# 修 士 論 文 概 要 書

Master's Thesis Summary

Date of submission: 07/18/2022 (MM/DD/YYYY)

| | | | | | |
|---|---|---|---|---|---|
| 専攻名 (専門分野) Department | Computer Science and Communications Engineering | 氏 名 Name | Siru Chen | 指 導 教 員 Advisor | Hiroshi Watanabe 印 Seal |
| 研究指導名 Research guidance | Research on Audiovisual Information Processing | 学籍番号 Student ID number | CD 5119FG29-1 | | |
| 研究題目 Title | Research on Face Recognition in Non-Ideal Scenes Based on Convolutional Neural Network | | | | |

## 1. Introduction

Facial recognition, like other biometrics, has the good features of being "unique" and "hard to copy," but also has the unique advantage of being "direct," "friendly," and "contactless." Real objects are very complex, and it is usually difficult to understand the essence of the object, and to know which features are important and what are the real object features. Studying the structure of biological neural networks allows us to infer the need for machines that can autonomously discover hidden knowledge rules, rather than simply instilling them into computers. It is very difficult for people to extract high level abstract data features from unprocessed data, but computers can solve the key problem of feature extraction by simulating the human brain to represent complex concepts through simplified models.

The problems studied in this paper mainly include the following aspects.

(1) Combined with specific practical applications, face recognition needs to consider the real-time nature of recognition, and face recognition sometimes needs to be judged by fuzzification processing, so it is not the more accurate the more it can meet the actual needs. Face recognition needs to solve the actual problem specifically according to different application scenarios.

(2) In the face of some non-ideal scenes, the recognition ability of face recognition technology and application is tested to different degrees, and face recognition will appear unrecognizable in the case of deepening difficulty of certain scenes. Therefore, the improvement of face recognition ability in non-ideal scenes is the main direction to be tackled in the next stage of face recognition technology and application research.

(3) The existing CNN models do not perform well in face recognition in non-ideal scenes, and there is much room for improvement in terms of accuracy, stability, robustness and other indicators. How to design an improved CNN model to cope with the non-ideal scenes prevailing in the face recognition process becomes an important issue.

## 2. Related Works

Identity verification and identification is a common security concern in modern society. Among the wide variety of identity verification tools, three main categories can be divided into possession verification (e.g. ID card), information verification (e.g. password) and biometric verification. It can be found that the verification method of biometric identification has obvious advantages in terms of security and convenience compared with other verification means [1][2]. Moreover, the biometric features selected after repeated testing have inherent properties such as non-replicability and stability of maintaining the same state within a certain period of time, which are highly suitable and applicable for identity verification and identification mainly in the field of security prevention and control.

In terms of personal information security, identity recognition has become a popular development direction for AI applications in information security in order to meet people's needs for information security. However, there are insecurity factors in several personal identity objects, such as ID cards and cell phones that can be easily lost or falsified. In contrast, human biometric features have the advantage of being inherent and irreproducible, such as facial features and fingerprint features, which can be better used for identity recognition because they are not repetitive and difficult to forge [3].

## 3. Proposed Approach

There are many typical neural network architecture models, such as the classic LeNet.5, which is mainly used for recognizing handwritten fonts; AlexNet network with deeper and improved model; VGGNet designed by Oxford University; GooLeNet developed by the Internet giant Google; and ZF Net, etc. Networks such as AlexNet are developed on the basis of LeNet.5 with deeper and wider, as well as improved. The structure of LeNet-5 is not complicated, but it is a high reference value for building your own CNN model, and the CNN model used in this paper refers to the single channel structure of AlexNet to build your own CNN network coarse model. In the process of building, the CNN network is further extended according to the network training recognition effect, and an improved CNN network fine model with better recognition effect and good generalization performance for the sample set in non-ideal scenes is built. [4]

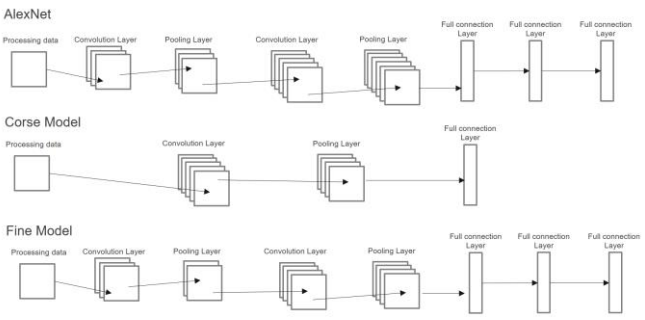


**Figure 1: AlexNet, Corse model, Fine model structure**

The coarse convolutional neural network model in this paper is designed as a ten-layer structure with five convolutional layers, four pooling layers and one fully connected layer. The model is also used to train and test on a pre-processed grayscale face image dataset. The face data set is trained by the model to optimize and tune the model parameters. The fine model then adds a set of

convolutional and pooling layers to the coarse model, and the ninth layer of the source network is connected to the newly added convolutional layer to extract deeper features during the fine training. This model is used to train the face dataset and to achieve optimal adjustment of the model parameters. Based on the trained network, face recognition tests are carried out using the test set.

## 4. Experiment

The experiment are the comparison of different CNN models for non-ideal scene face recognition. The experiment training tests are conducted on a total of 600 pairs of face images, of which 300 pairs belong to the same person. The selected comparison algorithms are fine model, coarse model, and AlexNet for comparison tests to quantitatively analyze the effect of different algorithms on face recognition in non-ideal scenes.

In this paper, we use convolutional neural network model to extract face features, and the limited scenes are some non-ideal scenes introduced above one by one. Therefore, for the qualification of "non-ideal scenes", the paper plans to select the sample sets that meet the conditions of "non-ideal scenes" that can be obtained from the CASIA WebFace database, FG.NET face dataset, CACD2000face dataset, etc. Usually, the images are inputted by cutting and resizing the collected images, and then the face database is grayed out and used as training input. As shown in Figure 1.

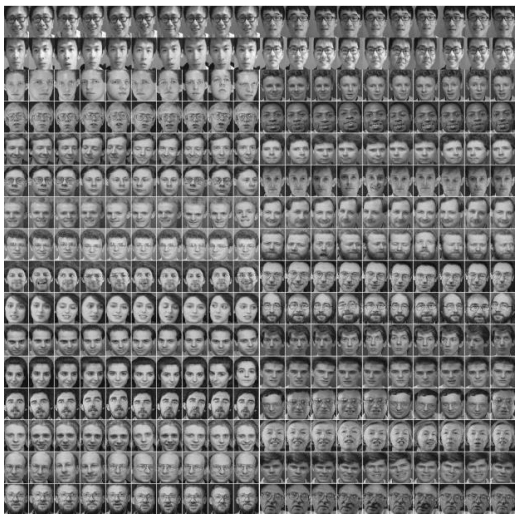

**Figure 2: Sample set of face recognition after grayscale processing (partial)**

## 5. Analysis

For comparison of experimental results in incomplete sampling scenarios. The coarse model has the best performance among the three CNN models in the distribution of face recognition results in different low-resolution scenes, in terms of accuracy, balance and stability of the accuracy distribution. In different wearing object scenarios, the fine model has the best performance, both in terms of accuracy and the balance and stability of accuracy distribution.

For comparison of experimental results in multi-pose scenarios, the experimental structure mainly consists of two parts, i.e., standard face photos and different posture face photos. Comparing the statistical indexes, we can find that among the face recognition results distribution of the three CNN models in the multi-pose scenario, the fine model has the best performance. In terms of accuracy, balance and stability of accuracy distribution, the refined model has the best performance among the three CNN models in the multi-pose scenario, and the leading advantage is obvious.

For comparison of experimental results under non-ideal lighting scenarios, the experimental architecture consists of two main parts, including the effects of light angle and light brightness changes on the application accuracy. It can be found that among the face recognition results distribution of the three CNN models in the non-ideal angle scenario, the coarse model and the fine model have similar recognition accuracy and have certain recognition ability, but the recognition stability is poor, and the recognition accuracy of AlexNet is the weakest.

In order to better compare the effectiveness of the three CNN models for face recognition, two core metrics, accuracy and stability, are selected to focus on the effectiveness characteristics of the three CNN models. Accuracy is defined as the mean value of data results, and stability is defined as the probability of achieving recognition accuracy of 85 and above, and the experimental comprehensive evaluation of the three CNN models is compiled as shown in Table 1.

**Table 1: Comprehensive comparison statistics of the three CNN models**

| | | Wearing object | Low Resolution | Multi-pose | Non-ideal brightness | Non-ideal angle |
|---|---|---|---|---|---|---|
| Fine Model | Accuracy | 95.40 | 98.17 | 83.83 | 97.36 | 82.50 |
| | Stability | 100% | 100% | 94.6% | 100% | 98.6% |
| Coarse Model | Accuracy | 93.96 | 99.33 | 65.52 | 98.74 | 79.44 |
| | Stability | 100% | 100% | 82.6% | 100% | 91.5% |
| AlexNet | Accuracy | 77.35 | 96.26 | 54.99 | 89.77 | 51.11 |
| | Stability | 93.2% | 100% | 62.9% | 100% | 81.5% |

## 6. Conclusion

In this paper, an improved CNN face recognition network structure is designed to expand the network model based on the original network with 10 layers, including 5 convolutional layers, 4 pooling layers and 1 fully connected layer, and to perform fine training and tuning. Based on the trained network, the face recognition test is conducted in both conventional and non-ideal scenes using the images in the constructed face database, and the recognition correct rate is 93.9%, which is better than the comparison algorithm.

## References


[1] Z. Ma and G. Zhang, Data Science in Face Recognition System Based on BP Neural Network, Journal of Physics: Conference Series, Vol. 1881, The 2nd International Conference on Computing and Data Science (CONF-CDS) pp.28-30 Jan. 2021

[2] Gary K. Y. Chan, Towards a Calibrated Trust-based Approach to The Use of Facial Recognition Technology, International Journal of Law and Information Technology, Vol. 29, No. 4, pp.305-331, Nov. 2021

[3] K. Giannou, K. Lander, and J. R. Taylor, Attentional Features of Mindfulness are Better Predictors of Face Recognition than Empathy and Compassion-Based Constructs, Psychological Reports, pp.1-35, Feb. 2022

[4] H. Vu, M. Nguyen, C. Pham, Masked face recognition with convolutional neural networks and local binary patterns, Applied intelligence (Dordrecht, Netherlands), vol. 52, no. 5, pp. 5497–5512, Aug 2021

**Research on Face Recognition in Non-Ideal Scenes Based on Convolutional Neural Network**

A Thesis Submitted to the Department of Computer Science and Communications Engineering,
the Graduate School of Fundamental Science and Engineering of Waseda University
in Partial Fulfillment of the Requirements for the Degree of Master of Engineering

Submission Date: July 18th, 2022

Siru Chen
(5119FG29-1)
Advisor: Prof. Hiroshi Watanabe
Research guidance: Research on Audiovisual Information Processing

# Acknowledgements

First and the most important, I would like to express my deepest grateful to my research supervisor Prof. Hiroshi Watanabe for offering me a position in his lab and give me support on my study and research in past three years. Not only for his academic guidance, but also give me support and confidence to face the pressure on research and work.

Secondly, I am glad to join the audio team of Watanabe lab and express my thanks to all members in the lab. Thank you for give me the advice and suggestion on my research. Also, the kindness of all the members made my school live more wonderful.

Thirdly, I am express my thanks to my roommate, who give me support and bring the joy to my school life.

Lastly, special thanks to my parents, who believe me and give unselfish love to me anytime.

# Abstract

With the progress of technology and the development of society, the security of identity verification and identification is attracting more and more attention. Different from traditional possession verification (such as ID card) and information verification (such as password), biometric identification technology extracts physiological or behavioral characteristics for identity identification. The object of biometric identification is the person itself, and no verification object and verification information are required, which directly determines that this authentication method is safer and more convenient. Face recognition technology is the one based on human facial features in the field of biometric identification, extracting facial features of human face through collection devices such as cameras and carrying out identity verification identification through feature matching. Facial recognition, like other biometrics, has the good features of being "unique" and "hard to copy," but also has the unique advantage of being "direct," "friendly," and "contactless." As a very challenging direction, face recognition in non-ideal scenes involves many core issues in the field of computer vision, and the application prospect of face recognition in non-ideal scenes is very wide. The study of 3D information is also crucial, and convolutional neural network technology plays an important role in both the field of face recognition and other directions in computer vision. Convolutional neural network, as an algorithmic framework that mimics biological neural networks, has great research value and is currently the most likely path to advanced intelligence.

In this paper, the basic convolutional neural network recognition framework is modified to improve the adaptability and reliability of the model in non-ideal scenes (occlusion, illumination change and angle change problems) by combining the latest research results of deep learning with the requirement of high accuracy for face recognition in non-ideal scenes, and finally the performance is verified by comparing the experiments with several classical network models. The research in this topic is carried out based on CNN for face recognition in non-ideal scenes. In this paper, the CNN-based face recognition model in non-ideal scenes is constructed, analyzed and trained. The study designs an improved CNN face recognition network structure diagram, and expands the network model based on the original design network with 10 layers, including 5 convolutional layers, 4 pooling layers and 1 fully connected layer, and adds a set of convolutional and pooling layers on top of the coarse model to form a new fine network. In the fine training, the ninth layer of the original network is connected with the newly added convolutional layer to extract deeper features to design the fine-training network. The coarse and fine models are trained on the face dataset in non-ideal scenes and finely tuned to achieve optimal adjustment of model parameters.

Due to the great development of related theories and technologies, the image recognition technology has undergone great changes in recent years. The size of the training dataset has become larger and larger, the feature extraction algorithms have gradually moving toward visual feature learning, and the training models have become more and more complex. The related technology, especially deep learning methods, has made great progress, but also faces many challenges. Further research is still necessary on how to design networks for the purpose of fast training of high-performance models.

**Keywords:** CNN; non-ideal scene; face recognition; accuracy enhancement

# List of Contents

# List of Figures

# List of Tables

# Chapter 1 Introduction

## 1.1 Research Background

Identity verification and identification is a common security concern in modern society, and with the achievements of modern science and technology, a series of modular tools have been developed to help with identity verification and identification. Among the wide variety of identity verification tools, three main categories can be divided into possession verification (e.g. ID card), information verification (e.g. password) and biometric verification. Among them, biometric verification refers to the detection and identification of biometric features by verifying specific biometric features of organisms, including fingerprints, iris, voice, DNA, portrait, etc., which can be used as the discriminatory criteria for identity verification. It can be found that the verification method of biometric identification has obvious advantages in terms of security and convenience compared with other verification means [1][2]. Moreover, the biometric features selected after repeated testing have inherent properties such as non-replicability and stability of maintaining the same state within a certain period of time, which are highly suitable and applicable for identity verification and identification mainly in the field of security prevention and control.

Facial recognition, like other biometrics, has the good features of being "unique" and "hard to copy," but also has the unique advantage of being "direct," "friendly," and "contactless." Therefore, with the gradual expansion of the identity verification field in recent years, face feature recognition has been more and more widely used in the field of identity verification and identification. A whole set of mature technology system has been initially formed around face recognition, and a considerable application scale has also been achieved[3].

Real objects are very complex, and it is usually difficult to understand the essence of the object, and to know which features are important and what are the real object features. Studying the structure of biological neural networks allows us to infer the need for machines that can autonomously discover hidden knowledge rules, rather than simply instilling them into computers [4]. It is very difficult for people to extract high level abstract data features from unprocessed data, but computers can solve the key problem of feature extraction by simulating the human brain to represent complex concepts through simplified models.

