Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/35/20230404

A study on the CNN-based face recognition across various facial angles

Zibin Chen

Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong, China

20100359d@connect.polyu.hk

Abstract. Ace recognition is a prominent computer technology with a broad range of applications, such as identity verification and face payment. Over the years, significant advancements have been made in the research and practical implementation of face recognition in academia. However, the variability of face images in real-world scenarios poses challenges to accurate face recognition. Specifically, the variation in facial angles affects the extraction of crucial face information by the model, yet there is a lack of research focusing on the angle factor's impact on face recognition. In this study, we collect and process face datasets, construct a convolutional neural network (CNN) model, capture images containing faces via a camera, detect and track faces from it automatically, and then recognize the detected faces. We investigate the accuracy of face detection and recognition under different angles. Experimental results demonstrate that when the angle ranges from 0 to 30, face detection and recognition achieve excellent accuracy. Within the angle range of 30 to 45, the performance remains within an acceptable range. However, when the angle exceeds 45 and reaches 60, face detection and recognition accuracy decline significantly, resulting in poor performance.

Keywords: ace recognition, multiple angles, CNN, LeNet-5.

1. Introduction

Face recognition is a biometric technology that obtains information by detecting facial feature information, which has become one of the most invested topics by scholars in biometrics and computer vision since the last century. The process of face recognition mainly contains face image acquisition, data preprocessing, feature extraction, identification [1].

Traditional methods rely on hand-crafted model features [2], such as edge texture descriptions, and methods that incorporate machine learning techniques such as principal component analysis [3], linear discriminant analysis [4], and support vector machines [5]. However, in real life due to huge variations in lighting conditions, facial expressions, and other factors, traditional methods may not adequately represent faces, so in recent years, they are gradually replaced by convolutional neural networks-based deep learning methods [6].

Its key advantage is that it utilizes very large datasets for training and thus learns the best features to characterize them. And with the large-scale face datasets available on the web today, face recognition methods based on CNN trained have achieved very significant results [7-8], because they are able to extract and learn the robust features in the images and thus are able to cope with the variability brought by the face images used in the training process.

DOI: 10.54254/2755-2721/35/20230404

Research by Belmonte and Allaert proposes that when deployed in unconstrained environments, face images are highly variable in their presentation in the real world [9]. The angle, as one of the essential factors affecting the deformability of the face image, varies in a way that affects the extraction of key information about the face by the model, which in turn affects the results of face recognition. At present, different scholars who study face recognition have proposed various approaches to improve the effect of the result in response to the challenges of pose change, light change, partial occlusion, etc., but there is still a lack of research on face detection and recognition under multiple angles. To explore the effect of angle variation on face recognition accuracy, this paper uses a convolutional neural network-based approach to explore the changes in face recognition accuracy under different angles.

2. Methodology

The methodology used in this study is divided into three main steps: obtaining a collection of face images, training a neural network model, and recognizing faces at different angles according to the model.

2.1. Introduction to the dataset

The dataset used during the experiment is LFW, which is popular to be used to investigate face recognition problem which is under unrestricted conditions. The images are mainly collected online, and contains 13233 images, each of which is consistent with the corresponding person's name, among which the face images are presented with multi-angle features, which is conducive to the training of the model in this study (Figure 1).

Figure 1. Part samples of LFW database.

2.2. Face detection

dlib is a powerful C++ toolkit that contains many features for machine learning and computer vision. A face detector is one of the toolkits that can be used to detect faces in images. The face detector used here is the default $\text{frontal} - \text{face} - \text{detector}$ provided by dlib, which is trained on faces by the training tool that comes with the dlib library. The detector is applied to the input image, and the return value is a rectangular feature box $(< \text{class}'\text{dlib.dlib.rectangles}' >)$, which has the following coordinate attributes (Figure 2).

