RESEARCH ARTICLE

EnhanceDeepIris Model for Iris Recognition Applications

SHOUWU HE AND **XIAOYING LI**

College of Computer Application, Guilin University of Technology at Nanning, Nanning 530001, China

Corresponding author: Shouwu He (heshou5@126.com)

ABSTRACT In this study, an iris recognition technique based on enhanced EnhanceDeepIris model is proposed in order to examine iris recognition at a deeper level. The process first uses convolutional features to extract human iris features. Then, in order to solve the deformation problem caused by the radial data of iris texture, the sequential metric model is introduced to realize the effective recognition of iris. Finally, the structure of the model is thoroughly examined to investigate its actual performance. The study found that after running the four algorithms on the ND-IRIS-0405 and CASIA-Lamp datasets, the loss function value of the research method began to approach its lowest value at the 14th and 9th iterations, while the other algorithms continued to slowly decrease. Additionally, at the 5th and 4th iterations respectively, the accuracy of the research method was nearly 91.00%.After 20 classification predictions, the average recognition accuracies of the research method proposed in this paper, iris segmentation method based on end-to-end multi-task segmentation network IrisST-Net, iris recognition method based on full complex-valued neural network, and iris recognition method with hybrid preprocessing and feature extraction were 99.88%, 98.72%, 97.47%, and 89.77%, respectively. These findings suggest that the study approach recognizes the iris the best and can identify its primary contour information with accuracy. The results of the application demonstrate that the research method provides the clearest detection of the iris edge, with less noise, and can accurately detect the main contour information of the iris. This provides a reference for optimizing related technologies in the field of image recognition.

INDEX TERMS EnhanceDeepIris model, deep learning, iris recognition, feature extraction.

I. INTRODUCTION

The discipline of information security has recognized biometric identification technology as a critical area of research due to the ongoing advancements in science and technology. Iris recognition (IR) is one of these technologies that has drawn the most interest and study because of its special benefits, which include high security and difficulty to forge. However, traditional IR methods have certain limitations in dealing with the complexity and diversity of Iris Images (I-I) [1], [2]. Therefore, designing an efficient and accurate information retrieval (IR) method has become a current research focus. In the last few years, the application of deep learning (DL) technology in the field of image recognition has achieved

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague.

remarkable results. Specifically, the emergence of convolutional neural networks (CNN) has provided new solutions for image recognition [3]. Although CNNs perform well in many tasks, they still face challenges when dealing with complex and diverse tasks such as IR [4], [5].To address these issues, the researchers proposed an IR method that utilizes the enhanced EnhanceDeepIris model. The approach involves using a deep convolutional network to extract I-I features, which can capture both local and global information, resulting in improved recognition accuracy. The EnhanceDeepIris model has been improved by introducing new optimization strategies and loss functions, resulting in further performance improvements. This is anticipated to yield fresh concepts and approaches for the advancement of infrared technology.

The experiment focuses on two innovative points. First, deep convolution is combined with the improved

EnhanceDeepIris model to improve the model recognition accuracy while improving the stability and robustness of the model. Second, a new optimization strategy is introduced in the method design process to improve the model's adaptability to different iris image characteristics, and ultimately improve the model's performance. The entire process was carried out in strict accordance with the experimental design, and research was carried out from two aspects: performance testing and application effect comparison to support the conclusions.

There are four sections to the article. An summary of recent domestic and international research is given in the first section. The second part introduces the research method, which utilizes deep convolution for feature extraction (FE) and ordinal metric modeling to analyze the iris and improve the EnhanceDeepIris model. The third section explores the practical performance and application effects of the proposed method. Finally, the fourth part summarizes the study's results.

II. RELATED WORKS

In the past decades, biometrics have made remarkable progress in the field of information security and personal authentication. IR, as a highly reliable and non-invasive biometric identification method, has attracted extensive research interest. M Choudhary et al. proposed three different models based on convolutional block and residual block in order to enrich the heterogeneous sensor IR technique. The performance of the constructed models was analyzed using score level fusion with two different superior FE methods. The results show that the model is able to accurately recognize minute patterns in the iris region [6].Y Safaa El-Din and other scholars summarized the methods for detecting face and iris and compared two different fine-tuning methods on six different datasets. The results showed that the existing DL models were able to discriminate different face and iris features with low error rates [7]. E Garea-Llano and other researchers proposed a YOLO network with image segmentation IR technique in order to effectively minimize the direct contact between human body and objects. The experiments focused on analyzing the images, iris detection and segmentation in videos. The practical feasibility of the proposed method was verified on different datasets [8]. For effective segmentation of I-I, R R Jha's team proposed a deep CNN based algorithm for iris ROI segmentation. The model was continuously optimized by applying techniques such as multiscale & multidirectional training. And the network stability was tested by rotating the image under aspect ratio with different contrast variations. The model was able to achieve good application outcomes and high recognition accuracy on several datasets, according to the final results [9]. To decrease the probability of error in locating the iris region in GT images, Y Chen and other scholars proposed a dense connected network model (DADCNet) based on dual attention modules with improved jump connections, and verified the actual performance of the constructed model by using the mask images after DADCNet segmentation. The outcomes demonstrated how well the built model performs and how precisely it can categorize various iris data [10].

The related techniques of DL have also been applied to other fields, and numerous scholars have conducted more in-depth research on them. A Ben Slama and other researchers and scholars proposed a recurrent disease detection system based on improved deep CNN in order to detect vestibular diseases in different patients. The performance of the model was validated on three different data categories during the process, and the results showed that the proposed method can effectively enhance the detection and analysis of dizziness factors in patients [11]. Two scholars, N Kumari and R Bhatia, in order to recognize human emotions from human facial images and videos, propose a DL model based facial emotion recognition method. The image recognition is enhanced by introducing an improved joint trilateral filter to remove the effect caused by impulse noise and finally an improved model is obtained. The proposed facial emotion recognition model has very good recognition results compared to the traditional model [12]. R Thangaraj et al. suggested a deep transfer learning technique based on artificial intelligence for the efficient diagnosis of tomato illnesses. The model detects the diseases suffered by tomato through real-time images and stored data. The results showed that tomato leaf diseases can be effectively detected and recognized with superior accuracy using the constructed model [13]. Two scientists, S Uğuz and N Uysal, suggested a detection and classification model based on a deep transfer learning architecture to successfully identify pests and illnesses impacting the growth of olive leaves.. The highest successful detection rate of the model was 95% in the experiments with data augmentation. In the experiment without data enhancement, the highest successful detection rate of the model was 88%. The data comparison revealed that the algorithm has a very superior overall performance [14]. For automated diagnosis and recognition of diabetic retinopathy, J D Bodapati et.al proposed a composite deep neural network architecture based on gated attention mechanisms. The experiment introduces spatial pooling cubes to obtain a simplified version of these data without losing more information. In addition the gated attention module in this model was able to reduce the attention to non-lesioned regions. This model earned a Kappa score of 79 and a prediction accuracy of 82.54% in comparison to the unimodal model [15].To solve the time-consuming annotation task in semantic segmentation, A. Amirkhani et al. proposed a semantic segmentation method that combines knowledge distillation technology and convolutional neural networks. Students' robustness and accuracy were assessed against the complexity of the training scenarios. It was found that the performance of this model was significantly better than existing results on clean data and corrupted data in semantic segmentation. The performance of the constructed multi-teacher framework was significantly improved by 32.18% [16]. In order to improve the clarity of lung disease images of COVID-19 patients, Mozaffari et al.

proposed using a deep learning model to improve image clarity. Throughout the process, technologies such as GAN and transfer learning are introduced, and these technologies are used to detect statistical data. Finally, through verification, it was found that the deep learning model has very superior performance [17].

