IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 31, 2018, accepted February 19, 2018, date of publication March 6, 2018, date of current version April 23, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2812208

Eye Recognition With Mixed Convolutional and Residual Network (MiCoRe-Net)

ZI WANG[1], CHENGCHENG LI[1], HUIRU SHAO[2], AND JIANDE SUN[2]
[1]Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville,TN 37996,USA
[2]School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China

Corresponding author: Jiande Sun (jiandesun@hotmail.com)

This work was supported in part by the Natural Science Foundation for Distinguished Young Scholars of Shandong Province under Grant JQ201718, in part by the Key Research and Development Foundation of Shandong Province under Grant 2016GGX101009, in part by the Natural Science Foundation of China under Grant U1736122, Grant 61603225, and Grant 61601268, and in part by the Shandong Provincial Key Research and Development Plan under Grant 2017CXGC1504.

ABSTRACT Although iris recognition has achieved big successes on biometric identification in recent years, difficulties in the collection of iris images with high resolution and in the segmentation of valid regions prevent it from applying to large-scale practical applications. In this paper, we present an eye recognition framework based on deep learning, which relaxes the data collection procedure, improves the anti-fake quality, and promotes the performance of biometric identification. Specifically, we propose and train a mixed convolutional and residual network (MiCoRe-Net) for the eye recognition task. Such an architecture inserts a convolutional layer between every two residual layers and takes the advantages from both of convolutional networks and residual networks. Experiment results show that the proposed approach achieves accuracies of 99.08% and 96.12% on the CASIA-Iris-IntervalV4 and the UBIRIS.v2 datasets, respectively, which outperforms other classical classifiers and deep neural networks with other architectures.

INDEX TERMS Eye recognition, iris recognition, deep learning, convolutional neural network, deep residual network, mixed convolutional and residual network (MiCoReNet).

I. introduction

Biometric recognition has been well studied in the recent decades, and plays a more and more important role in the area of information security, personal identification, etc. Among all kinds of biometric information, iris has stableness and uniqueness of features [1]. In recent years, researchers have achieved significant progresses in iris recognition and the best algorithms can reach the accuracy of more than 99% [2], [3]. However, iris recognition has disadvantages, mainly in the following two aspects: (1) Daugman etc. suggested that if we want to get a satisfied accuracy of recognition, the radius of iris should be at least 80-130 pixels. Moreover, internal factors such as eyelid and eyelash, and external factors, such as eyeglasses, can decrease the quality of the iris images. [4]. (2) The iris has to be precisely segmented from the eye/face image with powerful segmentation algorithms [4]. Besides the interference factors mentioned above, the fact that pupil is not strictly a circle makes the segmentation and the upcoming procedures, such as normalization, even more difficult.

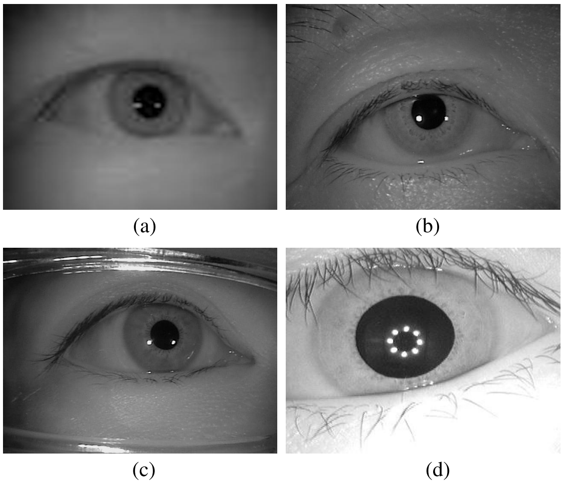


FIGURE 1. **Challenging Samples in iris recognition systems. (a) Sample with low resolution. (b) Sample occluded by eyelid and eyelash. (c) Sample interfered by eyeglasses. (d) Sample with deformation.**

Considering these factors, we propose an eye recognition framework with mixed convolutional and residual network. Compared with iris, the whole eye region consisting of the eyelash, pupil, iris, sclera, even with wrinkles and eyeglasses which makes data collecting much easier. Moreover, the eyelash, pupil, iris, sclera, even with wrinkles and eyeglasses included in the eye region, which are negative factors in iris recognition, are transfered into positive factors in our proposed method and even can provides helpful information for recognition features. In this paper, with these concerns we employ the powerful capability of multi-level feature learning in deep neural network and propose a novel recognition framework, which takes the whole eye image, rather than the segmented and normalized iris ones, as a kind of biometrics. Such system unburdens us the demanding on segmentation and relaxes the acquisition requirements.

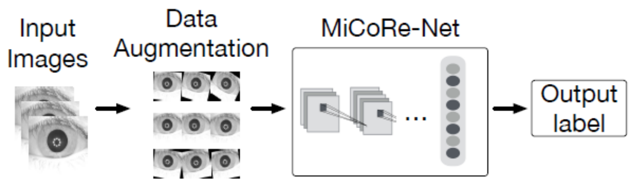


FIGURE 2. Eye Recognition with MiCoRe-Net.