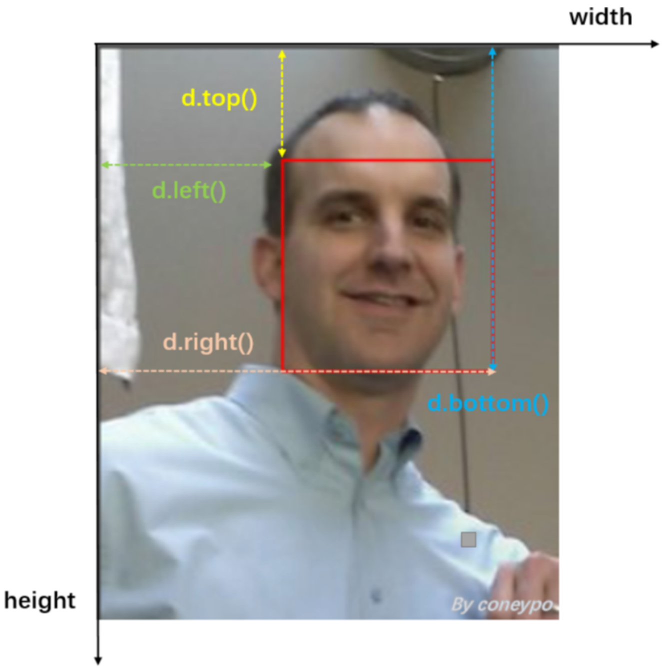


Figure 2. Attributes of the return value.

2.3. Face recognition and identification

CNNs usually contain an input part and a hidden part. The input part is the layer with pre-processed data input. Because of the gradient descent used for learning, normalization needs to be conducted on the input features. Besides, before feeding the learning data, normalization of the input data in channel or time/frequency dimensions is required as well. The hidden layer is the core of the entire structure. In a common neural network, it consists of Conv layer, activation function layer, Pool layer, and FC (fully connected) layer.

The role of the Conv layer is to do feature extractions by processing the data, the main thing is that there are multiple convolutional kernels, where the elements of each kernel corresponds to a weight coefficient (w) and a bias vector $[wX + b]$.

As follows, the dimension of the input data is calculated as N and the dimension of the output is given by the formula:

$$output = \frac{N-f+2p}{s} + 1 \tag{1}$$

Where f, s, p are the parameters, corresponding respectively to the size of the convolutional kernel, the convolutional step and the number of padding layers. If the input dimension is not N (not a square, but a rectangle), the height and width of the rectangle are substituted into the formula, and the two outputs are multiplied together as the dimension of the output.

The role of the activation function layer is to assist the convolutional layer in expressing complex features. Rectified Linear Unit (ReLU) is usually used as shown below:

$$f(x) = \begin{cases} 0 \; if \, x < 0 \\ x \; if \, x \geq 0 \end{cases} \quad f'(x) = \begin{cases} 0 \; if \, x < 0 \\ 1 \; if \, x \geq 0 \end{cases} \tag{2}$$

The Pool layer is responsible for taking the feature map passed from the convolutional layer and then perform feature selection and information filtering.

The role of the FC layer is to classify the incoming data or images, usually by reducing the dimensionality of the vectors transmitted from the pooling layer to one dimension.

This study employs CNN techniques, specifically relying on the LeNet-5 model [10]. LeNet-5 is an eight-layer neural network comprising an input layer, three Conv and Pool layers, a FC layer, and an output layer. The Conv layer is responsible for extracting significant features, the Pool layer aids in reducing dimensionality, and the FC layer contributes to classification. Two categories are identified: faces that correspond to expected individuals (true) and faces that do not correspond (false).

The specific structure is as follows: the 1st and 2nd layer input face image size of 64*64*3, output image size of 32*32*32. The 3rd and 4th layer input 32*32*32, output 16*16*64. The 5th and 6th layer input 16*16*64, output 8*8*64. The 7th layer (the FC layer), the convolutional output of the image is squashed into a one-dimensional vector, input size of 8*8*64, reconstruct to 1*4096, output 1*512. The 8th layer inputs 1*512, output 1*2, add again, output a number (Figure 3).

Figure 3. CNN architecture.

3. Results

3.1. Training details

The neural network back propagation optimization algorithm was used in this experiment. Firstly, the prediction results of the model are obtained by the forward propagation algorithm, the cross entropy is calculated, and then the neural network parameters are updated using the back propagation algorithm. This training is an iterative process, i.e., the back-propagation optimization algorithm is run repeatedly on the training data, and the neural network model is finally generated. Get the accuracy of the test data and train until the accuracy reaches 98% stop training save and exit.

For the setting of hyperparameters, this experiment sets the learning rate to 0.01, the optimizer used is the Adam optimizer, and cross-entropy as the loss function.

