Proceedings of the 3rd International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/135/2025.21081

Current Status, Challenges and Prospects of Face Fraud Detection

Yibin Wang[1,a,*]

[1]College of Computer Science, Institute of Beijing University of Technology, Beijing, China
a. martin_2003@emails.bjut.edu.cn
**corresponding author*

Abstract: Face fraud detection is an important technology to ensure the security of face recognition systems and is widely used in identity authentication, financial payment, smart security and other fields. With the increasing sophistication of fraud attack methods, especially the development of high-quality deep fakes and 3D mask technology, face fraud detection faces severe challenges. This paper systematically reviews the current status of face fraud detection research, from traditional methods to advanced technologies based on deep learning, and analyzes mainstream datasets and performance evaluation indicators. At the same time, this paper summarizes the main challenges in this field, including insufficient data diversity, generalization ability of models, and real-time issues. In response to these challenges, current solutions and technical trends such as data augmentation, adversarial training, and lightweight model design are explored. Finally, this paper looks forward to future development directions, including cross-modal data fusion, more efficient detection algorithms, and the establishment of legal and ethical norms, to provide a comprehensive reference for relevant researchers and promote the development and application of face fraud detection technology.

Keywords: face fraud detection, machine learning, deep learning, adversarial attack, biometric recognition.

1. Introduction

With the rapid development of artificial intelligence and biometric recognition technology, face recognition has become a core technology in the fields of identity authentication, financial payment, and intelligent security. Face recognition systems have the advantages of being contactless, convenient, and user-friendly, and are widely used around the world. However, with the popularization of technology, fraud attacks on facial recognition systems have become increasingly sophisticated, evolving from simple photo attacks to advanced fraud methods such as deep fakes and 3D masks, posing a major threat to information security and personal privacy.

As a key line of defense to ensure the security of face recognition systems, face fraud detection technology is becoming increasingly important[1]. This technology analyzes the authenticity of input images or videos, identifies and prevents fraudulent attacks, and ensures the reliability and robustness of face recognition systems. In high-risk scenarios such as financial transactions, border control, and judicial evidence collection, effective fraud detection mechanisms have become an indispensable security guarantee. Especially in the context of the rapid development of deep fake technology, fraud detection technology faces unprecedented challenges.

The basic process of face fraud detection includes four key steps: data collection, feature extraction, model analysis and result judgment[2]. During the data collection stage, the system obtains the facial data to be detected through various sensor devices; the feature extraction stage focuses on analyzing facial texture, geometric structure and dynamic features; the model analysis stage uses machine learning or deep learning algorithms to distinguish authenticity; and finally, security decisions are made based on the analysis results. Each link requires precise technical support and strict quality control to ensure the accuracy and real-time performance of the detection.

In recent years, face fraud detection technology has made significant progress. Traditional methods based on image quality analysis and manual feature extraction are gradually developing towards intelligent methods based on deep learning. The introduction of new technologies such as multimodal fusion, adversarial training and transfer learning has greatly improved the performance and robustness of the detection system. However, in practical applications, fraud detection technology still faces many challenges, including insufficient diversity of data sets, limited model generalization capabilities, and limited computing resources, and innovative solutions are urgently needed.

Strengthening research on facial fraud detection will not only help improve the security of biometric recognition systems, but also promote technological innovation in related fields such as computer vision, deep learning, and adversarial sample research. At the same time, the development of this field has important reference value for improving cybersecurity laws and regulations and formulating ethical norms. Through in-depth research on fraud detection technology and exploring safe and reliable protection mechanisms, it has far-reaching significance for building a more secure and trustworthy digital society.

Faced with increasingly severe security challenges, the research and application of face fraud detection technology will continue to deepen. Starting from the basic types of face fraud, this article deeply analyzes the application of traditional methods, machine learning methods, and deep learning methods in fraud detection, and reviews public data sets and benchmarks. In response to key challenges such as insufficient data diversity, rapid development of fraud technology, real-time requirements, and generalization capabilities, the corresponding solutions and technological progress are discussed. By comprehensively combing through existing research and exploring future development trends, this paper provides references for relevant researchers and promotes the development and application of face fraud detection technology. The structure of this paper follows a logical framework from basic to cutting-edge, from theory to practice, and strives to be comprehensive and focused.

2. Types and characteristics of facial fraud

2.1. Static Attack

Static attacks refer to attacks that use static images to deceive face recognition systems [3]. They are divided into two forms: photo attacks and replay attacks. These attacks are based on the characteristics of static images and attempt to deceive face recognition systems by forging or tampering with input images.

