The Development of Face Recognition Technology Based on Deep Learning

Xiaoxun Huang

School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou, 510006, China
S1690628446@outlook.com

Abstract. Nowadays with the constantly moving forward of the society, the development of facial recognition technology and the improvement of people's safety awareness, human beings have begun to conduct more in-depth research on facial recognition technology. Continuous research on deep learning technology can bring about new breakthroughs in facial recognition technology. The article compares artificial neural networks and convolutional neural networks to combine the advantages and disadvantages of the two neural networks. The relevant researchers carried out the integration of multiple networks and multiple channels to achieve a technological breakthrough. The article reviews the current research status and methods in four major directions: thermal infrared technology, facial feature recognition technology, age factor interference recognition technology, and stealth counterattack technology. At the same time, the article also introduces the application scenarios of facial recognition technology, such as in the fields of public security, identity verification, security systems, and smart campuses and so on. Through these three major parts, it comprehensively clarifies the development direction of facial recognition technology. In the future, researchers will continue to make progress in improving the accuracy of face recognition, overcoming the interference of age and expressions on the face recognition system, and overcoming covert attacks. At the same time, facial recognition technology will continue to develop in the direction of greater convenience and multi-modal integration.

Keywords: Face recognition technology, deep learning, neural networks, multimodal fusion

1 Introduction

With the continual development of science and technology, safety awareness has been given importance by people. Unlike traditional authentication methods such as passwords, keys, cards, fingerprints, facial recognition technology can enhance the efficiency of authentication. For identity recognition, by using facial recognition technology, the authentication process can be completed in just milliseconds so that it can improve the efficiency. Due to the fact that everyone has different biological characteristics and facial features vary greatly, it is difficult to modify or copy facial

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_34

information. Therefore, facial recognition technology can significantly enhance security. However, at present, there are still many problems and challenges associated with facial recognition technology. For instance, although the human face is relatively unchangeable, it undergoes alterations in appearance as a person ages. This could potentially cause the system to fail to recognize the same individual. Shifeng Wu et al.investigated whether age and gender have an impact on face recognition. Using deep learning methods for facial feature classification, it was found that the recognition effect of men compared to women varies significantly with age, and the recognition effect in middle age is lower than that in young and elderly people[1]. For instance, every person's facial expression changes constantly in accordance with their mood at any given moment. How can facial recognition technology overcome the problem of facial expression changes becoming a huge challenge. M. Senthil Sivakumar et al.addressed the issue that current facial recognition technology struggles to achieve efficient recognition of facial expression features. They proposed the CNN model using techniques such as data augmentation, max pooling and batch normalization. At the same time, they expanded the detection range of facial emotion classification and improved the model's performance and generalization ability[2]. For another example, how can facial recognition technology overcome the subtle changes brought about by makeup, wearing masks, and facial scars becoming a big problem. ShanzhongLei et al. have proposed a method for detecting forged faces and defending against adversarial attacks, which combines multiple features for decision-making, to address the issue that current face recognition systems are vulnerable to interference from deepfake technology. They have successfully achieved effective defense against deepfake face attacks[3].

This article will commence with the discussion of deep learning technologies related to face recognition and the research on face recognition technology. The role of deep learning neural networks in facial recognition technology is summarized. The existing flaws of facial recognition technology and the future trends of its application scenarios are discussed.

2 Deep Learning-Related Technologies

2.1 Deep Belief Network

Deep Belief Network(DBN) was an early development direction of face recognition technology. Before the rapid development of Convolutional Neural Network (CNN). DBN was the first deep learning model that outperformed the performance of traditional technical methods. The research and development of DBN laid the foundation for the development of CNN. The "unsupervised pre-training" and "supervised fine-tuning" of DBN offers solutions to the training problems of deep networks.

Huang Zhaohao focused on addressing the issue that DBN is prone to being disturbed by facial noise information when extracting facial features for learning. By using the method of multi-technology integration, the facial feature information extracted from the MB-CSLBP operator and the HOG operator is used as the input of the DBN for network training. At the same time, the Dropout technique was incorporated, which enhanced the generalization ability of the network and thus enabled the classification and recognition of face images[4].

Xu Weijun focused on the problem of temporal and spatial modeling in facial video expression recognition. The method of using DBN to introduce the end of the two-stream convolutional neural network for modeling thus improves the problem of the lack of temporal information in spatial features. The researchers set up a two-stream convolutional neural network that integrates DBN structures of different depths. They jointly concluded that a deeper network is better at capturing the highly nonlinear relationships of the fused features[5].