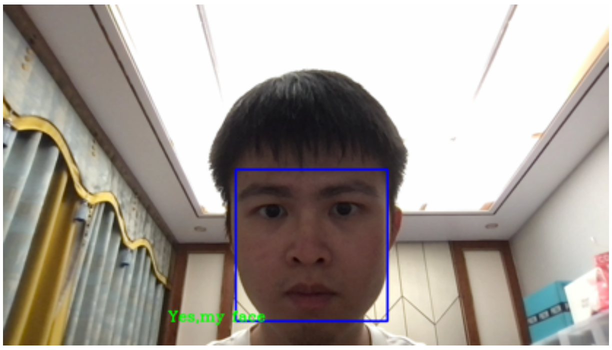
3.2. Comparison of experimental results

The face recognition used in this study is done by capturing the images of the face through the camera and automatically detecting and tracking faces that appear and then the detected face will be recognized. For this experiment, we capture face images at various angles, specifically degree 0, degree 15, degree 30, degree 35, degree 40, degree 45, degree 50, degree 55, degree 60. 30 images are collected for each angle, resulting in a total of 300 images.

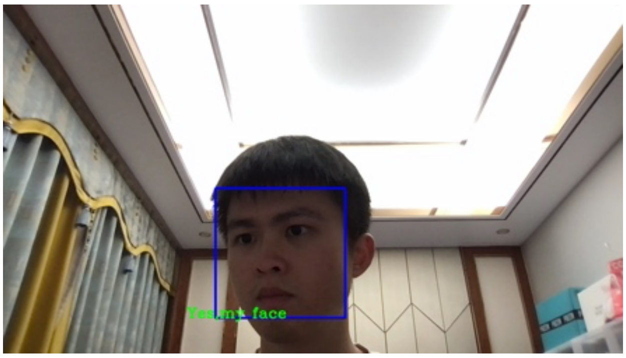
Since the method used in this study first detects the face and further recognizes the detected face, the results of the experiment can be categorized into three cases: yes (Detected and recognized successfully), no (Successful detection but failed recognition), Can't detect faces (Detection failure). The first case is recognized successfully, the last two cases are recognition failure.

When the angle is degree 0, degree 15, and degree 30, all the photos under each angle are detected and recognized successfully, which means that the detector can extract the facial features of the face well when the angle is less than 30 degree and recognize the face successfully (Figure 4). When the angle is degree 35, degree 40, degree 45, there is a detection failure (Successful detection but failed recognition, Detection failure), which indicates that as the angle increases, the extraction effect of the detector decreases, and sometimes even fails to detect the face. When the angle is degree 50, degree 55, degree 60, there is a sharp increase in the detection failure, when the angle is larger, the key information of the face is lost, and the face cannot be detected more often, so the detection and recognition effect is poor.

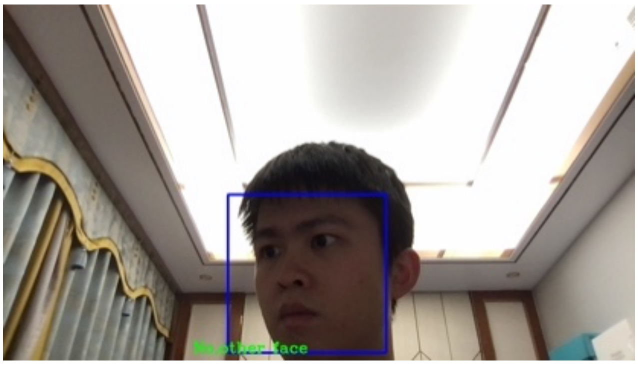
DOI: 10.54254/2755-2721/35/20230404



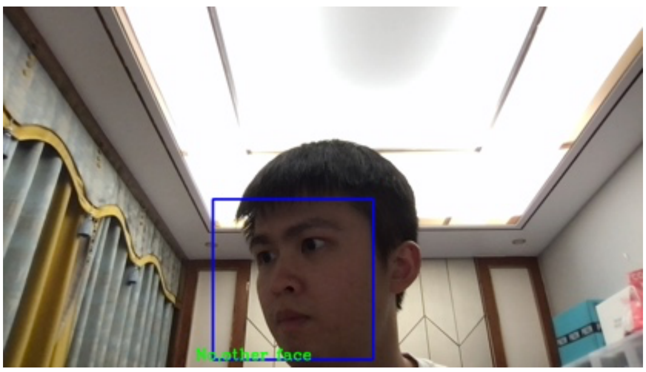
degree 0: Yes(Detected and recognized successfully)



degree 15: Yes(Detected and recognized successfully)

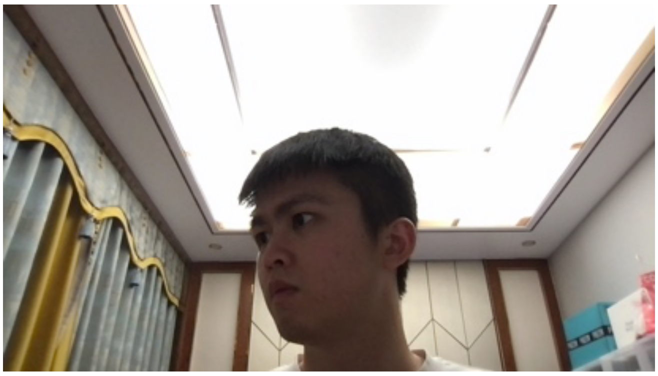


degree 40: No (Successful detection but failed recognition)



degree 45: No (Successful detection but failed recognition)

degree 50: Can’t detect faces (Detection failure)



degree 60: Can’t detect faces (Detection failure)

Figure 4. Results from different degrees.

