RESEARCH ARTICLE

Deep fusion of iris and periocular regions based on multi-modal biometric recognition technology

Yudong Chen[1, 2], Huanxiang Zheng[3], Qinying Ye[4, *]

[1]Department of Ophthalmology and Optometry, The First Dongguan Affiliated Hospital of Guangdong Medical University, Dongguan, Guangdong, China. [2]The Second School of Clinical Medicine, Guangdong Medical University, Dongguan, Guangdong, China. [3]International School, Jinan University, Guangzhou, Guangdong, China. [4]Department of Ophthalmology, The Second Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong, China.

Received: January 6, 2026; **accepted:** June 22, 2026.

The multi-modal fusion recognition of iris and eye area can effectively improve the performance of biometric recognition in complex scenes by integrating texture and structural information. To address the difference in multi-modal heterogeneous feature expression and insufficient information utilization in existing fusion methods, this study proposed an iris eye depth fusion model based on adaptive Gabor feature weighted fusion. The model constructed dual-stream convolutional network architecture, introduced a Gabor filter with learnable parameters in the feature extraction front-end to enhance texture representation capabilities, and combined a bidirectional cross-modal attention mechanism to achieve adaptive fusion of the two modal features. The results showed that the recognition accuracy on the dataset from Iris Image Database (version 4) of the Institute of Automation, Chinese Academy of Sciences (CASIA-IrisV4) reached 99.5%, and the equal error rate was reduced to 0.5%, which was better than the single-modality recognition method. In degraded scenarios such as blur and occlusion, the equal error rate of the fusion model was controlled within 1.0%, showing excellent robustness. Compared with existing research, this model effectively addressed heterogeneous feature expression differences and insufficient information utilization through the learnable Gabor filter and attention fusion mechanism.

Keywords: iris recognition; periocular features; multi-modal fusion; Gabor filter; attention mechanism.

[*]**Corresponding author:** Qinying Ye, Department of Ophthalmology, The Second Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong 524023, China. Email: zjyqy209@126.com.

Introduction

In today's rapid development of the information age, reliable identity authentication technology has become the cornerstone of ensuring national security, social order, and personal privacy [1]. Biometric identification technology has been widely used in fields such as security monitoring, financial payment, entry and exit management, and smart device unlocking due to its inherent uniqueness, stability, and resistance to theft or loss [2]. Among many biological characteristics, iris recognition is often considered to be one of the most accurate biometric technologies due to its extremely high uniqueness and stability over the life course [3].

Many studies have been done on this topic. He *et al*. proposed an enhanced heterogeneous iris verification model and verified that its recognition accuracy reached 99.88% on the dataset, which outperformed the comparison algorithm [4]. Li *et al*. developed a testing method and optimization strategies to address the image quality fluctuations in high-throughput iris recognition, providing valuable insights for algorithm development [5]. Zambrano *et al*. proposed an iris recognition method that combined low-pass filter and adaptive threshold coding, which significantly improved the performance [6]. However, single iris recognition is susceptible to occlusion and deception. Some studies complement each other by fusing two or more biometric features. The periocular area is anatomically adjacent to the iris and can be more easily obtained when partially obscured by the iris, therefore, periocular recognition is often combined with iris recognition [7]. Gupta *et al*. developed a real-time system based on machine learning for driver drowsiness detection to generate alerts through eye tracking [8]. Pi *et al*. improved facial recognition under mask occlusion with the recognition accuracy increased by 4 - 6% and more efficient [9]. Multi-modal and Other Multi-biometric Fusion (MOMF) technology provides new solutions to improve the reliability, security, and applicability of identification systems by integrating complementary information from multiple biometrics [10, 11]. Many researchers have launched investigations into it. Singh *et al*. proposed a security method based on MOMF, encryption, and watermarking in response to the urgent need for copyright protection and privacy leakage in digital image transmission and storage. The method showed high security and strong attack resistance with its concealment and robustness reaching 62.83% and 25.16% higher than those of the existing solutions [12]. Ryu *et al*. studied the problems inherent in a single biometric system, as well as the insufficient research on the fusion of facial and keystroke biometrics and the lack of heterogeneous multi-modal adaptability. The study first compared feature/score level fusion and then proposed hybrid strategies such as dual-path adaptation. Verified by three virtual datasets, dual-path adaptation had the highest accuracy, highlighting the key role of the adaptation strategy [13]. Since its sensitivity to local texture features, Gabor transform can effectively alleviate the expression differences between heterogeneous features by extracting detailed information at different scales and directions. Muzaffar *et al*. proposed a texture classification method based on Gabor filter to improve the anti-noise capability by rearranging filter responses, energy threshold binarization, and histogram feature extraction. The verification results were better than other advanced methods [14]. Zhao *et al*. proposed a lightweight grouped separable convolutional network combined with Gabor filter to address the large number of parameters and insufficient adaptability of semantic features of the Convolutional Neural Network (CNN) in hyperspectral image classification. On four real-world datasets, the model showed low training and memory overhead and competitive classification accuracy with few samples [15]. Fan *et al*. developed an aligned multi-level Gabor convolutional network to integrate insights from traditional methods and build a new database. On six benchmark datasets, its performance was better than other popular methods [16].