The training ability of AI convolutional neural network models keep enhance, and the exactness of recognition also keep improve. Deep learning is in rapid development, but still in the initial stage of development and its applications in many fields are not fully developed by humans. The research in this paper evaluates the recognition efficiency of each CNN model by analyzing the measurement results of different face recognition CNN models for face photos in different scenarios in order to derive an

efficient CNN model for face recognition in specific dimensions under specific evaluation criteria.

## 1.2 Motivation

With the development of artificial intelligence, more and more aspects of life have been involved in different intelligent applications. In terms of personal information security, identity recognition has become a popular development direction for AI applications in information security in order to meet people's needs for information security. Relying on personal identity characteristics is an important way of identification in the past, such as personal documents, cell phones and other objects that mark their identity. However, there are insecurity factors in these personal identity objects, such as ID cards and cell phones that can be easily lost or falsified. In contrast, human biometric features have the advantage of being inherent and irreproducible, such as facial features and fingerprint features, which can be better used for identity recognition because they are not repetitive and difficult to forge[5].

Face recognition is an important means of identification. It is a very innovative subject in the field of identity recognition and has much value for research and exploration. At the same time, because face recognition is widely used in life and has very high application value, it is a popular direction for research in information security practice projects today. Experts and scholars at home and abroad have been tirelessly exploring face recognition technology and have now harvested many successes and breakthroughs. However, there are still many difficulties in face recognition in practical applications. There are not many differences between faces, and the accuracy of face recognition is affected by a great number of factors as expressions, make-up, decorations and various other reasons may lead to changes in faces. Moreover, combined with specific practical applications, face recognition needs to consider the real-time nature of recognition, and face recognition sometimes needs to be judged by fuzzy processing, so it is not the more accurate the more it can meet the actual needs. Face recognition needs to solve the actual problem specifically according to different application scenarios, which is an important research trend at present.

CNNs use shared parameters between convolutional layers, which make the algorithm less dependent on memory size. Moreover, shared parameters can reduce the number of parameters for training, thus improving the performance of the algorithm. Compared to other machine learning algorithms that require image pre-processing or extraction of image features, CNN algorithms can omit these steps when processing images, which is an advantage for users[6].

Although several algorithms have been applied to deep learning, the application of deep learning is still limited by some shortcomings: for example, the construction of deep models needs to be built on the basis of many samples. In practical applications, the application scenarios of the algorithm have been launched with great success in areas such as face recognition and license plate character recognition, and many

excellent successes have been launched. The research of this topic is carried out based on CNN face recognition technology.

## 1.3 Outline

Face recognition technology has the characteristics of change and diversity because the face image used is not fixed, and the face expression, makeup, decoration, lighting, and resolution in non-ideal scenes can affect the accuracy of recognition. This paper focuses on the CNN-based face recognition technology for non-ideal scenes. The advantages of this technique over traditional face recognition techniques are the high recognition accuracy, the ease of user recognition and the lack of focus on feature specifics. The main research structure of this paper is in below:

Chapter 1 introduces the background, motivation and framework of this paper to lay the foundation of this research.

Chapter 2 introduces the core technologies and concepts involved in this study, focusing on the fundamentals and implementation of CNNs. By studying the basic principles of CNN models, the convolutional and downsampling functions included in the MatConvnet framework are used to process the images, and the results obtained are used to construct the convolutional and downsampling layers of the convolutional neural network model, etc., so as to build the required CNN training model. All implementations of the model are done using the MATLAB-based MatConvnet framework, and its library files, along with GPU acceleration.

Chapter 3 is devoted to the construction and training of a convolutional neural network model for face recognition. The coarse convolutional neural network model in this paper is designed as a ten-layer structure with five convolutional layers, four pooling layers and one fully connected layer. The model is also used to train and test on a pre-processed grayscale face image dataset. The face data set is trained by the model to optimize and tune the model parameters. The fine model then adds a set of convolutional and pooling layers to the coarse model, and the ninth layer of the source network is connected to the newly added convolutional layer to extract deeper features during the fine training. This model is used to train the face dataset and to achieve optimal adjustment of the model parameters. Based on the trained network, face recognition tests are carried out using the test set.

Chapter 4 is about CNN-based face recognition experiments and results reporting in non-ideal scenes. This chapter mainly investigates face recognition in conventional and non-ideal scenes, and uses the trained CNN model to recognize face information in non-ideal scenes.

The last chapter is a summary and outlook.

# Chapter 2 Previous Work

## 2.1 Face recognition

### 2.1.1 Classification of face recognition

Based on the broad application scenario and huge application value of face recognition technology, there are more technology expansion applications around face recognition technology, and several categories of face recognition technology classification have been formed accordingly. According to the different core application technologies of face recognition, face recognition can be divided into the following categories [7].

(1) Template based face recognition method

A face template is a pre-prepared comparison face without any obscuring object. When face recognition verification is carried out, the face template is placed at the center of the recognition subject's graphic, positioned according to coordinate planning, and the appropriate scaling size, angle and direction are selected. The key is to find the value of the center coordinate of the face template, and after the center coordinate value is located, the correlation and match between the face template and the subject face image are compared, which is used as the basis for face recognition verification.

(2) Face recognition method based on a priori geometric knowledge

A priori geometric knowledge refers to the geometric model of the important feature parts of the face, which is extracted based on the geometric feature values of specific feature parts. Specifically, the important parts of the face, such as eyes, nose, eyebrows, mouth and other feature points, are extracted inductively for their geometric feature values such as relative distance, proportion and spatial structure, and the geometric parameters of spatial feature vectors of multiple dimensions are obtained, and the one with the smallest value of the before-and-after comparison is taken as the result of face recognition.

(3) Face recognition method based on statistical theory

Facial features of human faces have statistical dimensional significance and can rely on statistical tools for standard quantification of facial features. Therefore, the face recognition based on statistical theory constructs the statistical model of recognition through the numerical statistics of the facial features of the face sample of the recognition subject, and uses the size of the statistical probability value as the basis for judging the accuracy of recognition. The advantage of face recognition based on statistical theory is that it has better robustness, i.e., it is less affected by light and image size. However, the disadvantage is that it is computationally intensive and requires a high degree of expertise for recognition, which is not conducive to widespread implementation.

(4) Neural network-based face recognition method

A neural network is a topological network containing a large number of processing nodes, which are connected to each other by weighted values. Neural network-based face recognition means that these weighted values are continuously adjusted in the learning process, and face recognition is performed by relying on back-propagation of primary colors to internally transform the input high-dimensional data into a low-dimensional subspace and thus preserve the important information in the source image data to form a discriminator for face recognition.

### 2.1.2 Face recognition process

The processing process of face recognition technology can be divided into four parts: face detection, image pre-processing, face feature extraction and face recognition, and each part is divided into specific operation steps, so the whole process of face recognition is complicated and systematic from the technical level, and each step has a decisive influence on the accuracy and matching degree of the final face recognition result. Therefore, the realization of face recognition function needs to rely on the comprehensive and standardized process operation mode of various technologies [8][9]. The process of face recognition is shown in Figure 2-1.

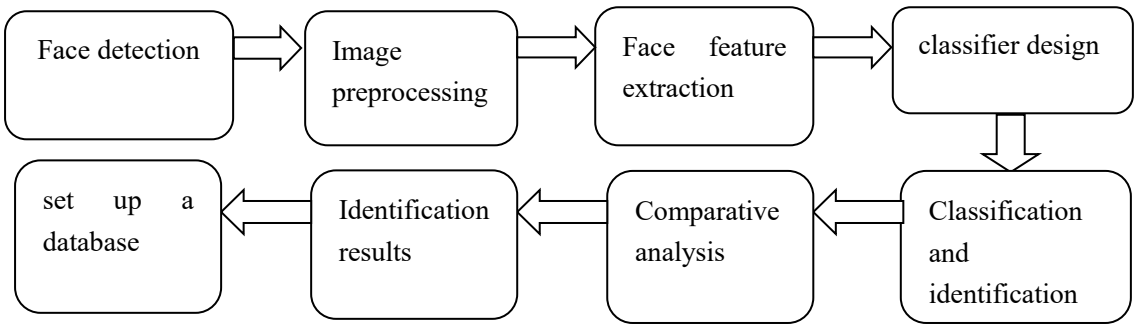


**Figure 2-1: Schematic diagram of face recognition processing flow**

1. Face detection

Face detection is to detect the exact location of the recognized face from the target image. Face detection can distinguish the whole image into face panels and non-face panels, so as to focus more on face panels for target recognition. Especially for face recognition in non-ideal scenes, the boundaries between face panels and non-face panels are often not very clear, and face detection is needed to detect and locate the face panels with unclear boundaries. The process of face detection is divided into two steps: first, it needs to detect whether the target image has a face part; second, if a face part that meets the recognition criteria is not detected, the whole face recognition process will be terminated, and if a face part that meets the recognition criteria is detected, it will enter the module recognition stage of face part detection, and the complete face will be detected and used in the next stage of processing. The non-ideal scenarios in the face detection stage, including lighting conditions, face angle, facial glasses hair, face size and other factors, have an important impact on the successful implementation of face detection[9]. The core technical means of face detection mainly includes two parts:

geometric features and image blocks. According to the specific situations of non-ideal scenes encountered in the face detection process, the selection of technical means of face detection is carried out in order to derive the most optimal face detection results.

2. Image pre-processing

Image pre-processing is a series of pre-processing of the face slab processed in the face detection stage before entering the formal face recognition stage to enhance the final presentation of face recognition. Image acquisition is an important part and the basic part of face recognition. However, image acquisition often has to face the interference of many non-ideal scenes, leading to the existence of distortion and noise pollution and other drawbacks of the collected images, at this time, if the collected original images are not pre-processed, it is likely to make the application of recognition substantially more difficult. Image pre-processing is divided into two parts: image enhancement and image restoration. Among them, image enhancement is to make the effective features more prominent and the irrelevant features weaker through targeted enhancement processing of the feature part of the captured image, which is more conducive to the development of face recognition. Typical representatives of image enhancement methods include histogram equalization, image smoothing and denoising, grayscale transformation, pseudo-color processing, etc. Image restoration is to restore the effective features lost in the image due to various scene constraints and non-ideal factors through a series of technical means. There are many scenes and situations for image restoration, which require specific and targeted technical processing. For example, a noisy image can be regarded as the final image after degradation, and the degradation process can usually be expressed as a combination of linear blur and Gaussian noise, using the a priori knowledge of the image to build a model for it, and solving the degradation process in reverse to obtain the essential image[10].

3. Face feature extraction

Face feature extraction is the key part of face recognition, and many index features that determine the accuracy of face recognition need to be determined in this part. The face has the property of high-dimensional sample, which makes the existing technical means for the processing of the original image have many limitations. Therefore, the extraction of target features for recognition and comparison becomes a reliable path for face recognition technology processing. Feature extraction is to decompose the high-dimensional human image into several low-dimensional feature slabs, and through the mapping from high-dimensional to low-dimensional, so that the subsequent face recognition only needs to compare several low-dimensional feature slabs to achieve the purpose of face recognition. The existing feature extraction methods are mainly divided into the following categories: geometric feature method, statistical feature method, support vector machine, and kernel technique. Each type of face feature extraction method has its own relative advantages, and the selection of face feature extraction method is carried out according to the different face scenes.

4. Classifier design

Classifier design is the result output link of face recognition. According to the demand orientation of face recognition, the recognition result output based on classifier design is divided into two categories: confirmation and recognition. Classifier design

is to compare the samples to be tested with the samples in the training set and derive the degree of similarity between them. The confirmation requirement is to compare the samples to be tested with the samples in the training set in a one-to-one comparison, according to a set threshold, to achieve the confirmation function; the recognition requirement is relatively more complex, which is to compare the samples to be tested with each sample in the training set, and locate the specific sample with the most matching result as the result output, so as to complete the recognition function. It can be found that the confirmation problem is a simplification of the recognition problem.

### 2.1.3 Application scenarios of face recognition

Face recognition technology is developed continuously in response to the needs of social development. Therefore, the specific application of face recognition can be divided into corresponding specific scenes according to the social needs and market demand in order to carry out more targeted face recognition application development. Especially in recent years, the reliance on face recognition technology in the field of public security has deepened, and a series of applications based on face recognition technology matching public security management have been developed one after another, greatly enriching the application scenarios of face recognition technology. Moreover, the application of face recognition technology is a top-down popularization mode. Firstly, application development is carried out in public security and government management departments, and then commercialization and civilization popularization is carried out continuously, which makes the present face recognition application scenario richer. The face recognition technology itself is widely researched, spanning more than ten disciplines, including computer technology, image processing, pattern recognition, artificial intelligence, human-computer interaction, neural network, database technology, optical devices, cognitive science, physiology and psychology, and so on. At the same time, the development of face recognition technology itself can also play a reverse compensating role to these supporting technologies and further promote the development and progress of basic science[11].

**Table 2-1: Overview of face recognition application fields and application scenarios**

| Applications | Application scenarios |
|---|---|
| Public Security | Public security patrol, searching for lost people, detecting wanted fugitives, etc. |
| Security | Anti-terrorism alarm, security check, immigration check, etc. |
| Transportation | Train station "face entry", traffic violations, etc. |
| Finance | Remote account opening, face payment, credit card and social security identification |
| Education | Verification of candidate information, attendance, campus and dormitory access management |
| Dining | Food ordering, etc.; KFC concept store used face recognition ordering system for the first time in December 2016 |
| Retail | Unmanned convenience store, shopping guide, etc. |
| Entertainment | Interactive games, face video splicing and segmentation, etc. |

## 2.2 Convolutional neural networks

### 2.2.1 Introduction of Convolutional Neural Network

Artificial neural network is a method of artificial intelligence that imitates the structure and units of human brain. However, in the process of image processing using traditional artificial neural network, despite the large amount of pre-processing work, there are still many noise factors in the image that affect the image recognition effect. In order to further improve the efficiency of image recognition, many different feature extraction algorithms and neural network models have been promoted one after another. Among them, convolutional neural network is the enhancement of artificial neural network, which is a special multilayer perceptron designed to automatically extract image features for recognizing two-dimensional images. The original image does not need much preprocessing to learn the invariant features of the image better. Currently, a typical convolutional neural network is a multilayer, trainable architecture. It includes input, convolutional layer (locally connected layer), sampling layer, normalization layer, fully connected layer, logistic regression layer and output layer, etc[12]. It becomes a hot topic of research for improving the recognition effect of convolutional neural networks on images, finding the most suitable network structure and parameter configuration for the dataset to be recognized, and the network structure with certain compatibility for different datasets.

Convolutional Neural Network has obvious advantages in processing and using voice files or recognizing images. Convolutional neural networks have many features, but the main one is weight sharing, similar to biological neural networks, which can reduce the difficulty of the network model by reducing the number of weights, especially for those very complex network models, and can directly input the image with the network. Compared to traditional algorithms, convolutional neural networks perfectly avoid the relatively tedious pre-processing work, such as extracting features and reconstructing

relevant data, and it shows a high degree of invariance, both in terms of tilt processing and in terms of translation processing operations, etc.

Among the multilayer network structures, convolutional neural networks successfully stand out among the many algorithms. Convolutional neural networks are able to effectively reduce the number of learned parameters in performing performance improvement training of forward BP algorithms. Since this network structure has been proposed, it has been able to reduce much of the preprocessing work related to the input data, significantly reducing the workload. From the very beginning of the input layer for data input, each layer is processed in a certain order, and then into the other levels of the structure, and in each level, there will be a convolution kernel will have significant features of the data analysis and selection, this method can be widely used to obtain the translation and rotation and other relatively invariant obvious features of the data up.