2.1.1. Photo attack

Photo attacks are the most common type of static attacks, where attackers use pre-prepared photos to simulate the facial features of the target object. When the photo quality is high, the facial recognition system may have difficulty judging the authenticity of the image. Attackers can use a variety of photos, including pictures uploaded on social media or public photos in news reports, to deceive the system by printing or displaying these pictures on the screen of smart devices. Modern facial recognition technology, especially those based on 2D images, is vulnerable to such attacks, especially when there are minor changes in lighting conditions and angles, and the system may not be able to identify it as a spoof attack.

2.1.2. Replay Attack

Replay attacks are another typical form of static attacks, where the attacker captures the victim's facial image and plays it back to deceive the recognition system. Unlike photo attacks, replay attacks are usually performed by re-presenting a static image electronically (such as through a screen or display device). In recent years, many facial recognition systems have begun to introduce liveness detection mechanisms to identify whether it is a real face. However, this method may still be threatened by replay attacks, especially when the system cannot accurately determine the dynamic features of the input image. In order to prevent such attacks, many new recognition technologies have tried to introduce multi-dimensional feature analysis based on deep learning, but still need to find a balance between performance and protection.

2.2. Dynamic Attack

Dynamic attacks are more complex because they involve not only the image itself, but also the simulation of dynamic features such as facial movements, expression changes, or lighting changes. This type of attack usually includes two forms: deep fakes and 3D mask attacks[3].

2.2.1. Deep fakes

Deep fake is an advanced fraud method based on deep learning technology, especially generative adversarial networks (GANs). By processing large-scale training data with deep learning algorithms, attackers can generate highly realistic fake videos that accurately simulate the target person’s facial expressions, voice, and movement characteristics[4]. This type of attack not only includes static image information, but also covers complete dynamic behavior simulation, which produces extremely misleading effects in short videos, live broadcasts, etc. The rapid progress of deep fake technology has continuously improved the quality of the videos it generates. Some high-quality fake content has reached a level that ordinary people cannot distinguish between true and false, posing a major threat to traditional facial recognition systems. To cope with this threat, facial recognition systems need to continuously upgrade their protection strategies, integrate advanced technologies such as multimodal recognition, behavioral analysis, and dynamic expression detection, and build a more complete defense system (Figure 1).

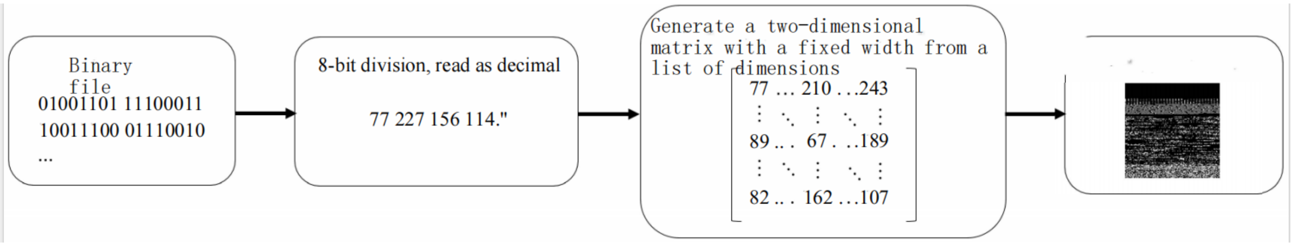

Figure 1: Defense system (Picture credit : Original)

2.2.2. 3D Mask Attack

3D mask attack is an attack method that simulates the target person's face by creating an accurate three-dimensional mask. Compared with traditional 2D photo attacks, 3D mask attacks can better simulate the geometric shape and depth information of the face, breaking through the limitations of static attacks. 3D mask attacks use high-precision 3D scanning technology to obtain the target's facial features and create realistic masks through 3D printing, aiming to deceive facial recognition systems. Due to its high precision and realism, 3D mask attacks are more threatening than traditional static attacks and can even bypass the protection of multiple cameras or depth sensors. Therefore, researchers are exploring the combination of 3D modeling and liveness detection to improve system defense capabilities.

2.3. Adversarial Example Attack

Adversarial sample attacks can be figuratively understood as a kind of "digital illusion." Imagine if people add some "digital noise" that is almost invisible to the naked eye to an ID card photo, it is like covering the photo with an extremely light veil. Although the photo does not look changed to the human eye, it can make the AI system make completely wrong judgments. This attack is particularly clever because it exploits an inherent weakness of deep learning models: over-reliance on pixel-level features. Just as humans can be fooled by visual illusions, AI systems can be fooled by carefully crafted tiny perturbations. To counter this threat, researchers have developed a range of defenses, such as “immunizing”the model through adversarial training so that it learns to recognize these perturbations. But the challenge of this problem is to ensure that the system can respond quickly while protecting the security of the system, and to find a delicate balance between security and performance.

