# A Face Recognition Pipeline Based on MTCNN for Detection and ResNet for Feature Extraction

**Xiangyu Shen**

Jincheng College, School of Computer and Software, Sichuan University, Chengdu, Sichuan 611731

**Abstract:** ***With the ongoing development of artificial intelligence and the continuous advancement of machine learning technologies, face recognition has become widely adopted in intelligent identification systems. In many campus environments, however, attendance systems continue to rely on traditional manual appointment and check-in methods, which incur significant time costs and negatively affect classroom efficiency. To address this issue, this study employs the MTCNN (Multi-Task Cascaded Convolutional Network) algorithm for face detection and the ResNet algorithm for face recognition, applying transfer learning to the classroom attendance task. Experimental comparisons were conducted with GoogLeNet, AlexNet, and VGG. The results demonstrate that ResNet achieves the highest face recognition accuracy among the models evaluated, indicating its strong potential for deployment in attendance systems.***

**Keywords:** MTCNN algorithm; ResNet algorithm; Face testing; Face recognition; Migration Learning.

## 1. INTRODUCTION

With the development of technology, the application of neural networks has gradually reached various fields, among which face detection and recognition has a leading edge. At present, there are various attendance systems. Looking at college campuses, teachers often use manual attendance, but there are shortcomings such as time consuming, inauthenticity, and inefficiency. If devices such as fingerprint scanning or iris recognition are used, the cost is often high. Face as a biological feature, it is unique and irreproducible. Therefore, face recognition can be well used in identification, and the use of neural network algorithms to achieve face recognition will greatly reduce cost and time to improve attendance efficiency. Pinyoanuntapong et al. [1] introduced Gaitsada, a self-aligned domain adaptation method for mmWave gait recognition, in 2023. Narouei et al. [2] examined the effects of germicidal far-UVC on ozone and particulate matter in a conference room in 2025. Xiao et al. [3] designed an ultrasmall $Fe_3O_4$-decorated polydopamine hybrid nanozyme for intensive wound disinfection in 2022. Zhao, Zhang, and Hu [4] proposed a smart warehouse track identification method based on Res2Net-YOLACT+HSV in 2023. Deng et al. [5] developed LLM-MVR, an LLM-guided multi-view reasoning distillation approach for sarcasm detection, in 2026. Wang et al. [6] presented QA-ReID, a quality-aware query-adaptive convolution leveraging fused global and structural cues for clothes-changing person re-identification, in 2026. Shan, Xu, Xia, and Lin [7] rethought wine tasting for Chinese consumers using a service design approach enhanced by multimodal personalization in 2025. Xia, Xu, and Shan [8] developed KOA-Monitor, a digital intervention and functional assessment system for knee osteoarthritis patients, in 2025. Ge and Wu [9] conducted an empirical study on the adoption of ChatGPT for bug fixing among professional developers in 2023. Wu et al. [10] introduced Tiny-Critic RAG, which empowers agentic fallback with parameter-efficient small language models, in 2026. Yan et al. [11] proposed PRISM, a pipeline for root-cause investigation via specialized multi-agents, in 2026. Finally, Yuan et al. [12] presented TA-Mem, a tool-augmented autonomous memory retrieval method for large language models in long-term conversational question answering, in 2026.

## 2. FACE DETECTION AND IDENTIFICATION

The process of face detection is basically divided into 6 steps: image acquisition; Face detection through neural network algorithms; The detected images are preprocessed; Extract the face features of the image; Match features to faces obtained from the image; Finally, we obtain the largest possible face and complete face recognition. The process of face detection and face recognition is shown in Figure 1 below:

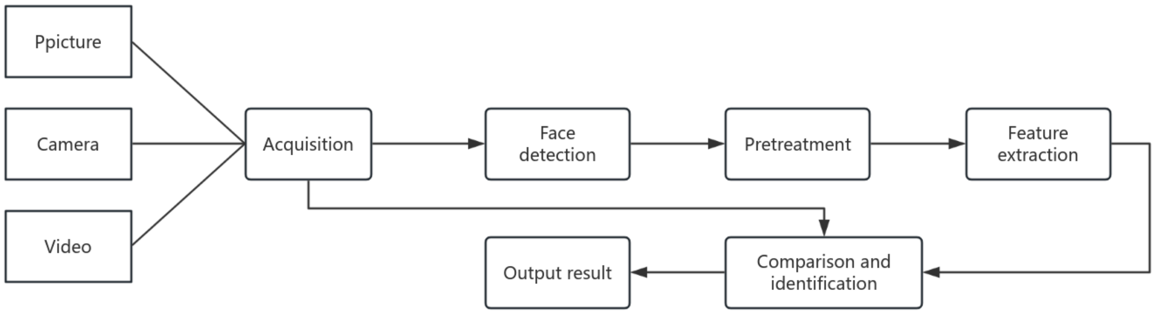


**Figure 1:** Process diagram for face recognition

This experiment uses MTCNN (Multi-task convolutional neural network) The multi-task convolutional Divine Network model is implemented to face detection, and the features are extracted by ResNet (Residue Network) deep residual network algorithm to realize face recognition.