### 2.2.2 Network structure of convolutional neural network

Convolutional neural network is one of the neural network originally designed for image recognition, and the main features are local perceptual domain and weight sharing, as shown in Figure 2-2. These feature can reduce the calculated amount of neural network and also improve the robustness of neural network. This kind of network structure take advantage of the spatial interconnection of images and has invariant to translation, scaling, tilting or co-deformation. The most important is the convolutional layer that considered as an adaptive feature extraction solution. The collection of a large number of convolutional layers can make a convolutional neural network to progressively upgrade features from low level to high level, even at the pixel level. And since the convolutional neural network combines feature extraction and classification in one neural network, it can be trained end-to-end[13][14].

Convolutional neural network is composed of three parts: hidden layer, input layer, and output layer, and it has an additional hidden layer compared with a normal neural network. The hidden layer is the core component of the convolutional neural network, which is composed of three structures: the convolutional layer, the pooling layer and the fully connected layer. The convolutional layer exists to enable the network to extract the relevant features of the input data, and it consists of many convolutional kernels inside. Each convolutional kernel has a large number of neurons, and the neurons in the same kernel have the same weights. Since the pooling layer does not have a weight factor w, the pooling layer is not recognized as a separate layer in some studies: without a convolutional layer, there is no pooling layer. Pooling can significantly reduce the number of parameters in the network without changing some image features of the original data, and in general there are two types of pooling: maximum pooling and mean pooling. Fully connected layers are used in convolutional neural networks for image features processed by convolutional and pooling processes, which allow probabilistic output of the processed image features by nonlinearity[15].

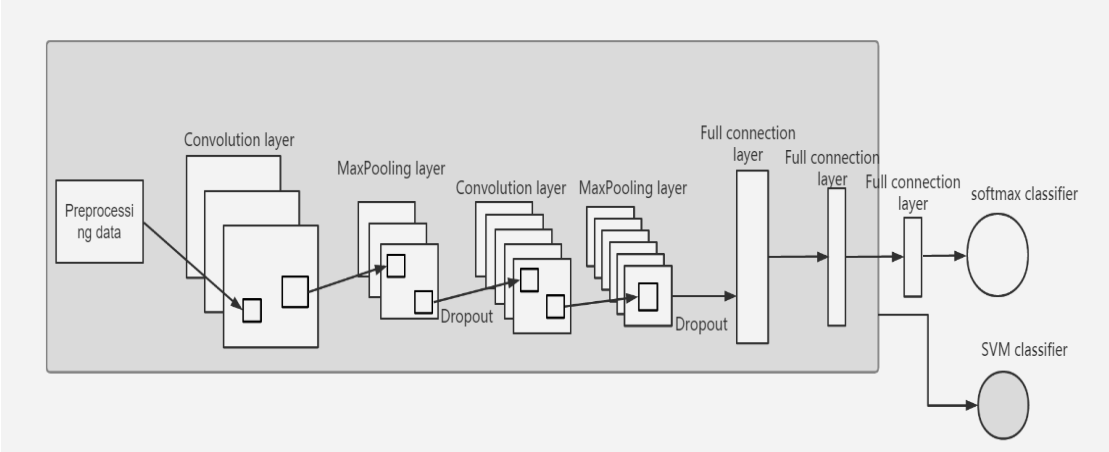


**Figure 2-2: Convolutional neural network framework**

Multiple convolutional kernels and pooling kernels are interconnected by array to form the framework of the whole convolutional neural network, and each convolutional kernel is connected together by many individual neurons, and the network structure is shown in Figure 2-3.

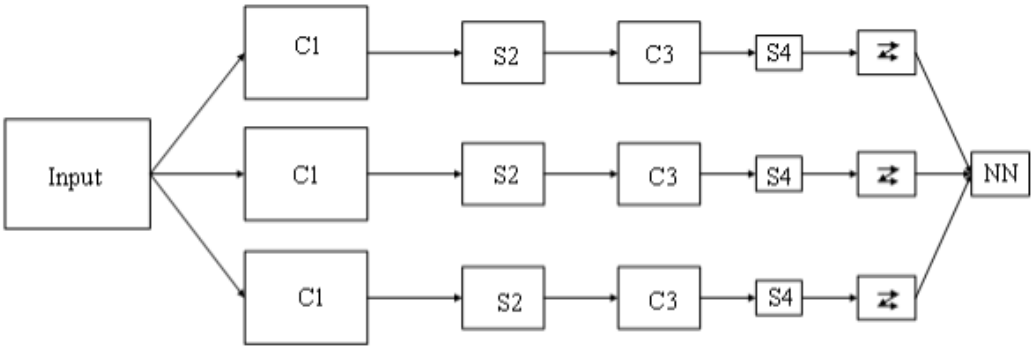


**Figure 2-3: Convolutional neural network structure diagram**

Usually, there are the data input for convolution neural network is the color image that have three color channels of RGB. So that we usually will convolving the original data with three convolutional kernel filters, which will train to get the mapping layer on top of C1. The convolved image features are then processed by the relu activation function to point out the complex feature information of the image to get the mapping layer on top of S2. Then Pooling operation of pooling kernel will be used to select and filter feature, which been processed, to get the feature information map on C3. After that, transmit the information form C3 into S4 layer, and transmit the S4 layer pixels into other neural networks through the non-spatial expansion structure of fully connected layer to get one-dimensional signals. In convolutional neural networks, the C layer is generally referred to as the convolutional and pooling layers, and the size of the convolutional kernel is the size of the perceptual field. Generally, the smaller the

size of the convolution kernel, the more detailed the final image features obtained and the more obvious the results of the features will be.

At the same time, since all the feature planes on the convolution layer use the same weight w, the number of parameters in the convolution process is smaller than that of traditional networks, which improves the complex structure of the network space. By convolving and pooling the input image data, each proposed feature can express the details of the original image in a more comprehensive way[16].

The number of internal parameters becomes large when the image features are extracted using convolution kernels in the convolution process. Therefore, in order to speed up the feature extraction, we add pooling kernels after each convolution kernel to reduce the number of features generated in each convolution kernel. In general, pooling kernels are expressed as max pooling and avg pooling. The max pooling kernel reduces the shift of the weight error in the convolution kernel when predicting the average value and also preserves more information texture of the image features. The mean pooling kernel reduces the variance of the prediction caused by the restricted domain size and preserves more background information of the image features.

### 2.2.3 How Convolutional Neural Networks Work

The most important element of experimenting with image data using convolutional neural networks is learning the weights w and biases b in the convolutional layer. Once the raw data is fed into the convolutional neural network, the convolutional kernel and the pooling kernel in the convolutional layer perform the extraction of features from the image. The extracted information is then processed using the relu activation function to obtain a new image feature map. The parameters in this map are calculated as follows:

$$u_j^i = \sum_{i \in M_j} x_i^{i-1} * W_{ij}^i + b_j^i \tag{2.1}$$

$$x_i^{i-1} = f(u_j^i) \tag{2.2}$$

Where $x_i^{i-1}$ represents the output of the activation function, $f(u_j^i)$ is the activation function, $b_j^i$ is the bias of the convolution kernel, and $W_{ij}^i$ is the weight of the convolution kernel.

Convolutional neural network experiments are dominated by the study of weights w and biases b. This study uses a back-propagation algorithm. If the trained output has a certain error between the real output value and the trained output value, the weights and biases are first recalculated by the trained output and the real input during the backpropagation of the error. Then the error signal is back-propagated from the last convolutional layer to the input layer through the classifier, and the weights w and biases b are relearned in this process. The cost function model J (W,b;x,y) for the point (x, y) is:

$$\mathrm{J}\ (\mathrm{W},\mathrm{b};\mathrm{x},\mathrm{y}) = \frac{1}{2}\left\|h_{W,b}(x) - y\right\|^2 \tag{2.3}$$

where y is the true input and $h_{W,b}(x)$ is the output function. The generation of overfitting of the model is reduced by adding a bias:

$$\mathrm{J}(\mathrm{W},\mathrm{b}) = J(\mathrm{W},\mathrm{b};\mathrm{x},\mathrm{y}) + \frac{\lambda}{2}W^T W \tag{2.4}$$

## 2.3 Non-ideal scenarios

### 2.3.1 Overview of non-ideal scenes of face recognition

The current face recognition technology has achieved eye-catching technical results after decades of development process. In terms of the actual usage effect of the existing face recognition technology and commercial grade face recognition applications, the correct rate of face recognition can already reach more than 90% under the conditions of small data scale, ideal imaging environment and user cooperation, which indicates that face recognition can accomplish the general functions under regular scenes. However, it does not mean that face recognition technology and application already have the functional attributes to cope with any scenes[17]. In fact, in the face of some non-ideal scenes, the recognition ability of face recognition technology and application is tested to different degrees, and face recognition will appear unrecognizable in certain scenes where the difficulty of the scenes deepens. Therefore, the improvement of face recognition ability for non-ideal scenes is the main attack direction for face recognition technology and application research in the next stage. An excellent face recognition application must have good robustness to these disturbances. Categorizing these uncontrollable factors, the main factors affecting the performance of recognition algorithms are lighting, posture, expression, occlusion, age, image quality, and makeup cosmetic.

**Table 2-2: Illustration of non-ideal scenarios of face recognition enumeration**

| | |
|---|---|
| Illumination | Too strong or too weak illumination or shadow due to angle changes can affect the appearance features of a face. The presentation of the same face under different illumination conditions can vary greatly, and this variation may even exceed the difference due to different faces. |
| Expression | Exaggerated face expressions can affect the structural information of faces to a greater extent, causing face template deformation and increasing the difficulty of feature extraction, which is also the reason for low recognition performance. |
| Posture | Pose changes affect the angle of face acquisition, and the face contour and the relative positions and distances of organs presented by the same face at different angles will change, which also increases the difficulty of feature extraction. |
| Age | The appearance presented at different ages will differ, such as skin texture, skin color, etc., which is also a problem face recognition must face. The human face will be changed with the age growth. So the algorithm of face recognition will have different recognition rate for same person in different age. Especially when in the growth period and aging severely, it will lead to the decrease of recognition rate. |
| Image quality | Inconsistent image sources, storage and transmission process will inevitably bring noise pollution to the image and affect the image quality, which also increases the difficulty of feature extraction and recognition. |
| Makeup and face-lifting | In the highly civilized modern society, people have become accustomed to wearing makeup when they go out, and the progress of medicine has brought more and more people the benefits that plastic surgery can enjoy, which undoubtedly also increases the difficulty of face recognition and requires comprehensive consideration from multiple aspects such as signal sensing and feature extraction to improve recognition performance. |
| Large data size | Usually a complete face recognition system often requires a large database set, and the performance of face recognition will gradually decrease as the database size increases. The decrease of recognition performance due to the increase of data size is also a problem that face recognition research has to face. |

### 2.3.2 Description of non-ideal scene classification for face recognition

(1) Face recognition under incomplete sampling scenarios

In most of time, image acquisition does not require contact with the subject and there is no human intervention, so face image acquisition in uncooperative scenarios is easily obscured. Especially in some tracking suspect videos, the suspects always wear hat, glasses, mask to cover their face. Which make the captured face image are not complete and clear. It will impact the feature extraction and matching even leading to the failure of the algorithm in serious cases[18].

(2) Face recognition in multi-pose scenarios

The face pose problem mainly refers to the change of facial expression and the tilt angle of the face during image acquisition, which is also a technical difficulty that needs

to be solved in face recognition research. If make the head of the face rotated around the three-dimensional vertical coordinate axis when the image is acquired, it will cause the partial absence of facial information, and the feature of face cannot be collect correctly, which will impact the recognition result. Likewise, some large facial changes like crying, laugh and so on. will also impact the result of facial recognition. Currently, most of the face recognition algorithms mainly target frontal and quasi-frontal face images, and the recognition rate of face recognition algorithms will drop sharply when there is a large degree of pitching or left and right-side tilting[19].

(3) Face recognition under non-ideal lighting scenes

Face recognition research is challenged by a series of uncontrollable factors, and among these uncontrollable factors, illumination variation is the primary challenge faced by face recognition research. For face recognition, Illumination is one of the most important challenge problem. Illumination affects the clarity of the captured image, which is inextricably linked to the accuracy of recognition. The light at different time such as day, night, etc. and different locations such as indoor, outdoor, etc. are completely different; even for the same light source, different irradiation angles will have different recognition results. The initial face recognition studies all assumed that the image set was collected under standard lighting environment, however, in practice, the lighting conditions are highly variable, and too bright, too dark, strong light and sidelight can cause image distortion and shadow, which seriously affects the appearance of faces in the images and causes the recognition rate to drop. Experiments show that drastic changes in illumination can lead to significant changes in the structure of face information, and in many cases, the differences in face imaging due to changes in illumination are even greater than those due to differences in people. There are two main reasons for the low performance of face recognition caused by illumination. Firstly, different illumination conditions can cause intra-class differences in faces, i.e. the same face will show external differences due to inconsistent illumination conditions; secondly, illumination itself has infinite dimensional characteristics, and it affects face images, both in relation to its own light source and also in relation to the geometric features of face external appearance, surface reflectance, etc[20][21].

Illumination changes are complex and irregular, but it is a problem that face recognition has to face. The lighting conditions are different at different times and on different occasions, and sometimes the difference is even huge, which will cause huge differences in the external appearance of the face when reflected on the face, and then lead to a sharp decline in the performance of the whole face recognition system. Therefore, it is extremely necessary to implement illumination pre-processing before feature extraction of faces, and whether the effect of illumination variation can be effectively removed is the key to put face recognition research into practical application.

# Chapter 3 Proposed Approach

## 3.1 Experimental design

There are many typical neural network architecture models, such as the classic LeNet.5, which is mainly used for recognizing handwritten fonts; AlexNet network with deeper and improved model; VGGNet designed by Oxford University; GooLeNet developed by the Internet giant Google; and ZF Net, etc. Networks such as AlexNet are developed on the basis of LeNet.5 with deeper and wider, as well as improved. The structure of LeNet-5 is not complicated, but it is a high reference value for building your own CNN model, and the CNN model used in this paper refers to the single channel structure of AlexNet to build your own CNN network coarse model. In the process of building, the CNN network is further extended according to the network training recognition effect, and an improved CNN network fine model with better recognition effect and good generalization performance for the sample set in non-ideal scenes is built[22].

