Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 7
TPCEE 2024

Overview of Mainstream Face Recognition Methods

Zehao Wu *

Department of applied physic, Hebei University of Technology, Tianjin, China

* Corresponding author: zehaowu@d2l.arizona.edu

Abstract. Human face is one of the most important features of human beings. In the information age, the face has personal information and emotional expression, social communication, public safety and other extremely important attributes. Due to the characteristics of inconspicuous facial features and small differences, traditional face recognition by manual training leads to low efficiency and high error rate. Based on the advantages of computer automation, face recognition technology has tended to be efficient, accurate and intelligent. Face recognition in a broad sense includes four stages: portrait image preprocessing, eigenface display, data training and data production. The use of CNN neural network can deal with complex abstract facial features. OpenCV open-source function libraries and Python simple and efficient machine language have advantages. This paper will focus on CNN deep neural network face recognition principle and OpenCV face recognition process two mainstream methods, as well as their defects comparison and conclusions.

Keywords: Face recognition; Neural network; Database; Open-Source Computer Vision Library.

1. Introduction

Facial recognition system is a new research hotspot in the field of computer vision. Facial recognition is based on extracted facial features. At present, facial features are transformed from geometric features to experience-driven artificial features, and finally to data-driven representation learning methods. Face recognition technology originated from Galton's paper in Nature in 1888, which proposed to represent face profile features by numbers, and constructed a framework for the method of data facial features. From 1965 to 1990, the research on human face has been stuck in the primary research stage of using facial geometric features and temple-based matching methods. Due to the influence of lighting environment, loss of facial information caused by occlusion, changes in facial expression and other factors, facial feature points cannot be accurately located, resulting in blurred face level recognition and stop judgment. After 1991, FERET portrait image database was established under the lead of the U.S. Department of Defense. In this stage, Eigenface algorithm recognition method and Fisherface algorithm based on subspace analysis were produced, but they were only suitable for ideal portrait images and small and medium-sized database conditions. From 1998 to 2013, in order to overcome the fundamental problems such as pixel gray value and lighting environment resulting in a wider variety of face recognition algorithms and models.

Linear discriminant analysis is mainly used to represent the linear modeling method and 3D face reconstruction method for manual feature extraction of pixel brightness or color value [1]. In 2012, AlexNet first proposed deep neural network learning methods such as convolutional neural network. Through multi-level model processing methods, it has stability and non-deformation in dealing with lighting environment, facial pose and human expression actions [2], which solves the historical problems of face processing and performs well in understanding abstract features (facial expression, smile, etc.). For example, DeepFace achieves 97.35% accuracy on the well-known LFW benchmark by training a 9-layer model on 4 million face images, which is the first time to approach human performance of 97.53% under unconstrained conditions [2]. The best conventional face recognition system at that time had an acceptance rate of 1% and an accuracy of only 50% on faces captured outdoors [3]. Nowadays, face recognition technology has been applied in access control, information payment, attendance system, camera photo positioning and other fields of life. Different from mainstream biometric recognition methods such as fingerprint recognition, palmprint recognition and

iris recognition, face detection is efficient and imperceptible. Objectively speaking, it meets the conditions such as visible light and face acquisition equipment, and can all collect the human face information, such as in the camera acquisition system. However, due to the lack of large-scale public datasets in the field of face recognition, most recent progress is still limited [4]. In addition, due to the fact that faces involve personal privacy, identity security and ethical issues, as well as legal restrictions, etc. (In 2019, California introduced a bill, and San Francisco became the first city in the world to restrict face recognition [5]), In addition, literature surveys show that the public's acceptance of face recognition technology is generally not high. For many reasons, the application of this technology is limited at the present stage and it has not been studied and applied in a wider field.

The most critical part of face recognition technology is the identification of facial features. Usually, in the component-based method, face recognition is based on the relationship features between faces, such as eyes, mouth, nose, contour and face boundary [6]. Because the face organ and other human organs have obviously different characteristics, modern technology can easily realize the rapid location of the face. However, the current face recognition is limited by the following:

1) Due to the limitation of equipment performance, the resolution of camera acquisition equipment is not high, so the facial information features are not obvious, and the recognition algorithm is difficult to distinguish.

2) The human facial features of different people are very similar. In addition, the human facial shape cannot be fixed and will change with age.

3) Identify environmental conditions that cause unstable recognition results, such as light in the day and night environment, or racial color reasons. The test results in ACLU show that the recognition of women with deep skin is the most unsuccessful, with an error rate of 34.7%[5], and facial occlusions (eye masks cause facial information loss).