### 2.1 MTCNN Network

MTCNN is a multi-tasking CNN network based on cascading architecture[2], where CNN is a convolutional divine network model, CNN neural network is similar to the process of human brain recognition of human images, From the biological brain vision neural network structure, input a picture, through the convolution layer neural network to extract the image features, using the activation function to map the data to get the image features, and CNN is widely used in computer vision and intelligent speech recognition. MTCNN is a deeper network structure based on CNN, which can simultaneously perform face detection and face alignment.

2.1.1 Three cascading neural network structures

MTCNN cascading architecture of the multi-task convolutional Divine network input is a picture, through the expansion of different sizes of the image generated, to build the image pyramid[3], the resulting image is passed into P-net (Proposal Network), R-net (Refined Network), O-net (Output Network) to get a more refined picture, the three cascading Divine network model is shown in Figure 2.

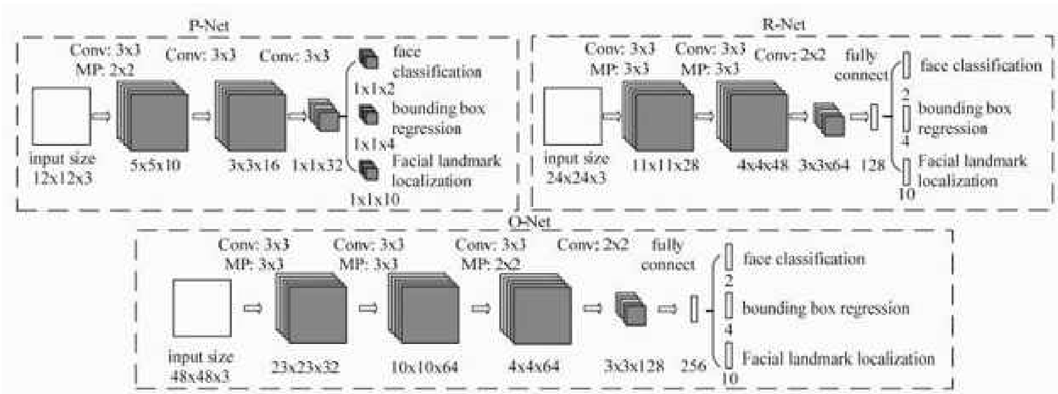


**Figure 2:** P-net, R-net, O-net network intent

Face candidate form, boundary regression and face candidate box generated by P-net network as shown[4]For comparison calibration, the overlaid borders are removed using non-maximum value suppression (NMS). The output of the upper layer is the input of R-net, further refined, the same as P-net, and finally NMS merge candidate boxes. More detailed processing is done in O-net. The difference is that the 5 key point coordinates of the face will eventually be output, and the more convolutional channels in this network, the deeper the network, the more accurate the detection of the face.

2.1.2 MTCNN Algorithm Core

Face feature point positioning and border regression problems: both are regression problems. The face feature point positioner problem ends with calculating the deviation of the predicted key point and the actual point. A border regression problem is a deviation between the calculated border and the actual border, which is usually used as a loss function of the regression problem using a European distance.

Face detection loss function: face detection belongs to the classification problem, cross entropy loss function (Cross Entropy), face detection cross entropy loss function is as follows:

Loss function weight problems in cascading architectures: In neural networks, P-net, R-net, and O-net loss functions all exist, but not all of them play a role, and the weights have different value emphasis on P-net and R-net and O-Net.

### 2.2 ResNet Network

As the network continues to deepen its stacking, it means that the features that can be extracted are more abundant. At the same time, it can better represent the image and get deeper features. In fact, the resulting model is not ideal. To solve the problem of deepening neural networks, network models deteriorate.[5], proposed by Kaiming He Deep Residual Network (ResNet), by which the final accuracy can be improved, the basic structure of the residual network is shown in Figure 3.

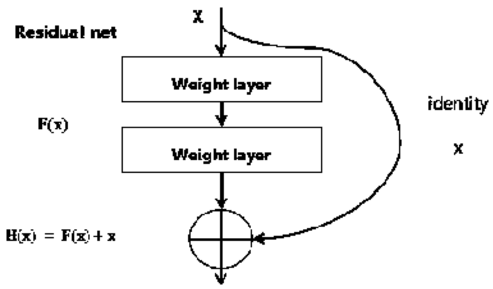


**Figure 3:** Sample diagram of the deep residual network

It can be seen that ResNet does not directly connect each convolution layer to each other, but introduces residual units.[6]The residue is the difference between the predicted value and the observed value, and the output consists of two parts, one is convolutional calculation, and one is to add the input as output, which retains the input information.