In summary, the DL algorithm is an advanced artificial intelligence technique that has been widely used in many fields, such as information retrieval and plant pest detection. The algorithm not only improves the recognition accuracy of the research object but also demonstrates strong adaptability in dealing with complex and changing environmental conditions. However, there are few studies on fusing deep convolutional features with ordinal metric modeling to improve the EnhanceDeepIris model for accurate recognition of the human iris. Therefore, an IR technique based on the enhanced EnhanceDeepIris model is experimentally proposed to further optimize the efficiency of the algorithm, improve the recognition of low-quality images, and enhance the recognition effect.

III. DESIGN OF A MOBILE END-TO-END IRIS RECOGNITION METHOD BASED ON AN IMPROVED ENHANCEDEEPIRIS MODEL

IR refers to individual identification based on the rich texture information of the iris of the human eye, in which human iris FE and coding modeling is the key to the recognition algorithm. Given that CNN has powerful feature learning capability and has been successfully applied in several computer vision tasks, the study takes two methods, deep convolutional features and ordinal metric modeling, as the basis to jointly improve the EnhanceDeepIris model, and completes the deployment on the mobile side to achieve excellent application results.

A. DESIGN OF IRIS RECOGNITION METHOD FOR HUMAN BODY COMBINING DEEP CONVOLUTIONAL FEATURES

Since AlexNet has been more successful on large-scale image dataset classification tasks, CNNs have begun to improve, resulting in the network needing to learn more and more parameters for the task. However, currently there is a great lack of relevant labeled datasets among IR tasks, and in order to adapt to small-scale iris datasets to reduce the likelihood of fitting, the experimentally designed network must be in the number of parameter settings, and therefore the perceptual domain of the last convolutional layer (CL) in the CNN architecture designed by the study on the input image should be reduced; and the spatial dimensions of

convolutional feature maps should be increased in order to extract more features. The research designed a lightweight network called Lw-IrisNet based on AlexNet, as shown in Figure 1.

In Figure 1, Lw-IrisNet and Lw-ReslrisNet are two CNN backbone networks. The entire process uses softmax loss and triplet loss to train the CNN network from scratch, and then performs ordinal modeling of the convolutional features and the design of the matching method. Unlike previous network architectures, the experimentally designed Lw-IrisNet lightweight network has spatially larger feature maps. The first CL had its step size set to 2, and the step sizes of the other CLs were set to 1. Furthermore, following the second and final CLs, a pooling layer with a size of 2×2 and a step size of 2 was accessed in order to further downsize the feature maps. Then for the result, the spatial size of the feature maps generated by the sixth CL (conv6) is 64×64 for an input image of 256×256 pixels, which is significantly larger compared to AlexNet's 14×14 size. This design allows the network to capture local features in greater detail [18], [19]. Meanwhile, in order to adapt to the small databases commonly used in IR tasks, Lw-IrisNet significantly reduces theoutput channels in each layer, and the final network parameters is only 9 megabytes, which is about one-seventh of the parameters in AlexNet. This lightweight network not only reduces the computational burden, but also improves the adaptability and recognition efficiency on small databases.

There have been two important tasks in the field of human biometrics recognition, namely authentication and identification. Authentication is the judgment of the input sample to correct whether it belongs to the same person or not and belongs to one-to-one matching task. Recognition, on the other hand, compares the sample with all other samples in the database and is a one-to-many matching task [20]. In order to complete the two different tasks in full, the experiments are proposed to choose the ternary loss and Softmax loss functions to train different CL networks. The computation of the ternary loss function is shown in equation (1).

In equation (1), as shown at the bottom of the page, x_j^i represents the jth image of the ith category. $f_\theta\left(x_j^i\right)$ represents the corresponding 128-dimensional embedding representation. The above equation can generate the corresponding CS triples, each of which contains an anchor image as well as a positive sample corresponding to the largest distance and a negative sample corresponding to the smallest distance on that batch of images. In order to effectively reduce the selection time, the C of the experiment is uniformly set to 16, while the S is set to 4. Then the Softmax loss function is

$$L_{BH}(\theta; X) = \sum_{i=1}^{C} \sum_{a=1}^{S} \left[a + \overbrace{\max_{p=1\ldots S}\left(\left\|f_\theta\left(x_a^i\right) - f_\theta\left(x_p^i\right)\right\|_2\right)}^{hardest\ positive} - \overbrace{\min_{j=1\ldots C; j\neq i; n=1\ldots S}\left(\left\|f_\theta\left(x_a^i\right) - f_\theta\left(x_p^i\right)\right\|_2\right)}^{hardest\ negative} \right] \tag{1}$$
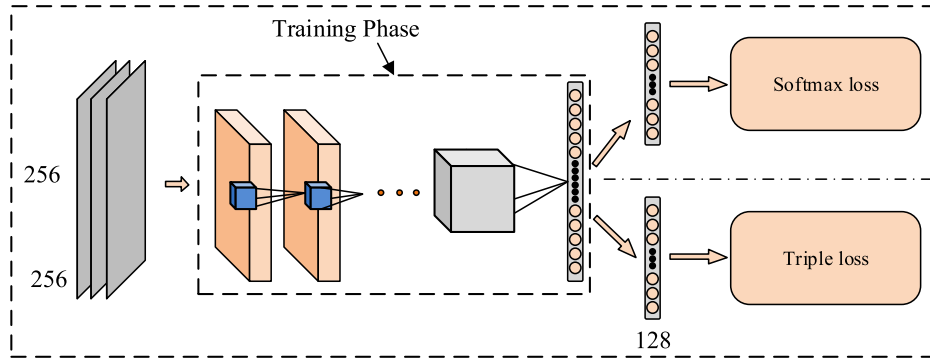
FIGURE 1. Lw IrisNet lightweight network based on AlexNet.

utilized to complete the recognition task, and the calculation of the corresponding loss function is shown in Equation (2).

$$loss = -\sum_{k=1}^{m} \hat{y}_k \log(y_k) = -\sum_{k=1}^{m} 1\left(\hat{y}_k = 1\right) \log(y_k) \tag{2}$$

Then an ordinal metric modeling approach is used to encode qualitative relationships between feature space regions to make more robust to changes such as noise and illumination. It is assumed that the convolutional feature at position z in the network structure is $f(z)$, and $\Omega(z)$ represents the feature region in the upper left corner of z. The calculation of the binary order encoding corresponding to position z is shown in equation (3).

$$b_i(z) = H\left(\sum_{t\in\Omega(z)} f_i(t) - \sum_{t\in\Omega(z+\Delta z)} f_i(t')\right), i = 1, \cdots, d \tag{3}$$

In equation (3), $b_i(z)$ represents the ith element in the sequence encoding of z. $H(\cdot)$ represents the step function; Δz represents the displacement between two feature regions. After image preprocessing, a binary mask image can be obtained to indicate whether each pixel in the unfolded iris image is the iris texture. During the process, a single channel image with the same size as the image feature space is generated in the last CL, which is called effective iris regions (EEIRs) [21]. Calculate the processed EEIRs to match the convolutional features of the iris unfolded image, as shown in equation (4).

$$w(z) = \frac{1}{|\Omega(z)|} \min\left(\sum_{t\in\Omega(z)} r(t), \sum_{t\in\Omega(z+\Delta z)} r(t')\right) \tag{4}$$

In equation (4), $|\Omega(z)|$ represents the elements in the $\Omega(z)$ region. $r(t)$ represents the EEIRs value at position t. The final mask image has the same spatial size as the sequence features of the iris. In practical application scenarios, due to factors such as rotation, it is difficult to align the image pairs that need to be matched, so it is necessary to find the position of the feature point BB. The calculation of this point is shown

in equation (5).

$$z^* = \arg\min_{z'\in\Re(z)}\left(\frac{1}{d}\sum_{i=1}^{d} b_i^{I_1}(z) \oplus b_i^{I_2}(z')\right) \tag{5}$$

In equation (5), $\oplus$ represents XOR operation. Considering the application of mask images, the similarity between image pairs is defined as $\rho(I_1, I_2)$. Eq. (6) presents the particular computation.

$$\rho(I_1, I_2) = \frac{\sum_Z \min\left(w^{I_1}(z), w^{I_2}(z^*)\right) \sum_{i=1} b^{I_1}(z) \oplus b_i^{I_2}(z^*)}{d\sum_z \min(w^{I_1}(z), w^{I_2}(z^*))} \tag{6}$$

B. DESIGN OF END-TO-END IRIS RECOGNITION METHOD BASED ON SEQUENTIAL METRIC MODELING STRATEGY

Although using pre trained models to extract convolutional features in Section A can obtain an encoded representation of I-I, the network training or feature learning and modeling involved in FE in this algorithm are two independent stages. Feature learning can effectively distinguish images of different categories to the maximum extent possible. During the testing phase, the focus is on modeling the features of the CL network, indicating a significant difference in focus between feature learning and feature modeling. The role of the two working separately is far inferior to the role of their fusion working together. Therefore, research proposes to combine feature learning and feature modeling in the network training stage to better distinguish high discriminative encoding as the optimization goal, in order to achieve better performance [20], [22]. Based on this, the experiment embeds order metric modeling after the last CL of the CNN network to achieve the fusion of the two. To achieve the correct embedding of order metric modeling, the experiment split the order metric modeling into three structural layers, and proposed a softened Hamming distance to calculate the similarity between the approximate binary encoding representations output by the metric softened encoding layer. Figure 2 displays the unique network design.

