

Research Article

An OCaNet Model Based on Octave Convolution and Attention Mechanism for Iris Recognition

Dong Zou,¹ Jianbing Feng ,^{1,2} Zhixin He,³ Liping Liu,³ Meijun Zhao,³ and Lihong Zheng³

¹Guangzhou Metro Group Co., Ltd., Guangzhou 510330, China

²South China University of Technology, Guangzhou 510640, China

³Guangzhou Metro Design & Research Institute Co., Ltd., Guangzhou 510010, China

Correspondence should be addressed to Jianbing Feng; fengjianbing@163.com

Received 3 July 2021; Revised 29 August 2021; Accepted 15 October 2021; Published 31 October 2021

Academic Editor: Pier Luigi Mazzeo

Copyright © 2021 Dong Zou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Iris recognition refers to identifying individuals based on iris patterns, which have been widely used in security systems, such as subway security and access control attendance, because everyone has a unique iris shape. In the study, we propose an OCaNet model for the iris recognition task. First, binarized threshold segmentation is used to locate the pupil and the pupil boundary is obtained; then, the Hough transform is applied to locate the outer edge of the iris; according to the located pupil and iris, the iris area image is obtained through image segmentation; finally, the iris image is normalized to adjust each original image to the same size and corresponding position, so as to eliminate the influence of translation, scaling, and rotation on iris recognition. Second, the normalized iris images are both input into the octave convolution module and attention module. The octave convolution module is used to extract the shape and contour features of the iris by decomposing the feature map into high and low frequencies. The attention module is applied to extract the color and texture characteristics of the iris. Finally, the two feature maps are concatenated and produce a distribution of output classes. Experimental results show that the proposed OCaNet model is significantly more accurate.

1. Introduction

Iris recognition has been widely used in many fields, especially in biometric pattern recognition [1]; at the same time, a lot of institutions and governments employ biometric technology due to its high accuracy over other human characteristics, such as handwriting and fingerprint [2]. Everyone has a unique pattern due to the significant difference between the complex textural pattern and color of the iris stroma [3]. According to a report from Daugman [4], more than one billion people in the world electronically register iris images in various databases around the world. Thus, iris recognition plays a key role in biometrics; it was first proposed by Burch [5], and amounts of iris recognition methods have emerged since then.

Conventional iris recognition methods require a high degree of user cooperation (including position and posture), and the illumination needs to be distributed. Currently, with

the development of deep learning, especially convolutional neural networks (CNNs), it has been widely applied in the task of object detection [6], image classification [7], and face recognition [8]; meanwhile, it has also led to breakthroughs in iris recognition, which shows good performance in recognition and classification scenarios. CNNs such as AlexNet [9], ResNet [10], and MobileNet [11] are widely employed in image classification. Liu et al. [12] applied CNNs to iris recognition for the first time; hierarchical convolutional neural networks and multiscale fully convolutional network were proposed for iris segmentation. Zhao et al. [13] proposed a method based on the capsule network architecture in iris recognition, and the dynamic routing method was added to the network. Gangwar et al. [14] proposed DeepIrisNet based on deep learning for iris recognition. Experiments on the ND-CrossSensor-Iris-2013 [15] and ND-IRIS-0405 datasets demonstrated that the proposed network exhibited superior performance. Nguyen et al. [16]

employed five different pretrained CNN networks on ImageNet [17] dataset, the multiclass SVM was used as a classifier for iris recognition, and the preliminary results demonstrate that off-the-shelf CNN features were extremely good at representing iris images and achieved promising recognition results on CASIA-Iris-Thousand and ND-CrossSensor-2013 datasets.

There are also some related research studies on iris recognition using the deep learning network [12, 18–23]. In 2015, He et al. [24] proposed a deep residual learning framework for image recognition; the characteristic of the residual network is that it is easy to optimize and can increase the accuracy by adding considerable depth. The internal residual block uses skip connections to alleviate the problem of gradient disappearance caused by increasing depth in the deep neural network. The network won the image classification and object recognition in the ILSVRC 2015 challenge with a top-5 error rate of 3.57%. In 2018, Woo et al. [25] proposed the convolution block attention module (CBAM), which represented the attention mechanism module for feed-forward convolution networks; for the feature map generated by the convolutional neural network, CBAM calculates the attention map of the feature map from the two dimensions of channel and space, and then, the attention maps are multiplied to the input feature map for adaptive feature learning. CBAM is a lightweight universal module, which can be incorporated into various convolutional neural networks for end-to-end training. In 2020, Wang [26] proposed an efficient channel attention (ECA) module to overcome the paradox of performance and complexity trade-off. Experimental results show that the ECA module is an extremely lightweight plug-and-play module. In 2019, Chen [27] introduced a new convolution method called octave convolution; by means of scale-space transformation and octave convolution operation, the network can deal with the high and low frequency components more clearly, and the computation of the low frequency components can be saved. In the study, we choose ResNet18 as the backbone of the proposed network, and the octave convolution and CBAM combined with ECA are used to improve the performance of ResNet18. ECANet will be used as the channel attention module, just as the fractal coding algorithm [28], which was first proposed for image compression. Many scholars have applied the algorithm to other fields, such as image denoising [29], image segmentation [30], and so on. Our proposed algorithm can recognize iris images efficiently and also can be used as a feature extractor for other visual tasks. However, these methods based on salient features and color spatial information and iris recognition mainly focus on extracting the shape and contour features.