2. Mainstream face recognition methods

2.1. Method based on Convolutional Neural Network (CNN) model

Deep machine learning includes CNN and traditional neural network (NN). Compared with traditional machine learning, it does not need manual processing of features in input feature processing, as long as it is output by the network. It is often used in image, natural language and translation processing, so it has great advantages in dealing with complex features of face images. Nowadays, the recognition error rate of deep learning network is far lower than that of human recognition and traditional recognition. At present, almost all computer vision tasks can be completed by CNN model training, such as image detection and tracking, graphics classification and retrieval, super-resolution reconstruction, which is now widely used in unmanned driving, font and face recognition.

2.2. Facial feature recognition process

CNN consists of input layer, convolutional layer, activation function, pooling layer and fully connected layer. The convolution layer is used to extract image features, pool compression features, and the fully connected layer is used to weight the weight. The main parameters of the convolution layer include: sliding window step size, convolution kernel size, filling edge, and the number of convolution kernels. How a convolution kernel works: For example, the input image is 32 * 32 * 3, 3 is its depth, that is, the three channels R, G and B, through the convolution kernel, in the figure, the convolution kernel is 5 * 5 * 3 filter, where the depth of the filter must be the same as the depth of the input image, that is, the corresponding 3, there can be many filters. A filter convolved with the input image results in a 28 * 28 * 1 feature map. For a 32 * 32 *3 image, we apply the convolution operation with 10 5 * 5 *3 and fliters with a step size of 1 and edge padding of 2.

The size of the output is 32 * 32 * 10, and the length and width of the feature map can also be kept unchanged after the convolution operation. The feature map is obtained by multiplying the input image and the corresponding position element of the filter, then summing, and finally adding b0(bias).

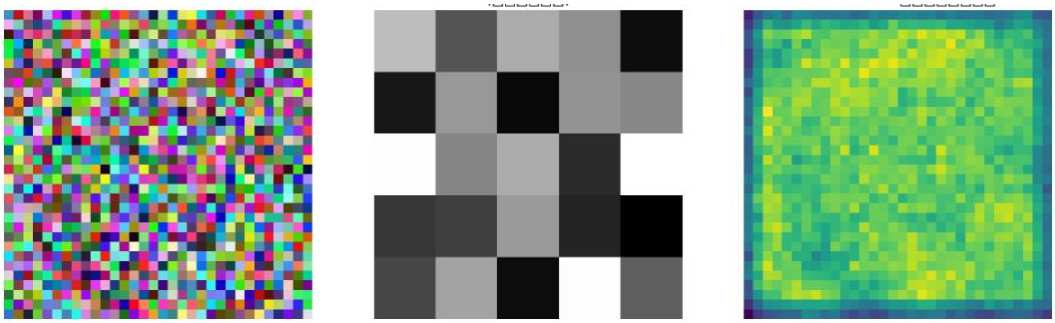

Figure 1. Facial feature recognition process

On the left is the input image (32x32x3) with three RGB channels in Fig.1. In the middle is a convolution kernel (5x5x3) with the same depth as the input image. On the right is the feature map after the convolution operation (32x32x1), using edge padding and a step size of 1.

The first layer and the corresponding elements in the box in the input image are multiplied and added to get 0, and the depth of the other two layers is 2 and 0 respectively, so 0+2+0+1 is 3, until the first element of the feature map on the right is 3. After the convolution operation, the blue box of the input image slides in step size 2. After completing the convolution operation, a 3x3x1 feature map is generated.

2.3. Edge filling

The principle of edge filling is to add a boundary to the image through the zero pad item, and the boundary elements are all 0, in order to better extract the features of each part of the boundary, so that the features of the boundary points can also be extracted many times, which is the role of padding. It is usually necessary to extract image features through multiple convolutions.

2.4. Convolution parameter sharing

The convolution operation can realize the local connection of the input image, thereby reducing the number of parameters in the network model. The number of parameters can be further reduced by exploiting another property of images. For example, if there are 10 convolution kernels of shape(5,5), each kernel corresponds to a parameter matrix in the original image, and the size is 5*5. If we didn't share weights, we'd get 5 * 5 * 3 * 32* 32* 10 +10 parameters. There will be 5 * 5 * 3 *10 + 10 parameters if we share them

2.5. Convolutional layers involve parameters

1) Step size of sliding window: the more it can move, the larger the feature map will be, and the more delicate the extracted features will be. The common step size is 1.

2) Convolution kernel size: select the size of the region - finally get the size of the number of results, generally 3×3 size.