In Figure 2, this framework is improved based on the research results of Mozaffari and other scholars [23]. The main innovation point is to jointly optimize feature learning and feature modeling in ConvOM when training neural
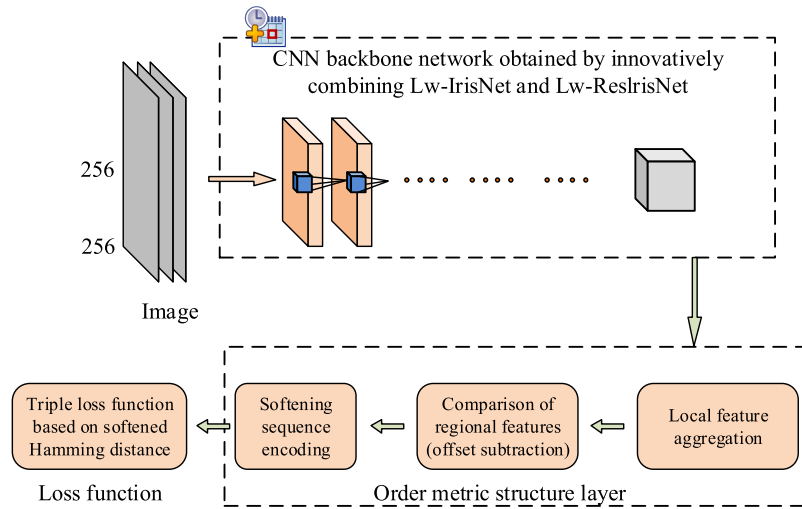
FIGURE 2. Overall architecture of end-to-end iris recognition method based on ordered metric embedding network.

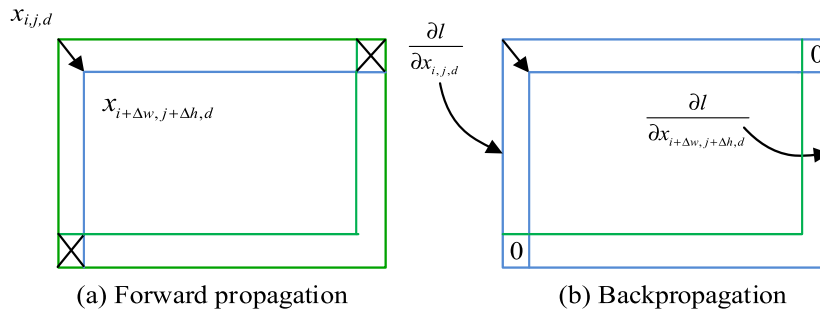(a) Forward propagation   (b) Backpropagation

FIGURE 3. Forward and back propagation calculation flow of regional feature comparison layer.

networks. The two parts complement each other and fully exploit the powerful learning ability of convolutional neural networks to obtain a more expressive model. The same scheme is used in the test phase. Since the test phase requires the use of mask images and the method in this chapter does not involve mask images during training, the method in this section is not a strict end-to-end iris recognition algorithm. The end-to-end mentioned can be understood as the encoding representation of the iris image that is almost the same as that in the testing stage can be obtained through the network itself during training. In this order metric structure layer, since the step function is not differentiable at point 0, but the smooth back propagation must be ensured, the experiment applies a class of sigmoid functions for approximate order encoding. And divide the order metric modeling part of convolutional features into three layers, namely the local feature aggregation layer, the regional feature comparison layer, and the softened order encoding layer [24].Considering that the radial data of iris texture will undergo deformation, the proportion of pooling kernel $\Omega(z)$ in the local feature aggregation layer is set to 8×2. To compensate for deformation. The regional feature comparison layer subtracts each element of the regional features obtained from the previous layer, in order to provide

effective data for subsequent sequential encoding layers. The calculation process for the forward and backward directions of this layer is shown in Figure 3.

As shown in Figure 3, the qualitative comparison operation is actually an independent operation between different channels. Assuming the output of position (i, j) in forward propagation is $y_{i,j,d} = x_{i,j,d} - x_{i+\Delta w, j+\Delta h, d}$, ignoring the parts beyond the boundary, the corresponding back propagation calculation is obtained as shown in equation (7).

$$\begin{cases} \dfrac{\partial l}{\partial x_{i,j,d}} = \dfrac{\partial l}{\partial y_{i,j,d}} \\ \dfrac{\partial l}{\partial x_{i+\Delta w, j+\Delta h, d}} = -\dfrac{\partial l}{\partial y_{i,j,d}} \end{cases} \tag{7}$$

In equation (7), d represents the channel. In order to complete the sequential metric modeling, it is necessary to quantify the output of the region feature comparison layer through a step function. Assuming that data greater than 0 is encoded as 1, and vice versa, it is 0 to achieve qualitative comparison between regional features. Due to the non differentiability of the step function at point 0, the experiment uses an approximate sequential encoding calculation as shown
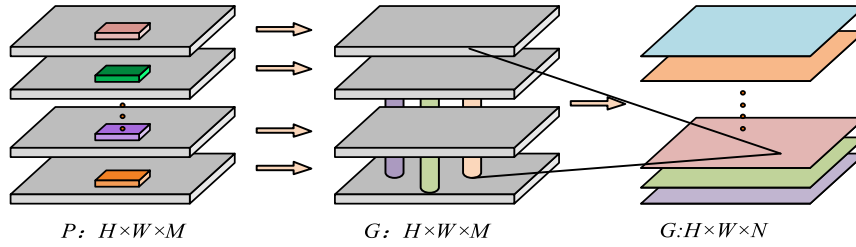
P：H×W×M   G：H×W×M   G:H×W×N

FIGURE 4. Deep separable convolutional layer structure.

in equation (8).

$$y = SoftOM(x) = \frac{1}{1 + e^{-\lambda x}} \qquad (8)$$

In equation (8), λ represents a hyper parameter, and when its value is 1, the *SoftOM* function becomes a Sigmoid function. The back propagation calculation of the corresponding layer can be further obtained through equation (8). Equation (9) illustrated the calculation.

$$\frac{\partial l}{\partial x} = \lambda y(1 - y)\frac{\partial l}{\partial y} \qquad (9)$$

The experiment employs a triplet loss function for efficient training in order to fully train the aforementioned network. In order to train the network, the loss function must be minimized; the precise computation is displayed in equation (10).

$$l = \max(D(f(x^a), f(x^p)) + a - D(f(x^a), f(x^n)), 0) \quad (10)$$

In equation (10), $f(x)$ is the feature representation of the image, and x^a represents the anchor image. x^p represents positive samples that belong to the same class as the anchor image, while x^n represents negative samples that do not belong to the same class as the anchor image. a is the minimum difference between the distance of negative and positive sample pairs. When the values of a, b are all 0-1 bits, the Hamming distance obtained is calculated using equation (11).