From table 1 it can be observed that when the angle varies in the range of 0-30 degree, the face detection and recognition effect are excellent, and the accuracy is high. When the angle changes in the range of 30-45 degree, the face detection effect is very good, and the face recognition effect decreases, but it is still at a high level and within the acceptable range. When the angle changes in the range of 45-60 degree, there are many cases where face detection is successful while recognition fails, and even the face cannot be detected in more cases, at this time identifiable features deformation is too large, the detector cannot effectively capture the identifiable features, the face identification performance decline seriously, and the effect is poor.

Table 1. The Results of the Experiment.

Degree	Yes (Detected and recognized successfully)	No (Successful detection but failed recognition)	Can’t detect faces (Detection failure)
0	30	0	0
15	30	0	0
30	30	0	0
35	26	4	0
40	25	5	0
45	25	4	1
50	9	15	6
55	8	15	7
60	2	16	12

4. Conclusion

This research is focused on the specific factor of angle, and it is worthwhile to further investigate other factors affecting face variability, such as illumination, occlusion, and so on.

In this study, we investigate the accuracy of face detection and recognition at different angles by constructing a convolutional neural network and conclude that face detection and recognition is very effective and accurate when the angle varies within the range of 0-30 degrees. When the angle varies within the range of 30-45 degrees, it is within the acceptable range. When the angle varies in the range of 45-60 degrees, the face detection and recognition accuracy decline seriously, and the effect is poor.

In terms of the improvement of the experimental method in this study, to maximize the angle within the acceptable range and improve the face recognition effect, an algorithm that combines face feature fusion with convolutional neural network for face recognition can be considered instead of the LeNet model used in this study. This method extracts more comprehensive face features, avoids the shortcomings of traditional convolutional neural network method that ignores local features when extracting face features, and can theoretically improve the model training effect and thus increase the angular fault tolerance.

References

[1] F. Deeba, H. Memon, F. Ali, A. Ahmed, and A. Ghaffar, “LBPH-based Enhanced Real-Time Face Recognition,” *International journal of advanced computer science & applications*, 10 (5), 2019.

[2] A. Althnian, N. Aloboud, N. Alkharashi, F. Alduwaish, M. Alrshoud, and H. Kurdi, “Face Gender Recognition in the Wild: An Extensive Performance Comparison of Deep-Learned, Hand-Crafted, and Fused Features with Deep and Traditional Models,” *Applied sciences*, vol. 11, no. 1, p. 89–, 2021.

[3] X. Chen and T. Huang, “Facial expression recognition: A clustering-based approach,” *Pattern recognition letters*, vol. 24, no. 9–10, pp. 1295–1302, 2003

[4] M. Li and B. Yuan, “2D-LDA: A statistical linear discriminant analysis for image matrix,” *Pattern recognition letters*, vol. 26, no. 5, pp. 527–532, 2005.

[5] Y. Wen, “An improved discriminative common vectors and support vector machine-based face recognition approach,” *Expert systems with applications*, vol. 39, no. 4, pp. 4628–4632, 2012.

[6] Tu X., Li J. and Hu Y., “Face recognition device and method based on reinforcement learning and deep convolutional neural network,” *Systems Science & Control Engineering*, 2021.

[7] D. Chen, X. Cao, F. Wen and J. Sun, "Blessing of Dimensionality: High-Dimensional Feature and Its Efficient Compression for Face Verification," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3025-3032, 2013

[8] Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z. Learning Multi-scale Block Local Binary Patterns for Face Recognition. In: Lee, SW., Li, S.Z. (eds) *Advances in Biometrics*. *Lecture Notes in Computer Science*, vol 4642, 2007.

[9] R. Belmonte and B. Allaert, Face analysis under uncontrolled conditions: from face detection to expression recognition., *NJ: John Wiley & Sons, Inc.*, 2022.

[10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEE*E, vol. 86, no. 11, pp. 2278–2324, 1998.