3. Research status

Face recognition technology forms the basic research framework for fraud detection. David J. Robertson's research [5] explored in depth the evaluation and optimization issues of face recognition in security-critical scenarios, and pointed out in particular that although super recognizers (SRs) have made progress in improving the accuracy of unfamiliar face recognition, the standardization of tests and the development of anti-fraud measures still need further research [1]. Related research has elaborated on the evolution and development prospects of facial recognition technology from multiple dimensions, highlighting the important position of face recognition as a future development direction. Although identity authentication based on face as a biometric feature has the advantages of being simple, fast, and contactless, the easy forgery of faces and the diversity of authentication scenarios make facial recognition systems face serious security threats. With the rapid development of deep learning and machine learning technologies [6], researchers have proposed a variety of methods to improve the accuracy and robustness of fraud detection. By systematically evaluating the research status of traditional methods, machine learning methods, and deep learning methods, combined with the analysis of public datasets and benchmarks, the performance evaluation of face fraud detection technology is of great significance.

3.1. Overview of Traditional Methods

Traditional methods are mainly based on image quality analysis and manual feature extraction, and the core idea is to use features designed by experts to distinguish between real and forged face images.

3.1.1. Image quality analysis

Image quality analysis is an important method in traditional fraud detection. Its basic idea is to determine whether the image conforms to the characteristics of a real face by analyzing the quality characteristics of the input image [1]. For example, photo attacks or replay attacks often result in noise, distortion, or compression traces in the image. Based on this, some methods try to determine the authenticity of images by detecting signs of forgery in them. Image quality analysis can be performed from multiple perspectives, including illumination uniformity, detail distortion, blurriness, noise, etc. These features are more effective when processing static images. Image quality analysis detects fraud by evaluating the following features:

$$Q = \{F\_texture, F\_blur, F\_illu\,min\,a\,tion\} \tag{1}$$

$$F\_texture\text{:Texture features; }F\_noise\text{:Noise Level; }F\_blur\text{:Blur degree} \tag{2}$$

3.1.2. Manual feature extraction

Manual feature extraction plays a key role in traditional face fraud detection. Essentially, it captures the most valuable features in face images through carefully designed mathematical operators. Researchers liken these feature extraction methods to the different perspectives of humans observing photos: Local Binary Patterns (LBP) observes subtle changes in skin texture and identifies texture features by comparing each pixel with its surrounding pixels; The Haar feature focuses on the contrast between light and dark on the face, similar to the shadows and lighting effects researchers notice when looking at a photo; the SIFT algorithm can find stable feature points no matter how the photo is rotated or scaled; The HOG feature focuses on capturing edge and contour information. These observation methods from different angles complement each other and together form the basis for identifying real faces and fraudulent images.

3.2. Deep Learning Methods

Deep learning methods have made breakthrough progress in face fraud detection, mainly including the following key directions:

3.2.1. Static feature extraction based on CNN

Static feature extraction based on convolutional neural networks (CNNs) has shown significant advantages in the field of face fraud detection, and has achieved effective recognition of forged faces through multi-level feature learning[7]. A typical CNN architecture contains multiple convolutional layers, pooling layers, and fully connected layers, which gradually extract image features from the bottom layer to the top layer. The shallow convolutional layers mainly extract basic features such as edges and textures, the middle convolutional layers capture facial structures and local semantic information, and the deep convolutional layers integrate global semantic features. During the feature extraction process, the multi-scale feature fusion mechanism integrates feature information at different levels through cross-layer connections, enhancing the model's ability to detect fraud patterns at different scales. The introduction of the attention mechanism enables the network to adaptively focus on key areas in fraud detection, such as eyes, lips and other parts that are easily forged. The application of residual connections effectively alleviates the gradient vanishing problem caused by the increase of network depth and improves the efficiency of feature extraction. The use of feature pyramid networks enhances the model's ability to process images of different resolutions and shows good adaptability in practical application scenarios.

$$\text{Feature Representation} = \alpha\cdot\text{Convolutional features}(X) + \beta\cdot\text{Pooling features}(X) + \gamma\cdot\text{Attention}(X)$$

X - Input face image,α - Convolution feature weight coefficient,β - Pooling feature weights,

$$\gamma\text{ - Attention weight,}\alpha + \beta + \gamma = 1\text{ (Weight Normalization Constraint)} \tag{3}$$