MOMF can effectively integrate multi-source biological feature information, but single iris recognition is susceptible to interference in complex scenarios. Existing multi-modal fusion methods still do not fully utilize the expression differences between iris and ocular heterogeneous features [17]. To address the heterogeneous feature expression differences and insufficient information utilization, this study proposed an adaptive Gabor CNN iris eye fusion model (AGCNN-IFPM) within the MOMF framework. A dual stream CNN architecture was constructed, and learnable Gabor filters were introduced at the front end to enhance texture representation, achieve adaptive weight allocation, and fuse feature layers through a bidirectional cross-modal attention mechanism, thereby more effectively extracting and fusing complementary information between the two modalities. This study provided an effective multi-modal fusion scheme for improving the robustness and security of identity recognition systems in complex scenarios.

Materials and methods

Single modality recognition of iris and periorbita
During the iris recognition process, the input image was first preprocessed. Then, based on Daugman's rubber sheet model, the circular iris was transformed from the Cartesian coordinate system to the polar coordinate system and normalized into a rectangular texture image. Gabor filters were used to extract texture features, and the recognition was completed by calculating the distance between the feature vectors. The core task of the preprocessing stage was iris positioning and normalization [18]. For the annular characteristics of the iris, this method first used calculus-based circle detection technology for preliminary positioning and then used the gradient integration method to calculate the gradient along the radial direction and perform circular integration. The circle corresponding to the maximum integral value was used as the iris boundary as follows.

$$\begin{cases} \max_{(r,x_0,y_0)} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds \right| \\ G_\sigma = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(r-r_0)^2}{2\sigma^2}} \end{cases} \quad (1)$$

where (x_0, y_0) was the coordinate of the pupil center and the reference point for iris positioning. r was the radius that defined the spatial extent of the iris boundary. $*$ was the convolution symbol, used to illustrate the Gaussian smoothing process. G_σ was a Gaussian smooth function, which smoothed the image to suppress noise and improve the robustness of gradient calculation. $I(x, y)$ was the human eye image. $\oint \frac{I(x,y)}{2\pi r} ds$ was the grayscale integral on the circular outline with (x_0, y_0) as the center and r as the radius. $\frac{\partial}{\partial r}$ was the partial derivative of the radius r, that was, calculating the rate of change of the average gray level on the circular path when the radius changed slightly. The algorithm traversed all possible circle center coordinates and radii in the image. For each group (r, x_0, y_0), the grayscale average of the circular path under the radius was calculated, and then the gradient of this average as the radius changes was observed. When the circular path moved from the inside of the pupil to the iris texture area, the grayscale would change drastically, and the gradient would reach a maximum value. The position corresponding to this maximum value was considered to be the boundary of the iris. After successfully locating the inner and outer boundaries of the iris, the annular iris area needed to be expanded into a rectangular image of a fixed size and normalized to eliminate geometric inconsistencies caused by translation, scaling, and rotation. This study used Daugman's rubber sheet model to convert the iris from the Cartesian coordinate system (x, y) to the polar coordinate system (r, θ). Its coordinate transformation formula was shown below.

$$\begin{cases} x(r,\theta) = x_0 + r \cdot cos(\theta) \\ y(r,\theta) = y_0 + r \cdot sin(\theta) \end{cases} \quad (2)$$

where r was normalized within the interval $[0, 1]$. 0 corresponded to the inner boundary, while 1 corresponded to the outer boundary. θ belonged to $[0, 2\pi]$. Through this mapping, the original annular iris was converted into a fixed-size rectangular texture image, which facilitated subsequent feature extraction and comparison. A Gabor filter was used to extract iris texture features, and iris recognition was conducted by calculating the distance between feature vectors. The periocular area referred to the surrounding area around the eyes including the upper and lower eyelids, canthus lines, eye sockets, *etc*. This area contained rich identification information and was often more accessible than the iris under partial occlusion or non-cooperative conditions. The periorbital recognition was divided into four stages including image acquisition and eye detection, extraction and preprocessing of the region of interest (ROI), feature extraction, and feature matching and recognition. In the image acquisition and eye detection stage, after the digital image containing the face was captured through the camera, the face detection algorithm was first used to determine the facial area. Then, the position coordinates such as the corner points of the eyes and the center of the pupil were accurately calibrated through sophisticated key point positioning technology. After successfully locating the eye area, the process entered the periorbita ROI extraction and preprocessing stage. Based on the located key points, the system automatically cut out a rectangular area containing the upper and lower eyelids, eye wrinkles, and surrounding skin texture according to preset rules, and performed illumination normalization on it as below.

$$I_{norm} = \frac{I-\mu}{\sigma+\varepsilon} \quad (3)$$

where I was the original ROI image. μ and σ were its mean and standard deviation. ε was a small constant set to prevent division by zero. The preprocessed image automatically learned the multi-level feature expression of the periocular area through a neural network based on Gabor filtering, thereby fully mining the subtle texture and geometric structure information. It extracted local texture, edge direction, and spatial context features step by step through multi-scale convolution layers, and combined batch normalization and activation functions to enhance feature discrimination capabilities. Finally, the high-dimensional feature vector was output, which effectively represented the uniqueness of the individual eye periphery and provided a reliable basis for subsequent matching and identification. In the feature matching and recognition stage, the similarity between the extracted feature vector and the pre-stored template in the database was calculated, and identity verification decisions were made based on the threshold comparison results.