Existing iris recognition methods can be divided into the key steps. (1) Iris acquisition: eye images are captured from an infrared camera. (2) Iris image processing: the iris region needs to be segmented from the captured image; meanwhile, in order to remove some inconsistencies in the size of the segmented iris regions because of eye ball movement, rotation of the camera, and pupil dilation, the segmented region must be normalized to be converted into a region of

consistent size. (3) Iris classification: it is worth noting that feature plays a key role in the image recognition. Therefore, iris features will be extracted for iris recognition.

This study conducts a study on iris recognition. First, binarized threshold segmentation is used to locate the pupil, and the pupil boundary is obtained; then, the Hough transform is applied to locate the outer edge of the iris; according to the located pupil and iris, the iris area image is obtained through image segmentation; finally, the polar coordinate transformation is used to normalize iris image. Second, the normalized iris images are both input into the octave convolution module and attention module. The octave convolution module is used to extract the shape and contour features of the iris by decomposing the feature map into high and low frequencies. The attention module is applied to extract the color and texture characteristics of the iris. The flowchart of our proposed method is shown in Figure 1.

The idea of the study is to use ResNet18 as the backbone network, incorporating CBAM, ECA, and the octave convolution modules, for iris recognition. It can improve the accuracy of the ResNet18 network due to the addition of CBAM and ECA modules. Incorporation of CBAM, ECA, and the octave convolution modules with ResNet for image classification is new for iris recognition. Our main contributions of the study are as follows:

- (1) To our best knowledge, there has been no standard method for iris recognition. This work is one of the few innovative investigations on iris recognition effectively.
- (2) In order to solve iris image classification problems, this study designed a model based on the octave convolution module and the attention module
- (3) We use ECANet as the channel attention module in the proposed method, which avoids dimensionality reduction, and increase the interaction of information between channels

2. Methods

We proposed a double branch network structure and named it OCANet, consisting of the octave convolution module and attention module. Compared with using a single network structure, the accuracy was improved. The proposed network is shown in Figure 2.

Iris images are input into both the octave convolution branch and attention module which perceive the same input in two different views. The octave convolution module is composed of four OCT blocks, and each OCT block is composed of two bottleneck cascades. Finally, the output of the feature maps extracted by the octave convolution module and attention module is concatenated and passed onto the fully connected layer, which produces a distribution of output classes. The network structural parameters are given in Table 1. The number of training epochs was generally kept constant at 300 epochs, learning rate is $1e-3$, and the momentum is 0.9. The model is trained using the Adam optimizer [31]. Weight decay is used to avoid overfitting of the model, and weight decay is $5e-4$.

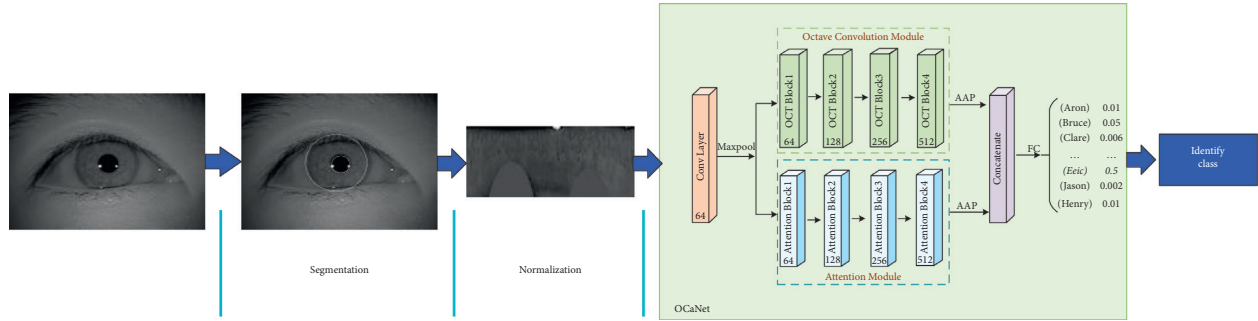


FIGURE 1: The framework for iris recognition. The iris image is segmented using binarized threshold and Hough transform and then geometrically normalized using polar coordinate transformation. Features are next extracted using the proposed OCaNet.