3) Edge padding: it is selected according to the step size, and some elements are repeatedly weighted to contribute. The inward points contribute more, the outward points contribute less, and the boundary points contribute more. Adding a circle of 0 outside can make up for the lack of some boundary features.

4) Number of convolution kernels: How many feature maps should we get? Note that each convolution kernel is different

5) Convolutional parameters sharing: the same set of convolutional sums is used to extract features of each region in the image

2.6. Pooling layer

The pooling layer is an important part of the common CNN component, which is called 'downsample'. Pooling layer retains the main information by extracting the main features of the feature map. In order to reduce the computation of the lower layer, this dimensionality reduction method retains the main information of the image with a smaller feature map for data dimensionality reduction. The operation mode of pooling layer includes maximum pooling, mean pooling, random pooling, median pooling, combination pooling and so on.

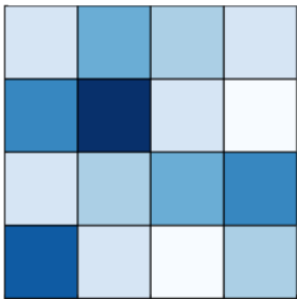

Figure 2. feature map (4x4)

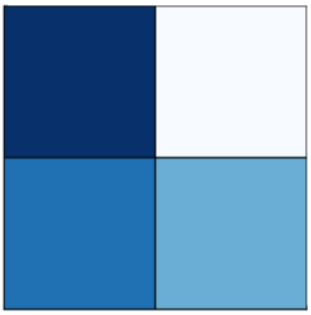

Figure 3. feature map (2x2)

The image shows the max pooling process. Fig. 2 is the input feature map (4x4) and on the right is the pooled feature map (2x2). The pooling layer slides (fig.3) in a 2x2 window and takes the maximum value in each window to generate smaller feature maps for compression and feature selection.

2.7. Activation function

The Activation Function is responsible for the weighting and nonlinear transformation of the input in the CNN, thus enabling the neural network to learn and represent complex patterns and features. If there is no activation function, the neural network is essentially a linear combination of the previous layer and the next layer, becoming a linear model. However, the activation function is used to introduce nonlinear factors so that the neural network can approximate any nonlinear function. This function improves the nonlinear modeling level of CNN, and the linear combination obtained from the original linear function approximates the linear function through the activation function to solve the complex task. Activation functions can also be used to stabilize the training process of the network by limiting the output to a certain range.

The Sigmoid function, known as the threshold of neural network, is a smooth and continuous S-shaped curve, which is conducive to gradient calculation and solving the binary classification problem through the input probability value. The output gradient of the function is infinitely close to 0 when the input value is extreme, which is easy to cause the gradient disappearance problem, and the output value between 0 and 1 leads to the non-0 centralization of the data and slows down the convergence speed

$$f(x) = \frac{1}{1+e^{-x}} \tag{1}$$

The Tanh function converges faster than Sigmoid at the 0 output point, and the output variable is between negative 1 and 1. The Tanh function is more widely used than the Sigmoid function, but the gradient disappearance problem still exists under extreme input values.

$$f(x) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} \tag{2}$$

The positive input region gradient of Relu function is always 1, which helps to alleviate the problem of gradient disappearance, and has a wider application range than Sigmoid and Tanh function.

$$\text{Relu} = \max(0, x) \tag{3}$$

2.8. Fully connected layer

The Fully Connected Layer is the basic component of artificial neural networks (Fully Connected Layer), also known as the dense layer. Different from the convolutional layer, pooling layer and activation function to extract features, in the fully connected layer structure, each input node is connected to each output node, through the weight of the connection, that is, the parameters that need to be adjusted in the neural network training project. After the convolution layer and the pooling layer, the fully connected layer is easy to process the extracted features, and the output results are finally obtained through combination and transformation. By using activation functions (ReLU, Sigmoid, Tanh), the fully connected layer can carry out more complex nonlinear mapping to improve the representation ability of the CNN model. However, because the parameters are affected by the size of the input and output dimensions, the output and input of larger dimensions have certain requirements for computer storage, and the number of parameters and calculation amount are very large, resulting in low efficiency. Methods such as regularization and Dropout are often used to deal with the shortcomings of large number of parameters and easy overfitting.

3. Method based on OpenCV database

OpenCV (Open-Source Computer Vision Library) is the most commonly used cross-platform computer vision library, which can be used for free in the commercial and visual fields. The underlying language is developed by C++, but it also provides C and Python language interfaces. OpenCV library can realize image processing functions, such as portrait and gesture recognition, motion pattern recognition and so on. As shown in Fig. 4, the conventional processing process includes face data import into OpenCV library, image data processing (gray scale conversion and image size correction, etc.), and then face detection data training.