3.2.2. Time series feature analysis based on RNN/LSTM

Temporal feature analysis based on recurrent neural network (RNN)/long short-term memory network (LSTM) provides powerful technical support for dynamic face fraud detection, identifying fraudulent behavior in videos by establishing the dependency relationship of sequence data[8]. The LSTM network, with its special gating mechanism, effectively solves the gradient vanishing problem of traditional RNN when processing long sequences and can capture long-term dependencies in video sequences. In the specific implementation, the network first extracts the spatial features of each frame of the image through the CNN encoder, and then inputs the feature sequence into the LSTM for time series modeling to analyze dynamic features such as changes in facial expressions, blinking frequency, and head movements. This combination of spatiotemporal features enables the model to accurately identify video replay attacks and deep fake videos, especially when detecting fraud features such as lip movements that do not match voice and unnatural facial expressions. Through end-to-end training, the model learns subtle temporal anomaly patterns in fraud videos(Figure 2).

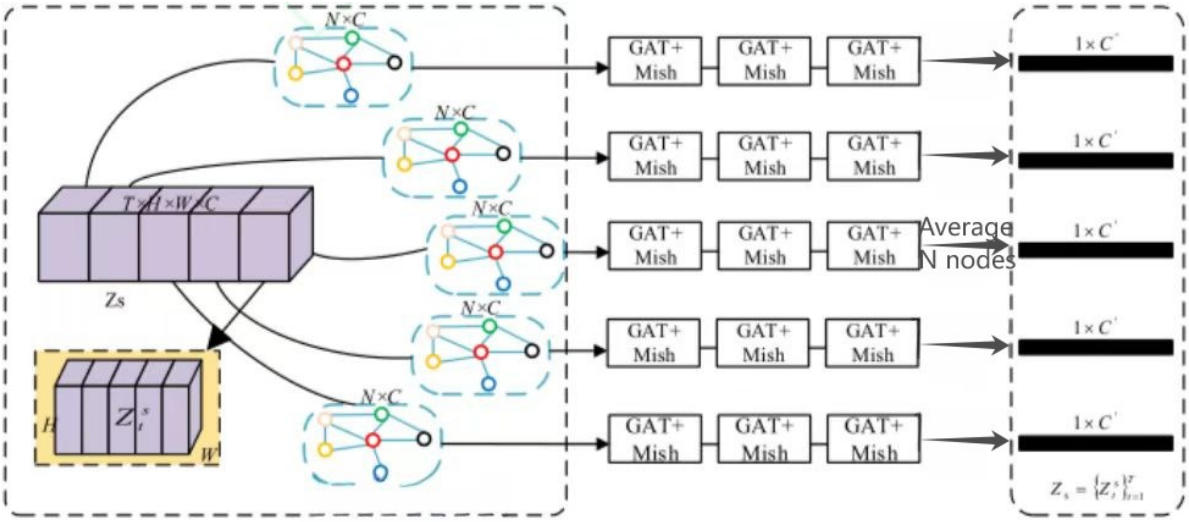

Figure 2: Feature extraction(Picture credit : Original)

3.2.3. Multimodal Fusion Method

Multimodal fusion methods significantly improve the performance and robustness of face fraud detection systems by integrating multi-source information such as vision, depth, infrared, and audio (Figure 3). In feature-level fusion, the features of each modality are fused after adaptive weight adjustment to form a unified feature representation[9]. Decision-level fusion combines the detection results of different models through voting or weighted averaging to make a final decision. The cross-modal attention mechanism realizes the dynamic alignment of visual and audio features and enhances the complementarity between modalities. The adaptive fusion strategy dynamically adjusts the weight distribution of each modality according to different scenarios, improving the environmental adaptability of the system. Physiological feature analysis introduces biometric features such as blink detection and lip movement analysis, and combines depth-guided 3D structure reconstruction and surface normal vector analysis to build a multi-level anti-counterfeiting system. This multi-dimensional feature fusion strategy significantly improves the system's defense capabilities against complex fraud attacks.