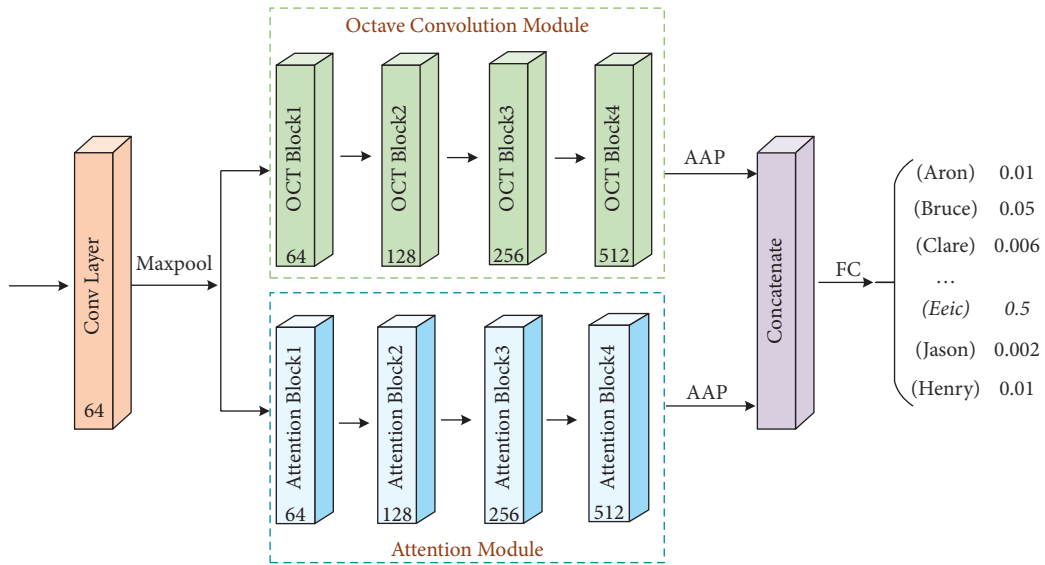


FIGURE 2: The proposed network architecture. It is divided into an attention module and an octave convolution module.

TABLE 1: Related parameters of different layers.

Network	Layer name	Layer type	Related parameters
—	Conv1	Conv2D	$7 \times 7, 64, \text{stride } 2$
—	Pool1	MaxPool2d	$3 \times 3, 64, \text{stride } 2$
OCT module	OCT block 1	Convolution	$\begin{bmatrix} 1 \times 1, 64, \text{stride } 1 \\ 3 \times 3, 64, \text{stride } 1 \\ 1 \times 1, 64, \text{stride } 1 \end{bmatrix} \times 2$
	OCT block 2	Convolution	$\begin{bmatrix} 1 \times 1, 128, \text{stride } 2 \\ 3 \times 3, 128, \text{stride } 2 \\ 1 \times 1, 128, \text{stride } 2 \end{bmatrix} \times 2$
	OCT block 3	Convolution	$\begin{bmatrix} 1 \times 1, 256, \text{stride } 2 \\ 3 \times 3, 256, \text{stride } 2 \\ 1 \times 1, 256, \text{stride } 2 \end{bmatrix} \times 2$
	OCT block 4	Convolution	$\begin{bmatrix} 1 \times 1, 512, \text{stride } 2 \\ 3 \times 3, 512, \text{stride } 2 \\ 1 \times 1, 512, \text{stride } 2 \end{bmatrix} \times 2$

TABLE 1: Continued.

Network	Layer name	Layer type	Related parameters
Attention module	Pool3	AdaptiveAvgPool2d	1, 1
	EA_1	Attention	$[3 \times 3, 64, \text{stride } 2] \times 4$
	Attention_2	Attention	$[3 \times 3, 128, \text{stride } 2] \times 4$
	Attention_3	Attention	$[3 \times 3, 256, \text{stride } 2] \times 4$
	Attention_4	Attention	$[3 \times 3, 512, \text{stride } 2] \times 4$
—	Pool2	AdaptiveAvgPool2d	1, 1
	FC	Inner product	—

The OCaNet model combines the advantages of the two modules by paralleling the two networks: OCT module and attention module. The multifaced characteristics of the iris will be extracted by the two modules. The octave convolution module is used to extract the shape and contour features of the iris by decomposing the feature map into high and low frequencies. The attention module is employed to extract the color and texture characteristics of the iris. The combination of the two effectively makes up for the lack of feature extraction with a single structure.

2.1. Octave Convolution. In this study, we use the existing CNN ResNet18 to extract features from iris images. ResNet has been widely used in areas such as image recognition, segmentation, and detection due to the advantage of simple and practical. The input image size is 64×64 pixels, and the “bottleneck design” is used in ResNet18 to reduce the number of parameters. The architecture of the bottleneck is shown in Figure 3.

Unlike existing multiscale methods, octave convolution can be directly used to replace the vanilla convolution and can be inserted to regular convolutional networks without special adjustment. Since octave convolution focuses on reducing the redundancy in the CNN spatial dimension, it is orthogonal and complementary to the existing methods that focus on the network topology design and reduce network features easily. The goal of octave convolution is to effectively process the high and low frequencies. The detailed process of the octave convolution is shown in Figure 4.

As stated in Chen et al. [27], let X and Y be the factorized input and output tensors. During convolution, $X \in \mathbb{R}^{C \times h \times W}$ can be decomposed into the low-frequency tensor $X^L \in \mathbb{R}^{\alpha \times C \times h \times W}$ and the high-frequency tensor $X^H \in \mathbb{R}^{(1-\alpha) \times C \times h \times W}$, where X^H denotes details, X^L denotes the overall structure, and $\alpha \in [0, 1]$ represents the ratio of channels allocated to the low-frequency part and the low-frequency feature maps. When calculating low-frequency output feature maps, it is necessary to use average pooling for downsampling of high-frequency input features. After that, the calculation formula is as follows:

$$\begin{aligned}
 f(H \rightarrow H) &= \text{Conv}(X^H), \\
 f(H \rightarrow L) &= \text{Conv}(\text{avgpool}X^H), \\
 f(L \rightarrow H) &= \text{upsample}(\text{Conv}(X^L)), \\
 f(L \rightarrow L) &= \text{Conv}(X^L),
 \end{aligned} \tag{1}$$