Deeper neural network model validation results not from poorer model performance due to overfitting and declining gradients, but rather from the failure to learn a constant mapping in the deeper network layer.[7]The effect of deepening the network disappeared. However, the addition of residual networks is equivalent to the addition of non-equivalence maps, so that the subsequent deep networks can begin to map equivalences, so that deep networks can function, and finally improve the model accuracy.

## 3. EXPERIMENTS AND RESULTS ANALYSIS

This experiment using python program, call OpenCv image processing, as well as access to the camera to write images, using the PyTorch framework released by Facebook. There are many ways to capture images. The most common ones are image uploading and camera capture. To get a more realistic picture, this experiment took pictures of faces by camera and uploaded pictures.

### 3.1 Image acquisition

The experiment used a camera to collect the face, side face, head up, wear glasses, wear a mask, Images such as skewed head and uploaded images obtain a data set, which is then processed into grayscale graphs by image conversion, and the data is divided into training sets and test sets by 8: 2. The file obtained with the data and labels is shown in Figure 4:

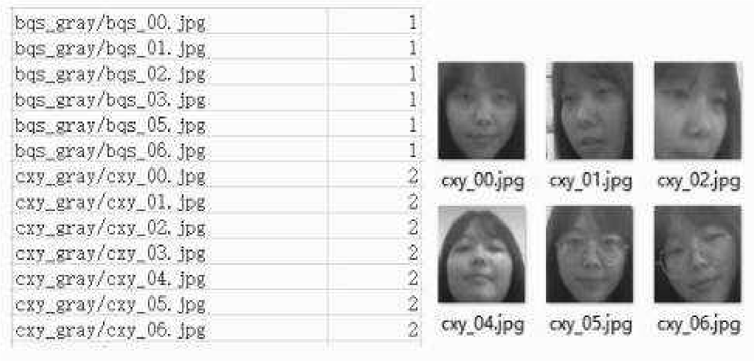


**Figure 4:** Images in the dataset and corresponding tags

### 3.2 MTCNN Face Detection Steps

1 Create face dataset: Create a face digital library by crawling the photo collection of network stars, and extract the face ROI region through MTCNN, including the left eye, the right eye, the nose, the left corner of the mouth and the right corner of the mouth.

In the classroom face attendance application, the face images of 5 students are added, and the same texture structure as the original document folder is maintained, and the data. csv file is formed.
Divided data set: divided data is used as training set, written to train.csv file with OpenCv, divided data as test set written to test.csv file.

Data preprocessing: Using transform to normalize the image to get the characteristic vector, and using MTCNN to detect the face, the result is as shown in Figure 5.

**Figure 5:** Image of MTCNN detection effect

As shown in Figure 3, MTCNN is able to detect the features of the face in the image where the person has the expression and the side face.

### 3.3 ResNet Model Training

The facial recognition neural network used in this experiment is ResNet-18 for migration learning, with a loss function of cross entropy, an optimizer of random gradient descent, a learning rate of 0.1, a weight loss of 0.05, and a batch size of 37, using Softmax for classification. And using ResNet, GoogLeNet, AlexNet, VGG to do a comparative experiment.

3.3.1 Migration Learning

Computer system network and telecommunications Transfer learning in deep learning refers to the problem that has been well solved in a field, The learned parameters can be transferred to similar problems, which can be solved quickly without the need to retrain a model, i.e. the ability to argue against each other. Because the image structure is similar and can be trained using the same parameters, it is often used in computer vision. This experiment performed migration learning on face datasets on the ResNet-18 model, which greatly saved the training time.

3.3.2 Experimental results

This experiment freezes the convolutional layer of the pre-trained model, modifies the number of neurons in the final fully connected layer to the number of face categories in this experiment, and then uses the modified ResNet-18 model for face recognition. Comparing the results of the experiments is as follows in Table 1:

**Table 1:** Results of four common network tests

| Network model | Accuracy% |
|---|---|
| ResNet | 90.62 |
| GoogLeNet | 84. 83 |
| AlexNet | 70.00 |
| VGG | 68.75 |

It can also be seen from Table 1 that the ResNet model recognizes faces more accurately than other models on the test set and can effectively recognize faces. The effect of identification is as shown in Figure 6:

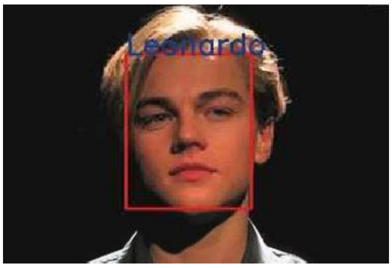


**Figure 6:** ResNet Face Recognition Leonardo

Using ResNet transfer learning model to recognize face pictures, as shown in the picture can be very good to recognize Leonardo, face recognition accuracy is high.

