. Compared with the traditional shallow learning model, the network depth of deep learning model is more than 6 layers or even 100 layers deeper. Through feature extraction and integration layer by layer, the information can be transformed and copied into the new feature space, which makes prediction or classification easier. In addition, the deep learning algorithm modifies the layer number constraint of the traditional neural network in theory and practice, and can change the number of layers and super parameters according to the needs of users. In addition, the training method has also changed. The traditional neural network often uses the gradient type algorithm to train the network, and the deep learning model adopts the training method of the store level, the gradual training and the subsequent fine adjustment after the training model is produced.

![Figure 1](image_url)

**Figure 1.** Expression recognition rate of CNN with different depths

As can be seen from Figure 1, with the increase of CNN depth, the expression recognition rate does not always rise. The possible reasons for this phenomenon are as follows: first, the RBM reconstruction error of each hidden layer accumulates with the increase of layers, which leads to the decrease of recognition rate; second, the gradient descent algorithm increases the error with the increase of layers. The experimental results show that when the depth of CNN is 4 layers, the expression recognition rate is the highest. Therefore, the network depth of CNN is 4 layers, namely, 3-layer RBM network and 1-layer BP network. The depth of CNN is 4 layers, and the training time is about 172 hours.
In order to verify the effectiveness and feasibility of this method in facial expression recognition, the test set is tested in the trained network, and the face recognition rate reaches 92.2% in the test set. The comparison of experimental results is shown in Figure 2. It can be seen from the table that the face expression recognition rate of this method is better than the other three methods in the overall expression recognition efficiency. CNN is an unsupervised learning method. Through unsupervised learning, CNN constantly seeks the interface with the most classification ability in high-dimensional space, and obtains more and more strong features of classification ability through unsupervised learning of multi-layer nonlinear structure. Finally, through continuous learning and classification of these features, the face recognition rate is greatly improved.

**4.2. Description of Face Recognition**

1. Face acquisition and detection
   
   The camera collects different facial images, for example, at different times, images, different scenes, different backgrounds and different lights to form a foreground face database. The face in the image is distinguished from the background, the face is detected, and its position and size are determined according to the facial features.

2. Image preprocessing
   
   Based on the results of face detection, geometry and gray scale are used to remove irrelevant information from facial images, extract useful and real information, enhance image recognition information, and simplify information.

3. Feature extraction
   
   Traditional facial image technology is mainly based on knowledge statistical method, which mainly includes geometric behavior method and template matching method; statistical facial expression method mainly includes key element analysis and independent body analysis. After using deep learning network model, the model can automatically capture face features.

4. Matching and recognition

   In this process, the similarity between the features stored on the computer and the features previously stored on the computer is used to determine whether the facial features are consistent. The most commonly used identification methods include statistical identification method and structure identification method. Statistical recognition method first forms a character space, and then uses it to classify and recognize statistical decision-making concepts; structure recognition method focuses on
the description of object structure features. Face recognition, also known as identity, is a one to many images matching process, which is used to compare the similarity between the currently collected images and all user images stored in the system.

5. Conclusions
In the research of face image feature extraction method based on deep learning algorithm, this paper discusses the local feature extraction and global feature extraction based on deep learning algorithm, and proposes a training classification method based on deep learning model, combining local pattern and GLQP representation feature extraction algorithm. In this paper, the local quantization method is used to input the data set preprocessed by the filter into the network. The depth of CNN network is selected as 4 layers, and the network is trained to produce high-resolution features. The results show that deep learning has more automatic feature extraction ability than shallow learning. The rich environment design function provides enough samples for the training of deep learning model, deepens the deep learning model, and helps to develop more unique features. This feature extraction method has clear structure and hierarchical structure.

References
[1] Dorpenyo, I. K. . (2019). Risky election, vulnerable technology: localizing biometric use in elections for the sake of justice. Technical Communication Quarterly, 28(4), 361-375.
[2] Budiman, P. A. , Tresnawati, T. , Alamsyah, A. T. , & Riandini. (2019). Study for integration of multi modal biometric personal identification using heart rate variability (hrv) parameter. Journal of Physics Conference Series, 1364, 012074.
[3] Wong, K. A. . (2020). The face-id revolution: the balance between pro-market and pro-consumer biometric privacy regulation. The Journal of High Technology Law, 20.
[4] Magdalena Sepúlveda Carmona. (2019). Biometric technology and beneficiary rights in social protection programmes. International Social Security Review, 72(4).
[5] Nanjun, H. , Paoletti, M. E. , Mario, H. J. , Leyuan, F. , Shutao, L. , & Antonio, P. , et al. (2019). Feature extraction with multiscale covariance maps for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, PP, 1-15.
[6] Kou, G. , Yue, J. , Ma, Y. , & Wang, Q. . (2019). Sar image invariant feature extraction by anisotropic diffusion and multi-gray level simplified pcnn. IEEE Access, PP(99), 1-1.
[7] Liu, Y. , Wang, F. , Liu, L. , & Zhu, Y. . (2019). Secondary signal-induced large-parameter stochastic resonance for feature extraction of mechanical faults. International Journal of Modern Physics B, 33(7), 1950157.
[8] Chavez-Badiola, A. , Farias, F. S. , Mendizabal-Ruiz, G. , Garcia-Sanchez, R. , & Garcia-Sandoval, J. P. . (2020). Predicting pregnancy test results after embryo transfer by image feature extraction and analysis using machine learning. Scientific Reports, 10(1), 4394.
[9] Halm, S. . (2019). Deep learning in marketing und service. Ibusiness Executive Summary, 29(1), 6-6.
[10] Dai, J. , Li, C. , Lan, T. , & Wang, Z. . (2020). Comparative study on iii-v mosfet and si-mosfet model parameters based on bp neural networks algorithm. IOP Conference Series Materials ence and Engineering, 768, 072040.