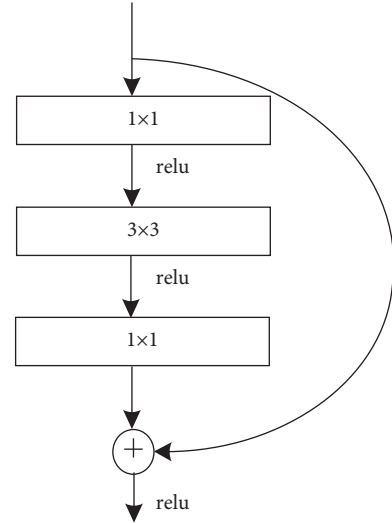


FIGURE 3: Bottleneck architecture.

where $f(H \rightarrow H)$ and $f(H \rightarrow L)$ represent the intra-frequency update, while $f(L \rightarrow H)$ and $f(L \rightarrow L)$ represent the interfrequency communication. The output feature maps of low-frequency and high-frequency can be calculated:

$$\begin{aligned}
 Y^H &= f(H \rightarrow H) + f(L \rightarrow H), \\
 Y^L &= f(H \rightarrow L) + f(L \rightarrow L).
 \end{aligned} \tag{2}$$

The vanilla convolutions of OCT blocks in the proposed network are replaced with octave convolutions. In the study, we set the first octave convolution input and the last octave convolution output to all high frequencies. For all other octave convolutions, weighing accuracy, and calculation amount, we set $\alpha = 0.5$.

2.2. Attention Block. CBAM represents the attention mechanism module of the convolution module. It is an attention mechanism module that combines spatial and channel, which has proven to be an effective network. Figure 5 shows the attention block.

As described in [25], given an intermediate feature map $X \in \mathbb{R}^{C \times W \times h}$ as input, where C is the number of channels, W and h are the width and height of the reduced image obtained after convolution. The attention model successively establishes a 1D channel convolution layer $M_C \in \mathbb{R}^{C \times 1 \times 1}$ and

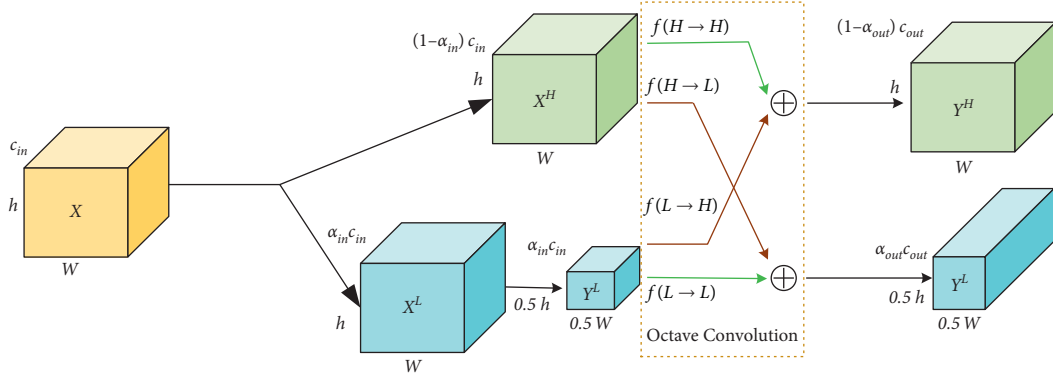


FIGURE 4: Detailed process of the octave convolution.

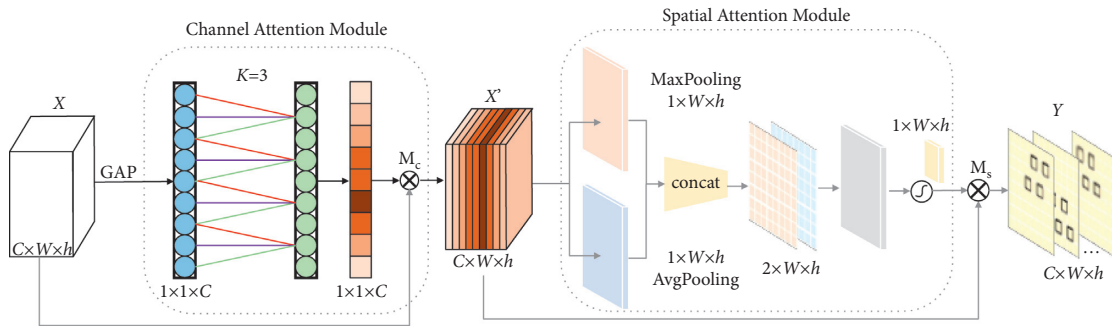


FIGURE 5: Attention block.

a 2D spatial convolution $M_S \in \mathbb{R}^{1 \times W \times h}$. The formulas are as follows:

$$\begin{aligned} X' &= M_C(X) \otimes X, \\ Y &= M_S(X') \otimes X', \end{aligned} \quad (3)$$

where \otimes represents element-wise multiplication. Y is the final refined output.

In the study, we use ECANet as the channel attention module, and each channel of a feature map is considered as a feature detector [32], given the aggregated feature maps M_C obtained by global average pooling (GAP); then, the kernel size is adaptively computed by a mapping of channel

dimension C . Finally, the channel weights are generated by performing a fast 1D convolution of size k , where $k = 3$.