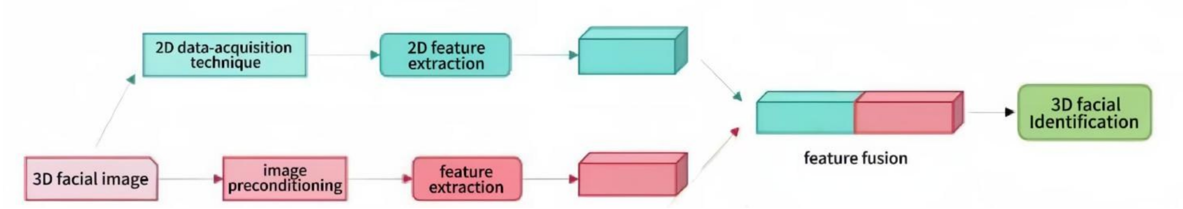

Figure 3: Multimodal fusion method(Picture credit : Original)

3.3. Performance Evaluation

3.3.1. Evaluation Metrics

The performance evaluation of face fraud detection systems is comprehensively measured using multiple standardized metrics [10]. The true positive rate (TPR) reflects the system's ability to correctly identify real faces and is calculated by the ratio of the number of real faces correctly classified to the total number of real samples. The false positive rate (FPR) characterizes the probability that the system mistakenly identifies a forged face as a real face, and is determined by the ratio of the number of forged samples that are misclassified to the total number of forged samples. The equal error rate (EER), as an important indicator of the overall performance of the system, represents the error rate value when the true positive rate is equal to the false positive rate. A lower EER indicates that the system has better recognition performance. The receiver operating characteristic curve (ROC) plots the relationship between TPR and FPR under different thresholds, intuitively showing the performance of the system at different operating points. The area under the curve (AUC) quantifies the overall discrimination ability of the system. The closer the AUC value is to 1, the better the system performance.

3.3.2. Main datasets

Table 1: Some main data sets

Dataset name	Sample size	Features	Main attack types
WMCA	72K	Multimodality	Print, replay, mask attack
CASIA-SURF	21K	Depth Information	Print and replay attacks
Celeb-Spoof	625K	Celebrity Images	Print, replay, deepfake attacks
SWFFD	200K	Wild scene	Comprehensive Attack

Table 1 shows the characteristics and main attack types of different datasets. The research shows that the multimodal fusion method has achieved significant performance improvements on various data sets, especially in dealing with complex scenarios and new attack methods. Future research will continue to focus on how to improve the generalization ability and real-time performance of the model, as well as how to deal with evolving attack methods.

4. Current Challenges

With the widespread application of face recognition technology in security, finance, smart devices and other fields, the importance of face fraud detection technology has become increasingly prominent. However, although existing technologies can effectively prevent some common frauds to a certain extent, they still face many challenges. The following will explore in depth the main challenges currently faced by face fraud detection technology, including the diversity and scale limitations of data sets, the rapid development of fraud attack technologies, real-time and device resource limitations, as well as generalization capabilities and cross-scenario application issues.

4.1. Diversity and size limitations of datasets

Datasets play a key role in modern face fraud detection research, but currently have significant limitations in two core dimensions: diversity and scale. The lack of dataset diversity makes it difficult for models to effectively deal with the diverse types of fraud attacks in real-world scenarios. Existing datasets mainly contain static photo attacks, replay attacks, and basic deep fake samples, and cannot fully cover emerging attack methods such as 3D mask attacks and advanced deep fakes. Researchers have expanded the coverage of attack types by building specialized datasets such as Face3D and deep fake video datasets, but the construction and popularization of these datasets still face major challenges [1]. At the same time, the limitation of dataset size also restricts the improvement of model performance, especially when dealing with complex real-world scenes. Obtaining high-quality fraud samples requires a complex collection process and a large amount of computing resources, and the data annotation process requires a lot of manpower and time. In addition, data privacy protection and ethical compliance requirements make it increasingly difficult to collect large-scale personal data worldwide. Building a fraud detection dataset with sufficient diversity and scale while ensuring data privacy and compliance has become a technical challenge that urgently needs to be overcome in this field[2].

4.2. Rapid development of fraud attack technologies (such as high-quality deepfakes)

The rapid development of fraud attack technologies, especially the breakthrough of deepfake technology based on generative adversarial networks (GANs), has posed a severe challenge to face recognition systems. Deepfake technology can generate highly realistic fake videos and images[11]. Through this technology, attackers can accurately synthesize a complete virtual image that matches the target person’s facial features and voice, and produce highly deceptive content in scenarios such as video calls, live broadcasts, and surveillance videos. As technology continues to mature, attackers have the ability to generate complex dynamic videos that can realistically simulate human expressions, lip shapes, and facial movements, rendering traditional static image fraud detection methods ineffective [12]. This high-quality deep fake attack places higher precision requirements on facial recognition systems. Attackers fine-tune the generated content to make fake videos more difficult to distinguish from real videos. Faced with these technical challenges, existing fraud detection systems need to break through the traditional static detection paradigm and develop new detection methods that can analyze video frame sequences and dynamic features. Developing more intelligent fraud detection technology and improving the ability to identify high-quality deep fake attacks have become important research directions in this field [7].