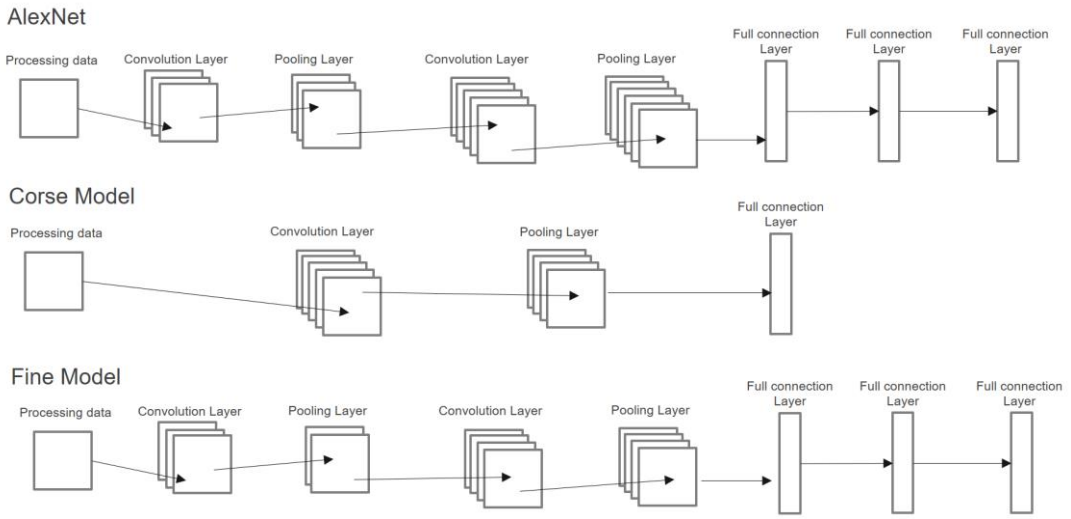


**Figure 3-1: AlexNet, Corse model, Fine model structure**

The research process of this paper will mainly use Matlab programming language, and MatConvnet deep learning framework of Matlab is used as a tool to conduct CNN learning research. For this research, MatConvnet provides a convenient and rich deep learning library, and also has powerful interface calls. The MatConvnet framework based on MatConvnet is also used to build convolutional neural network models quickly. With the above advantages, many programs and models in this research can be written and built easily and quickly. In addition, during the experimental execution, the intervention of a large number of existing libraries is required. Table 3-1 presents the core libraries and their functional features utilized in this experimental program[23].

**Table 3-1: Introduction to the core libraries for running the experimental program**

| Core Library | Features |
|---|---|
| MatConvnet | MatConvnet is a very simplified, modular degree of learning framework for neural networks very well. Configuration is very simple, written in Matlab, easy to view and modify the code. Based on Matlab + MatConvnet development, the modules can be packaged into a user-friendly API, the direct use of MatConvnet's own framework can reduce a lot of detail program writing, it is modular, from design to produce results is a very rapid process. |
| CuDNN | CuDNN (CUDA Deep Neural Network) is an acceleration library for deep neural networks in the CUDA framework. It is a standard library from NVIDIA to accelerate the deep learning training process with high performance, ease of use, and low memory overhead. At this stage, deep learning is inseparable from GPU acceleration, and CuDNN is not necessary when using GPU to train models, but it is generally used more often. |
| Numpy | Numpy is a matrix-based mathematical computing library that provides multi-dimensional array support, and can be visualized as a matrix processor by combining with MatConvnet. In this experimental design, it is mainly used for image data processing. |
| CUDA | CUDA, the general-purpose parallel computing model, can be thought of as a library, but rather as a parallel computing framework introduced by NVIDIA based on its own GPU. It requires that the computation to be performed is massively parallelizable to be useful. cUDA is transparent, scalable, etc., and depends mainly on the hardware graphics card. |
| Opencv | To recognize a face in a non-ideal scene, it is first necessary to detect and locate the position of the face in the image. this design uses the face detector designed by Opencv library for face detection. And after the detection, the face is cropped and saved to realize the pre-processing of face recognition. |

The face database for this experiment is mainly collected from the field, and according to the needs of the experimental environment and parameter settings, the undercomplete scene, multi-pose scene and non-ideal lighting scene are set up respectively, and each scene is subdivided into corresponding scenes according to the needs of scene parameters and dimensions. The under-complete scene was divided into scene parameter settings under different resolutions and wearers; the non-ideal lighting scene was divided into scene parameter settings under different lighting angles and brightness; the multi-stance scene was set according to different common postures, including low head, raised head, left half side head, left side head, right half side head, right side head, right side face, etc. After completing the experimental scene settings, the initial settings of the experimental architecture were made to unify the resolution, contrast, recognition accuracy, compensation mechanism and other relevant settings that may affect the experimental results.

## 3.2 Face data samples

Convolutional neural network-based image processing requires a large collection of images for the computer to learn. There are many databases dedicated to faces with rich database size and variety. The widely used ones are CASIA WebFace face database designed by CAS, LFW face comparison dataset, FG.NET face dataset and MORPH face dataset, CACD2000 face dataset, etc[24][25].

In this paper, we use convolutional neural network model to extract face features, and the limited scenes are some non-ideal scenes introduced above one by one. Therefore, for the qualification of "non-ideal scenes", the paper plans to select the sample sets that meet the conditions of "non-ideal scenes" that can be obtained from the above datasets. Usually, the images are inputted by cutting and resizing the collected images, and then the face database is grayed out and used as training input. As shown in Figure 3-1.

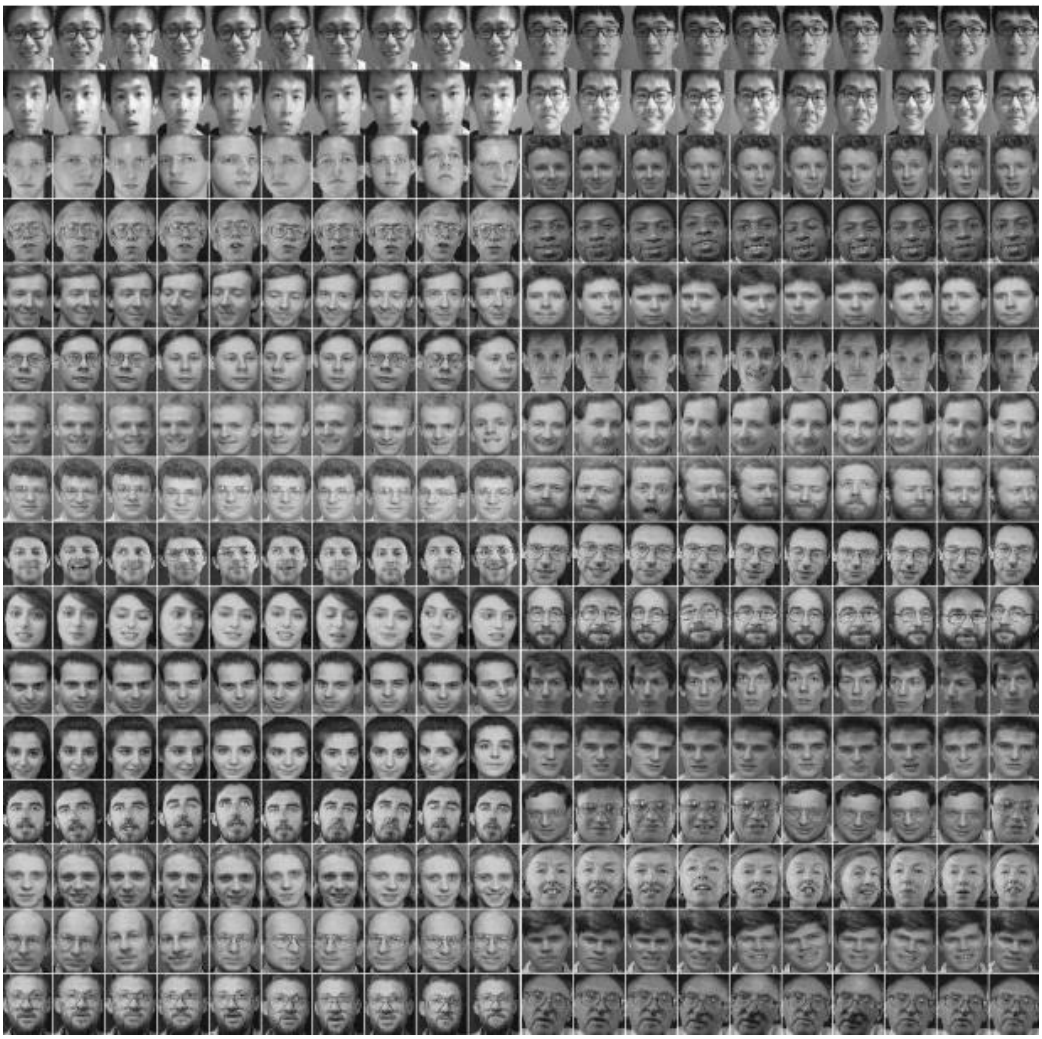

**Figure 3-2: Sample set of face recognition after grayscale processing (Source: From reference [7], uder Creative Commons Attribution 3.0 License)**

### 3.2.1 Experiments of CNN models in under-complete sampling scenarios

As shown in Figure 3-2, the experimental architecture of the comparison under the incomplete sampling scenes has two main components, which are divided into two categories with wearers and low-resolution scenes. Among them, each group of wearables is one standard front face photo, three photos with eyes (glass1, glass2, glass3) and three photos with hats (hat1, hat2, hat3), totaling five groups. Each group of low-resolution photos was divided into one standard face photo and three low-resolution photos, for a total of five groups tested.

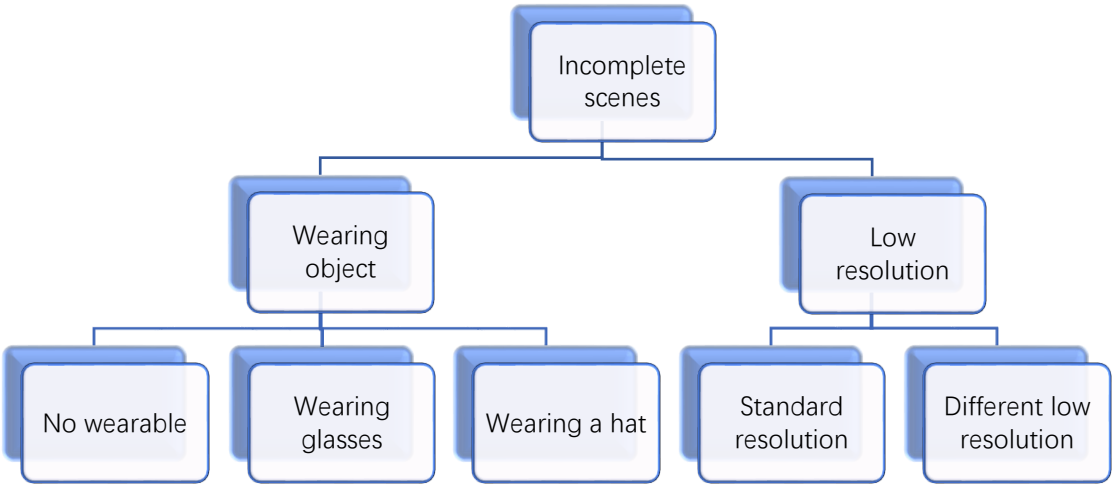


**Figure 3-3: Experimental design architecture for the undercomplete scenario**

The experimental process in the undercomplete sampling scenario is based on the experimental design idea of shooting experimental samples without wearer, with wearer (glasses, hat) and standard subscale rate and low resolution (three groups with different low resolution) respectively. The data is the portrait samples wearing different styles of glasses, different styles of different angles of hats, and without any wearing object. Among them, the uncovered portrait samples are used as training set samples and the covered portrait samples are used as samples to be detected, which are imported into the three face recognition applications to obtain the results and record them. On the other hand, for the standard resolution and three low resolution portrait inspection materials, the standard resolution portrait samples are used as training set samples, three different sets of low-resolution portrait samples are used as samples to be detected, which are imported into the three face recognition applications, and the results are obtained and recorded, while the other sets of portrait samples are compared according to the same procedure.

### 3.2.2 Experiments of CNN models in multi-pose scenes

As shown in Figure 3-3, the experimental architecture of the comparison in multi-pose scenes consists of two main parts, i.e., standard face photos and different pose face photos. A total of five groups were evaluated, and each group was divided into one standard face photo, and ten different pose photos (low head, raised head, left half side head, left side head, left side face, right half side head, right side head, right side face, etc.).

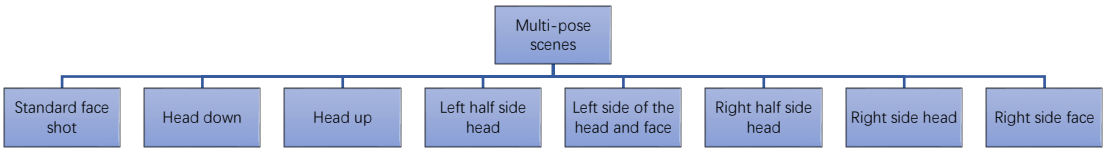


**Figure 3-4: Experimental design architecture of multi-pose scenario**

The experimental process in the multi-pose scenario is based on the experimental design idea of taking experimental samples of standard pose face photos and different pose face photos respectively. The data have the different pose portrait samples and the standard pose portrait samples. Among them, the standard pose portrait inspection material is used as the training set sample and the multi-pose portrait inspection material is used as the sample to be detected, which are imported into three face recognition applications to obtain and record the results.

### 3.2.3 Experiments of CNN models under non-ideal illumination scenarios

As shown in Figure 3-4, the experimental architecture of the comparison under non-ideal lighting scenes consists of two main parts, including the effects of light angle and light brightness changes on the application accuracy. Among them, each group of photos in the angle part is a standard front face photo and five photos under different lighting conditions (forward lighting, left 30-degree lighting, left 45-degree lighting, right 30-degree lighting, right 45-degree lighting); each group of photos in the luminance part is a standard front face photo and six photos with different lighting intensities (-30, -20, -10, +10, +20, +30).

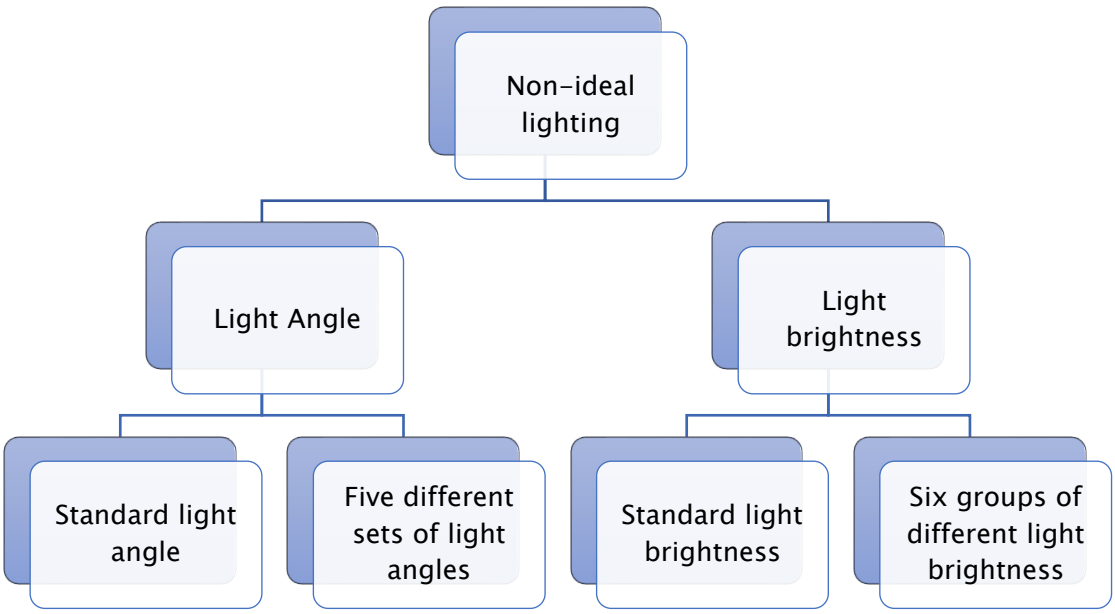


**Figure 3-5: Experimental design architecture for non-ideal lighting scenarios**

The experimental process in the non-ideal lighting scenario was based on the experimental design idea of taking experimental samples with standard light angle, five groups of different light angles and standard light intensities, and six groups of different light intensities, respectively. For non-ideal illumination scene with light angle portrait inspection material, the data is the portrait samples under different lighting angles. Among them, the standard-lighting-angle portrait samples are used as training set samples and the different-lighting-angle portrait samples are used as samples to be detected, which are imported into the three face recognition applications respectively to obtain the results and record them. For non-ideal lighting scene with light brightness portrait inspection material, the portrait samples under standard illumination are used as training set samples and the portrait samples under different illumination are used as samples to be detected, which are imported into the three face recognition applications respectively, and the results are obtained and recorded.