The spatial attention module is used to extract the relationship of the internal space and calculate in which a small part should be focused on. The input of the module is the data processed by channel attention. Average pooling and max pooling methods are applied to compress the values of multiple channels. To a single channel, go through a 2-dimensional convolution in the spatial layer and calculate the weight of the space. The specific method is shown in the following formula:

$$M_s(X) = \sigma(f^{3 \times 3}([\text{AvgPool}(X); \text{MaxPool}(X)])) = \sigma(f^{3 \times 3}([X_{\text{avg}}^s; X_{\text{max}}^s])), \quad (4)$$

where σ represents the sigmoid function, and $f^{3 \times 3}$ denotes a convolution operation with the filter size of 3×3 , which is composed of convolution, normalization, and ReLU.

Therefore, we use ECANet as the channel attention module in the proposed method, which avoids dimensionality reduction, and increase the interaction of information between channels, while reducing complexity while maintaining performance (through one-dimensional convolution).

2.3. Fusion and Decision Layers Design. Shervan [33] proposed local ternary patterns and a multilayer neural network, which provides a new network topology. At the same time, the author points out that multilayer perceptrons are often used to visual pattern identification. The last bottleneck architecture of the ResNet18 network is connected to the average pooling layer and output 512-dimensional features, then the output of the feature maps extracted by the octave convolution module and attention module is concatenated. The

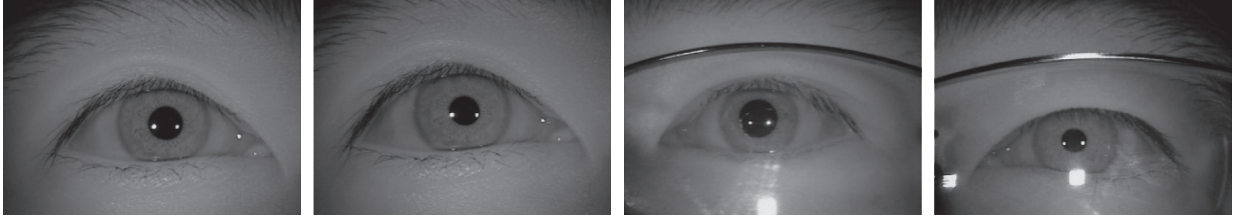


FIGURE 6: Sample images from CASIA-Iris-Thousand.

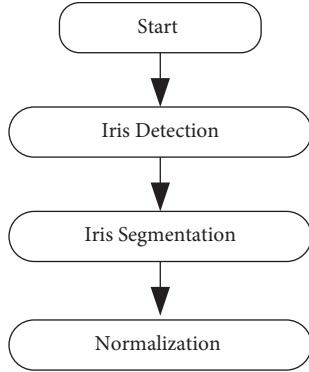


FIGURE 7: Flowchart of the iris recognition.

concatenated features are fed into the fully connected layer. We used cross-entropy loss as the loss function:

$$\text{Loss}(x, \text{label}) = -w_{\text{label}} \log \frac{e^{X_{\text{label}}}}{\sum_{j=1}^N e^{X_j}} = w_{\text{label}} \left[-X_{\text{label}} + \log \sum_{j=1}^N e^{X_j} \right], \quad (5)$$

where $w \in \mathbb{R}^N$ is the activation value with Softmax, N is the dimension of X , $\text{label} \in [0, C - 1]$ is the corresponding label, and $w \in \mathbb{R}^C$ is a vector with dimension C , which is applied to represent the weights of labels.

3. Experiments and Results

3.1. Dataset. In the study, we select CASIA-Iris-Thousand database. CASIA-Iris-Thousand [34] is a subset of the CASIA-IrisV4 database, with the total number of 20000 images, consisting of left and right eye images of 1000 individuals. The database is with images of size 640×480 . Examples of the database are shown in Figure 6.

3.2. Image Preprocessing. In general, the obtained images contain not only the useful information but irrelevant information, such as the eyelashes and whites of the eyes, and in a highly noninvasive system, the position and size of the iris in the image will change due to no requirements on the subject. Therefore, before performing iris recognition, it is necessary to determine the position of the iris in the image and perform operations such as normalization. The specific process is shown in Figure 7.

We need to isolate the actual iris region from the original images to remove the influence of eyelids, eye fluid, and tiny tissues, eliminate noise interference, use translation

transformation and image alignment methods to eliminate shift, scale, and rotate, and remove interference due to noise such as reflections.

The part of the iris that contains the texture is the part between the two approximately circular borders inside and outside. However, these two circles are not completely concentric, and we need to deal with the inner and outer boundaries separately. Iris positioning is a key link in the iris recognition system. The accuracy of the positioning directly affects the follow-up work, and the iris positioning is also the most time-consuming part of the whole process. In the study, first, the binarization threshold segmentation is used to locate the pupil, and the pupil boundary can be obtained. Then, the Canny operator is applied to extract the edge of the original image, and the Hough transform is used to locate the outer edge of the iris. According to pupil positioning and iris positioning, image segmentation is performed to obtain an image of the iris area.

After the segmentation image is obtained, normalization is performed to adjust each original image to the same size and corresponding position, thereby eliminating the impact of translation, scaling, and rotation on iris recognition. In this step, all of the input images were just resized.