4.3. Real-time performance and device resource limitations

Real-time performance has become a key challenge for face fraud detection technology, especially in application scenarios such as financial payment, mobile device unlocking, and security monitoring that require extremely high response speed[4]. Modern fraud detection methods, especially deep learning-based algorithms, require a large amount of GPU computing resources during training and inference, which is in sharp conflict with the demand for real-time processing[12]. Researchers use technical means such as model quantization and pruning optimization to reduce computational complexity, and use edge computing to migrate tasks from the cloud to local devices to reduce latency and improve response speed[13]. At the same time, real-time processing also requires efficient integration and analysis of data streams from multiple sensors such as RGB cameras and depth cameras to achieve rapid decision-making while ensuring detection accuracy. This requires that the fraud detection system achieve an optimal balance between accuracy and speed in algorithm design and engineering implementation[14].

4.4. Generalization and cross-scenario application issues

The challenge of generalization ability essentially reflects the significant gap between laboratory environments and real-world applications. From the perspective of environmental factors, the lighting, angles, and backgrounds in real scenes are all dynamically changing and far more complex than the conditions in the training data. At the same time, there are obvious differences in imaging characteristics between cameras of different brands and models. High-end devices may be equipped with HDR technology and better photosensitive elements, while entry-level devices have relatively poor image quality. This hardware difference will affect the detection effect. In addition, training data often comes from specific scenarios and lacks sufficient diversity. Just like teaching a model to identify fraud in front and in full light does not mean it can identify fraud in the side and in dim light. To solve this challenge, it is necessary to build a richer training data set and develop an algorithm architecture that can adapt to different environments and device characteristics.

5. Solutions and technological progress

5.1. Improved data augmentation and generation techniques

Improved data augmentation and generation techniques provide practical solutions to the problem of insufficient dataset diversity. By combining multiple image transformation operations such as random rotation, scaling, flipping, color jittering, etc., the scale of training data is successfully expanded. When processing deep fake attack samples, methods such as brightness adjustment, contrast enhancement, and noise injection are used to simulate feature performance under various lighting and imaging conditions. The face synthesis model based on conditional generative adversarial network (cGAN) has outstanding performance in data generation. By introducing identity preservation loss and feature consistency constraints in network design, the generated face images show rich expressions, postures and lighting changes while maintaining identity information. After adopting these improved data augmentation and generation techniques, the detection ability and generalization performance of the model are significantly enhanced.

5.2. Adversarial Training Methods

Adversarial training enhances the robustness of models by constructing high-quality adversarial examples [8]. In practice, a multi-step adversarial attack based on projected gradient descent (PGD) is used to generate deceptive samples that are visually similar to the original images but contain carefully designed perturbations. During the training process, the generated adversarial samples are dynamically added to the training set so that the model can learn to resist these attacks (Figure 4). Experiments show that the model trained through adversarial training has stronger defense capabilities when facing unknown attacks, and the detection accuracy is improved.

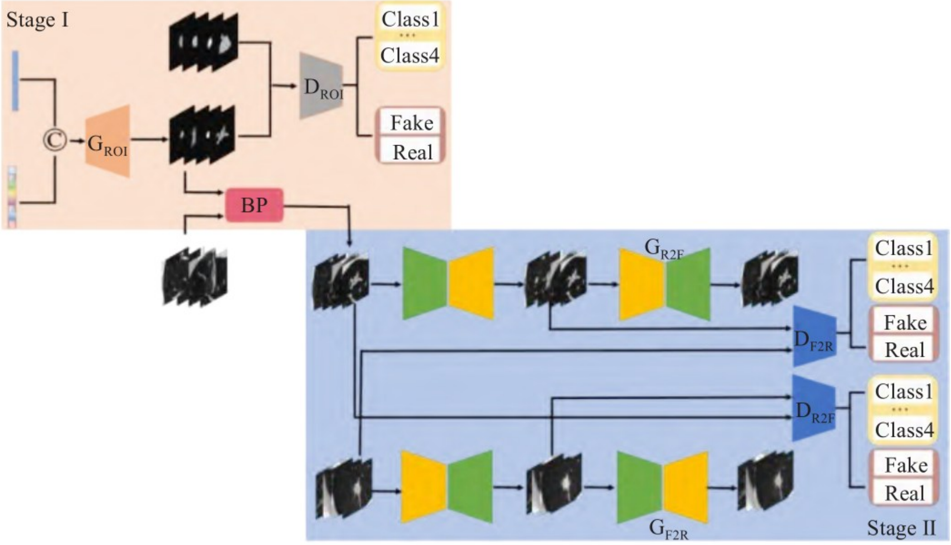

Figure 4: Adversarial training(Picture credit : Original)

