ISSN: 2664-9640

Research on Face Recognition Method Based on Deep Learning

Tiantian Chen[1,2], Hongrong Jing[1,2] and Hongjie Zhang[1, 2]

[1] Artificial Intelligence Key Laboratory of Sichuan Province, Zigong 643000, China;

[2] School of Automation & Information Engineering, Sichuan University of Science & Engineering, Zigong 643000, China.

Abstract

This With the popularization and development of information technology and big data technology, face recognition technology has gained more technical support, its technology continues to improve and mature, and there will be more and more applications related to face recognition. By analyzing the related technologies of face recognition and deep learning, this paper explains the traditional face recognition algorithms, focusing on the PCA and LDA algorithms based on dimensionality reduction and the LBP operator and HOG features based on hand-designed design, and briefly summarizes them. Basic principles and advantages and disadvantages. In academia, convolutional neural networks have always been a research hotspot and have been widely used. In the second half of this article, we mainly explain deep learning algorithms, and summarize several types of classic face recognition algorithms, such as the famous residual network that reduces network parameters. In different level， these algorithms have improved face recognition to varying degrees.

Keywords

Face recognition; Convolutional neural network; Residual network.

1. Introduction

Face recognition technology is not only an earlier technology, but also an emerging technology field that is full of vitality and full of academic research charm. With the increase in the technological innovation of artificial intelligence, big data, and cloud computing and the acceleration of technological change in recent years, face recognition, as an important application of artificial intelligence, has also caught up with these three "express trains". A series of products based on face recognition technology have achieved large-scale landing. Deep learning is also called deep neural network. Compared with general neural networks, its network is deeper, more computationally intensive, more complex, and more difficult to train. However, when the training data is extremely rich, deep learning can find the inherent characteristics and laws of the data, which is incomparable to traditional neural networks. As a branch of machine learning, deep learning has a very important position in the field of artificial intelligence. Its goal is to establish a neural network model with thinking ability like human beings, which can make corresponding intelligent responses to events.

Convolutional Neural Network (CNN), as a typical representative of deep learning technology, is different from other network models. It has two special designs: ① Convolutional layer, which uses local connections and weight sharing This greatly reduces the number of network parameters that need to be trained, and at the same time makes the network model simpler. ② Reduce the dimensionality through the pooling layer, reduce the amount of calculation, and also make the network robust and prevent over-fitting. Convolutional networks are mainly used to recognize zoom, displacement and other forms of distorted two-dimensional graphics, such as handwritten numbers [1], license plate and house number recognition.

2. Face Recognition Algorithm Based on Dimensionality Reduction

2.1. Principal Component Analysis

In 1901, Pearson et al. proposed Principal Component Analysis (PCA), which is the most widely used data dimensionality reduction algorithm. The main idea of PCA is to map n-dimensional features to k-dimensions. This k-dimension is a brand new orthogonal feature, also called principal component, which is a k-dimensional feature reconstructed on the basis of the original n-dimensional feature. Combining the pixel information of each column of a picture into an mX1 matrix in turn will cause the dimensionality to be too high (increased interference information), so the idea of dimensionality reduction is needed [2]. Optimization goal: The input picture has the largest covariance matrix, that is, the largest variance at the same latitude, and the correlation between different dimensions is zero. Meet the requirement of maximum variance, that is, the eigenvalue of the original covariance matrix is the largest. Therefore, choosing the largest eigenvalue corresponding to the eigenvector is the best way of this kind of projection. Disadvantages: In the case of complete ignorance of the data, PCA transformation cannot obtain better data information. When each column of the image should be connected to make the image dimension too large, the speed of feature extraction with a large amount of calculation is too busy to achieve the effect of timeliness [3].

2.2. Linear Discriminant Analysis

The method proposed by R.A. Fisher in 1936 is to minimize the intra-class distance and maximize the inter-class distance after the sample data is mapped to another feature space. It is a dimensionality reduction technique of supervised learning. Let u mean the value of each sample.

Inter-class dispersion matrix:

$$S_b = \sum_{i=1}^{m} (u_i - u)\,(u_i - u)^T \tag{1}$$

Intra-class dispersion matrix:

$$S_w = \sum_{i=1}^{m} \sum_{x \in s_i} (x - u_i)\,(x - u_i)^T \tag{2}$$

Restrictions:

$$\begin{cases} max\,J\,(w) = \dfrac{w^T S_b w}{w^T S_w w} \\ s.t., w^T S_w w = 1 \end{cases} \tag{3}$$

Lagrangian function:

$$J(w) = w^T S_b w - \lambda(w^T S_w w - 1) \tag{4}$$

$$S_w^{-1} S_b w = \lambda w \tag{5}$$

Therefore, selecting the eigenvector corresponding to the largest eigenvalue is the best direction of projection.