## 3.3 First stage: initial CNN model construction

As shown in Figure 3-5, a CNN structure diagram is designed in this paper with reference to the LeNet-5 network structure, which has 10 layers, including 5 convolutional layers, 4 pooling layers and 1 fully connected layer. This model is used as the first coarse network model for training and testing.

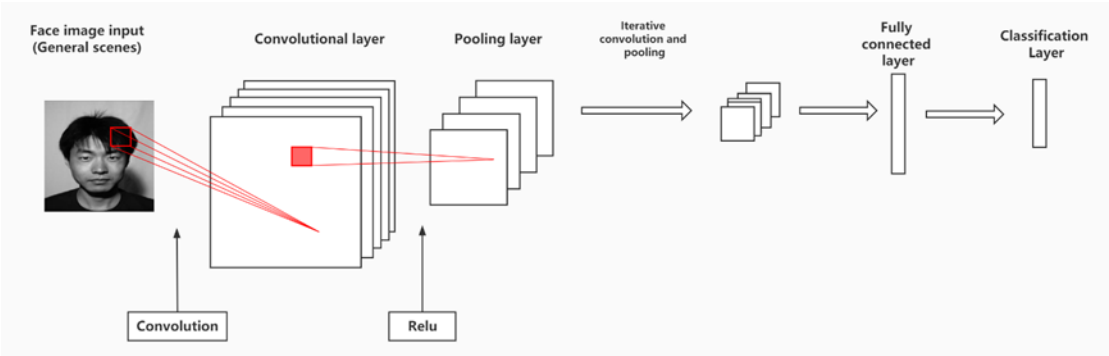


**Figure 3-6: Schematic diagram of the initial CNN model structure**

(1) Data layer

The main role of the data layer is to input relevant information, and often to be able to distinguish the differences between the data, to add labels to the data. When the data with labels is input, it will be further processed in the whole CNN network structure to the final model to process the data into a specific feature vector. The data labels as images are then compared in the network model to form a specific calculation of the loss function. The data layer in this stage covers both regular and non-ideal scenes, i.e., the face dataset in the full scene.

(2) Convolutional layer

In the CNN network structure model designed in this paper, a total of five convolutional layers exists. Convolutional layers 1 and 3 are used to compute the convolution of the image information after receiving it, and can activate the input information in time for normalization. The other three convolutional layers are basically the same in the process of convolutional computation of the image information. When the image information from the previous layer enters the convolutional layer, as shown in Figure 3-6, the convolutional layer 5 will perform three activations and three convolutional calculations on the input information, using a cascade of convolutional forms, using small-sized convolutional kernels (1×1, 3×3, 5×5), and finally a maximum pooling; convolutional layers 7 and 8 will convolve the image information again. The purpose of the multiple convolution operation is to reduce the parameters of the deep learning network, so that the depth of the network can be guaranteed. At the same time, the gradient disappearance will affect the whole CNN output result, so in the convolution layer 5, the convolved image is added to the full connection and regression again, and the dimensionality is set to 2010 according to the image data set to be trained[26][27][28].

The purpose of the normalization process is to prevent the data information of the input image set from being scattered, which makes the generalization ability of the whole CNN model insufficient. In this paper, a batch normalization of the data set is used to aggregate the data and transform it into a data set with mean 0 and variance 1.

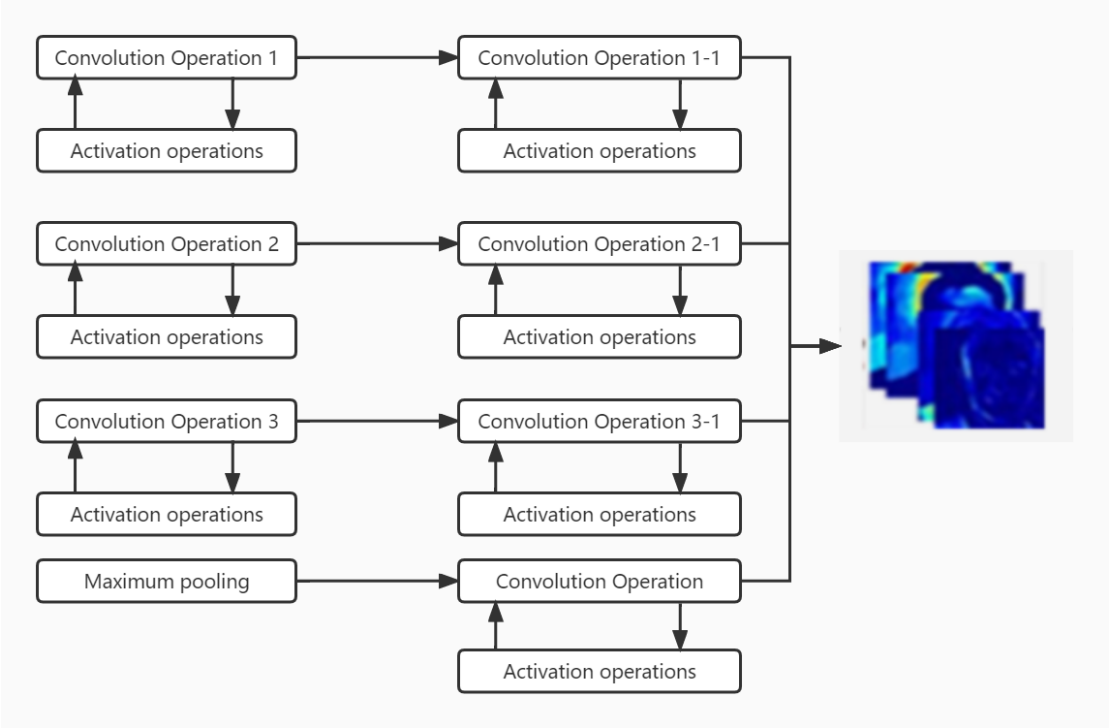


**Figure 3-7: Schematic diagram of C5 convolutional layer operation**

(3) Activation function

The activation of the data used in the CNN network structure model adopted in this paper is inspired by the development of neuroscience. By activating the neurons of the data, this nonlinear function can be constrained to the input and output data, thus effectively reducing the overfitting of the network. As shown in Figure 3-7, the commonly used activation functions are shown. Among them, Figure 3-7(a) is the Sigmoid function, Figure 3-7(b) is the Tanh function, Figure 3-7(c) is the Relu (Rectified Linear Units) function, and Figure 3-7(d) is the schematic diagram of the Softplus function.

Among them, the Sigmoid type function can be seen in the figure, mainly the information related to the middle region is processed for gain, while the data will be suppressed twice. There are similarities with the processing in neuron science. However, the Sigmiod type function will be activated about 50% of the time during the activation of the input signal in the model, which is different from the relevant theory and will cause great trouble to the training of the data set. Softpuls function can be regarded as a smoother version of Relu function, which is different from Relu function in that it does not have the effect of sparse signal. By comparison, it is found that the Relu function can process the data faster and is more efficient for training the input data set, so the Relu function regression method is used in this designed model[29][30].

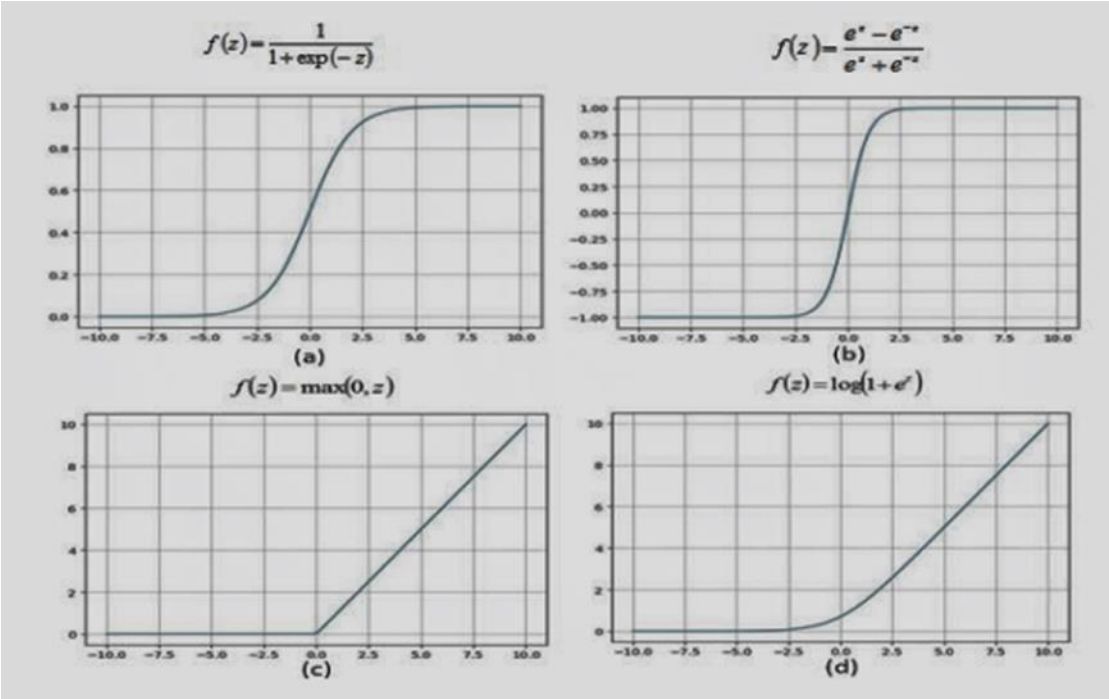


**Figure 3-8: Schematic diagram of the activation function**

(4) Batch normalization and Dropout

Another improvement of the convolutional network structure designed in this time, compared with the traditional CNN, is the method and parameter settings, where a normalization operation is added to each convolutional layer before inputting the next layer, which on the one hand allows for faster convergence and at the same time controls overfitting to reduce the regularization operation. In addition, it also allows the use of a larger learning rate during training. Also due to the large number of CNN layers and the corresponding number of network parameters, the database face images designed in this paper are limited and the data set is relatively small, in order to avoid the overfitting process, parameter tuning is used in which the Dropout strategy is used, and due to the use of batch normalization, the Dropout ratio is set slightly lower (0.2 to 0.5) than the 0.5 of the traditional method.

(5) Pooling layer

As described in Chapter 2, the network is constructed using a maximum pooling approach with a step size of 2. This means that the size and step size of the pooling kernel are defined at the beginning, so that the maximum value is found in a specific region.

(6) Fully connected layer

The fully connected layer is, in general, at the end of the whole network. This layer is able to transform the digital images input from the data layer into feature vectors after convolutional pooling and to do normalization of the feature vectors. The specific structure and parameters have been introduced in Chapter 2.

## 3.4 Second stage: improved CNN model construction

In the subsequent experiments, it was found that the CNN model constructed in Section 3.3, which tested well in the face library, began to decline in recognition rate when the model was applied to face image datasets in non-ideal scenes, and could not perform the recognition task well. Therefore, further improvement of this network is considered. Considering the complexity of non-ideal scenes, a deeper CNN network model is considered in this section. This model combines the convolutional neural network model of the previous section as one of the components of the new network structure: Network 1. Based on the recognition training results of non-ideal scenes, a convolutional layer and a pooling layer after tuning the parameters are added respectively to form another component of the new network: Network 2. In order to fully consider the impact of non-ideal scenes on face recognition accuracy, the expanded network model structure is used, and the specific network structure is shown in Figure 3-8.

In the fine training process, the convolutional layer in the newly added network 2 is connected to the last pooling layer of the coarse model network, and then after a pooling and full connection, the last layer is still a regression layer, except that the vector dimension of the regression changes at this time according to the different data sets. The main purpose of the fine training process is to adjust the parameters in network 2, and to obtain a new model, as can be seen from the subsequent experiments, network Therefore, in the fine training process, the parameters of the trained model of network 1 are still used in the coarse training phase, i.e., the parameters of the pooling layer and all network layers before it is fixed, and only the convolutional layer and pooling layer of network 2 are further trained and tuned. When the improved model is used for training, the original training database is still used for coarse training, while the new finely trained network with the addition of Network 2 requires the inclusion of a homemade face image dataset in a non-ideal scene in the training set for more in-depth feature extraction.

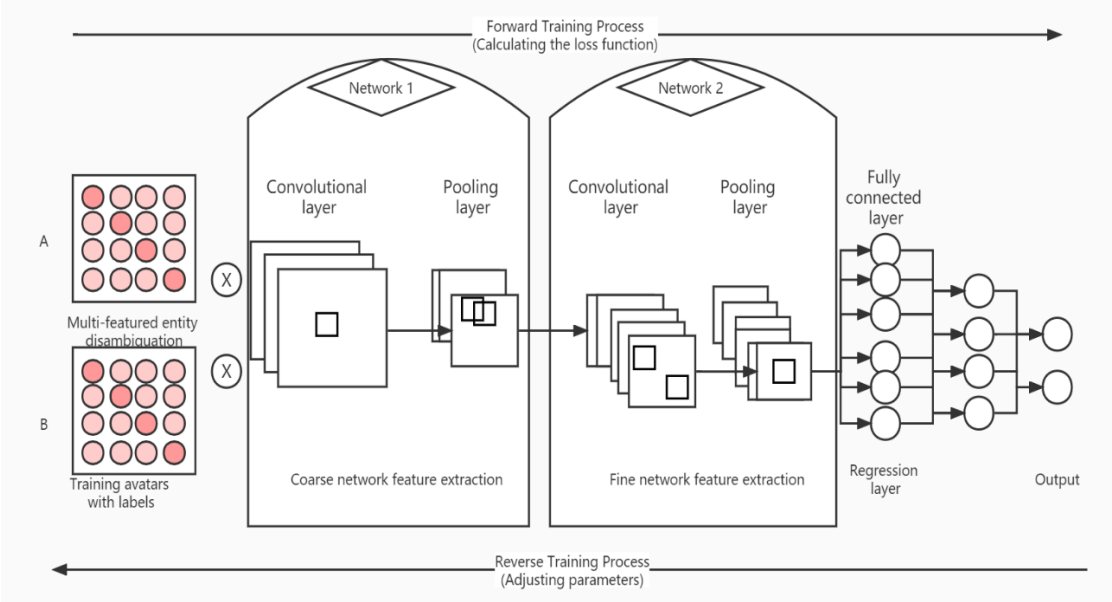


**Figure 3-9: Schematic diagram of the structure of the improved CNN model**

# Chapter 4 Experiments and Results

## 4.1 Experimental setup and environment configuration

In the process of CNN training and testing, there are many parameters to choose and set. In the adjustable parameters, summarizing the experience and information, there are certain reference standards for setting initialization of some parameters, which will be briefly introduced in this section.