Multi-modal iris and periocular depth fusion model construction

This study explored improved methods to overcome the current shortcomings of iris recognition and periocular area by proposing an Adaptive Gabor CNN-Iris-Periocular Fusion Model (AGCNN-IPFM). By combining the Gabor filter with biological vision mechanism and the powerful deep CNN and fully exploiting the complementary information of the two modalities through adaptive deep fusion at the feature layer, the texture features of the iris and periocular area could be more effectively extracted and fused. The proposed model was divided into four parts including data input, Gabor-enhanced dual-stream feature extraction network, multi-modal deep fusion module, and joint recognition output module (Figure 1). In the data input stage, the normalized iris and periocular ROI images were input. The iris image was accurately positioned, and polar coordinate was transformed to a unified size of 64 × 512 pixels, ensuring that the iris texture was completely preserved. The periocular ROI area was expanded and cropped with the iris as the center, and the size was standardized to 224 × 224 pixels, completely including the upper and lower eyelids, eye corners, and other discriminative feature areas. All input images were grayscale normalized, and histogram equalized to eliminate the effects of illumination differences. In the Gabor-enhanced dual-stream feature extraction network, the Gabor filter with biological rationality was introduced to enhance the CNN's ability to extract texture features. The dual-stream parallel structure was used to process biological features of different modalities. The Gabor enhancement module was divided into two independent feature extraction branches. Its first convolutional layer was enhanced to a Gabor convolutional layer. The Gabor filter had direction selectivity and frequency selectivity and could simulate the response characteristics of the human visual system to texture. The calculation was shown below.

$$G(x,y;\lambda,\theta,\psi,\sigma,\gamma) = exp(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}) cos(2\pi\frac{x'}{\lambda}+\psi) \quad (4)$$

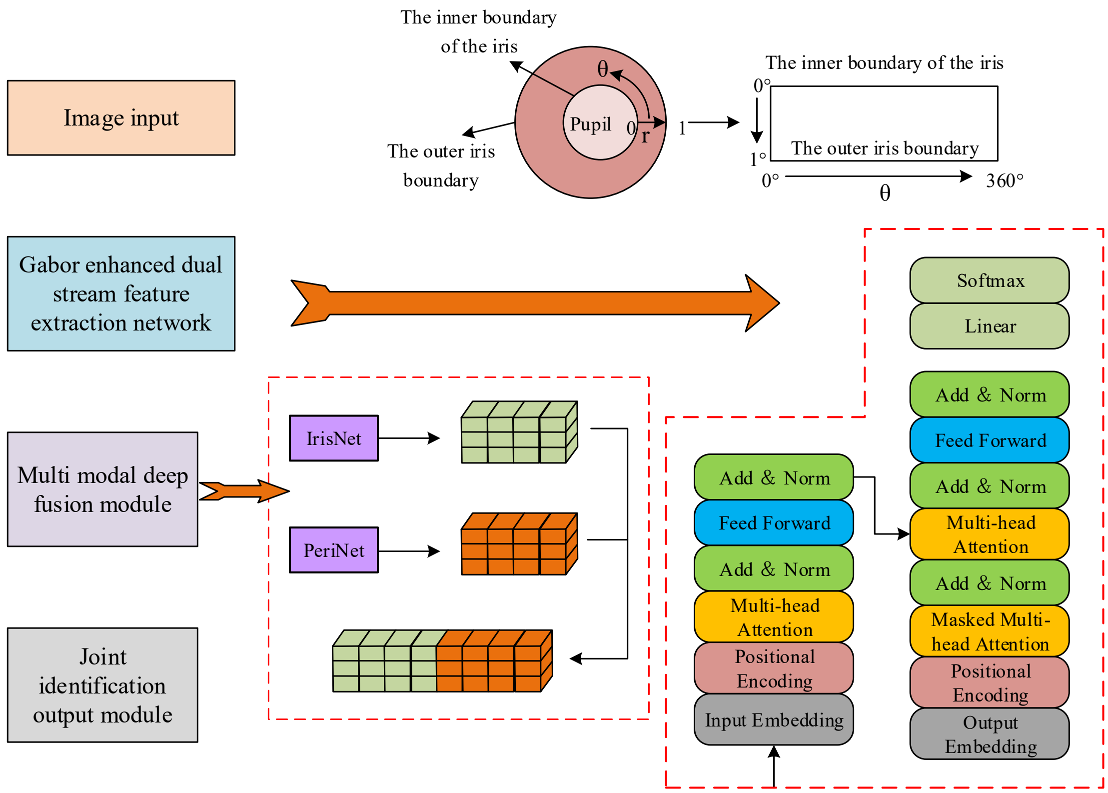

Figure 1. Model flowchart.

where $(x^{'}, y^{'})$ was the rotated coordinate. λ was the wavelength of the sine function. θ was the direction. ψ was the phase offset. σ was the standard deviation of the Gaussian envelope. γ was the spatial aspect ratio. Initial convolution kernel was multiple Gabor filters of different directions and scales, and their parameters were fixed during the training process. The parameters of the Gabor filter were used as network trainable variables, which allowed the network to adaptively optimize filters based on iris and periocular recognition, allowing it to learn the most effective texture representation from the data. The iris branch was built based on the lightweight Residual Network (ResNet) and used the inverted residual bottleneck module to ensure feature expression capabilities while controlling computational complexity. The periocular branch also used the ResNet-18 architecture, using residual connections to ensure effective propagation of deep features. The two branches were re-calibrated in different scale spaces. The modulation coefficient β_s was generated by the fully connected layer. The calibration formula was shown as follows.

$$\beta_s = FC(GAP(F_s)) \quad (5)$$

where F_s was the feature map at the s-th scale. GAP was global average pooling. FC was the fully connected layer. The modulated multi-scale features were fused through weighted summation as below.