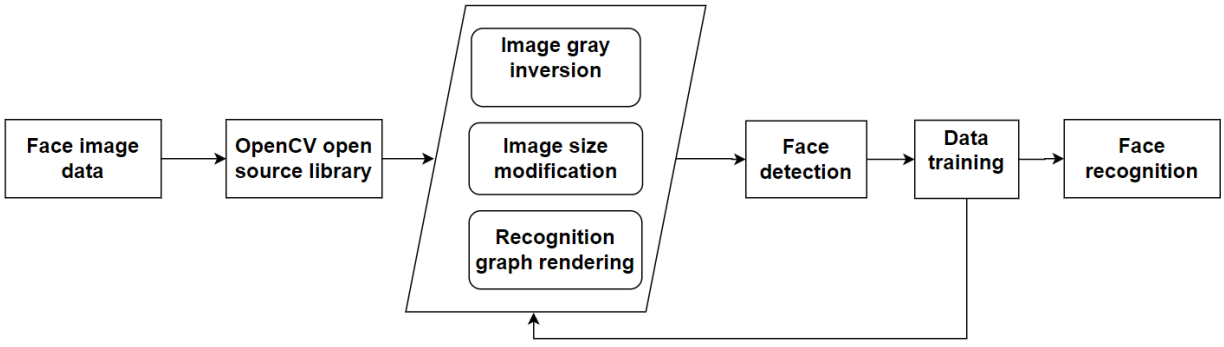

Figure 4. Face recognition process based on OpenCV

3.1. Gray transformation

Color image is a combination of three colors: red, green and blue. Gray transformation is a method to transform the image into gray color by removing the color information of the image itself, which can highlight the essential information of the image and enhance the features of the image. In essence, the computer can not intuitively understand the picture information like human beings, so it is necessary to digitize the image information to allow the computer to process the data.

Cv.cvtcolor (img, cv.color_bgr2gray) function in OpenCV library is used to process gray data, and then cv.imshow() function is used to generate gray portrait, and finally cv.imwrite() function is used to save gray image. Fig. 5 and Fig. 6 were obtained by using CMU Multi-PIE Face database and

FDDB (Face Detection Data Set and Benchmark) face data set through gray transformation processing.

Figure 5. CMU Multi-PIE [7]

Figure 6. FDDB [8]

3.2. Location area

To realize the rapid recognition of face and face, we must accurately locate the collected face image, obtain the rough face area through planning and positioning, and then coordinate the portrait area. The rectangle (frame, (x, y), (x+w, y+h)) function is used to generate a rectangle with length x and height h. cv.circle(img,center=(x,y),radius=r) means to generate a circular positioning frame with center coordinates x and y and radius r. For example, the circular and rectangular positioning in fig 7.

Figure 7. Locate the face area

3.3. Image processing phase

Processing flow consists of the following steps in fig. 8:

1) Load the pre-trained face detection model: use the Haar cascade classifier for face detection.

2) Load a video stream or image: You can choose to capture the video stream from the camera in real time or load a local video file.

3) Face detection: Convert the image to a grayscale image, and then use the detectMultiScale method for face detection.

4) Draw the detected face frame: Draw a rectangular box on the original image to identify the detected face.

5) Face recognition: recognize faces through pre-trained face recognition models.