For the coarse training network, the activation function has been introduced in Section 3.3, this experiment uses the currently commonly used Relu function; each layer uses the zero-mean Gaussian function to initialize the weights; the batch size is set to 80; to prevent training interruptions, every 20 epoch to save the network parameters, the validation period is set to 8; the dropout ratio is set to 50%, the training The initial learning rate is set to 0.01, and the learning rate is adjusted using the STEP strategy, which reduces the learning rate by a factor of ten for every 65 iterations; this experiment uses a single GPU and utilizes CUDA acceleration for the network to improve the training speed[31] [32].

For the fine-training network, which aims to further extract the depth information of the data images to obtain the features that extract better robustness to face images in non-ideal scenes, the parameter settings are the same as those of the coarse network, except that the number of training cycles is set to 200 during the further fine-training process. As shown in Table 4-1, the parameter configurations for the two training stages.

**Table 4-1: Parameter configurations for each training stage**

| Test | Training cycle | Test cycle | Save cycle | Batch | Initial learning rate | Processor | Training Momentum | Weight decay |
|---|---|---|---|---|---|---|---|---|
| Rough training | 100 | 8 | 100 | 80 | 0.01 | GPU | 0.9 | 0.0005 |
| Fine training | 200 | 8 | 200 | 60 | 0.01 | GPU | 0.9 | 0.0005 |

The environment configuration used for the experiments is shown in Table 4-2.

**Table 4-2: Configuration table of the running environment**

| Item Type | Configuration |
|---|---|
| CPU | AMD A4 APU (tm)2.6GHz |
| GPU | GTX HD970 4G VM |
| Memory | 8G RAM |
| Hard Disk | 500GB HDD |
| Operating System | Windows 10 Basic (64 bit) |
| Software Version | MATLAB R2014b |

## 4.2 CNN face recognition training experiments

In this section, only the CNN coarse and fine models described in the previous section are trained and tested. The applied face dataset training sets are mainly external face datasets in regular scenes and homemade face datasets in non-ideal scenes, where the images are all manually extracted from some high-quality face data images as well as the face images of some celebrities crawled on the Internet to form the dataset.

(1) Network 1 training test experiment

The network 1 trained in this experiment mainly tests the recognition of common faces in regular scenes, and the dataset is selected as 50% of the total training dataset. The ratio of training set to validation set is unchanged. The training processes each parameter configuration is configured according to the coarse training network parameters in the previous section, and 200 is set as the training period. The test accuracy is the one-dimensional vector of the final output regression of the network, which is transformed into a probability vector to determine the probability that the input image belongs to the label class.

From the experimental results, it can be seen that the final network is converged, and as the training proceeds, the loss values of the test set and the training set keep decreasing, and the accuracy of the test set keeps increasing. The final test accuracy reaches 96.4%. It can be seen that the training effect of network model 1 is good on this dataset.

Then the face dataset in regular scenes is selected for network testing experiments. The test was conducted according to the testing protocol provided by this database, and a total of 300 face pairs were selected. The database testing method is actually to input a pair of images separately on the already trained network, get two each feature vectors through the network, and determine that the two images are not yet one person by calculating the cosine similarity of the vectors according to the set threshold. After the test, the ROC (receiver operating characteristic curve) curve commonly used for face recognition was obtained to be the most evaluation index, as shown in Figure 4-1. In the ROC curve, the larger the area included in the lower right of the curve, the better the model effect, and the final test accuracy was 91.8%. It can be seen that the coarse model of network 1 is tested well on the face dataset in conventional scenes.

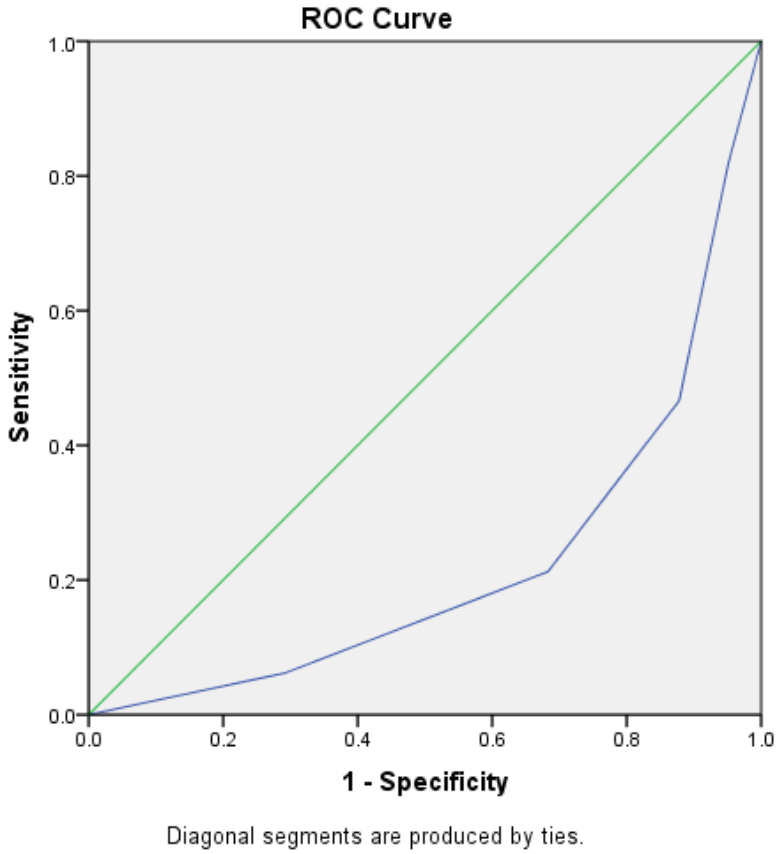


**Figure 4-1: Network 1 face recognition ROC graph (conventional scene)**

In order to further experiment, the trained model of Network 1 is also used and the same criteria are used to test the experiment on the data set under non-ideal scenes. Therefore, on the basis of network 1, this paper proposes an improved "fine model" by adding network 2.

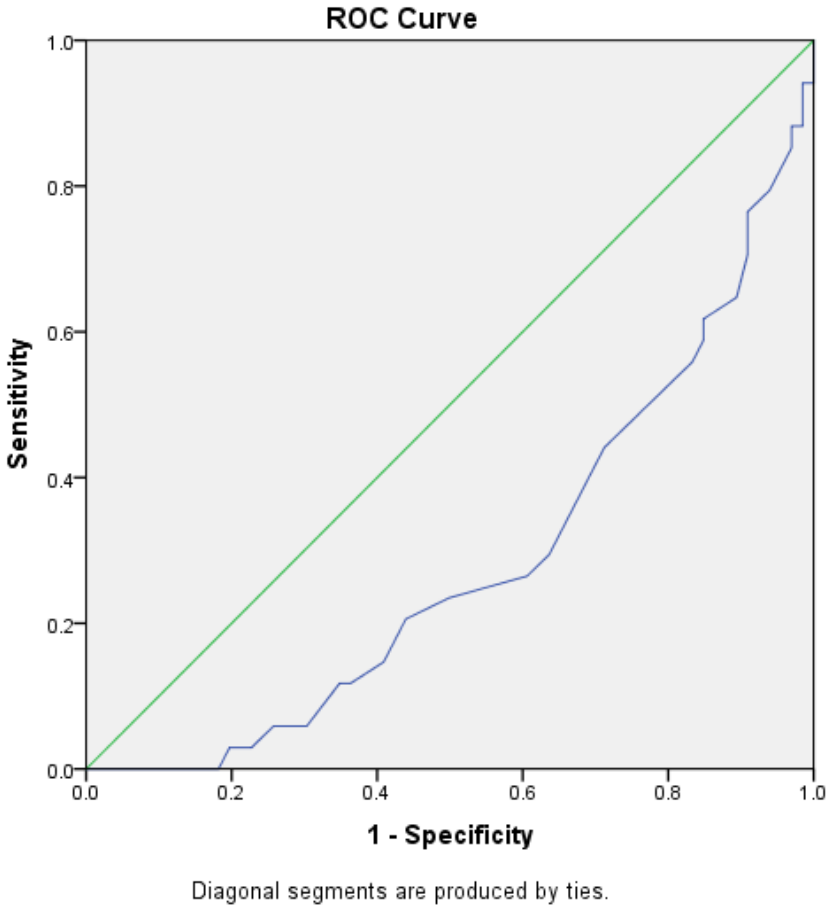


**Figure 4-2: Network 2 face recognition ROC graph (non-ideal scenario)**

## 4.3 Comparison experiments of different CNN models for non-ideal scene face recognition

In order to compare the superiority of the improved CNN model used in this paper, this section will use other face recognition algorithms in a non-ideal scene face library for testing experimental comparison. Given the size of the network structure and hardware equipment conditions, training tests are conducted on a total of 600 pairs of face images, of which 300 pairs belong to the same person. The selected comparison algorithms are fine model, coarse model, and AlexNet for comparison tests to quantitatively analyze the effect of different algorithms on face recognition in non-ideal scenes. Among them, the AlexNet network uses the model encapsulated in MatConvnet.

### 4.3.1 Comparison of experimental results in incomplete sampling scenarios

The experiments in the under-complete sampling scenes are divided into two categories: wearing objects and low-resolution scenes.

A total of five groups of face recognition under low-resolution scenes are tested, and

Table 4-3 shows the data statistics about the low-resolution face recognition results of the three CNN models. According to the descriptive statistics in Table 4-4, it can be found that in the cross evaluation of the maximum value index item, the fine model is the first, the coarse model is the second, and AlexNet is the lowest, indicating that in the single face recognition result output, the fine model can achieve the maximum accuracy of recognition in the low resolution scene; in the cross evaluation of the minimum value index item, AlexNet is the lowest, the fine model is the second, and the coarse model is the highest, indicating that in the cross-sectional evaluation of the mean value metric, the coarse model is the highest, the fine model is the second highest, and AlexNet is the lowest, indicating that the coarse model performs the best, the fine model is the second, and AlexNet is the worst in the overall comparison of the accuracy of face recognition results in different low-resolution scenes. In the cross-sectional evaluation of the standard deviation index, the coarse model is the lowest, the fine model is the second, and AlexNet is the largest. The coarse model has the best performance among the three CNN models in the distribution of face recognition results in different low-resolution scenes, in terms of accuracy, balance and stability of the accuracy distribution. The recognition accuracy performance of the fine model is in the middle level, and the recognition accuracy of AlexNet is the weakest.

**Table 4-3: Statistics of face recognition results of three CNN models in low-resolution scenes**

| Resolution | Characters 1 | | |
|---|---|---|---|
| | Resolution1 | Resolution2 | Resolution3 |
| Fine Model | 99.68 | 99.17 | 97.57 |
| Coarse Model | 99.68 | 99.64 | 99.1 |
| AlexNet | 98.26 | 98.39 | 94.23 |
| Resolution | Characters 2 | | |
| | Resolution1 | Resolution2 | Resolution3 |
| Fine Model | 99.81 | 99.36 | 96.2 |
| Coarse Model | 99.68 | 99.61 | 99.06 |
| AlexNet | 99.32 | 98.38 | 94.79 |
| Resolution | Characters 3 | | |
| | Resolution1 | Resolution2 | Resolution3 |
| Fine Model | 99.7 | 98.94 | 96.24 |
| Coarse Model | 99.67 | 99.59 | 98.82 |
| AlexNet | 98.65 | 97.63 | 90.91 |
| Resolution | Characters 4 | | |
| | Resolution1 | Resolution2 | Resolution3 |
| Fine Model | 99.4 | 98.43 | 94.08 |
| Coarse Model | 99.68 | 99.56 | 97.67 |
| AlexNet | 98.84 | 97.45 | 89.36 |
| Resolution | Characters 5 | | |
| | Resolution1 | Resolution2 | Resolution3 |

| | | | |
|---|---|---|---|
| Fine Model | 99.6 | 99.08 | 95.3 |
| Coarse Model | 99.68 | 99.61 | 98.86 |
| AlexNet | 98.58 | 98.38 | 90.68 |

**Table 4-4: Descriptive statistics of face recognition results of three CNN models in low-resolution scenes**

| | Maximum | Minimum | Mean | Std |
|---|---|---|---|---|
| Fine Model | 99.81 | 94.08 | 98.17 | 1.78 |
| Coarse Model | 99.68 | 99.1 | 99.33 | 0.54 |
| AlexNet | 99.32 | 89.36 | 96.26 | 3.28 |

A total of six groups were tested for face recognition in the wearing object scenario, and Table 4-5 shows the data statistics about the wearing object face recognition results of the three CNN models. According to the descriptive statistics in Table 4-6, the coarse model is the first, the fine model is the second, and AlexNet is the lowest in the cross evaluation of the maximum value index, which means that the coarse model can achieve the maximum recognition accuracy in the single face recognition result output; in the cross evaluation of the minimum value index, AlexNet is the lowest, the fine model is the second, and the coarse model is the highest, which means that in the cross-sectional evaluation of the mean value metric, the fine model is the highest, the coarse model is the second, and the AlexNet is the lowest, indicating that the fine model performs the best, the coarse model is the second, and the AlexNet is the worst in the overall comparison of face recognition accuracy in different wearer scenarios. In the cross evaluation of the standard deviation index, the fine model is the lowest, the coarse model is the second, and AlexNet is the largest, indicating that the fine model has the most balanced recognition accuracy distribution and the most stable output, the coarse model is the second, and AlexNet is the worst in the distribution of face recognition results under the wearing object scenario. Comparing the statistical indexes, we can find that among the face recognition results distribution of the three CNN models in different wearing object scenarios, the fine model has the best performance, both in terms of accuracy and the balance and stability of accuracy distribution. The recognition accuracy of the coarse model is in the middle level, and the recognition accuracy of AlexNet is the weakest, and in the recognition of some inspection images, AlexNet even appears to be unable to recognize the extreme situation.

**Table 4-5: Statistics of face recognition results of three CNN models in the wearing object scene**

| Wearing object | Characters 1 | | | | | |
|---|---|---|---|---|---|---|
| | glass1 | glass2 | glass3 | hat1 | hat2 | hat3 |
| Fine Model | 97.29 | 96.66 | 95.9 | 94.9 | 94.48 | 94.4 |
| Coarse Model | 97.8 | 98.27 | 96.57 | 88.27 | 77.44 | 72.96 |
| AlexNet | 87.26 | 86.49 | 86.24 | 83.8 | 83.5 | 81.76 |

| Wearing object | Characters 2 | | | | | |
|---|---|---|---|---|---|---|
| | glass1 | glass2 | glass3 | hat1 | hat2 | hat3 |
| Fine Model | 91.93 | 91.45 | 85.72 | 91.37 | 91.33 | 91 |
| Coarse Model | 89.19 | 81.75 | 81.56 | 92.57 | 86.85 | 75.11 |
| AlexNet | 78.37 | 0 | 0 | 79.76 | 80.49 | 79.71 |
| Wearing object | Characters 3 | | | | | |
| | glass1 | glass2 | glass3 | hat1 | hat2 | hat3 |
| Fine Model | 95.76 | 95.59 | 95.32 | 97.04 | 96.1 | 96.18 |
| Coarse Model | 98.14 | 96.74 | 97.45 | 97.3 | 97.16 | 97.32 |
| AlexNet | 87.81 | 85.2 | 82.87 | 0 | 78.27 | 77.47 |
| Wearing object | Characters 4 | | | | | |
| | glass1 | glass2 | glass3 | hat1 | hat2 | hat3 |
| Fine Model | 97.64 | 97.74 | 93.3 | 95.46 | 96.34 | 94.39 |
| Coarse Model | 98.89 | 98.98 | 95.77 | 95.06 | 97.75 | 96.03 |
| AlexNet | 87.47 | 86.97 | 83.53 | 75.46 | 81.81 | 80.02 |
| Wearing object | Characters 5 | | | | | |
| | glass1 | glass2 | glass3 | hat1 | hat2 | hat3 |
| Fine Model | 97.91 | 97.71 | 95.21 | 97.88 | 96.35 | 97.11 |
| Coarse Model | 98.67 | 98.83 | 93.69 | 98.74 | 97.32 | 97.94 |
| AlexNet | 86.3 | 84.06 | 79.86 | 92.65 | 85.82 | 88 |
| Wearing object | Characters 6 | | | | | |
| | glass1 | glass2 | glass3 | hat1 | hat2 | hat3 |
| Fine Model | 98.48 | 96.65 | 96.71 | 96.71 | 98.06 | 98.22 |
| Coarse Model | 99.09 | 98.52 | 98.57 | 98.05 | 98.98 | 99.17 |
| AlexNet | 89.78 | 88.82 | 87.45 | 83.73 | 92.01 | 91.92 |

**Table 4-6: Descriptive statistics of face recognition results of three CNN models under wearing object scenes**

| | Maximum | Minimum | Mean | Std |
|---|---|---|---|---|
| Fine Model | 98.48 | 85.72 | 95.40 | 2.61 |
| Coarse Model | 99.09 | 95.06 | 93.96 | 7.26 |
| AlexNet | 92.65 | 0 | 77.35 | 23.37 |

### 4.3.2 Comparison of experimental results in multi-pose scenarios

The experimental structure of the multi-pose scenario comparison mainly consists of two parts, i.e., standard face photos and different posture face photos. A total of five groups were evaluated, and each group was divided into one standard face photo and ten different posture photos (head down, head up, left half side head, left side head, left side face, right half side head, right side head, right side face, etc.). The experimental procedure for the multi-pose scenario is to take the standard and different pose face photos respectively according to the experimental design.