$$F_{ms} = \sum_{s=1}^{S} \beta_s \cdot Up(F_s) \quad (6)$$

where S was the number of scales. Up was the up-sampling operation. The multi-modal deep fusion module adopted a Bidirectional Cross-Modal Attention Mechanism (BiCMAM), taking into account the guidance of iris features to periocular features and the guidance of periocular features to iris features. BiCMAM first input the iris features and the periorbital features separately, enhancing the local context modeling ability through horizontal and vertical neighborhood perception. Then, it sent them to IrisNet and PeriNet for deep encoding, respectively. Finally, it passed through a linear layer and output through Softmax. Given the iris features and periocular features, the cross-modal attention weight was calculated as follows.

$$\begin{cases} A_{iris\to peri} = Soft\,max(\frac{(F_{iris}W_Q)(F_{peri}W_K)^T}{\sqrt{d_k}}) \\ A_{peri\to iris} = Soft\,max(\frac{(F_{peri}W_Q)(F_{iris}W_K)^T}{\sqrt{d_k}} \end{cases} \quad (7)$$

where F_{iris} was the iris feature. F_{peri} was the eye periphery feature. W_Q and W_K were the projections of the query and key. d_k was the dimension of the key vector. The feature enhancement based on attention weight was shown below.

$$\begin{cases} \tilde{F}_{iris} = F_{iris} + A_{peri\to iris}(F_{peri}W_V) \\ \tilde{F}_{peri} = F_{peri} + A_{iris\to peri}(F_{iris}W_V) \end{cases} \quad (8)$$

where W_V was the projection matrix of values. Based on the deep fusion of features, the system further introduced a meta-decision mechanism as follows.

$$w = Sigmoid(FC_2(Re\,L\,U(FC_1(GAP(F_{fused}))))) \quad (9)$$

where w was the modal weight vector. By analyzing the modal reliability under different environmental conditions, the system could adaptively adjust the final decision-making strategy. The joint recognition output module was the final decision-making link of the multi-modal deep fusion system, which converted the deep fusion feature representation into reliable identity recognition results. It used fully connected layers and activation functions to eliminate redundant information and retained the most discriminative feature elements.

Experimental setup and evaluation methods

To comprehensively evaluate the performance of iris and periocular recognition technology, this study conducted an in-depth analysis from multiple dimensions including recognition accuracy, F1 score, Area Under the Curve (AUC), Equal Error Rate (EER), False Acceptance Rate (FAR), and True Positive Rate (TPR). Recognition accuracy referred to the proportion of correctly predicted samples to the total sample size. F1 value was the harmonic average of precision and recall, used to measure the comprehensive recognition ability under class imbalance conditions. AUC value reflected the overall ability of the model to distinguish positive and negative samples within the entire threshold range. EER was the error rate when the false acceptance rate and false rejection rate were equal, used to measure the balance between system security and convenience. FAR referred to the proportion of unauthorized samples accepted incorrectly among all unauthorized samples. TPR referred to the proportion of correctly accepted authorized samples to all authorized samples. The experiment used high-precision acquisition equipment and diverse datasets to ensure comprehensiveness and accuracy of evaluation, while following standardized testing protocols to minimize interference from external variables. The hardware environment of the experiment was Windows 10 operating system, Intel Core i7-13700KF processor, and 32 GB of RAM. Python 3.6 (https://www.python.org/) and Global Mapper 24 (64-bit) (Blue Marble Geographics Company, Harlowe, Maine, USA) were used in the research. Data processing was carried out using Pandas 2.1.0 (https://pandas.pydata.org/). These configurations provided a reliable benchmark for subsequent performance comparisons. This study utilized a subset of the fourth edition of the Iris Image Database of the Institute of Automation, Chinese Academy of Sciences (CASIA-IrisV4) (http://english.ia.ac.cn/db/201610/t20161026_169399.html), which contained 600 iris images, covering samples under different individuals, lighting conditions, and collection equipment to ensure data diversity and representativeness. All images were preprocessed, unified in size of 224 × 224 pixels, and divided into training set, validation set, and test set at a ratio of 8:1:1. To verify the contribution of each module to the model performance, this study designed ablation experiments with five sets of experimental conditions. Experiment 1 was a baseline dual stream network that used simple feature cascade fusion without using Gabor filters. Experiment 2 introduced a fixed parameter Gabor filter at the front end of the baseline network for feature extraction. Experiment 3 further set the Gabor filter parameters to trainable. Experiment 4 only added bidirectional cross-modal attention fusion mechanism on the basis of baseline network, without using Gabor layer. Experiment 5 introduced both learnable Gabor filters and attention fusion mechanisms, namely the complete AGCNN-IFPM model. By comparing EER and other indicators in the five experiments mentioned above, the learnability of Gabor filters and the gain of attention fusion strategy on recognition performance were quantitatively analyzed. Subsequently, to evaluate the robustness in degraded scenarios, this study constructed four types of degraded subsets including blur, occlusion, pose changes, and noise based on the CASIA-IrisV4 test set. The fuzzy subset was obtained by processing the image with Gaussian blur kernel (σ = 3.0). The occlusion subset randomly covered a 50 × 50 pixel black square around the eye area to simulate occlusion. A random rotation of ± 15° was applied to the image based on the posture change subset. The salt and pepper noise with a density of 0.05 was added to the noise subset. Each degraded subset maintained the same sample distribution as the original test set, used to test the performance of the model under non-ideal conditions.