2.2 Convolutional Neural Network

Compared with traditional technologies, Convolutional Neural Network (CNN) can automatically learn the multi-level and deep features of human faces. It can precisely capture the details of the human face, thus achieving a stronger ability for feature extraction. Meanwhile, CNN has deep network robustness and better data processing capabilities, which enables it to withstand external interference. Also， CNN can achieve end-to-end learning, which simplifies the system process and is more efficient. However, the unimproved traditional convolutional neural networks may have problems such as failing to achieve the expected accuracy and needing to improve in terms of efficiency.

Wang Xinyu addressed the issue that CNN has a low accuracy rate in extracting facial emotion features. By adopting lightweight design and incorporating batch normalization and Dropout in the network, the accuracy issue of the original CNN was improved.The researchers found that the improved DSCNN network achieved a higher accuracy rate compared to the original VGG16 network (an increase of 3.04% - 5.68%). Based on the above research, the accuracy and robustness have been significantly improved[6].

Anselme Atchogou et al aimed to address the shortcomings of traditional cameras in dealing with the interference caused by weak light and shadow reflections.The researchers employed CNN, GWO, and the multi-classifier combination method, as well as the GWO-enhanced classifier so that the thermal FER index was improved, enabling accurate emotion recognition in dim and noisy environments[7].

2.3 Multimodal Fusion

Facial feature recognition is an important direction for the development of face recognition technology. Due to the limitations of single-modal methods in recognizing facial expressions, most researchers are now shifting towards multi-modal approaches. As a result, various multimodal fusion technologies emerged, mainly manifested in the facial recognition system which can identify faces and changes in facial expressions based on audio and surveillance videos[8].

Lan Chaofeng et al. focused on addressing the issues of poor integration of current audio-visual technologies and insufficient correlation between facial and language information. They employed the DCNN-U-Net speech separation method and DCNN to extract lip and audio features. Then they repeatedly combined the results multiple times. Using PESQ, STOI, and SDR as the three evaluation metrics for speech separation, the results show that the DCNN-U-Net model achieves more efficient speech separation[9].

Farshad Safavi et al. have been addressing issues such as how to further improve the facial emotion recognition system. They employed a new type of transformer-based multimodal deep network. By integrating the comprehensive features from neurophysiology with facial expression features, the classification of emotions was achieved, which enhanced the arousal degree, valence, preference degree and dominance degree.Based on the above research, they have achieved human-machine empathy[10].

3 Research Direction of Facial Recognition Technology

3.1 Thermal Infrared Technology

Wei Ying pointed out that currently, the accuracy and robustness of facial recognition technology in complex environments are relatively weak, and infrared technology can address the issues of complex environments. She designed an infrared-specific feature extraction block and established an architecture of deep residual feature fusion network (IFR-ResNet). She enhanced the feature representation ability and recognition accuracy in the infrared domain, achieving an improvement in accuracy rate to 0.9989. This also demonstrated the stronger robustness of infrared images in complex environments [11].

Wang Fangbin et al. pointed out that the current thermal infrared technology has problems such as low resolution, unclear details and blurred boundaries. They utilized the Gaussian Difference (DoG) edge feature image to correct the color space channel mapping weights. Using the direction gradient histogram (HOG), the features of the long-wave infrared polarized images of human faces were analyzed. They proposed a framework for integrating the RGB space of facial polarization thermal images and a support vector machine (SVM) for face recognition, thereby achieving an improvement in the accuracy of polarization infrared face thermal image recognition to 75.6% [12].

Marcin Kowalski et al. have pointed out the current situation where research in thermal infrared and near-infrared fields is still insufficient. They adopted a research approach that integrates visible light domain technology, high-quality imaging and thermal infrared low-light imaging capabilities. They tested multiple public databases using four different algorithm architectures and thus proposed a triplet face verification method. Therefore, they achieved an accuracy of 90.61% for TAR @ FAR 1% [13].