## 4. CONCLUSION

On campus, the attendance system is still traditional for manual appointments and check-in. To reduce the time cost and reduce the impact of appointments, face detection and face recognition are implemented based on MTCNN and ResNet algorithms. MTCNN extracts the face ROI region through a cascade architecture neural network, and in comparison experiments, ResNet is better than other face recognition algorithms, and can be well applied to classroom attendance systems. Face recognition is widely used and can be extended to other application scenarios by establishing a large database, optimizing the neural network, and improving the accuracy of face recognition to be applied to intelligent recognition systems.

## REFERENCES


[1] Pinyoanuntapong, Ekkasit, et al. "Gaitsada: Self-aligned domain adaptation for mmwave gait recognition." 2023 IEEE 20th International Conference on Mobile Ad Hoc and Smart Systems (MASS). IEEE, 2023.

[2] Narouei, F. H., Tang, Z., Wang, S. I., Hashmi, R. H., Welch, D., Sethuraman, S., ... & McNeill, V. F. (2025). Effects of germicidal far-UVC on ozone and particulate matter in a conference room. Plos one, 20(8), e0328224.

[3] Xiao, J., Hai, L., Li, Y., Li, H., Gong, M., Wang, Z., ... & He, D. (2022). An Ultrasmall Fe3O4 - Decorated Polydopamine Hybrid Nanozyme Enables Continuous Conversion of Oxygen into Toxic Hydroxyl Radical via GSH - Depleted Cascade Redox Reactions for Intensive Wound Disinfection. Small, 18(9), 2105465.

[4] Zhao, X., Zhang, L., & Hu, Z. (2023). Smart warehouse track identification based on Res2Net-YOLACT+HSV. Innovation & Technology Advances, 1(1), 7–11. https://doi.org/10.61187/ita.v1i1.2

[5] Deng, X., Yang, Y., Wang, B., Zhang, X., Huang, S., Zhang, Y., & Lu, Q. (2026). LLM-MVR: LLM-Guided Multi-View Reasoning Distillation for Sarcasm Detection. Available at SSRN 6795979.

[6] Wang, Y., Jiang, K., Zhang, T., Tian, K., & Jiang, G. (2026). QA-ReID: Quality-Aware Query-Adaptive Convolution Leveraging Fused Global and Structural Cues for Clothes-Changing ReID. arXiv preprint arXiv:2601.19133.
[7] Shan, X., Xu, Y., Xia, T., & Lin, Y. S. (2025, October). Rethinking Wine Tasting for Chinese Consumers: A Service Design Approach Enhanced by Multimodal Personalization. In 2025 International Conference on Content-Based Multimedia Indexing (CBMI) (pp. 1-5). IEEE.
[8] Xia, T., Xu, Y., & Shan, X. (2025, May). KOA-Monitor: A Digital Intervention and Functional Assessment System for Knee Osteoarthritis Patients. In International Conference on Human-Computer Interaction (pp. 388-405). Cham: Springer Nature Switzerland.
[9] Ge, H., & Wu, Y. (2023). An Empirical Study of Adoption of ChatGPT for Bug Fixing among Professional Developers. Innovation & Technology Advances, 1(1), 21–29. https://doi.org/10.61187/ita.v1i1.19
[10] Wu, Y., Liang, P., Xiang, Y., Yuan, M., Liu, J., Yang, J., ... & Yan, W. (2026, March). Tiny-Critic RAG: Empowering Agentic Fallback with Parameter-Efficient Small Language Models. In 2026 9th International Conference on Advanced Algorithms and Control Engineering (ICAACE) (pp. 2577-2580). IEEE.
[11] Yan, W., Wu, Y., Liang, P., Yuan, M., Liu, J., Yang, J., & Li, X. (2026, March). PRISM: Pipeline for Root-cause Investigation via Specialized Multi-agents. In 2026 International Conference on Generative Artificial Intelligence and Information Security (GAIIS) (pp. 709-712). IEEE.
[12] Yuan, M., Liu, J., Yang, J., Li, X., Yan, W., Wu, Y., & Liang, P. (2026, March). TA-Mem: Tool-Augmented Autonomous Memory Retrieval for LLM in Long-Term Conversational QA. In 2026 9th International Conference on Advanced Algorithms and Control Engineering (ICAACE) (pp. 2684-2688). IEEE.