$$a \oplus b = a\bar{b} + \bar{a}b = a(1 - b) + (1 - a)b = -2ab + a + b \qquad (11)$$

Next, set the output of the order metric structure layer of the triplet image to (X^a, X^p, X^n), and the size of X^a to $h \times w \times C$. By integrating, the corresponding loss function can be calculated as $l = \max(D(X^a, X^p) + a - D(X^a, X^n), 0)$. When the loss function is not 0, the back propagation calculation obtained is shown in equation (12).

$$\begin{cases} \dfrac{\partial l}{\partial X^a} = \dfrac{2X^n - 2X^p}{h, w, C} \\ \dfrac{\partial l}{\partial X^p} = \dfrac{2X^p - 2X^a}{h, w, C} \\ \dfrac{\partial l}{\partial X^n} = \dfrac{2X^a - 2X^n}{h, w, C} \end{cases} \qquad (12)$$

C. DESIGN OF IRIS RECOGNITION METHOD FOR INTEGRATING AND IMPROVING ENHANCEDEEPIRIS MODEL

To effectively reduce the parameters of the network, greatly reduce computational complexity, and ultimately achieve compression and acceleration, the experiment proposes the introduction of depthwise separable convolution to divide the standard convolution layer into two sub layers: channel wise convolution and point wise convolution [25], [26]. The output feature map can be obtained through calculation, as shown in equation (13).

$$G_{k,l,n} = \sum_{i,j,m} F_{i,j,m,n} \cdot P_{k+i-1,l-j-1,m} \qquad (13)$$

In equation (13), the computational cost of the standard CL is set to $K \times K \times M \times H \times W \times N$, which depends on the input channels M and can output N channels. The convolution kernel's size and the output feature map's spatial size are both $H \times W$. Figure 4 depicts the structure of depthwise separable CLs.

Using the channel by channel convolution layer in the architecture of Figure 4, independent operations are performed on each channel of the input feature map P to obtain the corresponding channel of the output feature $\hat{G}$, as shown in equation (14).

$$\hat{G}_{k,l,m} = \sum_{i,j} \hat{F}_{i,j,m} \cdot P_{k+i-1,l-j-1,m} \qquad (14)$$

In equation (14), $\hat{F}$ represents the number of M filter banks with a size of $K \times K$ in each channel CL. The computational cost of channel by channel CLs is expressed as $K \times K \times H \times W \times M$. However, channel by channel CLs can only filter different channels in the input feature map and cannot fuse information to generate new features. Therefore, depthwise separable convolution needs to be compensated through channel by channel CLs [27]. In summary, replacing the standard CL with a depthwise separable CL reduces the proportion of parameters and computational costs as shown in equation (15).

$$\frac{K \times K \times H \times W \times M + 1 \times 1 \times M \times H \times W \times N}{K \times K \times M \times H \times W \times N}$$
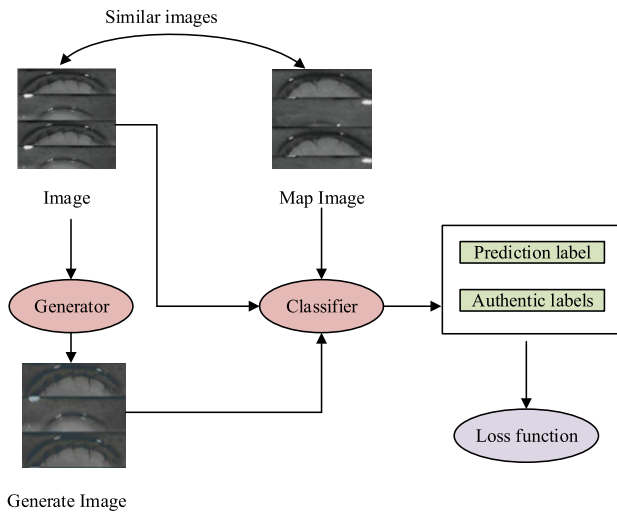$$= \frac{1}{N} + \frac{1}{K^2} \qquad (15)$$

FIGURE 5. The EnhanceDeepIris model framework constructed by the research institute.

For the 3×3 CL, using equation (15) results in an 8–9 reduction in both the number of parameters obtained and the computational cost. Using the deep convolutional neural network framework described above, it is possible to better train the model's classification network, finish the image classification and model convergence, and ultimately create a dual network model called EnhanceDeepIris. Finally, a dual network model, EnhanceDeepIris, is designed. The generative network in the model is responsible for generating expanded data based on a small sample iris dataset, enabling the classification network in the model to train better and obtain a more accurate and stable classification model. A generator and a classifier are the two primary components of the EnhanceDeepIris paradigm. Based on the training set, the generator creates enhanced images, and the classifier—the first trained classification network—is further trained using both enhanced and real images [28]. The specific architecture is shown in Figure 5.

In Figure 5, the generator in the model is actually an image generation model, mainly used to generate augmented data, essentially an autoencoder. The generator mainly consists of 5 residual blocks, among which the down sampling layer mainly includes two CLs; The down sampling process can reduce the feature mapping size input to the residual unit, increase the receptive field of the residual unit, reduce computational complexity, and improve the model's tolerance for input noise, reducing the possibility of overfitting. Two transposed CLs make up the up sampling layer. They are used to increase the feature map before the network's final output, producing a picture that is the same size as the original input image [29], [30]. The up sampling process and down sampling process are symmetrical to each other. The transpose convolution operation is shown in Figure 6.

An image classification network is used in the classifier portion of the research institute's model. Based on deep CNN,

I-I is recognized and classified to evaluate the adaptability of the generated model to different image classification networks. The overall framework for improving the IR method is shown in Figure 7.

IV. PERFORMANCE TESTING AND APPLICATION ANALYSIS OF THE ENHANCEDEEPIRIS MODEL IN IRIS RECOGNITION

Experiments were carried out to examine the real performance of the enhanced EnhanceDeepIris model in IR and detecting systems in order to fully illustrate the higher performance of the suggested method in IR. Firstly, introduce the dataset library and experimental setup, and then systematically evaluate the performance of the constructed system. To facilitate the design and training of the network, the iris area is usually expanded into a square image with a size of 256×256 in the preprocessing stage of the experiment. The number of system iterations was controlled within 160, and the Adam optimizer was used. The initial learning rate was set to 0.001. The cosine annealing scheduler was used to adjust the learning rate. To prevent excessive Fitting and weight decay, using L2 regularization. The PyTorch framework was used for all studies, and Table 1 displays the pertinent simulation environment parameters.

TABLE 1. The experimental basic environmental parameters.

Parameter variables	Parameter selection
Simulation tools	Simulink
Operating system	Windows10
Open source framework	Caffe
Toolkit	MatConvNet
System PC side memory	24G
CPU main frequency	2.62Hz
GPU	RTX-2070
Central Processing Unit	Intel i7-3770k CPU
Graphics card	NVIDIA GTX 1060
Data storage	MySQL
Data regression analysis platform	SPSS 26.0
Initial learning rate	0.01
Optimizer	Adam optimizer
Learning rate scheduler	Cosine annealing scheduler
Number of iteration rounds	<200
Loss function target value	Approaches 0
Norm regularization	L2 regularization