3.2 Expression Feature Recognition

Baiqiang Gan et al. addressed the issue that current facial expression recognition is prone to overfitting and gradient vanishing, resulting in low accuracy. Firstly, they replaced the fully connected layer with the average pooling layer and integrated the VGG16 network with the long short-term memory network approach. Thus, a deep learning convolutional network capable of recognizing expressions (DCNNBEA) was designed. The ability to process facial expression images has been enhanced and the accuracy rate has reached 49.8%. Then they used the TAFMN activation function and the multi-layer network TaFMN activation function to optimize the parameters for expression recognition. They reduced the risks of gradient vanishing or gradient explosion, achieving a face recognition accuracy rate of 68.4% [14].

Wang Xiaofeng et al. addressed the issue that current technical methods are unable to accurately capture the subtle changes in facial expressions and excessive computation can lead to overfitting. They employed the Efficient Local Attention Mixture Feature Network (ELA-MFN) and the Multi-Head Mixture Attention Mechanism (MHAtt).At the same time, the method of combining multi-scale convolution kernels with the ELA attention mechanism and generating attention maps through multiple attention heads was adopted. Therefore, the problem of overfitting has been solved, and better robustness has been achieved[15].

Ashi Agarwal et al. addressed the issue where the presence of obstructions covering the face led to the loss of facial information, making traditional facial expression recognition methods inadequate. They employed the MobileNetV2 architecture that integrates the attention mechanism and SE. Thus, they designed a new MobileNetV2 architecture. The researchers enhanced the feature extraction capability and achieved an accuracy rate of up to 97.94%[16].

Morteza Najmabadi et al. addressed the issue that facial expression recognition requires a high level of feature extraction capability. They employed local edge decoding binary patterns and lightweight one-dimensional convolutional neural networks. By adopting the weighted kernel representation strategy to fuse high-order and low-order features simultaneously, the anti-interference ability has been enhanced. The stability of the feature extraction capability of the face recognition system in complex situations has been achieved [17].

3.3 Age factor Interferes with Recognition

Wu Chenmou et al. addressed the issue that the current age-related changes in facial features result in low accuracy rates for face recognition. The researchers employed the innovative Age Semantic Group Face Network (ASGFN) approach. At the same time, an age semantic enhancement module was introduced, which integrated this information with the latent facial information extracted by the general face recognition network. Then, they constructed approximate curves for different ages. This has thereby enhanced the noise resistance capability and the accuracy rate of recognition for different age groups and achieved an accuracy rate as high as 97% [18].

Liu Ermiao pointed out that the current one-dimensional vector method has problems such as the easy loss of key identity feature information, which affects the accuracy and requires reliance on complex models. They used multi-feature fusion and hybrid convolutional attention-based cross-age Transformer network model and a cross-age recognition network model with deep expanded attention mechanism. At the same time, they adopted a feature extraction network that integrates shallow texture and deep semantic information and combined it with convolutional attention. This thus reduces the loss of information such as hair and wrinkles. The recognition rate has reached as high as 99.61% [19].

Anders Hast et al. aimed to address the issue that age variation poses a significant challenge to facial recognition technology. They utilized the InsightFace and DeepFace models and simultaneously employed the facial feature vectors generated by

InsightFace as stable biometric identifiers. Therefore, the acquisition performance was improved, enabling large-scale age recognition [20].

3.4 Invisible Counterattack Attack

Feng Anjin addressed the issue that current facial recognition systems are vulnerable to sample attacks. The researchers employed the diffusion model for generating adversarial samples and the attention mechanism diffusion denoising defense method. They employed text-guided techniques to conceal adversarial perturbations. At the same time, by using face masks to limit the perturbations and combining the attention mechanism with the diffusion model. Therefore, the transferability of adversarial samples was enhanced and the optimal defense time was shortened, thereby achieving dual guarantees for adversarial defense [21].

Wang Yibo et al. were addressing the issues of insufficient samples and limitations of adversarial attacks in current facial recognition technology. They employed the symmetrical network recognition algorithm and used a small sample symmetric training method to obtain feature vectors and trained an SVM classifier based on these feature vectors, while also introducing the PGD algorithm. This has enhanced the anti-attack capability of the face recognition system and enabled face recognition in scenarios where there are insufficient samples and with adversarial attacks [22].

Hu Weitao aimed to address the issue that the facial recognition system would make significant errors when it was subjected to attacks. He employed the down-conversion attack algorithm based on pre-trained samples. At the same time, an iterative approach was adopted to search for perturbations and design an incomplete physical adversarial attack. Finally, the attack algorithm was deployed on the embedded device. They have increased the generation speed of adversarial samples and the success rate of attacks. Moreover, by simulating physical attacks through incomplete physical counterattacks, the attack process was simplified, providing new ideas for counterattacks [23].