Five groups of face recognition were tested in the multi-pose scenario, and Table 4-7 shows the statistics of the multi-pose face recognition results for the three CNN models. According to the descriptive statistics in Table 4-8, it can be found that in the cross evaluation of the maximum value index item, the fine model is the first, the coarse model is the second, and AlexNet is the lowest, which indicates that in the single face recognition result output, the fine model can achieve the maximum accuracy in the multi-pose scene. In the cross-sectional evaluation of the mean value metric, the fine model is the highest, the coarse model is the second, and the AlexNet is the lowest, indicating that the fine model performs the best, the coarse model is the second, and the AlexNet is the worst in the overall comparison of the accuracy of face recognition results in different multi-pose scenes. In the cross-sectional evaluation of standard deviation indexes, the fine model is the lowest, the coarse model is the second and AlexNet is the largest, indicating that the fine model has the most balanced recognition accuracy distribution and the most stable output in the distribution of face recognition results in multi-pose scenes, the coarse model is the second and AlexNet is the worst. Comparing the statistical indexes, we can find that among the face recognition results distribution of the three CNN models in the multi-pose scenario, the fine model has the best performance. In terms of accuracy, balance and stability of accuracy distribution, the fine model has the best performance among the three CNN models in the multi-pose scenario, and the leading advantage is obvious. The recognition accuracy of the coarse model is in the middle level, and the recognition accuracy of AlexNet is the weakest, and in the recognition of some inspection images, AlexNet even appears to be unable to recognize the extreme situation.

**Table 4-7: Statistics of face recognition results of three CNN models in multi-pose scenes**

| Multi-pose | Characters 1 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pose 1 | Pose 2 | Pose 3 | Pose 4 | Pose 5 | Pose 6 | Pose 7 | Pose 8 | Pose 9 | Pose 10 |
| Coarse Model | 98.4 | 97.55 | 88.11 | 1.59 | 97.43 | 89.16 | 99.47 | 49.64 | 97.79 | 22.09 |
| AlexNet | 84.81 | 85.44 | 78.66 | 86.54 | 55.17 | 80.58 | 94.36 | 77.86 | 0 | 88.29 |
| Fine Model | 96.61 | 92.71 | 94.99 | 92.92 | 56.55 | 94.22 | 91.54 | 60.71 | 99.12 | 95.93 |
| Multi- | Characters 2 | | | | | | | | | |

| pose | Pose 1 | Pose 2 | Pose 3 | Pose 4 | Pose 5 | Pose 6 | Pose 7 | Pose 8 | Pose 9 | Pose 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Coarse Model | 88.21 | 86.83 | 78.12 | 0 | 84.01 | 75.6 | 0 | 93.35 | 85.95 | 64.89 |
| AlexNet | 22.09 | 96.91 | 1.81 | 92.56 | 94.12 | 82.73 | 0.8 | 98.98 | 77.6 | 3.55 |
| Fine Model | 96.1 | 95.21 | 93.07 | 72.5 | 94.57 | 91.28 | 50.45 | 98.55 | 95.67 | 78.95 |
| Multi-pose | Characters 3 | | | | | | | | | |
| | Pose 1 | Pose 2 | Pose 3 | Pose 4 | Pose 5 | Pose 6 | Pose 7 | Pose 8 | Pose 9 | Pose 10 |
| Coarse Model | 87.94 | 87.02 | 86.5 | 29.41 | 82.08 | 91.78 | 21.69 | 1.35 | 99.27 | 96.96 |
| AlexNet | 81.41 | 81.74 | 0 | 0 | 77.22 | 55.63 | 93.87 | 87.46 | 93.96 | 0 |
| Fine Model | 92.22 | 92.45 | 86.43 | 70.42 | 90.15 | 70.81 | 43.32 | 98.72 | 95.38 | 98.61 |
| Multi-pose | Characters 4 | | | | | | | | | |
| | Pose 1 | Pose 2 | Pose 3 | Pose 4 | Pose 5 | Pose 6 | Pose 7 | Pose 8 | Pose 9 | Pose 10 |
| Coarse Model | 96.92 | 95.34 | 80.54 | 1.91 | 95.75 | 85.35 | 1.08 | 99.19 | 93.29 | 79.15 |
| AlexNet | 81.27 | 79.55 | 67.2 | 0 | 81.33 | 0 | 46.39 | 83.59 | 46.37 | 0 |
| Fine Model | 94.91 | 92.7 | 86.11 | 50.05 | 88.84 | 94.63 | 54.07 | 98.63 | 95.43 | 90.05 |
| Multi-pose | Characters 5 | | | | | | | | | |
| | Pose 1 | Pose 2 | Pose 3 | Pose 4 | Pose 5 | Pose 6 | Pose 7 | Pose 8 | Pose 9 | Pose 10 |
| Coarse Model | 85.98 | 75.57 | 43.27 | 11.09 | 48.62 | 3.6 | 1.71 | 99.26 | 94.48 | 1.5 |
| AlexNet | 80.07 | 0 | 0 | 0 | 0 | 0 | 52.49 | 83.58 | 93.38 | 80.07 |
| Fine Model | 93.26 | 93.46 | 54.21 | 79.2 | 58.82 | 23.1 | 98.85 | 78.86 | 92.68 | 93.26 |

**Table 4-8: Descriptive statistics of face recognition results of three CNN models in multi-pose scenes**

| | Maximum | Minimum | Mean | Std |
|---|---|---|---|---|
| Fine Model | 98.55 | 43.32 | 83.83 | 17.58 |
| Coarse Model | 98.4 | 1.08 | 65.52 | 36.81 |
| AlexNet | 96.91 | 0 | 54.99 | 37.98 |

### 4.3.3 Comparison of experimental results under non-ideal lighting scenarios

The experimental architecture of the comparison under non-ideal illumination scenes consists of two main parts, including the effects of light angle and light brightness changes on the application accuracy.

A total of five groups of face recognition under non-ideal luminance scenes were tested, and Table 4-7 shows the data statistics about the non-ideal luminance face recognition results of the three CNN models. According to the descriptive statistics in Table 4-8, it can be found that in the cross evaluation of the maximum value index item, the coarse model is the first, the fine model is the second, and AlexNet is the lowest, indicating that in the single face recognition result output, the coarse model can achieve the maximum accuracy of recognition in the non-ideal luminance scene; in the cross evaluation of the minimum value index item, AlexNet is the lowest, the fine model is the second, and the coarse model is the highest, indicating that in In the cross-sectional evaluation of the mean value metric, the coarse model is the highest, the fine model is the second, and AlexNet is the lowest, indicating that the coarse model is the best, the fine model is the second, and AlexNet is the worst in the overall comparison of the accuracy of face recognition results in different non-ideal brightness scenes. In the cross-sectional evaluation of the standard deviation index, the coarse model is the lowest, the fine model is the second and AlexNet is the largest, indicating that the coarse model has the most balanced recognition accuracy distribution and the most stable output in the distribution of face recognition results under non-ideal brightness scenes, the fine model is the second and AlexNet is the worst. The coarse model has the best performance among the three CNN models in the distribution of face recognition results in non-ideal luminance scenes, in terms of accuracy, balance and stability of accuracy distribution. The recognition accuracy performance of the fine model is slightly worse than that of the coarse model, and the recognition accuracy of AlexNet is the weakest.

**Table 4-9: Statistics of face recognition results of three CNN models under luminance change scenes**

| Luminance Change | Characters 1 | | | | | |
|---|---|---|---|---|---|---|
| | -30 | -20 | -10 | 10 | 20 | 30 |
| Fine Model | 96.83 | 97.6 | 97.78 | 99.01 | 98.04 | 96.52 |
| Coarse Model | 98.82 | 99 | 99.13 | 99.51 | 99.08 | 98.78 |
| AlexNet | 88.56 | 90.42 | 92.28 | 95.42 | 91.93 | 89.86 |
| Luminance Change | Characters 2 | | | | | |
| | -30 | -20 | -10 | 10 | 20 | 30 |
| Fine Model | 96.81 | 97.14 | 97.26 | 98.92 | 98.11 | 96.39 |
| Coarse Model | 98.41 | 98.44 | 98.99 | 99.5 | 99.27 | 98.2 |
| AlexNet | 87.16 | 87.34 | 90.98 | 93.22 | 92.54 | 88.37 |
| Luminance Change | Characters 3 | | | | | |
| | -30 | -20 | -10 | 10 | 20 | 30 |
| Fine Model | 96.5 | 96.6 | 96.76 | 98.83 | 98.29 | 97.52 |
| Coarse Model | 98.16 | 98.06 | 98.36 | 99.41 | 99.19 | 98.7 |
| AlexNet | 85.26 | 85.06 | 86.37 | 95.63 | 92.09 | 88.76 |
| Luminance | Characters 4 | | | | | |

| Change | -30 | -20 | -10 | 10 | 20 | 30 |
|---|---|---|---|---|---|---|
| Fine Model | 97.07 | 97.11 | 97.98 | 98.02 | 95.81 | 95.28 |
| Coarse Model | 98.5 | 98.54 | 98.63 | 99.03 | 98.28 | 97.81 |
| AlexNet | 86.22 | 86.07 | 89.59 | 91.81 | 89.09 | 86.02 |
| Luminance Change | Characters 5 | | | | | |
| | -30 | -20 | -10 | 10 | 20 | 30 |
| Fine Model | 96.87 | 96.91 | 98.29 | 98.68 | 97.31 | 96.55 |
| Coarse Model | 98.17 | 98.3 | 98.89 | 99.46 | 98.98 | 98.51 |
| AlexNet | 87.08 | 87.03 | 89.94 | 95.02 | 93.32 | 90.52 |

**Table 4-10: Descriptive statistics of face recognition results of three CNN models under luminance change scenes**

| | Maximum | Minimum | Mean | Std |
|---|---|---|---|---|
| Fine Model | 99.01 | 96.5 | 97.36 | 0.90 |
| Coarse Model | 99.51 | 98.3 | 98.74 | 0.45 |
| AlexNet | 95.02 | 85.06 | 89.77 | 2.98 |

A total of five groups were tested for face recognition in the non-ideal angle scenario, and Table 4-11 shows the data statistics about the non-ideal angle face recognition results of the three CNN models. According to the descriptive statistics in Table 4-12, it can be found that in the cross evaluation of the maximum value index item, the coarse model is the first, the fine model is the second, and AlexNet is the lowest, indicating that in the single face recognition result output, the coarse model can achieve the maximum accuracy in the non-ideal angle scenario. In the cross-sectional evaluation of the mean value index, the fine model is the highest, the coarse model is the second highest, and AlexNet is the lowest, indicating that the fine model performs the best, the coarse model is the second highest, and AlexNet is significantly behind in the overall comparison of the accuracy of face recognition results in different non-ideal angle scenes. However, the standard deviation between the coarse model and the fine model is smaller, but the overall value is larger, indicating that in the distribution of face recognition results under non-ideal angle scenarios, the coarse model and the fine model have the most balanced distribution of recognition accuracy, but the output is less stable, and AlexNet is the worst. Comparing the statistical indicators, it can be found that among the face recognition results distribution of the three CNN models in the non-ideal angle scenario, the coarse model and the fine model have similar recognition accuracy and have certain recognition ability, but the recognition stability is poor, and the recognition accuracy of AlexNet is the weakest.

**Table 4-11: Statistics of face recognition results of three CNN models under non-ideal angle scenes**

| Angle Change | Characters 1 | | | | |
|---|---|---|---|---|---|
| | Left side | Right side | Smooth | Left 30 | Right 30 |

| | light | light | light | degrees | degrees |
|---|---|---|---|---|---|
| Fine Model | 94.34 | 93 | 92.31 | 90.79 | 93.29 |
| Coarse Model | 96.86 | 96.9 | 80.64 | 0 | 94.07 |
| AlexNet | 77.56 | 75.43 | 66.03 | 58.63 | 76.93 |
| Angle Change | Characters 2 | | | | |
| | Left side light | Right side light | Smooth light | Left 30 degrees | Right 30 degrees |
| Fine Model | 80.33 | 82.51 | 79.17 | 90.22 | 79.13 |
| Coarse Model | 74.1 | 83.06 | 83.41 | 74.54 | 93.49 |
| AlexNet | 0 | 67.45 | 0 | 0 | 0 |
| Angle Change | Characters 3 | | | | |
| | Left side light | Right side light | Smooth light | Left 30 degrees | Right 30 degrees |
| Fine Model | 78.79 | 95.18 | 27.31 | 90.75 | 90.65 |
| Coarse Model | 96.67 | 66.83 | 85.11 | 74.94 | 49.62 |
| AlexNet | 59.85 | 77.99 | 72.14 | 78.19 | 72.44 |
| Angle Change | Characters 4 | | | | |
| | Left side light | Right side light | Smooth light | Left 30 degrees | Right 30 degrees |
| Fine Model | 93.02 | 91.17 | 79.7 | 93.26 | 90.56 |
| Coarse Model | 83.6 | 89.48 | 72.85 | 94.01 | 89.6 |
| AlexNet | 0 | 76.01 | 68.07 | 59.71 | 0 |
| Angle Change | Characters 5 | | | | |
| | Left side light | Right side light | Smooth light | Left 30 degrees | Right 30 degrees |
| Fine Model | 8.58 | 91.55 | 90.36 | 78.16 | 88.27 |
| Coarse Model | 88.49 | 91.4 | 82.87 | 66.4 | 77.09 |
| AlexNet | 74.3 | 74.44 | 74.46 | 0 | 68.2 |

**Table 4-12: Descriptive statistics of face recognition results of three CNN models under non-ideal angle scenarios**

| | Maximum | Minimum | Mean | Std |
|---|---|---|---|---|
| Fine Model | 95.18 | 78.16 | 82.50 | 20.01 |
| Coarse Model | 96.9 | 0 | 79.44 | 19.70 |
| AlexNet | 77.56 | 0 | 51.11 | 32.32 |

### 4.3.4 Comprehensive comparison

The above quantitative statistical analysis of the experimental results of the three CNN models was conducted separately, and the recognition performance of the three CNN models can be initially judged based on the visual display of statistical data and the indexed display of descriptive statistics. In order to better compare the effectiveness of the three CNN models for face recognition, two core metrics, accuracy and stability, are selected to focus on the effectiveness characteristics of the three CNN models[33][34]. Accuracy is defined as the mean value of data results, and stability is defined as the probability of achieving recognition accuracy of 85 and above, and the experimental comprehensive evaluation of the three CNN models is compiled as shown in Table 4-13.