In the study, we use polar coordinate changes to normalize the iris. The iris region $I(x, y)$ can be mapped from the Cartesian coordinate (x, y) to the polar coordinate (r, θ) , and the process can be expressed as

$$I(x(r, \theta), y(r, \theta)) \longrightarrow I(r, \theta), \quad (6)$$

where $r \in [0, 1]$, $\theta \in [0, 2\pi]$ is an angle. The inner and outer boundaries of the iris are both circular; starting from the center of the pupil, the coordinates of the intersection with the inner and outer boundaries of the iris are $(x_p(\theta), y_p(\theta))$ and $(x_s(\theta), y_s(\theta))$, respectively. Then, map each point in the iris image to the polar coordinate (r, θ) , and the following formulas are used:

$$\begin{cases} x(r, \theta) = (1 - r)x_p(\theta) + rx_s(\theta) \\ y(r, \theta) = (1 - r)y_p(\theta) + ry_s(\theta) \end{cases} \quad (7)$$

Finally, the normalized iris image is fed into the OCANet model to recognize the iris image.

3.3. Effect of Different Models. In order to verify the efficiency of the model proposed in the study, we compared the accuracy of different models on the iris image database. The batch size is 32. The model was trained for, at most, 300 epochs. The loss and accuracy curves of different models on

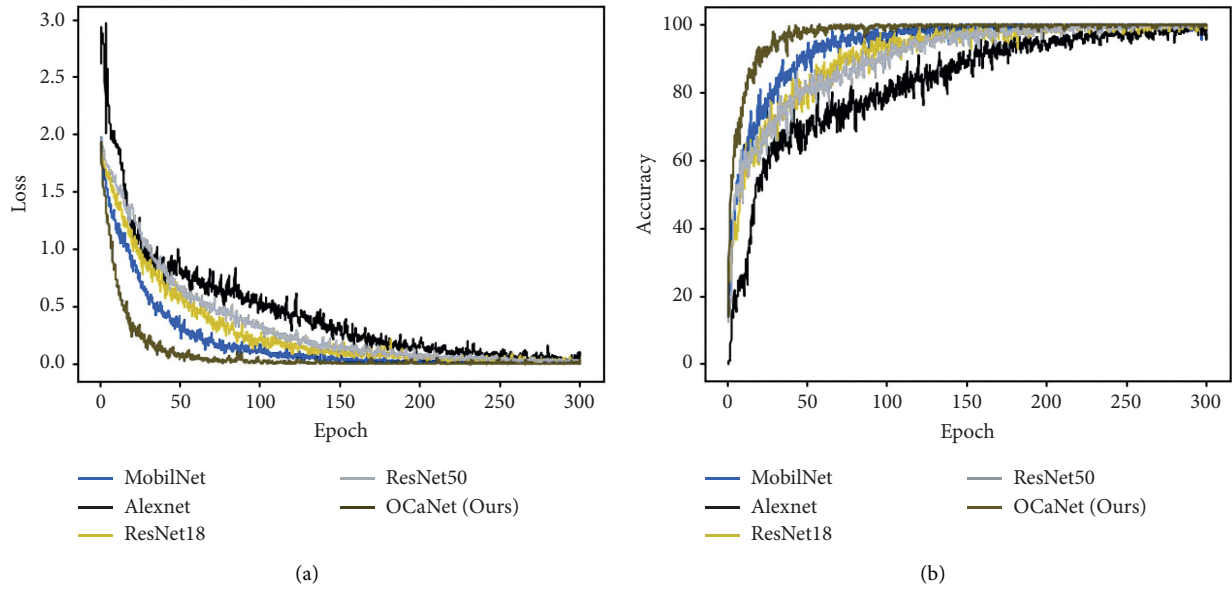


FIGURE 8: Accuracy and loss curves of different models on the iris dataset. (a) Loss curve of different models. (b) Accuracy curve of different models.

TABLE 2: Comparison of performance between different models on iris image.

Model	Accuracy (%)	Number of misjudgments
AlexNet	62.46	1235
ResNet50	84.34	515
ResNet18	86.07	458
MobileNet	87.87	399
OCaNet (ours)	91.52	279