Disadvantages: Large limitations, limited by sample types, the dimension of the projection space is at most $N - 1$ dimensions of the number of samples.

2.3. The Similarities and Differences between PCA and LDA

The same point: Both can reduce the dimensionality of the data. When reducing the dimensionality, both use the matrix dimensionality reduction method, and both assume that the data conforms to the Gaussian distribution.

Difference: Assuming that a piece of data is n-dimensional, LDA can only be reduced to n -1 dimensions, while PCA is not restricted. From a mathematical point of view, LDA selects the projection direction with the best classification feature performance, while PCA selects the maximum variance of the sample projection. LDA is actually a supervised dimensionality reduction technology, while PCA is actually an unsupervised dimensionality reduction technology. PCA can only be used for dimensionality reduction, LDA can be used not only for dimensionality reduction, but also for classification. section headings are in boldface capital and lowercase letters.

3. Face Recognition Algorithm Based on Manual Design

3.1. LBP

Local binary pattern (LBP) [4,5,6], LBP was proposed by Ojala et al. in 1996. It is an operator used to describe the local texture features of an image. The original LBP algorithm mainly determines the binary code by comparing the gray value of the center pixel of the window with the gray value of the surrounding pixels in a 3 × 3 window.

The algorithm process is described as follows:

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c) \tag{6}$$

$$sign(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{7}$$

Advantages: This method reduces the error caused by not fully aligning the face area within a certain range. Different weights can be assigned to different regions. Disadvantages: Only the size of the pixel values in the local area is compared, and the relationship between them-the transformation of the image gray level is not considered. Partial key information may be lost, so the extraction of feature information is not comprehensive.

3.2. Hog

HOG feature [7] is better than the previous manual feature extraction methods in its feature extraction efficiency. The HOG operator is used to extract image features. Descriptors describing the gradient direction and intensity distribution. HOG is suitable for grayscale images. Its algorithm flow: (1) Grayscale the input image, convert the RGB image to a grayscale image with a ratio of 3: 6: 1. When the image photo is blurred, it can be corrected by Gamma, and its function is to increase the overall brightness of the image or reduce. After the input image color space is normalized, the advantages of using HOG to extract image features: it can effectively describe the feature information of the local area of the image; it has rotation and illumination without distortion. (2) Divide the picture into several cells, and calculate the gradient histogram of each cell. Need to calculate the horizontal and vertical gradient and calculate the angle. Using gradient calculation, you can find the gradient in the x direction and the y direction. Then use the formula to find the amplitude and direction of the corresponding pixel.

Horizontal gradient:

$$g_x = \sqrt{(f(x-1, y) - f(x+1, y))^2} \tag{8}$$

Vertical gradient:

$$g_y = \sqrt{(f(x, y+1) - f(x, y-1))^2} \tag{9}$$

Angle:

$$\theta = \frac{g_x}{g_y} \tag{10}$$

(3) Divide the cell to form multiple blocks into one block, and normalize the features, which can make the features more robust. (4) Combine all the feature vectors, count the HOG features of all overlapping blocks in the window, and connect them in sequence to obtain a complete HOG feature.

3.3. Haar

The Haar [8] feature is the Haar-like feature, also known as the Viola-Jones recognizer, which is a commonly used feature description operator in the computer vision field, and is mostly used for face detection, pedestrian detection, and other target detection. There are only white and black rectangles in the Haar-like feature template. When the template is used to extract different areas in the image, the difference between the sum of the gray values of the pixels in the black area and the white area in the template is taken as the template extracted feature. Different templates are used to extract the features of the pictures, and finally the more representative features are screened out and then classified using a strong classifier. The Haar eigenvalue reflects the gray level changes of the image.

4. Face Recognition Based on Deep Learning

4.1. AlexNet

AlexNet[9] is a classic neural network designed by Professor Hinton's student Alex Krizhevsky in 2012. The neural network achieved the best results in the ImageNet competition that year. The difference between AlexNet and traditional recognition methods is that it uses deep learning ideas and uses GPUs to achieve accelerated calculations. In the design of the network structure, an 8-layer neural network is adopted, including 5-layer convolutional neural network and 3-layer fully connected network. The convolution layer is the most important layer in the entire neural network. The core part of this layer is the filter, or convolution kernel. The process of convolution is to use these weight values to continuously multiply the RGB values of these pictures. To extract picture data information. Convolution not only extracts image information, but also achieves dimensionality reduction effects. The pooling layer can effectively reduce the size of the matrix, thereby reducing the parameters in the final full chain layer. Using the pooling layer can speed up the calculation speed and prevent over-fitting problems. The role of the full chain layer is to perform correct image classification.