Results and discussion

Iris and periorbita recognition effect

This study compared AGCNN-IPFM with Visual Geometry Group-16 (VGG16) [19], Alex Deep CNN (AlexNet) [20], and Lecun CNN (LeNet) [21] on single-modality recognition. The effectiveness of iris recognition demonstrated that, as the number of iterations increased, the accuracy of each model showed an upward trend. AGCNN-IPFM achieved the highest accuracy of 86.2% after 300 iterations, which was the best performance followed by AlexNet and VGG16. LeNet performed the worst with slow improvement in accuracy and low overall accuracy (Figure 2a). The loss of the TI model decreased the fastest and the early convergence was obvious. Although VGG16 declined slowly in the early stage, it stabilized at a low level in the later stage. The losses of AlexNet and LeNet decreased slowly, especially LeNet still had certain fluctuations in the later period (Figure 2b). The trend of periorbital recognition was roughly the same as that of iris recognition. AGCNN-IPFM had an accuracy of 88.3%, which was the best performance followed by AlexNet and VGG16, while LeNet performed the worst with slow improvement in accuracy and low overall accuracy (Figure 2c). The loss rate of each model decreased with the number of iterations during the training process. The loss of AGCNN-IPFM decreased the fastest and the early convergence was obvious. Although VGG16 declined slowly in the early stage, it stabilized at a low level in the later stage. The losses of AlexNet and LeNet decreased slowly, especially LeNet still had certain fluctuations in the later period (Figure 2d). The results confirmed that AGCNN-IPFM had optimal convergence and accuracy in both iris and periorbita recognition tasks, verifying its advantages in feature extraction and model structure design. The F1 value of iris recognition reached 0.985 (Figure 3a), and the AUC was 0.992 (Figure 3b), indicating that it was highly balanced between Precision and Recall (PR) and had a strong ability to distinguish true and false samples within the full threshold range, making it suitable for high security scenarios. The F1 value of periorbita recognition was 0.952 (Figure 3c), and the AUC was 0.976 (Figure 3d). Although it was slightly lower, under degradation conditions such as blur, occlusion, or posture changes, its F1 value and AUC decreased less, reflecting stronger environmental robustness. Specifically, iris was suitable for accurate authentication under high-quality images due to its stable texture and significant differences between classes. Based on

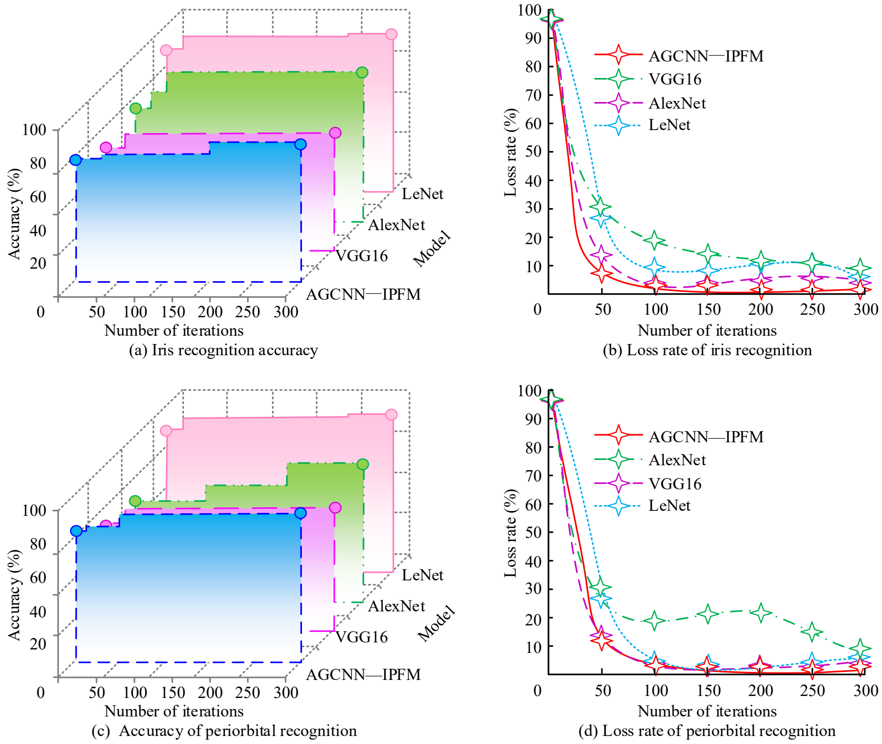

Figure 2. Iris recognition and periorbital recognition effectiveness.