The experiment first selects two publicly applicable I-I datasets for experimentation. The first one is the ND-IRIS-0405 dataset, which includes a total of 64980 grayscale images captured by 356 people using the LG2200 iris camera, each containing images of the left and right eyes, totaling 712 categories. The age range of the collected individuals is between 18 and 75 years old, including 158 females and 198 males. The second dataset is the CASIA IrisV4 Lamp dataset (hereinafter referred to as CASIA Lamp), which is one of the largest number of images in the iris dataset publicly available to the Chinese Academy of Sciences. CASIA Lamp uses a handheld iris sensor for image acquisition, consisting of a total of 411 people, 819 categories, and 16212 grayscale images. Next, the performance of iris segmentation and contour parameter extraction methods based
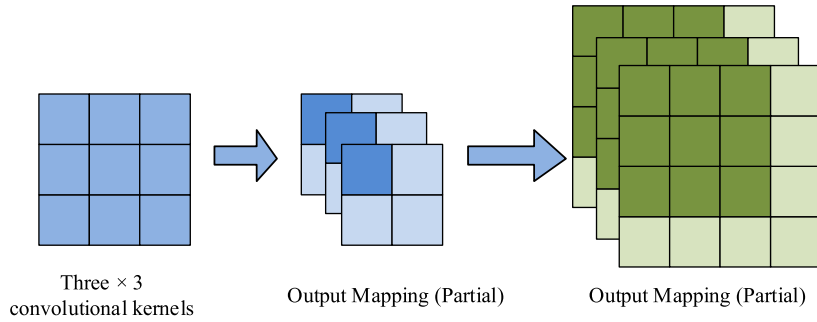
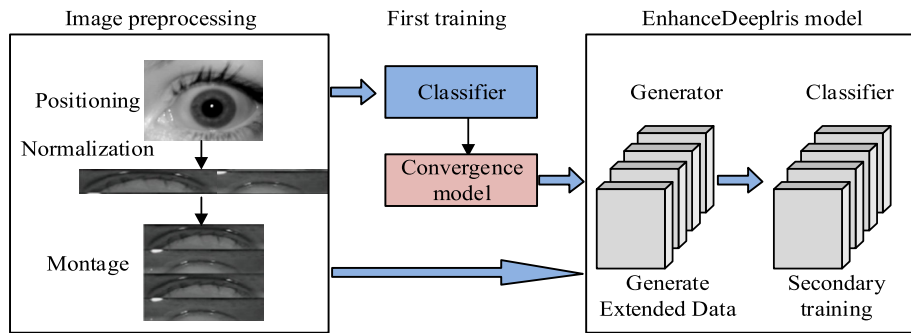**FIGURE 6.** Up sampling transpose convolution process.

FIGURE 7. Overall framework for improving iris recognition methods.
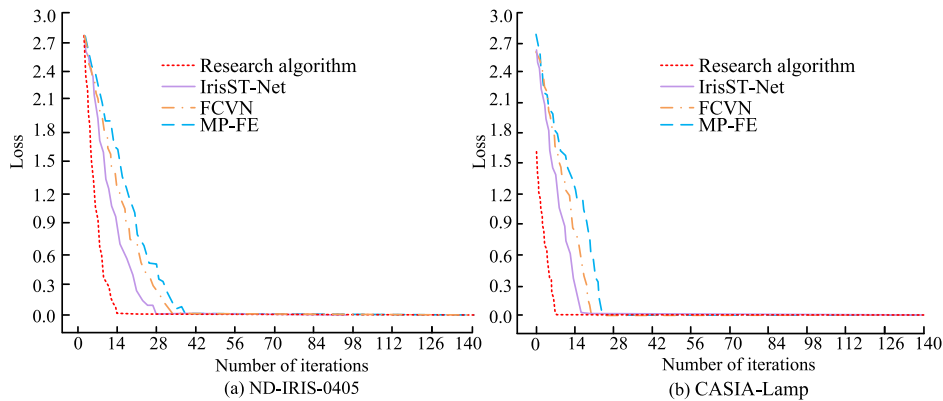

FIGURE 8. Changes in loss function values.

on full complex valued neural network (FCVN), mixed pre-processing and feature extraction (MP-FE), and end-to-end multi task segmentation network IrisST-Net were compared with those constructed by the research institute [31], [32], [33]. To ensure the smooth progress of the entire experiment and avoid experimental randomness, all algorithms are operated under unified conditions. Firstly, compare the loss function values obtained from training four algorithms on two datasets, as shown in Figure 8.

Figure 8 (a) shows the variation of loss function values for four algorithms on the ND-IRIS-0405 dataset. On the training iris dataset, it is noticed that the loss function of the network model steadily lowers with an increase in the

number of training iterations. The loss function under the study technique can rapidly drop to the lowest value when the number of system iterations reaches the fourteenth, and the downward trend is highly stable. The other three algorithms have a slower descent speed and may exhibit curve oscillation during the process. IrisST-Net only began to reach the target value after 28 iterations of the system. Figure 8 (b) shows the variation of loss function values for different algorithms on the CASIA Lamp dataset. Upon reaching nine training iterations, the research method's loss function value starts to converge towards the minimal value, whilst the values of the other three methods continue to gradually decline. When the iteration coefficients are 19, 21, and 26 times respectively,
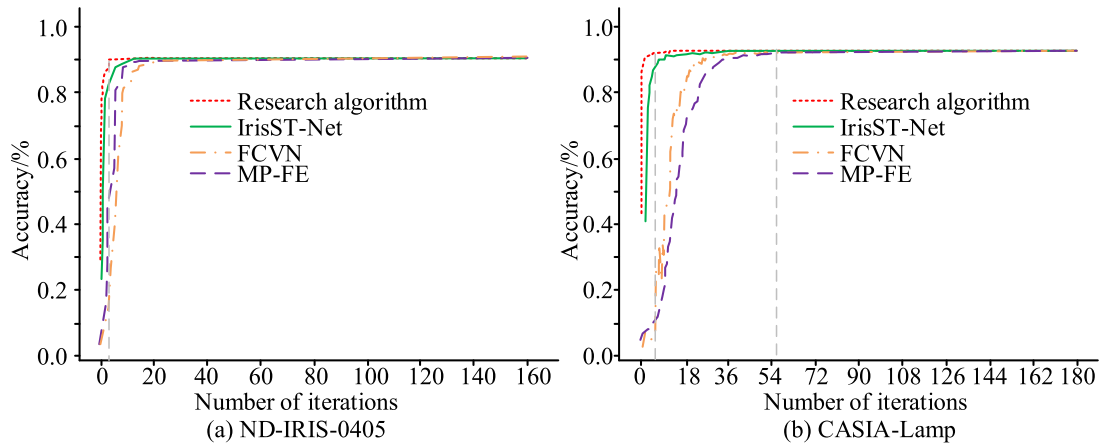
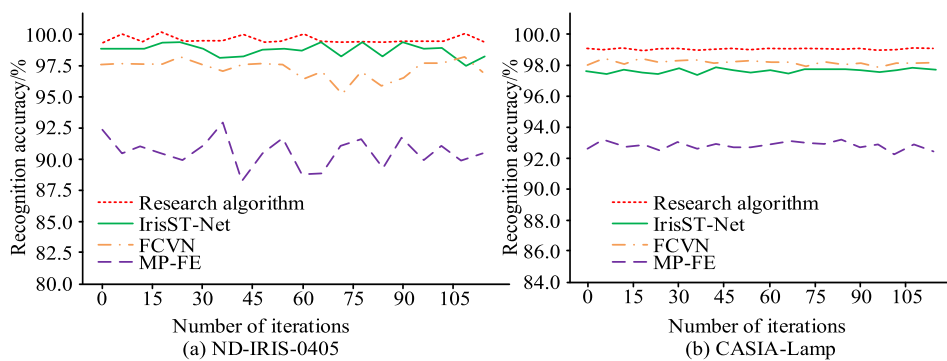**FIGURE 9.** Comparison of accuracy changes of different algorithms.

FIGURE 10. Predicted CRR curve changes of four models running 20 times on two datasets.

the loss functions of IrisST-Net method, FCVN method, and MP-FE method start to approach the minimum value. The above results indicate that the loss function value of the research method decreases more rapidly and has the smallest value, indicating good robustness of the model. Analyze the accuracy changes of IR on two datasets, as shown in Figure 9.