4 Applications of Facial Recognition Technology

4.1 The Field of Public Safety

The facial recognition technology has enabled the police to make significant breakthroughs in criminal investigations. It captures facial information through surveillance cameras and microphones, and then analyzes the information and compares it with the database. Thus, it can quickly identify the identity and location of the suspect. For instance, the police can quickly identify the fugitives through surveillance in public places and set up surveillance in the relevant locations. Therefore, the police can achieve rapid apprehension. The emergence of facial recognition technology has significantly enhanced the efficiency of case handling, the accuracy of case processing, and the efficiency of police apprehension.

In the future, facial recognition technology will be widely applied in the field of public security. It is highly integrated with deep learning neural networks, enabling more efficient and rapid identification of criminal suspects and capable of handling complex situations such as diverse environments, multiple angles, age variations, and disguises. It integrates with biometric technologies such as iris and fingerprint recognition and it is applied in scenarios with high security requirements to achieve an increase in accuracy and multi-modal recognition. It is integrated with AR technology, which enables security personnel to collect information in real time and conduct database comparisons, thereby achieving more efficient and accurate results. The above-mentioned research is conducive to maintaining social stability and enhancing the happiness of the people.

4.2 Identity verification field

Face recognition technology has rapidly developed and been widely adopted in various fields such as government services, police investigations, security checks at railway stations and airports, key area control, and community screenings. It verifies the identities of individuals through the rapid extraction, analysis and recognition of facial information, thereby improving the efficiency and accuracy of the verification process. This ensures that phenomena such as impersonating others' identities, forging documents, and losing control of personnel are avoided.

In the future, banks, securities firms and other financial institutions will be widely integrated with facial recognition technology. Through remote control, numerous business operations can be carried out and the risk of identity fraud is also reduced. This has thus achieved higher service efficiency and a better user experience. At the same time, the integration of facial recognition technology with healthcare can enable online identity verification, thereby enhancing convenience. The high accuracy of facial recognition can also prevent issues such as confusion of identities, thus forming an accurate, efficient and convenient medical system. The integration of facial recognition technology with urban construction, such as in railway stations, airports and the Internet of Things, can enable the identification of pedestrians who run red lights, thereby enhancing the service level and safety of the community.

4.3 The Field of Security Systems

Face recognition technology is widely used in various entrances and exits of residential areas, dormitories, office buildings, banks, etc. Identity verification through facial recognition has replaced the traditional key system, reducing the risks of loss and misuse. And they can extract facial information in real time through surveillance cameras. If any suspicious person enters a sensitive area, it can serve as an alarm and preventive measure.

In the future, as deep learning technology continues to develop and progress, the face recognition technology will also be continuously optimized. The speed of face recognition can reach millisecond level. At the same time, facial recognition technology is integrated with emerging technologies such as the Internet of Things, big data, and cloud technology. This leads to the formation of a networked and intelligent system. It is integrated with the community security system to achieve cardless access control, reducing the occurrence of theft incidents and ensuring safety. It is combined with campus security management, and can trace the students' entry and exit records to ensure their personal safety.

5 Conclusion

This article conducts a progressive analysis and comparison by combining two major types of deep learning methods. This article also analyzes the development direction of deep learning and lists the current status of multimodal fusion. It conducts an analysis of the advantages and disadvantages in mainstream areas such as thermal infrared technology, facial expression recognition, cross-age face recognition and stealth counter-attack, and summarizes the current research directions and achievements. Finally, a summary review of the application of facial recognition technology is presented. At present, the facial recognition technology still has a very broad prospect for development. For instance, in complex situations, issues such as micro-expression changes, facial feature variations caused by age, and counter-attack systems have still not been better resolved. Even in dim conditions, the accuracy rate of face recognition remains relatively low. Furthermore, the current level of ease with which facial recognition is carried out still needs to be improved. It is expected that in the future, both watches and glasses that can be worn on the body will be capable of facial recognition. How to achieve greater convenience, privacy and efficiency while improving the aforementioned issues remains a direction that researchers need to strive for.