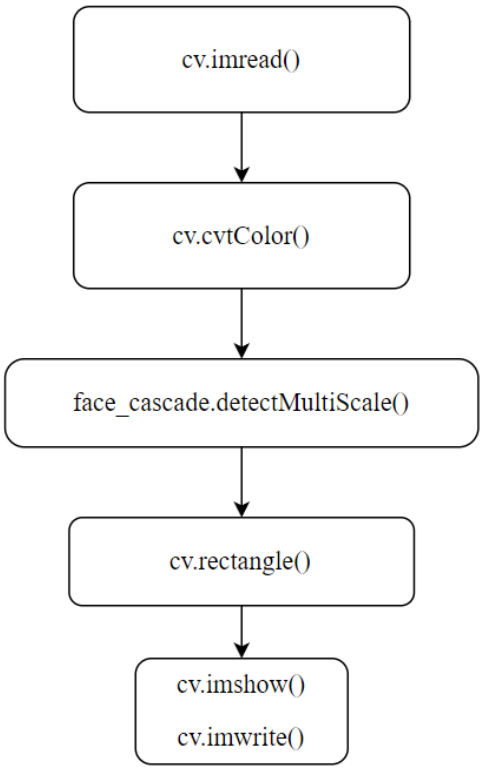

Figure 8. Processing flow

3.4. Defect comparison and trend

In terms of local perception, convolutional neural network CNN can use convolutional check images for local perception, which means that CNN can better capture local features, process detailed information while maintaining spatial invariance, and pooling operation can effectively reduce the computational load of feature images while retaining the main feature information, reflecting the high efficiency of CNN. It can be used efficiently on computer GPU hardware. Unlike NN neural network, its process is divided into multiple stages, which is reflected in the high computational complexity caused by the large number of layers, high requirements and high dependence on data training, and the lack of certain training data will lead to overfitting or insufficient model generalization ability. As a result, CNN model performs poorly in the real field environment (face occlusion and illumination influence) test. The difficulty lies in the different sizes and shapes of different locations, which makes it difficult to fit the model and reduces the model accuracy [6], and cannot effectively recognize facial expressions. However, the improved Convolution Neutral Network with Attention Mechanism (ANN) based on CNN can interpret the issue of facial image occlusion, so the CNN model has obvious defects in facial feature occlusion. OpenCV has advantages as a highly optimized and real-time application library for computer vision task processing, but it is not always suitable for complex tasks, especially in the areas of deep computer vision and deep learning. OpenCV is not suitable for all scenarios, although it is rich in features and provides multiple language interfaces. But not all tasks are suitable for OpenCV. In some cases, other specialized tools or libraries may be better

suited to specific needs. In addition, face recognition technology also needs to conduct research on the accurate positioning of faces in videos and the related development of face 3D modeling [9].

4. Conclusion

This paper describes the development and mainstream of face recognition technology. The development of face recognition technology began in 1888 and went through the initial feature extraction stage until AlexNet proposed a method based on deep neural networks. Deep learning solves the limitations of traditional methods in complex environments, and is the mainstream trend in the future face recognition field. Facial expression recognition task generally consists of four parts: face image acquisition, image preprocessing, feature extraction and feature classification Convolutional neural network (CNN) -based method: extracting complex facial abstract features is widely used in computer vision tasks such as image detection and classification. CNN includes input layer, convolution layer, activation function, pooling layer and full connection layer. Image features are extracted through convolution kernel, pooled compressed features, and weighted by the fully connected layer. It is a facial feature recognition process that uses input images to extract features through the convolution layer, then performs pooled compression, and finally performs classification recognition through the fully connected layer. The OpenCV library method is relatively simple, the main process is divided into face data import, image processing stage including gray transformation, size correction, data training several stages. Face recognition technology still faces various challenges such as equipment performance limitations such as facial information features are not obvious, recognition algorithms are difficult to distinguish, environmental factors such as lighting lead to unstable recognition results, and facial occlusion will also lead to facial information loss resulting in recognition. In the future, face technology will develop in the direction of more intelligent, safe and fair in the fields of deep learning to promote technological progress, cross-field integration application, attention to privacy ethical issues, and improvement of cross-racial and cross-gender identification accuracy.

References

[1] Jing C K, Song T, Zhuang L, Liu G, Wang L, Liu K. A review of face recognition technology based on deep convolutional neural networks. Computer Applications and Software, 2018, 35 (1): 223 - 231.

[2] Wang M, Deng W. Deep face recognition: A survey. Neurocomputing, 2021, 429: 215 - 244.

[3] PHILLIPS P J, Grother P, MICHEALS R J, BLACKBURN D M, Tabassi E, Bone M. Overview and Summary, 2003.

[4] Parkhi O, Vedaldi A, Zisserman A. Deep face recognition. BMVC 2015 - Proceedings of the British Machine Vision Conference, 2015, 1 – 12.

[5] Xing H Q. Legal regulation of face recognition. Comparative Law Research, 2022, 5: 51 - 63.

[6] Karamizadeh S, Abdullah S M, & Zamani M. An overview of holistic face recognition. IJRCCT, 2013, 2 (9): 738 - 741.

[7] CMU multi-PIE Face database. http://www.flintbox.com/public/project/4742/, 2024/8/1.

[8] Face Detection Data Set and Benchmark face data. https://vis-www.cs.umass.edu/fddb/, 2024/8/1.

[9] Yan Y, & Zhang Y J. Advances in video-based face recognition. Acta Computer Sinica, 2009, 32 (5): 878 - 886.