Figure 9 illustrates how the accuracy of all network models on two datasets progressively increases with network training as the number of iterations increases. Whether during the training process or after the network reaches the fitting state, the research method did not exhibit any oscillation on both datasets, and the accuracy remained stable. In Figure 9(a) and Figure 9(b), when the system is trained to the 4th and 6th times respectively, the accuracy of the research method is infinitely close to 0.91, while at this time the accuracy of the other three algorithms Always in a state of constant change. On the CASIA Lamp dataset, the accuracy of the MP-FE method only reaches its maximum value when the system is iterated 56 times, and there is a continuous occurrence of tortuous changes during the training process. The above results all indicate that the recognition effect of the research method is superior. This may be because the research method also takes into account the spatial information in I-I during training, and the extracted images also contain spatial features
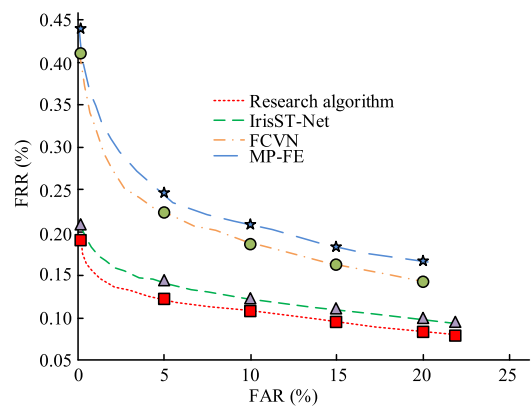

FIGURE 11. ROC curves of different models.

from the training dataset, eliminating noise features, resulting in a higher accuracy of IR. The accuracy rate approaching 91% is obtained because the data set used in the experiment contains iris images under different lighting conditions, different resolutions, and different backgrounds. At the same time, in order to prevent the model from overfitting, the experiment has adopted L2 regularization in the model training process. Regularization operations can effectively reduce the
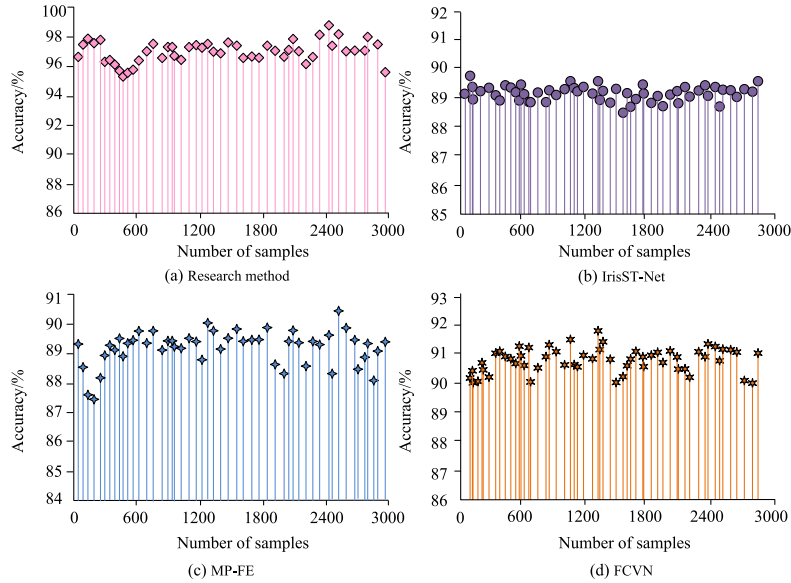
FIGURE 12. Comparison of recognition accuracy of four models in self built datasets.
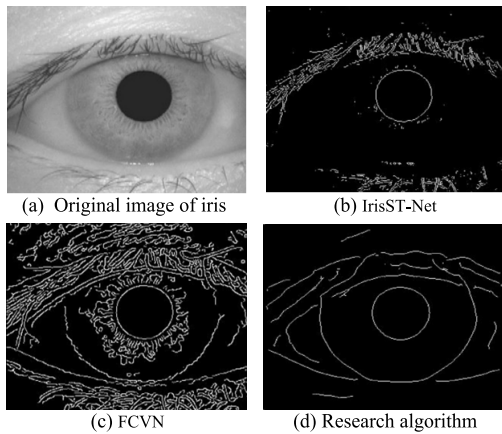

FIGURE 13. Iris edge detection effect under different algorithms.

model's dependence on training data and improve the model's generalization ability on unseen data.

Then, 20 classification predictions were performed on the two dataset images on all the obtained classification models, and the images were drawn based on the correct recognition rate (CRR) of each time to verify the classification performance and stability of the model. The results obtained are shown in Figure 10.

After 20 classification predictions, Figure 10(a) displays the four algorithms' average recognition accuracy on the ND-IRIS-0405 dataset. It can be observed that the average recognition accuracy of the system has remained at the highest value throughout the entire recognition process under the operation of the research method, with a value of 99.88%. The average recognition accuracy of IrisST-Net method, FCVN method, and MP-FE method are 98.72%, 97.47%, and 89.77%, respectively. Figure 10 (b) shows the

average recognition accuracy change on the CASIA Lamp dataset after 20 classification predictions. The average recognition accuracy of research methods, IrisST-Net method, FCVN method, and MP-FE method has been fluctuating steadily, with numerical ratios of 98.88%, 98.12%, 97.89%, and 93.55%, respectively. By comparison, the constructed method achieved the highest average recognition accuracy in both datasets and exhibited excellent predictive stability in various classification networks. Then, the ND-IRIS-0405 dataset images were predicted on all the obtained classification models, and the error acceptance rate (FAR) and error rejection rate (FRR) of the models were calculated. The ROC curves of all classification models on the dataset were plotted, as shown in Figure 11.

Figure 11 shows the ROC curves of four algorithms. The algorithm's recognition performance improves with increasing curve proximity to the coordinate axis. It can be observed that the ROC curve of the MP-FE method is located at the top of the coordinates and has the lowest recognition performance. The ROC curve is closer to the coordinate axis and has the best recognition performance for iris. Select 2000 high-quality images (200 categories, 10 images per category) from two iris databases. Compare the correct recognition time and number of recognition for four different algorithms, as illustrated in Table 2.

Table 2 shows that the research method performed preprocessing, convolutional FE, and mask image processing on the selected images, which took 337ms, 408ms, and 15ms respectively, and was able to correctly detect and recognize all the selected images. Compared to other algorithms, the research method has the highest recognition efficiency and accuracy for images under operation. To validate the recognition performance of the constructed model again, the four algorithms were finally applied to a self built iris dataset, which collected
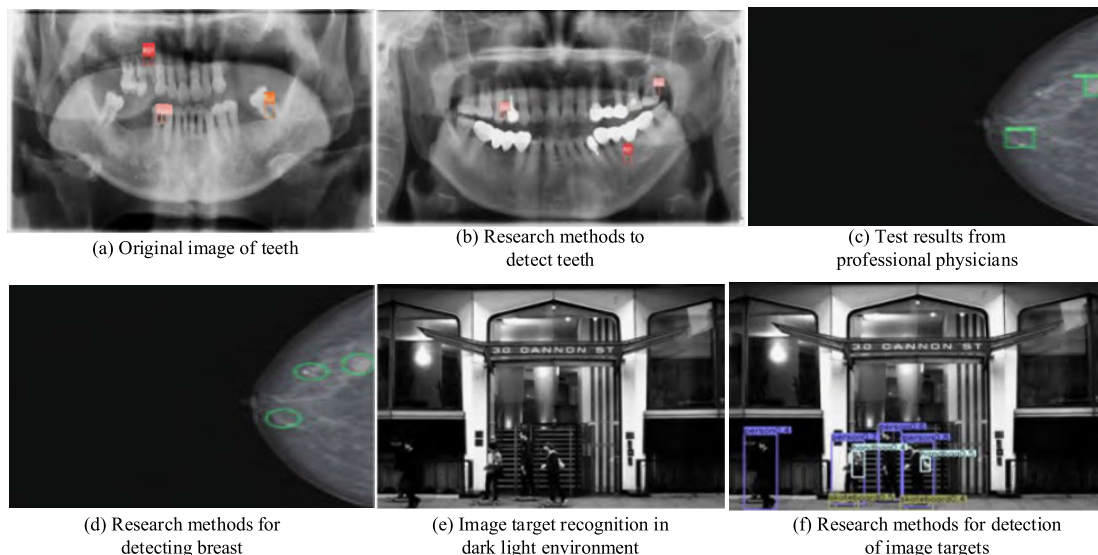
(a) Original image of teeth

(b) Research methods to detect teeth

(c) Test results from professional physicians

(d) Research methods for detecting breast

(e) Image target recognition in dark light environment

(f) Research methods for detection of image targets

FIGURE 14. Recognition and detection results of different images.