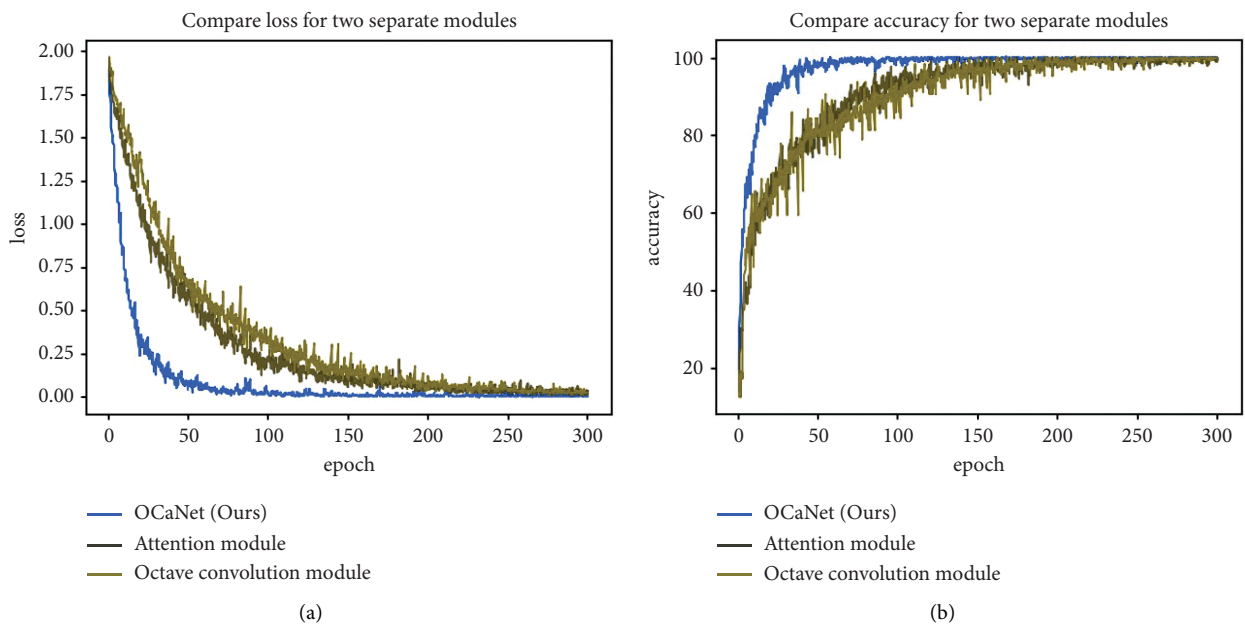


FIGURE 9: Accuracy and loss curves of different modules on the iris dataset. (a) Loss curve of different modules. (b) Accuracy curve of different modules.

TABLE 3: Comparison of performance between different modules on the iris dataset.

Model	Accuracy (%)	Number of misjudgments
ResNet18	86.07	458
CBAM + ResNet18	84.10	523
Oct + ResNet18 (Oct module)	89.27	353
ECBAM + ResNet18 (A module)	89.57	343
OCaNet (ours)	91.52	279

the iris dataset are shown in Figure 8. It can be seen from the change of accuracy and loss with epoch in Figure 8 that the changes in the accuracy and loss during the entire training process are relatively smooth.

In the entire network training stage, the fastest and smoothest convergence is the proposed network, the slowest convergence is AlexNet. After 29 epochs, the loss quickly dropped from the initial 1.8 to less than 0.2. After 59 epochs, the network convergence begins to show a more obvious slowdown, and the curve shows small fluctuations in the process, the network loss decreased from 0.2 to 0.05. After 94 epochs, the loss of the network gradually stabilizes, and there is no loss in the subsequent training process.

In the early stage of training, the accuracy of our network reaches more than 96%, and after 59 epochs, the accuracy of the network gradually stabilized, the curve fluctuations are very small, and the entire network tends to be saturated, and there is no loss in the subsequent training process.

The comparison of performance between different models on iris image is given in Table 2. From Table 2, we can observe that compared with other models, the proposed model achieves the best result, which has an accuracy of 91.52%.

3.4. Effect of Different Modules. In this part, we compared the network with two separate modules on the iris image database. The batch size is 32. The model was trained for, at most, 300 epochs. The loss and accuracy curves of different modules on the iris dataset are shown in Figure 9.

It can be seen from the change of accuracy and loss with epoch in Figure 9 that for the training dataset, the loss values of the network with attention module and the network with the octave convolution module are slightly lower than those of OCaNet. As such, OCaNet showed a slightly better performance than the network with the attention module and the network with the octave convolution module on the training dataset. Therefore, our proposed model provided the highest accuracy and the lowest loss among all models on the training dataset.

We compare the accuracy of the network with a single module, and the performance of different models on the iris database is given in Table 3. From the results, we can observe that ResNet18 with CBAM has the lowest accuracy. The models with the modules (octave convolution and attention module) clearly outperformed the original models, which indicates that ResNet18 with the two modules are suitable for recognizing iris. Compared with ResNet18, the accuracy of ResNet18 with the octave convolution module and ECA + CBAM(ECBAM) module have increased by 3.2% and

TABLE 4: Comparison of performance between the single modal network and multimodal network.

Model	Accuracy (%)	Number of misjudgments
ResNet50	84.34	515
OCa-ResNet50	85.35	482
MobileNet	87.87	399
OCa-MobileNet	89.85	334
ResNet18	86.07	458
OCaNet (ours)	91.52	279

3.5%, respectively. However, the accuracy of ResNet18 with CBAM has reduced by 1.97%, which indicates that ResNet18 with the CBAM module is not suitable for recognizing iris. However, the proposed model in the study achieves the highest accuracy, and these results demonstrate the effectiveness of the ECBAM for iris recognition.

3.5. Results Evaluation. To verify the feasibility of the network proposed in this study, we also compared the accuracy of single modal networks and the networks with the OCa module. The results are given in Table 4. From the results, we can observe that the OCa module achieves notable gain over ResNet50, MobileNet, and ResNet18. The OCa module using ResNet18 as backbone achieves 1.67% and 6.17% gains over using ResNet50 and MobileNet as backbone models, respectively. These results verify that the OCa module has a good ability to enhance ResNet50, MobileNet, and ResNet18 for iris recognition. In summary, the experimental results show that the OCa module can well improve the performance of the network, which is suitable for iris recognition.

4. Conclusions

In this study, we propose an OCaNet model based on octave convolution and attention mechanism for iris recognition. The iris images are simultaneously fed into the octave convolution module and attention module, and the octave convolution module is used to extract the shape and contour features of the iris by decomposing the feature map into high and low frequencies. The attention module is applied to extract the color and texture characteristics of the iris. Finally, the two feature maps are concatenated and produces a distribution of output classes. The proposed method has obvious advantages over other methods; this indicates that the proposed method is more suitable for iris recognition.