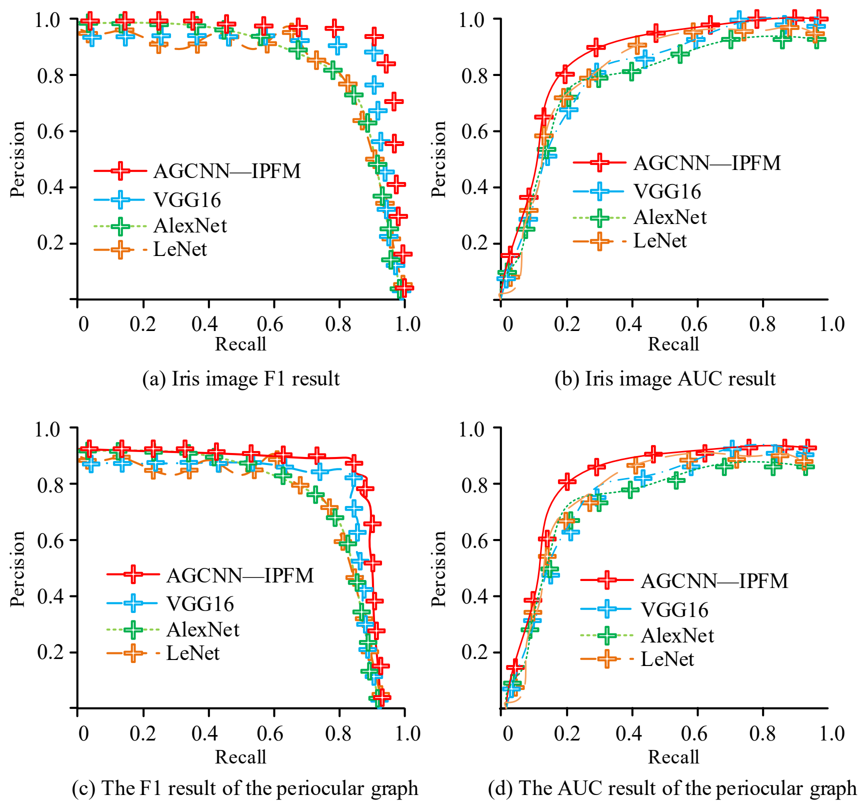

Figure 3. Comparison of F1 and AUC values.

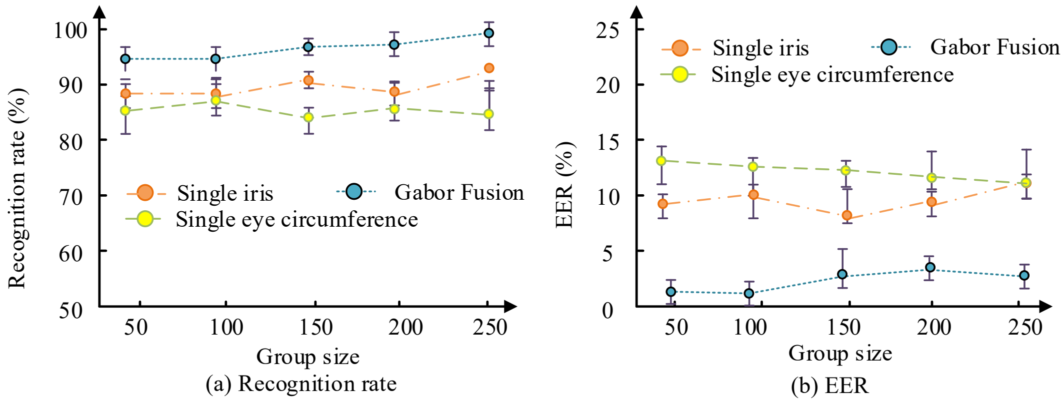

Figure 4. Comparison of different modal performance.

structural redundancy and spatial scalability, periorbita could still maintain effective discrimination in non-ideal acquisition environments.

Multi-modal iris-periocular fusion model effect

Compared with a single modality, the fusion model used complementary information of iris and periorbita features to improve the recognition accuracy and robustness of the system. Considering the challenges faced by multi-modal recognition such as environmental noise, illumination changes, and individual differences, using a fusion strategy could significantly reduce the limitations of a single modality, thereby improving the overall recognition effect. The recognition accuracy of a single iris and a single periorbita on the test set reached 88.7% and 86.3%, respectively, while the recognition accuracy of Gobar after fusing iris and periocular features was 99.5% (Figure 4a). The EERs of single iris and single periorbita recognition were 11.3% and 13.7%, respectively. After fusing iris and periocular features, the EERs of Gabor algorithm and feature-level fusion were 0.5% and 0.6%, respectively (Figure 4b). The Gabor algorithm demonstrated the best recognition performance under multi-feature fusion conditions, which was better than other comparison models. When the FAR was 1%, all models could maintain a high TPR, of which single iris recognition was 99.5%, single periorbita recognition was 98.2%, and the recognition algorithm reached the optimal 99.9%. As the security level increased, the FAR dropped to 0.1% and 0.01%, the performance of each model declined to varying degrees, and the TPR remained at 99.7% and 99.4%, respectively. The TPR of single periorbita recognition dropped to 90.3% while FAR was 0.01% (Figure 5). The recognition algorithm still maintained the highest pass rate under extremely low FAR, proving that its features had extremely high discriminability and reliability, and were very suitable for high-security identity authentication. This study constructed a subset containing blur, occlusion, pose change, and noise to simulate degradation scenarios. Under blurry and occlusion conditions, the recognition accuracy of single iris recognition dropped significantly with the EER under the fuzzy subset rising to 2.8% and the EER under the occlusion subset reaching 4.1% (Figure 6a). Single periorbita recognition performed more robustly under posture changes and noise interference with the EER of the posture change subset reaching 3.5% and the EER of the noise subset reaching 2.9% (Figure 6b). The EER of the fusion strategy under blur, occlusion, attitude, and noise conditions were only 0.9%, 1.0%, 0.7%, and 0.8%, respectively, which fully reflected the robustness of the model in degraded scenarios (Figure 6c). The results further verified the adaptability of its adaptive mechanism to complex environments. The results of ablation experiments demonstrated that, when the Gabor filter with fixed parameters was used for feature