TABLE 2. Time consumption of each module for different recognition algorithms (ms).

Network architecture	Preprocessing	Convolutional feature extraction	Mask Image Processing	Correctly identify the number of items
FCVN	350	451	19	1900
MP-FE	348	421	18	1912
IrisST-Net	351	416	16	1924
Research method	337	408	15	2000

a large amount of iris data from different populations in four countries: China, the United States, the United Kingdom, and Japan. The dataset was relatively diverse, with a total of 3000 high-quality images. Figure 12 displays the recognition accuracy of the four algorithms in the self-built dataset.

Figures 12 (a) to (d) show the recognition performance of the research method, IrisST-Net method, MP-FE method, and FCVN method, respectively. The number of I-I samples is represented by the horizontal axis, and the recognition accuracy is represented by the vertical axis. IrisST-Net's recognition accuracy varies from 88% to 90%, with the majority falling between 89% and 90%, as Figure 12 illustrates. The recognition accuracy of MP-FE is the highest at 90.48%, mostly within the range of 87% to 91%. The FCVN models are all above 90%, but most are concentrated around 90.5%. The recognition performance of the research methods is above 95%, with the vast majority remaining stable at 96.0%. They have higher recognition accuracy, stronger stability, and better practical application effects. Then select an original image on the ND-IRIS-0405 data set to compare the iris edge detection effect. The results are shown in Figure 13.

From Figure 13, it can be observed that the IrisST-Net model and FCVN model show significant loss of information

on the outer boundary of the iris, as well as high levels of noise, resulting in lower applicability for IR and detection. The research method has the clearest detection results for iris edges and contains less noise. It can detect the main contour information of the iris, which is conducive to further analysis and processing of I-I for different populations and has more advantages.Finally, the research method was used to detect different images, and the results are shown in Figure 14.

Figure 14 (a) and (b) are the results of using research methods to identify patients' dental diseases. It can be seen that the research method can accurately identify inflammation in the patient's tooth root tips and detect dental diseases more accurately. Figure 14 (c) and (d) are the results of testing the patient's breast using research methods. It can be seen from the observation that the breast detection results of the research method are consistent with the professional detection and identification results given by doctors, and additional tumor masses were also detected. Figure 14 (e) and (f) show the results of identification and detection of people in dark light environments using research methods. The comparison shows that the research method can accurately identify pedestrians in dark light environments. Based on the above results, it can be seen that the research method can be applied to many different fields and the detection results are more accurate. Has superior generalization ability.

V. CONCLUSION

The IR has emerged as one of the essential technologies for offering high security and accuracy as a result of the biometric field's quick development. To address the limited sample size of iris datasets in experiments, a data augmentation dual network model based on the improved EnhanceDeepIris is proposed. The study conducted experiments on the ND-IRIS-0405 dataset and the CASIA-IrisV4-Lamp dataset. Indicators

like recognition time and accuracy were used to assess the model's performance. The data indicated that the IrisST-Net method had the highest recognition accuracy on the ND-IRIS-0405 dataset at 99.88%, followed by the FCVN method at 98.72%, the MP-FE method at 97.47%, and the lowest accuracy was achieved by the method at 89.77%. Similarly, on the CASIA Lamp dataset, the IrisST-Net method had the highest recognition accuracy at 98.88%, followed by the FCVN method at 98.12%, the MP-FE method at 97.89%, and the lowest accuracy was achieved by the method at 93.55%. The ROC curve of the IrisST-Net method was closer to the coordinate axis, indicating superior IR performance. The selected I-I underwent three operations: preprocessing, convolutional FE, and mask image processing, with corresponding time consumption of 337ms, 408ms, and 15ms. Research methods were used to detect the edges of the human iris, resulting in the clearest detection results with less noise in the image. This method can detect the main contour information of the iris. Based on the above, it can be concluded that the research method has a higher recognition accuracy for iris detection and obtains more stable test results. This provides effective reference and practice for the application of DL algorithms in IR. However, the model still requires further optimization and adjustment when faced with practical deployment challenges such as data diversity and environmental changes. Future research could investigate the adaptability of models to different races and age groups, as well as their performance in a wider range of application scenarios.

REFERENCES

[1] K. Bhosle and V. Musande, "Evaluation of deep learning CNN model for recognition of devanagari digit," *Artif. Intell. Appl.*, vol. 1, no. 2, pp. 114–118, Feb. 2023, doi: 10.47852/bonviewaia3202441.

[2] P. Preethi and H. R. Mamatha, "Region-based convolutional neural network for segmenting text in epigraphical images," *Artif. Intell. Appl.*, vol. 1, no. 2, pp. 119–127, Sep. 2022, doi: 10.47852/bonviewaia2202293.

[3] S. Zhao and B. Zhang, "Joint constrained least-square regression with deep convolutional feature for palmprint recognition," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 1, pp. 511–522, Jan. 2022, doi: 10.1109/TSMC.2020.3003021.

[4] Y. Kumar and S. Gupta, "Deep transfer learning approaches to predict glaucoma, cataract, choroidal neovascularization, diabetic macular edema, DRUSEN and healthy eyes: An experimental review," *Arch. Comput. Methods Eng.*, vol. 30, no. 1, pp. 521–541, Jan. 2023, doi: 10.1007/s11831-022-09807-7.

[5] S. Makowski, P. Prasse, D. R. Reich, D. Krakowczyk, L. A. Jäger, and T. Scheffer, "DeepEyedentificationLive: Oculomotoric biometric identification and presentation-attack detection using deep neural networks," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 3, no. 4, pp. 506–518, Oct. 2021, doi: 10.1109/TBIOM.2021.3116875.

[6] M. Choudhary, V. Tiwari, and U. Venkanna, "Enhancing human iris recognition performance in unconstrained environment using ensemble of convolutional and residual deep neural network models," *Soft Comput.*, vol. 24, no. 15, pp. 11477–11491, Aug. 2020, doi: 10.1007/s00500-019-04610-2.

[7] Y. S. El-Din, M. N. Moustafa, and H. Mahdi, "Deep convolutional neural networks for face and iris presentation attack detection: Survey and case study," *IET Biometrics*, vol. 9, no. 5, pp. 179–193, Jul. 2020, doi: 10.1049/iet-bmt.2020.0004.

[8] E. Garea-Llano and A. Morales-Gonzalez, "Framework for biometric iris recognition in video, by deep learning and quality assessment of the iris-pupil region," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 6, pp. 6517–6529, Jun. 2023, doi: 10.1007/s12652-021-03525-x.

[9] R. R. Jha, G. Jaswal, D. Gupta, S. Saini, and A. Nigam, "PixISegNet: Pixel-level iris segmentation network using convolutional encoder–decoder with stacked hourglass bottleneck," *IET Biometrics*, vol. 9, no. 1, pp. 11–24, Jan. 2020, doi: 10.1049/iet-bmt.2019.0025.

[10] Y. Chen, H. Gan, Z. Zeng, and H. Chen, "DADCNet: Dual attention densely connected network for more accurate real iris region segmentation," *Int. J. Intell. Syst.*, vol. 37, no. 1, pp. 829–858, Jan. 2022, doi: 10.1002/int.22649.

[11] A. Ben Slama, A. Mouelhi, H. Sahli, A. Zeraii, J. Marrakchi, and H. Trabelsi, "A deep convolutional neural network for automated vestibular disorder classification using VNG analysis," *Comput. Methods Biomechanics Biomed. Eng., Imag. Visualizat.*, vol. 8, no. 3, pp. 334–342, May 2020, doi: 10.1080/21681163.2019.1699165.