5.3. Federated learning and privacy protection techniques

The federated learning framework solves the data privacy problem through distributed training[15]. In the specific implementation, each terminal device trains the model based on local data and only uploads the encrypted gradient information to the central server for aggregation. Combined with the differential privacy mechanism, random noise is added to the uploaded gradients to effectively prevent the leakage of private information. In actual deployment, this scheme not only protects user data security, but also achieves continuous optimization of model performance, with an accuracy rate reaching that of centralized training [16].

5.4. Efficient lightweight model design

The lightweight model design significantly reduces computational complexity through depth-wise separable convolution and channel pruning techniques[17]. The "teacher-student" knowledge distillation framework is adopted in the network architecture to transfer the knowledge of the large model to the small network. At the same time, it combines a variety of optimization methods such as model quantization and pruning, as well as a special architecture design for mobile devices, so that the model can run efficiently in a resource-constrained environment, taking into account both performance and efficiency requirements.

6. Future prospects

As a key link in the security protection of biometric recognition systems, face fraud detection technology has made significant progress in algorithm design, feature extraction, and system deployment. In response to the key challenges currently faced, future development will continue to evolve in the direction of greater efficiency, intelligence, and security.

In view of the challenges of generalization and computational efficiency of current fraud detection algorithms, future algorithm optimization will focus on model architecture innovation and training strategy improvement[18]. In terms of architecture design, the transformer-based multi-head attention mechanism provides a new idea for capturing fine-grained forgery features. By constructing a multi-scale feature pyramid network (FPN) and combining it with a cross-level feature fusion module, the model's ability to perceive local forgery traces is enhanced[11]. At the same time, adversarial training and knowledge distillation techniques are introduced, and a "teacher-student" network structure is used to achieve model lightweighting, significantly reducing computational complexity while maintaining detection accuracy. The meta-learning framework has shown unique advantages in improving the generalization ability of the model. By designing a fast learning algorithm based on task adaptation, the model can quickly learn the feature representation of new attack patterns from a small number of samples. Combined with the contrastive learning strategy, positive and negative sample pairs are constructed to enhance the model's adaptability to different scenarios and devices by maximizing the distance between real samples and forged samples in the feature space. This method, which combines multiple advanced learning paradigms, provides a new technical path to break through the performance bottleneck of existing detection algorithms.

Cross-modal data fusion technology significantly improves the performance of fraud detection systems by integrating multi-source data such as facial texture, depth information, and infrared spectra[19]. The feature alignment mechanism based on deep neural networks achieves effective fusion of information from different modalities and dynamically adjusts the importance of each modal feature by designing an adaptive weight allocation strategy[20]. Especially when dealing with advanced fraud attacks, dynamic detection methods that integrate temporal information and physiological features show obvious advantages. By analyzing temporal features such as blinking frequency and micro-expression changes, deep fakes and 3D mask attacks can be effectively identified. The introduction of the multi-task learning framework provides an innovative paradigm for feature extraction [21]. By learning fraud detection, facial segmentation, expression recognition and other related tasks in parallel, a knowledge transfer channel between tasks is established. A feature sharing module based on the attention mechanism is designed to achieve selective fusion of different task features and improve the model's ability to express fraud features. This multi-objective optimization strategy not only enhances the generalization ability of the model, but also reduces the dependence on large-scale labeled data. By constructing a hierarchical loss function and balancing the learning objectives of each task, the overall detection performance can be improved. This method of integrating multi-task learning has opened up a new direction for the development of fraud detection technology.

In the future, it is crucial to establish a standardized evaluation system for the development of face fraud detection technology. A hierarchical test dataset can be constructed to cover a variety of fraud types from simple photo attacks to advanced deep fakes, taking into account influencing factors such as lighting, angles, and device characteristics. Then, the test samples are graded through the difficulty evaluation mechanism based on adversarial samples to achieve accurate measurement of the performance of the detection algorithm. The evaluation index system can be roughly considered from the dimensions of accuracy and real-time performance, using ROC curve analysis and EER to evaluate detection capabilities, designing a comprehensive scoring mechanism for computing delay, memory usage and energy consumption to establish a robustness evaluation framework and test the algorithm's resistance to new attacks.

However, the widespread use of face fraud detection technology raises profound legal and ethical issues[5]. On the technical level, it is crucial to improve system transparency by designing interpretable detection models.