**Table 4-13: Comprehensive comparison statistics of the three CNN models**

| | | Wearing object | Low Resolution | Multi-pose | Non-ideal brightness | Non-ideal angle |
|---|---|---|---|---|---|---|
| Fine Model | Accuracy | 95.40 | 98.17 | 83.83 | 97.36 | 82.50 |
| | Stability | 100% | 100% | 94.6% | 100% | 98.6% |
| Coarse Model | Accuracy | 93.96 | 99.33 | 65.52 | 98.74 | 79.44 |
| | Stability | 100% | 100% | 82.6% | 100% | 91.5% |
| AlexNet | Accuracy | 77.35 | 96.26 | 54.99 | 89.77 | 51.11 |
| | Stability | 93.2% | 100% | 62.9% | 100% | 81.5% |

Combining the comprehensive comparison statistics in Tables 4-13, the performance of the three models in this experiment is summarized and evaluated as follows.

Fine model: under the wearer scene, the precision and stability of the fine model performed the best, in which the precision reached a high level of 95.40 and the stability maintained 100%, meaning that the fine model was able to stably recognize images in any wearer scene under this experimental design; under the low-resolution scene, the precision of the fine model maintained a high level and the stability also maintained 100%; under the multi-pose In the non-ideal luminance scene, the accuracy of the fine model maintained a high level and the stability was 100%; in the non-ideal angle scene, the accuracy of the fine model decreased significantly, but the stability of the recognition was high. Overall, the fine model has the best output results in the experiments of three specific scenes of wearing objects, multiple poses, and non-ideal luminance, respectively, and the difference with the other two applications is more subtle in terms of low resolution and non-ideal angle, etc. The comprehensive evaluation can conclude that the fine model has the strongest comprehensive ability of face recognition in several non-ideal scenes in this experiment.

The accuracy and stability of the coarse model are robust in the wearable scenario, where the accuracy reaches a high level of 93.96 and the stability remains 100%, meaning that the coarse model can stably recognize images in any wearable scenario under this experimental design; in the low-resolution scenario, the accuracy of the

coarse model remains high and the stability also remains 100%; in the multi-pose In the non-ideal brightness scene, the accuracy of the coarse model is maintained at a high level and the stability is 100%; in the non-ideal angle scene, the accuracy of the coarse model decreases significantly and the stability of recognition is good. Overall, the coarse model has the best output results in the experiments of two specific scenes of resolution and non-ideal angle, respectively, but the recognition ability in two scenes of multi-pose and non-ideal angle has obvious shortcomings. Comprehensive evaluation can conclude that the recognition ability of the coarse model in this experiment under different categories of non-ideal scenes shows a trend of strong and weak polarization, and the overall recognition ability is at a medium level.

In the low-resolution scenes, the accuracy of AlexNet is maintained at a high level, and the stability is also maintained at 100%; in the multi-pose scenes, the accuracy of AlexNet decreases significantly to 54.99, and the stability is also maintained at 93.2%. In the non-ideal luminance scenes, AlexNet's accuracy reaches a medium level and its stability is 100%; in the non-ideal angle scenes, AlexNet's accuracy decreases significantly and its stability is poor. Overall, AlexNet has the worst face recognition ability in several non-ideal scenes of this experimental design framework, and in some specific non-ideal scenes, AlexNet has repeatedly failed to recognize the problem.

# Chapter 5 Conclusion

## 5.1 Summary

Currently, the field of artificial intelligence is developing deep learning, in particular, is an important direction in the field of artificial intelligence, developing in areas such as image recognition and data processing. In the direction of image processing recognition, the requirements for high precision, high speed and high complexity are growing[35][36]. In this paper, the CNN-based face recognition technology for non-ideal scenes is an important application and attempt of deep learning. At the same time, this study aims to understand the structure and operation of convolutional neural and to implement its function for recognition applications[37].

In this study, the basic structure of CNN and existing CNN models are introduced, and the basic principles of CNN implementation for image recognition are analyzed. An improved CNN face recognition network structure is designed to expand the network model based on the original network with 10 layers, including 5 convolutional layers, 4 pooling layers and 1 fully connected layer, and to perform fine training and tuning. This model is used to train the face dataset and to optimize the model parameters. Based on the trained network, the face recognition test is conducted in both conventional and non-ideal scenes using the images in the constructed face database, and the recognition correct rate is 93.9%, which is better than the comparison algorithm.

With this improved CNN network model, an experimental study on the direction of CNN for face image recognition in non-ideal scenes is carried out in this topic, which simplifies a lot of image pre-processing work. The generalization performance of CNN and the high recognition rate are also verified on different databases.

## 5.2 Outlook

Deep learning is in full swing this year, and deep convolutional neural networks perform exceptionally well in the direction of image processing and recognition. However, deep learning still lacks perfect theoretical support and macro framework[38]. At the same time, a major development trend of convolutional neural network is to expand the network depth and width with the development of GPU and other hardware, and to achieve better network performance by relying on the training of a large amount of labeled data. This will certainly be accompanied by the human and material consumption of data collection, high training cost, and more restrictions[39].

For this research paper, the face recognition problem itself is highly complex as well as restrictive due to its limitations. The CNN model described in this paper, as well as the extraction and processing of image data, still has many aspects that can be improved. The recognition method can be further improved and studied in the following aspects:

1. The training of deep convolutional neural networks requires high experimental

hardware, and improving the hardware performance can increase the training speed of the network, thus achieving further expansion of the training data and increasing the complexity of the network, thus improving the network performance. For example, using multi-GPU parallel acceleration operations, etc[40].

2. The CNN designed in this paper is only a preliminary improved model, and a certain number of experiments have been done on a limited database, and the image database can be continuously expanded in the future to further improve the effectiveness of the network.

3. Deep learning requires a large amount of labeled data. Combined with unsupervised learning methods, the study of unsupervised convolutional neural network models can reduce the work of labeling a large number of data sets.

# References


[1] Z. Ma and G. Zhang, Data Science in Face Recognition System Based on BP Neural Network, Journal of Physics: Conference Series, Vol. 1881, The 2nd International Conference on Computing and Data Science (CONF-CDS) pp.28-30 Jan. 2021

[2] Gary K. Y. Chan, Towards a Calibrated Trust-based Approach to The Use of Facial Recognition Technology, International Journal of Law and Information Technology, Vol. 29, No. 4, pp.305-331, Nov. 2021

[3] Y. Huang, H. H. Chen, Deep face recognition for dim images. Pattern recognition, vol. 126, pp.14-25 Jun. 2022

[4] S. Chaabane, M. Hijji, R. Harrabi, H. Seddik. Face recognition based on statistical features and SVM classifier. Multimedia tools and applications. vol. 81, no. 6, pp. 8767–8784. Feb, 2022

[5] K. Giannou, K. Lander, and J. R. Taylor, Attentional Features of Mindfulness are Better Predictors of Face Recognition than Empathy and Compassion-Based Constructs, Psychological Reports, pp.1-35, Feb. 2022

[6] G. Lokku, G. H. Reddy, M. N. G. Prasad, OPFaceNet: OPtimized Face Recognition Network for noise and occlusion affected face images using Hyperparameters tuned Convolutional Neural Network, Applied soft computing, vol. 117, article 108365, pp.1-21, Mar. 2022

[7] P. Li, Q. Zhang, Face Recognition Algorithm Comparison based on Backpropagation Neural Network, Journal of physics. Conference series, Vol 1865, no. 4, 2021, vol. 1865, no. 4, pp.1-9, Apr 2021

[8] X. Bajrami, B. Gashi, Face recognition with Raspberry Pi using deep neural networks, International journal of computational vision and robotics, vol. 12, no. 2, pp. 177–193, Feb 2022

[9] G. Levakov, O. Sporns, G. Avidan, Modular Community Structure of the Face Network Supports Face Recognition, Cerebral cortex (New York, N.Y. 1991), pp. 1-14, Dec 2021.

[10] T. Liu, B. Yang, Y. Geng, S. Du, Research on Face Recognition and Privacy in China-Based on Social Cognition and Cultural Psychology, Frontiers in psychology, vol. 12, pp. 1-12, Dec 2021

[11] L. Shurui, Face Recognition Algorithms Based on Deep Learning, Learning & Education, vol. 4, Jun 2021.

[12] C. Guo, C. Xu, Y. Xin, Intelligent Community Access Method based on Face Recognition under Epidemic Conditions of COVID-19, International Core Journal of Engineering, pp. 712-720, Jul 2021

[13] V. Sundaresan, S. S. Amala, Monozygotic twin face recognition: An in-depth analysis and plausible improvements, Image and vision computing, vol.116, Article 104331, pp. 1-10, Dec 2021

[14] S. D. Lin, L. Chen, W. Chen, Thermal face recognition under different conditions, BMC bioinformatics, vol. 22, no. Suppl 5, pp. 1-18, Nov 2021

[15] P. Modi, S. Patel, A State-of-the-Art Survey on Face Recognition Methods, International journal of computer vision and image processing, vol. 12, no. 1, pp. 1–19. 2021
[16] T. Wang, L. Chen, Research on Automated Information System of Non-sense Attendance Using Face Recognition and Large Database, Journal of physics, Conference series, vol. 2083, no. 4, pp. 1-7, Dec 2021
[17] D. Rawat, Deepanshu, V. Gupta, M. Attri, A. Upadhyay, Face Recognition with Python, IITM Journal of Management and IT, pp24-28, Jun 2021
[18] N.K. Mishra, D. Mainak, S.K. Singh, Multiscale parallel deep CNN (mpdCNN) architecture for the real low-resolution face recognition for surveillance, Image and vision computing, vol. 115, article 104290, pp. 1-11, Nov. 2021
[19] J. Govind, G. C. Zacharias, M.S. Nair, J. Rajan, An empirical study of the impact of masks on face recognition, Pattern recognition, vol. 122, article 108308, pp. 1-17, Feb 2022
[20] J. Coe, M. Atay, Evaluating Impact of Race in Facial Recognition across Machine Learning and Deep Learning Algorithms, Computers (Basel), vol. 10, no. 9, pp. 1-25, Sep 2021
[21] Y. Ma, H. Wang, J. Wan, Z. Wang, Y. Yang, C. Huang, Design and implementation of face recognition system based on convolutional neural network, Journal of physics, Conference series, vol. 2029, no. 1, pp. 1-7, Sep 2021
[22] N. Abudarham, I. Grosbard, G. Yovel, Face Recognition Depends on Specialized Mechanisms Tuned to View-Invariant Facial Features: Insights from Deep Neural Networks Optimized for Face or Object Recognition, Cognitive science, vol. 45, no. 9, e13029, pp. 1-25, Sep 2021
[23] H. Vu, M. Nguyen, C. Pham, Masked face recognition with convolutional neural networks and local binary patterns, Applied intelligence (Dordrecht, Netherlands), vol. 52, no. 5, pp. 5497–5512, Aug 2021
[24] A. Dhamija, R. B. Dubey, Analysis of age invariant face recognition using quadratic support vector machine-principal component analysis, Journal of intelligent & fuzzy systems, vol. 41, no. 1, pp. 683–697, Jul 2021
[25] X. Wang, Z. Zhen, S. Xu, J. Li, Y. Song, J. Liu, Behavioral and neural correlates of social network size: The unique and common contributions of face recognition and extraversion, Journal of personality, vol. 90, no. 2, pp. 294–305, Aug 2021
[26] Q. Tong, J. Su, L. Ma, Z. Wang, Intricate Face Recognition Based On Virtual Sample Generation. Journal of physics, Conference series, vol. 1992, no. 4, pp. 1-8, Aug 2021
[27] L. Li, Z. Lan, Z. Liu. Face Recognition Algorithm based on Image Gradient Compensation, International Core Journal of Engineering, vol. 7, no. 8, pp. 553-558, 2021
[28] Y. Ren, X. Xu, G. Feng, X. Zhang, Non-Interactive and secure outsourcing of PCA-Based face recognition, Computers & security, vol. 110, article 102416, pp. 1-11, Nov 2021
[29] W. Eva, Face Recognition Technology Based on Partial Facial Features, Science insights (Winston-Salem, N.C.), vol. 37, no. 5, pp. 292–297, Jul 2021

[30] C. Karri, O. Cheikhrouhou, A. Harbaoui, A. Zaguia, H. Hamam, Privacy Preserving Face Recognition in Cloud Robotics: A Comparative Study, Applied sciences, vol. 11, no. 14, pp. 1-26, Jul 2021
[31] H. Jing, H. Zhang, T. Chen, L. Chen, Q. Hu. Research on Face Recognition Method based on Deep Learning, International Core Journal of Engineering, vol. 7, no. 7, pp. 165-170, Jul 2021
[32] T. Chen, H. Zhang, and H. Jing, A Survey of Research on Deep Face Recognition based on Gabor Features, International Core Journal of Engineering,vol. 7, no. 7, pp.246-250, July 2021
[33] A.G. Musikhin, B. S. Yu, Face recognition using multitasking cascading convolutional networks, IOP conference series, Materials Science and Engineering, vol. 1155, no. 1, pp 1-6, Jun 2021
[34] M. Chen, Research on Responsive Information Interaction Design Based on Facial Recognition Technology, Journal of physics. Conference series, vol. 1952, no. 2, pp. 1-7, Jun 2021
[35] H. Imaoka, H. Hashimoto, K. Takahashi, A.F. Ebihara, J. Liu, A. Hayasaka, Y. Morishita, K. Sakurai, The future of biometrics technology: from face recognition to related applications, APSIPA transactions on signal and information processing, vol. 10, no. 1, pp. 1-13, May. 2021
[36] L. Zhou, M. Gao, C. He, Study on face recognition under unconstrained conditions based on LBP and deep learning, Journal of computational methods in sciences and engineering, vol. 21, no. 2, pp. 497–508, Sep 2020
[37] P. Lu, B. Song, L. Xu, Human face recognition based on convolutional neural network and augmented dataset, Systems science & control engineering, vol. 9, no. S2, pp. 29–37, May, 2021
[38] W. Liu, L. Zhou, J. Chen, Face Recognition Based on Lightweight Convolutional Neural Networks. Information (Basel), vol. 12, no. 5, pp. 1-19, Apr. 2021
[39] M.M. Ahsan, Y. Li, J. Zhang, M.T. Ahad, D. G. Kishor, Evaluating the Performance of Eigenface, Fisherface, and Local Binary Pattern Histogram-Based Facial Recognition Methods under Various Weather Conditions, Technologies (Basel), vol. 9, no. 2, pp. 1-13, Apr 2021
[40] D. Zeng, R. Veldhuis, L. Spreeuwers, A survey of face recognition techniques under occlusion, IET biometrics, vol. 10, no. 6, pp. 581–606, Apr 2021.