[12] N. Kumari and R. Bhatia, "Efficient facial emotion recognition model using deep convolutional neural network and modified joint trilateral filter," *Soft Comput.*, vol. 26, no. 16, pp. 7817–7830, Feb. 2022, doi: 10.1007/s00500-022-06804-7.

[13] R. Thangaraj, S. Anandamurugan, and V. K. Kaliappan, "Automated tomato leaf disease classification using transfer learning-based deep convolution neural network," *J. Plant Diseases Protection*, vol. 128, no. 1, pp. 73–86, Feb. 2021, doi: 10.1007/s41348-020-00403-0.

[14] S. Uguz and N. Uysal, "Classification of olive leaf diseases using deep convolutional neural networks," *Neural Comput. Appl.*, vol. 33, no. 9, pp. 4133–4149, May 2021, doi: 10.1007/s00521-020-05235-5.

[15] J. D. Bodapati, N. S. Shaik, and V. Naralasetti, "Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 10, pp. 9825–9839, Jan. 2021, doi: 10.1007/s12652-020-02727-z.

[16] A. Amirkhani, A. Khosravian, M. Masih-Tehrani, and H. Kashiani, "Robust semantic segmentation with multi-teacher knowledge distillation," *IEEE Access*, vol. 9, pp. 119049–119066, 2021, doi: 10.1109/ACCESS.2021.3107841.

[17] J. Mozaffari, A. Amirkhani, and S. B. Shokouhi, "A survey on deep learning models for detection of COVID-19," *Neural Comput. Appl.*, vol. 35, no. 23, pp. 16945–16973, May 2023, doi: 10.1007/s00521-023-08683-x.

[18] N. A. Ismail, C. W. Chai, H. Samma, M. S. Salam, L. Hasan, and N. H. A. Wahab, "Web-based university classroom attendance system based on deep learning face recognition," *KSII Trans. Internet Inf. Syst. (TIIS)*, vol. 16, no. 2, pp. 503–523, Feb. 2022, doi: 10.3837/tiis.2022.02.008.

[19] R. U. Haque, A. L. Pongos, C. M. Manzanares, J. J. Lah, A. I. Levey, and G. D. Clifford, "Deep convolutional neural networks and transfer learning for measuring cognitive impairment using eye-tracking in a distributed tablet-based environment," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 1, pp. 11–18, Jan. 2021, doi: 10.1109/TBME.2020.2990734.

[20] H. S. Nogay, T. C. Akinci, and M. Yilmaz, "Detection of invisible cracks in ceramic materials using by pre-trained deep convolutional neural network," *Neural Comput. Appl.*, vol. 34, no. 2, pp. 1423–1432, Jan. 2022, doi: 10.1007/s00521-021-06652-w.

[21] N. Luo, H. Yu, Z. You, Y. Li, T. Zhou, Y. Jiao, N. Han, C. Liu, Z. Jiang, and S. Qiao, "Fuzzy logic and neural network-based risk assessment model for import and export enterprises: A review," *J. Data Sci. Intell. Syst.*, vol. 1, no. 1, pp. 2–11, Jun. 2023, doi: 10.47852/bonviewjdsis32021078.

[22] J. Saeed and S. Zeebaree, "Skin lesion classification based on deep convolutional neural networks architectures," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 41–51, Mar. 2021, doi: 10.38094/jastt20189.

[23] J. Mozaffari, A. Amirkhani, and S. B. Shokouhi, "ColonGen: An efficient polyp segmentation system for generalization improvement using a new comprehensive dataset," *Phys. Eng. Sci. Med.*, vol. 47, no. 1, pp. 309–325, Jan. 2024, doi: 10.1007/s13246-023-01368-8.

[24] S. Navaneethan, P. S. S. Sreedhar, S. Padmakala, and C. Senthilkumar, "The human eye pupil detection system using BAT optimized deep learning architecture," *Comput. Syst. Sci. Eng.*, vol. 46, no. 1, pp. 125–135, 2023, doi: 10.32604/csse.2023.034546.

[25] A. Sungheetha and R. Sharma, "Design an early detection and classification for diabetic retinopathy by deep feature extraction based convolution neural network," *June*, vol. 3, no. 2, pp. 81–94, Jul. 2021, doi: 10.36548/jtcsst.2021.2.002.

[26] A. R. Ran, C. C. Tham, P. P. Chan, C.-Y. Cheng, Y.-C. Tham, T. H. Rim, and C. Y. Cheung, "Deep learning in glaucoma with optical coherence tomography: A review," *Eye*, vol. 35, no. 1, pp. 188–201, Jan. 2021, doi: 10.1038/s41433-020-01191-5.

[27] L. Gaur, U. Bhatia, N. Z. Jhanjhi, G. Muhammad, and M. Masud, "Medical image-based detection of COVID-19 using deep convolution neural networks," *Multimedia Syst.*, vol. 29, no. 3, pp. 1729–1738, Jun. 2023, doi: 10.1007/s00530-021-00794-6.

[28] Z. Xu, W. Wang, J. Yang, J. Zhao, D. Ding, F. He, D. Chen, Z. Yang, X. Li, W. Yu, and Y. Chen, "Automated diagnoses of age-related macular degeneration and polypoidal choroidal vasculopathy using bi-modal deep convolutional neural networks," *Brit. J. Ophthalmol.*, vol. 105, no. 4, pp. 561–566, Apr. 2021, doi: 10.1136/bjophthalmol-2020-315817.

[29] S. W. Khadim, H. K. Ibrahim, and A. M. Shadhar, "The finger vein recognition using deep learning technique," *Wasit J. Comput. Math. Sci.*, vol. 1, no. 2, pp. 1–7, Jun. 2022, doi: 10.31185/wjcms.43.

[30] P. Payal and M. M. Goyani, "A comprehensive study on face recognition: Methods and challenges," *Imag. Sci. J.*, vol. 68, no. 2, pp. 114–127, Mar. 2020, doi: 10.1080/13682199.2020.1738741.

[31] K. Nguyen, C. Fookes, S. Sridharan, and A. Ross, "Complex-valued iris recognition network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 182–196, Jan. 2023, doi: 10.1109/TPAMI.2022.3152857.

[32] A. Ullah, A. Salam, H. El Raoui, D. Sebai, and M. Rafie, "Towards more accurate iris recognition system by using hybrid approach for feature extraction along with classifier," *Int. J. Reconfigurable Embedded Syst. (IJRES)*, vol. 11, no. 1, p. 59, Mar. 2022, doi: 10.11591/ijres.v11.i1.pp59-70.

[33] Y. Liu, W. Shen, D. Wu, and J. Shao, "IrisST-Net for iris segmentation and contour parameters extraction," *Int. J. Speech Technol.*, vol. 53, no. 9, pp. 11267–11281, May 2023, doi: 10.1007/s10489-022-03973-8.

SHOUWU HE was born in Hequ County, Shanxi, China, in July 1979. He received the bachelor's degree in computer science and technology from Xinzhou Normal University in 2004 and the master's degree in computer application technology from Jiangxi Normal University in 2007.

From 2007 to 2016, he worked as a full-time Teacher at the College of Computer Application, Guilin University of Technology at Nanning, where he has been the Deputy Director, since 2017. He had led and completed a research project at provincial and ministerial level, and participated in two research projects at or above the municipal level. He has published five scientific research papers, two of which are included in the EI database, obtained a utility model patent and a software copyright. His research interests include big data processing and machine learning.

XIAOYING LI was born in Fenyang, Shanxi, in July 1981. She received the bachelor's degree in computer science and technology from Xinzhou Normal University in 2004 and the master's degree in computer software and theory from Jiangxi Normal University in 2007.

From 2021 to 2023, she worked as the Deputy Director at the Teaching and Research Office, Guilin University of Technology at Nanning. From 2007 to 2021, she worked as a full-time Teacher at Guilin University of Technology at Nanning, in recent years, she had led and participated in three municipal level scientific research projects, four provincial and ministerial level educational reform projects. She has published 13 academic papers, including two indexed by EI. She have obtained one utility model patent and one software copyright. Her research interests include machine learning, image processing, and computer vision.

● ● ●