Experimental results demonstrate that a deep network with an octave convolution module and attention module is efficient for iris recognition. In the test of the CASIA-IrisV4 database, OCa + ResNet18 has the highest accuracy of

91.52%, compared with the existing state-of-the-art methods. In future, we intend to apply the proposed method to other image classification areas, such as natural scene images, TV news channel video frames, and other textual images.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key R&D Program of China (2017YFB1201203).

References

- [1] A. M. Sarhan, "Iris recognition using discrete cosine transform and artificial neural networks," *Journal of Computer Science*, vol. 5, no. 5, pp. 369–373, 2009.
- [2] A. Albadarneh, I. Albadarneh, and J. Alqatawna, "Iris recognition system for secure authentication based on texture and shape features conference," in *Proceedings of the 21 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, Amman, Jordan, November 2015.
- [3] H. Proença and L. A. Alexandre, "Iris recognition: analysis of the error rates regarding the accuracy of the segmentation stage," *Image and Vision Computing*, vol. 28, no. 1, pp. 202–206, 2010.
- [4] J. Daugman, *Major International Deployments of the Iris Recognition Algorithms: A Billion Persons*, University of Cambridge, Cambridge, U.K., 2014.
- [5] K. D. L. Irsch and J. H. M. Guyton, "Anatomy of eyes," *Encyclopedia of Biometrics*, pp. 1212–1217, Springer, Boston, MA, USA, 2009.
- [6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 580–587, Columbus, OH, USA, June 2014.
- [7] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, June 2015.
- [8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, Boston, MA, USA, June 2015.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [11] A. G. Howard, M. L. Zhu, B. Chen et al., "Mobilenets: efficient convolutional neural networks for mobile vision applications," <http://arxiv.org/abs/1704.04861>.
- [12] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan, "Accurate iris segmentation in non-cooperative environments using fully convolutional networks," in *Proceedings of the 2016 International Conference on Biometrics (ICB)*, pp. 1–8, Halmstad, Sweden, June 2016.
- [13] T. Zhao, Y. Liu, G. Huo, and X. Zhu, "A deep learning iris recognition method based on capsule network architecture," *IEEE Access*, vol. 7, pp. 49691–49701, 2019.
- [14] A. Gangwar and A. Joshi, "DeepIrisNet: deep iris representation with applications in iris recognition and cross-sensor iris recognition," in *Proceedings of the 2016 International Conference on Image Processing (ICIP)*, pp. 2301–2305, Phoenix, AZ, USA, September 2016.
- [15] P. J. Phillips, W. T. Scruggs, A. J. O'Toole et al., "FRVT 2006 and ICE 2006 large-scale experimental results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 831–846, 2010.
- [16] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan, "Iris recognition with off-the-shelf CNN features: a deep learning perspective," *IEEE Access*, vol. 6, pp. 18848–18855, 2018.
- [17] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. F. Li, "Imagenet: a large-scale hierarchical image database," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, Miami, FL, USA, June 2009.
- [18] E. Jalilian and A. Uhl, "Iris segmentation using fully convolutional encoder-decoder networks," *Deep Learning for Biometrics*, pp. 133–155, 2017.
- [19] E. Severo, R. Laroca, S. Bezerra et al., "A benchmark for iris location and a deep learning detector evaluation," in *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, Rio de Janeiro, Brazil, July 2018.
- [20] D. Menotti, G. Chiachia, A. Pinto et al., "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 864–879, 2015.
- [21] J. Tapia and C. Aravena, "Gender classification from NIR iris images using deep learning," *Deep Learning for Biometrics*, pp. 219–239, 2017.
- [22] L. He, H. Li, F. Liu, N. Liu, Z. Sun, and Z. He, "Multi-patch convolution neural network for iris liveness detection," in *Proceedings of the 2016 IEEE International Conference Biometrics Theory, Applications Systems (BTAS)*, pp. 1–7, Niagara Falls, NY, USA, September 2016.
- [23] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, "A deep learning approach for iris sensor model identification," *Pattern Recognition Letters*, vol. 113, pp. 46–53, 2018.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," <http://arxiv.org/abs/1512.03385>.
- [25] S. H. Woo, J. C. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," <http://arxiv.org/abs/1807.06521>.
- [26] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: efficient channel attention for deep convolutional neural networks," <http://arxiv.org/abs/1910.03151>.
- [27] Y. Chen, H. Fan, Z. Yan et al., "Drop an octave: reducing spatial redundancy in convolutional neural networks with octave convolution," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3434–3443, Seoul, South Korea, November, 2019.

- [28] A. E. Jacquin, "Image coding based on a fractal theory of iterated contractive image transformations," *IEEE Transactions on Image Processing*, vol. 1, no. 1, pp. 18–30, 1992.
- [29] M. Ghazel, G. H. Freeman, and E. R. Vrscay, "Fractal-wavelet image denoising revisited," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2669–2675, 2006.
- [30] R. Ferrell, S. Goeason, and J. Tobin, "Application of fractal encoding techniques for image segmentation," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5132, pp. 69–77, 2003.
- [31] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," <http://arxiv.org/abs/1412.6980>.
- [32] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, September 2014.
- [33] F. E. Shervan, "Bark texture classification using improved local ternary patterns and multilayer neural network," *Expert Systems with Applications*, vol. 158, pp. 1–8, 2020.
- [34] *CASIA Iris Image Database, V-1.0*, <http://www.sinobiometrics.com>, 2010.