7. Conclusion

This article summarizes the current status of research on facial fraud detection and analyzes the main challenges and solutions at present. The widespread use of facial fraud detection technology has also raised many legal and ethical issues. Existing technologies have made some progress in improving detection accuracy and system transparency. The introduction of mature technologies such as rad-CAM visualization enhances the interpretability of the system and helps users and regulators understand model judgments. The feature extraction method based on homomorphic encryption protects the security of biometric data, while the introduction of federated learning framework and differential privacy technology further reduces data leakage and privacy risks.

In the future, with the advancement of privacy protection technology and laws and regulations, the security and compliance of biometric data will be further guaranteed, improving the accuracy of the detection system and user experience.

References

[1] *Liu, B., You, Z. C., Jiang, F. L. (2024). Research on ethical governance technology of biometric technology - based on technology track theory. Chinese Medical Ethics, 37(10), 1125-1132.*

[2] *Zeng, M., Yang, X. (2025). Control-source legal response to the crisis of non-essential spread of facial data. Lanzhou Journal, 1-21.*

[3] *Li, L. X., Mu, X. H., Li, S. Y., Peng, H. P. (2020). A Review of Face Recognition Technology. IEEE Access, 8, 139110-139120.*

[4] *Xiao, J. L. (2024). Design and implementation of access control interlocking system in electromagnetic radiation places. Anhui University of Science and Technology, Engineering Technology II.*

[5] *Qin, W. K. (2024). Research on deep fake video detection method based on spatiotemporal features. Chinese People's Public Security University Social Sciences I; Information Technology.*

[6] *Hu, Y. J., Wang, Y. F., Liu, B. B., et al. (2021). A Survey on the Latest Development and Typical Methods of Face Anti-Spoofing. Journal of Siganl Processing, 37(12), 2261-2277.*

[7] *Zhang, W. X., Wang, X. L., Wang, X. Y., et al. (2024). A Deepfake Face Detection Method with Enhanced Focus on Forgery Regions. Journal of Graphics, 1-13.*

[8] *Zhang, Z. M. (2024). Mobile phone user fatigue monitoring and evaluation based on facial video information. University of Science and Technology Beijing from https://kns.cnki.net/kcms2/article/abstract.*

[9] *Meng, Y., Chen, J. B., Zhang, Z., et al. (2024). Knowledge and data driven intelligent interpretation of remote sensing images: progress and prospects. Journal of Remote Sensing, 28(11), 2698-2718.*

[10] *Zhao, J. H., Li, H. C., Wang, D. M., et al. (2024). Model pruning technology in the Internet of Things: current status, methods and prospects. Journal of Internet of Things, 8(04), 1-13.*

[11] *Rossler, A., et al. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. Proceedings of the IEEE International Conference on Computer Vision, 1-11.*

[12] *Liu, Y., et al. (2020). Multi-Modal Deepfake Detection. IEEE Transactions on Image Processing.*

[13] *LI, G. L. (2023). Criminal Law Regulation of Face Recognition Authentication Cracking. Trends in Social Sciences, 11,43-51.*

[14] *Pizlo, F., et al. (2019). Fast and Scalable Image Processing on Edge Devices. ACM International Conference on Computing Systems.*

[15] *Yu, J. F., Hua, C. J., Jiang, Y. (2024). Deep Learning. Chemical Industry Press, 202407.440.*

[16] *Huang, Y., Zhang, X. X., Hu, S. L., et al. (2024). Abnormal data identification method of photovoltaic array based on two-step Pair-Copula[J]. Acta Energiae Solaris Sinica, 5(12),10-21.*

[17] *Ling, X. M., Chen, H. Y., Zhang, X. Y., et al. (2025). Speaker recognition algorithm based on ASP-SERes2Net. Journal of Beijing University of Technology, 51(01), 42-50.*

[18] *Ren, Q. Y., Wang, Y. D., Shi, J. (2024). Research progress of convolutional neural network target detection algorithm. Science Technology and Engineering, 24(32), 13665-13677.*

[19] *Wang, Z. Y., Zhang, Z. Y., Zhu, R. J., et al. (2024). Research on the application of machine learning in cognitive enhancement. Psychological Science, 47(06), 1519-1529.*

[20] *Chen, H. Y., Hu, R. X. (2024). Research on the application of deep learning technology in criminal investigation image processing. Journal of Hunan Police College, 36(04), 75-84.*

[21] *Qu, H. R., Yang, Z. L., Zhang, H., et al. (2024). A review of image depth perception SLAM based on deep learning. Navigation Positioning & Timing, 11(06), 11-27.*