References

1. Wang, D.: Effect of subject's age and gender on face recognition results. Journal of Visual Communication and Image Representation 60, 116 – 122 (2019)
2. Senthil Sivakumar, M., Gurumekala, T., Megalan Leo, L., Thandaiah Prabu, R.: Expert System for Smart Virtual Facial Emotion Detection Using Convolutional Neural Network. Wireless Personal Communications 133(4), 2297 – 2319 (2023)
3. Lei, S., Song, J., Feng, F., Yan, Z., Wang, A.: Deepfake Face Detection and Adversarial Attack Defense Method Based on Multi-Feature Decision Fusion. Applied Sciences 15(12), 6588 (2025)
4. Huang, Z.: Research on Face Recognition Algorithm Based on Fusion Features and Deep Belief Network. Northeast Petroleum University (2021)
5. Xu, W.: Research on Facial Video Expression Recognition Based on Dual-stream Convolutional Neural Network. Hebei University (2023)
6. Wang, X.: Research on Facial Emotion Recognition Technology Based on Convolutional Neural Network. Dalian Jiaotong University (2024)
7. Atchogou, A., Tepe, C.: Robust emotion recognition in thermal imaging with convolutional neural networks and gray wolf optimization. Signal Processing: Image Communication, 117363 (2025)
8. Hong, H., Shen, G., Huang, F.: A Review on Facial Expression Recognition Technology. Computer Science and Exploration 16(8), 1764 – 1778 (2022)
9. Lan, C., Guo, R., Zhang, L., Wang, S., Zhang, M.: Research on DCNN-U-Net speech separation method based on Audio-Visual multimodal fusion. Signal, Image and Video Processing 19(4), 269 (2025)

10. Safavi, F., Venkannagari, V.R., Parikh, D., Vinjamuri, R.K.: Deep Fusion of Neurophysiological and Facial Features for Enhanced Emotion Detection. IEEE Access (2025)
11. Wei, Y.: IFR-ResNet: An Infrared Face Recognition Network Based on Deep Residual Feature Fusion. Intelligent Computer and Applications, 1 – 10 (2025)
12. Wang, F., Jin, C., Zhu, D., et al.: Infrared Polarization Face Recognition Based on RGB Color Space. Laser and Optoelectronic Engineering Progress 59(12), 535 – 543 (2022)
13. Kowalski, M., Grudzień, A., Mierzejewski, K.: Thermal – visible face recognition based on cnn features and triple triplet configuration for on-the-move identity verification. Sensors 22(13), 5012 (2022)
14. Gan, B., Zhang, C.: Target Detection and Network Optimization: Deep Learning in Face Expression Feature Recognition. Journal of Sensors 2022(1), 5423959 (2022)
15. Wang, X., Han, T., Liu, S., Ajmal, M.S., Chen, L., Zhang, Y., Liu, Y.: MHAN: Multi-head hybrid attention network for facial expression recognition. Pattern Recognition, 112015 (2025)
16. Agarwal, A., Susan, S.: Attention-augmented squeeze-and-excitation enhanced mobile network for occluded facial expression recognition in resource-constrained environments. Signal, Image and Video Processing 19(9), 687 (2025)
17. Najmabadi, M., Masoudifar, M., Hajipour, A.: Weighted classification of deep and traditional histogram-based features with kernel representation for robust facial expression recognition. Applied Soft Computing, 113630 (2025)
18. Wu, C., Lee, H.J.: Learning age semantic factor to enhance group-based representations for cross-age face recognition. Neural Computing and Applications 34(15), 13063 – 13074 (2022)
19. Liu, E.: Research on Cross-Age Face Recognition Based on Discrimination Model. Inner Mongolia Normal University (2024)
20. Hast, A., Zhou, Y., Lai, C., Blohm, I.: Analysis of Age Invariant Face Recognition Efficiency Using Face Feature Vectors. In: International Conference on Robotics, Computer Vision and Intelligent Systems, pp. 47 – 65. Springer, Cham (2024)
21. Feng, A.: Research and Application of Anti-Attack and Defense Technology for Face Recognition Algorithms. University of Electronic Science and Technology of China (2025)
22. Wang, B., Yang, Y., Lou, B., et al.: Research on Sample-Based Face Recognition Algorithm for Anti-Interference. Internet of Things Technology 15(4), 33 – 36 (2025)
23. Hu, W.: Research on Individual Adversarial Attacks for Face Recognition and Embedded Implementation. Donghua University (2025)