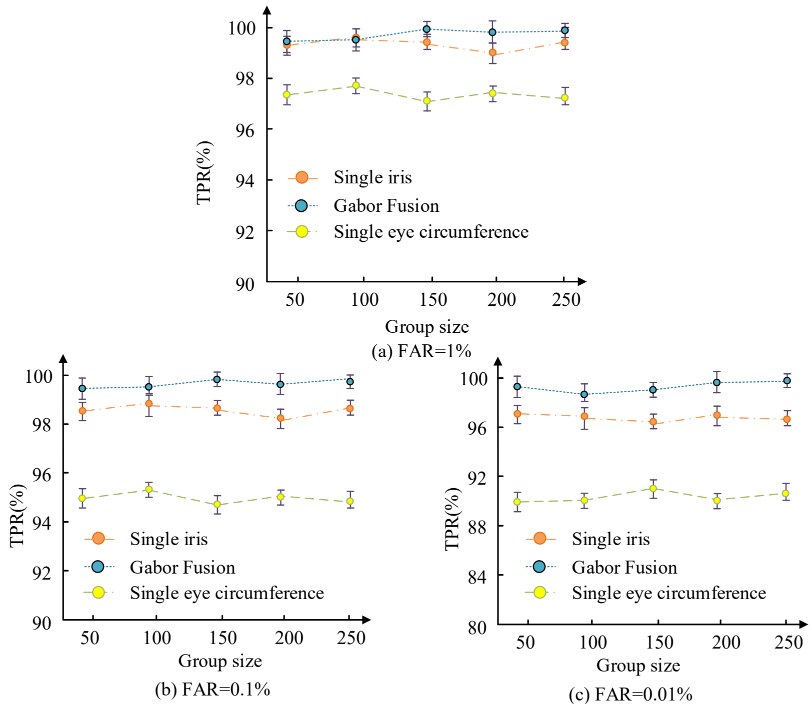

Figure 5. Comparison of true rates.

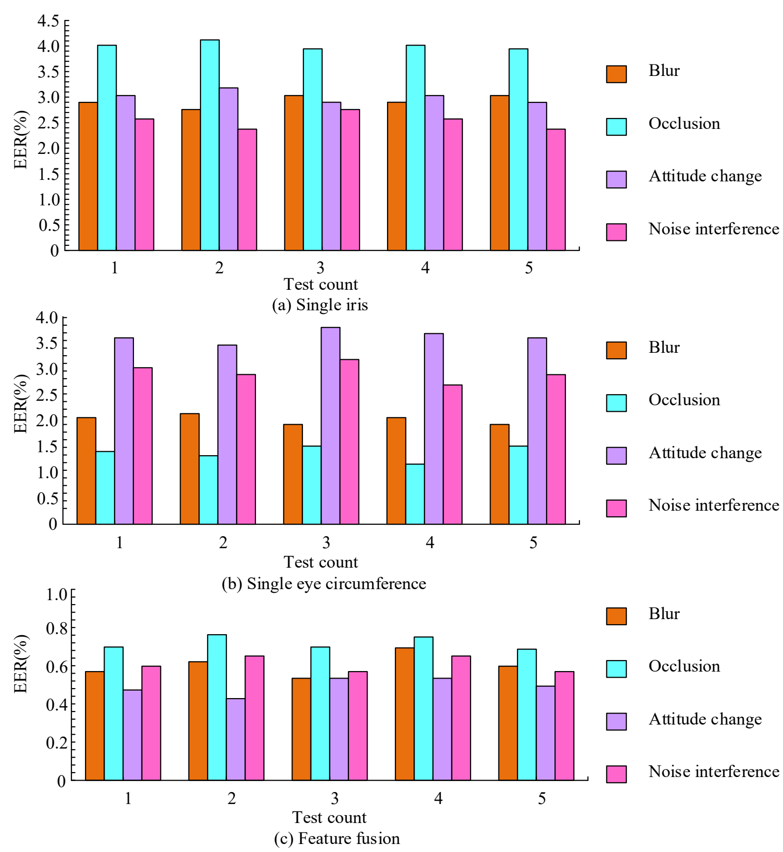

Figure 6. Robustness test.

Table 1. Comparison of ablation experiments.

Experiment Number	Model Configuration	Gabor convolution	Integration policy	EER (%)	Accuracy (%)
1	Baseline model	×	Feature concatenation	0.71	80.3
2	Gabor fixed initialization	√ (Fixed parameter)	Feature concatenation	0.67	84.4
3	Learnable Gabor	√ (Learnable parameters)	Feature concatenation	0.63	90.5
4	Attention fusion	×	Attention weighted	0.59	90.5
5	Complete integration	√ (Learnable parameters)	Attention weighted	0.52	99.7

extraction in experiments 1 to 3, the EER dropped from 0.71% to 0.63%. When the Gabor parameters were trainable, the performance was further improved, and the EER was reduced to 0.59%. The results showed that, by fine-tuning the filter parameters, the network could adaptively learn the specific frequencies and directions that were most suitable for distinguishing iris and periocular textures, thereby extracting more discriminative features. From Experiments 1 and 4, without using the Gabor layer and only replacing the fusion strategy from simple feature concatenation to attention fusion, the EER dropped from 0.69% to 0.55%, which indicated that the adaptive fusion strategy could avoid the interference of irrelevant features in series fusion and intelligently balance the contributions of the two modalities according to the quality of specific samples. In the comparison between experiments 4 and 5, after the learnable Gabor filter was introduced, the EER further dropped to 0.51%, indicating that the Gabor layer with learnable parameters and the attention fusion mechanism had a synergistic gain effect (Table 1). The Gabor layer provided more relevant input features for the attention mechanism, which could maximize the benefit of Gabor enhanced features.