4.2. VGGNet

VGGNet [10] is a deep-level CNN architecture [11]. VGG is an algorithm for image classification that the Oxford University team participated in in 2014, and the accuracy of image classification in winning the championship is very high. A total of six different models have been developed for the VGGNet model. The different suffix values connected after VGG represent different network layers. The outstanding contribution of VGG16 introduced in this section lies in the use of a 3X3 convolutional layer. Increasing the depth of the network can help improve the recognition rate of the model. The most used VGG-Net models are VGG-16 and VGG-19. For example, there is a 16-layer convolutional neural network VGG-16. The main feature is that the network structure has 13 convolutional layers, all of which are 3X3 convolutions and the number of channels increases sequentially, all using maximum pooling, and finally composed of three fully connected layers.

4.3. GoogLeNet

GoogLeNet was proposed by Google engineer Christian Szegedy in a paper published with CVPR2015. The network participated in the ImageNet competition in 2014 and surpassed VGGNet to win the championship. GoodLeNet is an important network in the development of deep learning, not only because it increases the number of network layers to 22, but also

because of the idea of the Inception module [12]. This module jumps out of the basic network structure proposed by AlexNet and has an important impact on the development of deep convolutional neural networks.

4.4. ResNet

With the increase in the number of network layers, it is inevitable that the gradient disappears and explodes, resulting in saturation or even decline in the accuracy of the training set. Therefore, the network depth cannot be increased arbitrarily. For this problem, Kaiming He [13] et al. proposed a network module—the residual module. This residual structure handles the deep degradation of the network well. When the network depth is up to 1,000 layers, the accuracy rate will not drop too much, that is, the residual network is easier to optimize, and the accuracy of the network will be greatly improved as the number of layers increases. For example, the residual depth of 152 layers is many times higher than that of VGG, but the complexity of the network is low, so the residual network structure is better than VGG.

5. Conclusion

In recent years, domestic and foreign scientific research scholars have made many substantial breakthroughs in the research of face recognition technology. Compared with other machine learning face technologies, deep learning has key advantages: first, it can learn from features. It can detect complex interactions in the middle; secondly, it can learn low-level features from the raw data that is hardly processed; moreover, it can process both high-cardinality data and unlabeled data. Furthermore, he is particularly good at big data processing related aspects. Therefore, deep learning can learn more useful data and build more accurate models. However, deep learning also has some imperfections. For example, it takes a long time to train the model, even up to several weeks. It requires continuous iteration to optimize the model, and it cannot guarantee the global optimal solution. The setting of hyperparameters needs to be determined based on experience and practice, and how to perform efficient numerical calculations under the background of large data volume and deep network structure.

References

[1] Jing Chenkai, Song Tao, Zhuang Lei, Liu Gang, Wang Le, Liu Kailun. Overview of Face Recognition Technology Based on Deep Convolutional Neural Networks. Computer Applications and Software, 2018, 35(1): 223-231.

[2] Li Mengxiao, Yao Shiyuan. Design and Improvement of Face Recognition System Based on PCA[J]. Computer Science, 2019, 46(z1): 577-579.

[3] Li Gang, Gao Zheng. A Review of Automatic Face Recognition Methods [J]. Computer Application Research, 2003, 20(8): 4-9, 40. DOI:10.3969/j.issn.1001-3695.2003.08.002.

[4] Ojala T, Pietikanen M, Maenpaa T. Multiresolution Gray-scale and Rotation Invariant Texturee Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.

[5] Zhang G, Huang X, Li SZ, Wang Y, Wu X. Boosting Local Binary Pattern(LBP)-Based Face Recognition. Chinese Conference on Advances in Biometric Person Authentication, 2004, 3338:179-186.

[6] Yang B, Chen S. A Comparative Study on Local Binary Pattern (LBP) Based Face Recognition: LBP Histogram Versus LBP Image. Neurocomputing, 2013, 120(10):365-379.

[7] Zhou H, Lin D J, Huang T S. Static Hand Gesture Recognition Based on Local Orientation Histogram Feature Distribution Model[C]//Computer Vision and Pattern Recognition Workshop, 2004: 161.

[8] Bhavana R. Maale, Dr. Suvarna Nandyal. Face Recognition Based on Haar Cascade Classifier[J]. Journal of Research in Science and Engineering,2021,3(5).

[9] Yang Xu,Shang Zhenhong. Facial Expression RRecognition Based on Improved AlexNet[J]. Progress in Laser and Optoelectronics,2020,57(14):243-250.

[10] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-scale Image Recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[11] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-scale Image Recognition[J]. Computer Science, 2014.

[12] Song Chunhe,Li Zexi,Yu Hongxia,Liu Yiyang,Feng Tieying,Zhang Xuejian. An Oil Well Fault Identification Method Based on Improved GoogLeNet[J]. Journal of Jiangsu University of Science and Technology (Natural Science Edition),2021,35(02):52-58.

[13] Wu S, Zhong S, Liu Y. Deep Residual Learning for Image Steganalysis[J]. Multimedia Tools and Applications, 2018, 77(9): 10437-10453.