The results of this study aligned with previous study that the BiCMAM could achieve more refined feature interaction and weighting rather than simple concatenation or decision superposition compared with the feature-level/score-level fusion [22]. Although Kalpana *et al*. verified the effectiveness of Gabor transform in feature extraction, most of its filter parameters were fixed [23]. This study made the Gabor parameters learnable, thereby adaptively optimizing the filter and extracting more discriminative biological features. To solve the single biometric identification being susceptible to interference in complex scenes and insufficient utilization of heterogeneous features in multi-modal fusion, a dual-stream iris-periocular deep fusion model based on adaptive Gabor feature weighted fusion convolutional network was constructed in this research to embed the Gabor filter with learnable parameters into the front-end of the dual-stream network to enhance texture feature extraction and introduce a BiCMAM to achieve adaptive weighted fusion of iris and periocular features at the feature layer. This study achieved effective results on static images but did not consider the input of dynamic images or videos. Future work should be expanded to multi-frame information fusion in dynamic scenarios to improve the system's performance and anti-spoofing capabilities in real-time authentication scenarios.

Acknowledgements

The research was supported by the First Dongguan Affiliated Hospital, Kicking off High-level Research (Grant No. GCC2023001).

References

1. Abdul-Al M, Kyeremeh GK, Qahwaji R, Ali NT, Abd-Alhameed RA. 2024. The evolution of biometric authentication: A deep dive into multi-modal facial recognition: A review case study. IEEE Access. 12(5):179010-179038.

2. Rukhiran M, Wong-In S, Netinant P. 2023. IoT-based biometric recognition systems in education for identity verification services: Quality assessment approach. IEEE Access. 11(3):22767-22787.
3. Gao R, Bourlai T. 2024. On designing a swiniris transformer based iris recognition system. IEEE Access. 12(1):30723-30737.
4. He S, Li X. 2024. Enhancedeepiris model for iris recognition applications. IEEE Access. 12(4):66809-66821.
5. Li M, Wang Y, Zhang K, He Z, Sun Z. 2025. Exploring near-infrared iris image sequences for high throughput iris recognition. IEEE T Inf Foren Sec. 20(6):5718-5731.
6. Zambrano JE, Pilataxi JI, Perez CA, Bowyer KW. 2024. Iris recognition using an enhanced pre-trained backbone based on anti-aliased CNNs. IEEE Access. 12(2):94570-94583.
7. Yeh CH, Shen WC, Chi HY, Chi HY, Lin CE, Chen JS. 2023. Enhancing online learning monitoring with novel image recognition method using Dlib for eye feature detection. 2023 12th International Conference on Awareness Science and Technology (iCAST). IEEE. 14(3):340-345.
8. Gupta K, Choubey S, William P, Kale CP. 2023. Implementation of motorist weariness detection system using a conventional object recognition technique. International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT). IEEE. 13(4):640-646.
9. Pi J, Zeng J, Lu Q, Jiang N, Wu H, Zeng L, *et al*. 2023. Adv-Eye: A transfer-based natural eye makeup attack on face recognition. IEEE Access. 11(4):89369-89382.
10. Sumalatha U, Prakasha KK, Prabhu S, Nayak VC. 2024. A comprehensive review of unimodal and multi-modal fingerprint biometric authentication systems: Fusion, attacks, and template protection. IEEE Access. 12:64300-64334.
11. Feng A, Han C, Gong J, Yi Y, Qiu R, Cheng Y. 2024. Multi-scale learnable Gabor transform for pedestrian trajectory prediction from different perspectives. IEEE Trans Intel Transp Syst. 25(10):13253-13263.
12. Singh HK, Baranwal N, Singh KN, Singh AK. 2024. Using multi-modal biometric fusion for watermarking of multiple images. IEEE T Consum Electr. 70(1):3487-3494.
13. Ryu R, Yeom S, Herbert D, Dermoudy J. 2025. From fusion to adaptation: Investigation on enhancing multi-modal biometric authentication systems. IEEE Access. 13(4):146169-146188.
14. Muzaffar AW, Riaz F, Abuain T, Abu-Ain WAK, Hussain F, Farooq MU, *et al*. 2023. Gabor contrast patterns: A novel framework to extract features from texture images. IEEE Access. 11(6):60324-60334.
15. Zhao Z, Xu X, Li J, Li S, Plaza A. 2023. Gabor-modulated grouped separable convolutional network for hyperspectral image classification. IEEE Trans Geosci Remote Sens. 61(14):1-17.
16. Fan D, Liang X, Zhang C, Jia W, Zhang D. 2023. AMGNet: Aligned multilevel Gabor convolution network for palmprint recognition. IEEE TCSVT. 34(6):4175-4189.
17. Mokayed H, Quan T, Alkhaled L, Sivakumar V. 2023. Real-time human detection and counting system using deep learning computer vision techniques. AIA. 1(4):221-229.
18. Sardar A, Umer S, Rout RK, Sahoo KS, Gandomi AH. 2024. Enhanced biometric template protection schemes for securing face recognition in IoT environment. IEEE IoT J. 11(13):23196-23206.
19. Simonyan K, Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations (ICLR).
20. Krizhevsky A, Sutskever I, Hinton GE. 2017. ImageNet classification with deep convolutional neural networks. CACM. 60(6):84-90.
21. LeCun Y, Bottou L, Bengio Y, Haffner P. 2022. Gradient-based learning applied to document recognition. Proc IEEE. 86(11):2278-2324.
22. Umer S, Sardar A, Rout RK, Tanveer M, Razzak I. 2023. IoT-enabled multi-modal biometric recognition system in secure environment. IEEE IoT J. 10(24):21457-21466.
23. Kalpana P, Sitaraman SR, Harakannanavar SS, Alsalami Z, Nagaraj S. 2024. Efficient multi-modal biometric recognition for secure authentication based on faster region-based convolutional neural network. 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON). IEEE. 23(4):1-5.