Our framework consists of two main parts: data augmentation and mixed convolutional and residual network (MiCoRe-Net) as shown in Fig.2. Since the images from datasets are usually collected with certain distances and perspectives, the representation of the feature appears relatively shallow to some extent. For example, in the UBIRIS.v2 database, images are captured in a range of four to eight meters with not only frontal perspective, but left and right off-angle perspectives as well. Data augmentation including rotation, cropping, rotation after cropping, and the combination of them, can greatly enrich the feature of each sample/class of the training set, which provides the deep neural network with more information. In the experiment, we will compare the performance between different data augmentation strategies. A system can be trained to fit different kinds of testing samples by feeding such augmented training samples to it.

In this study, we implement a six-layer MiCoRe-Net followed by a flatten and two fully connected layers. The MiCoRe-Net combines the ConvNet and ResNet by starting with a convolutional layer and inserting a convolutional layer between every two residual layers. This framework gains advantages from both architectures, i.e., the advantage of fast convergence from the convolutional network and of the non-saturation feature from the residual network. Our eye recognition system achieves an accuracy of 99.08% on the CASIA iris dataset. Even under a much wilder situation of data collection, our approach can still be comparable to the best performance of iris recognition algorithms. The main contributions of this study are as follows.

(1) We propose a novel eye recognition system that breaks the limitations in the iris recognition system, which makes the biometric recognition task much more convenient. The new framework is with great practicability and efficiency comparing with traditional iris recognition systems.

(2) We propose a MiCoRe-Net based recognition system, which utilizes the raw data, rather than the features extracted by the local descriptors. The pre-processing procedure is simplified and the presentation of features is self-learned. The new architecture takes advantages from both convolutional neural network (ConvNet) and deep residual network (ResNet) and achieves satisfied performance.

The rest of this paper is organized as follows. In Section 2, we briefly introduce related work about iris and eye recognition, along with the researches related to convolutional neural network. Then the proposed eye recognition framework with MiCoRe-Net is presented in Section 3. Experimental results and discussions are presented in Section 4. Finally, we conclude the paper in Section 5.

II. RELATED WORK

Traditional iris recognition frameworks usually need features extraction through some local descriptors. Filter-based methods play a dominant role in early stage. Daugman proposed a Gabor filter-based method for iris recognition [5]. Besides, wavelet filter-based methods can also achieve high performance [6]. Another popular branch of feature extraction method is based on local descriptors, such as Local Binary Patterns (LBP) [7]–[9], Dual-Cross Patterns (DCP) [10], [11], Local Phase Quantization (LQP) [12]–[14], and Daisy [15] etc. Among all these local descriptors, LBP, which compares the relationship between the center pixel and its surroundings pixel in a small window and generates a binary code as the feature, is the most famous one because of its computational simplicity, uniqueness in feature description and tolerance against various kinds of interferences. A recent study named ALBP [2], is an approach based on improved LBP, achieves an impressive accuracy of 99.91% on the CASIA-Iris-IntervalV4 database.

Convolutional neural network (CNN) and deep residual network (ResNet), achieve great success in the field of object recognition. Such deep learning based algorithms have already been largely utilized on face recognition and related applications [16], [17], but few in iris or eye recognition. DeepIris [18] is one among these studies, which focuses on iris verification rather than recognition. However, recognition is a more challenging task than verification.

The field that utilizes the eye as the biometric information for the identification/recognition purpose has not been well studied yet. Some researchers combined the iris and other information to improve the performance. For example, Z.Zhou etc. introduced a multimodal eye recogniton system where the information of the iris and the sclera can be obtained at the same time [19]–[21]. Although they claimed that the study is an eye recognition approach, however, a combination of the iris and the sclera information is treated as the feature feeding into the classifier. Such approach is more like a data fusion application, rather than using the whole eye images as the input. Therefore, this study is still in the scope of traditional biometric recognition frameworks. One of the earliest study that uses the entire eye image

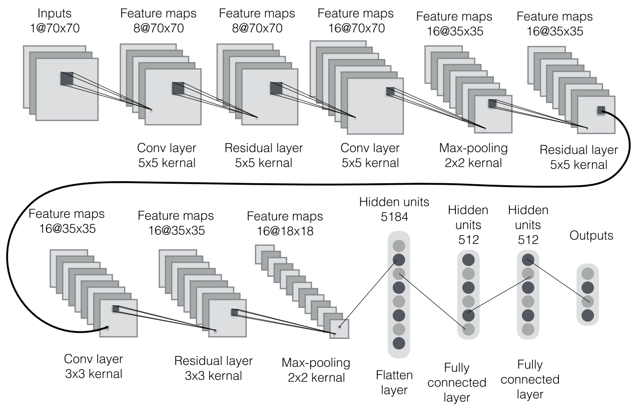


FIGURE 3. **The MiCoRe-Net framework.**

as the feature/input is [22] to detect the driver's driving state.

III. PROPOSED METHOD

The overall framework and the MiCoRe-Net architecture of our eye recognition system are depicted in Fig. 2 and 3 respectively. In this section, we first give a detailed description of the architecture, and then present the novel features of the eye recognition system, including mixed convolutional and residual network (MiCoRe-Net), data augmentation, and ReLU activation function.

A. THE ARCHITECTURE

Our framework contains 11 layers in total, including 3 convolutional layers, 3 residual layers, 2 max-pooling layers, 2 fully connected layers, and 1 flatten layer. In the residual layer, the building block contains one layer. At the beginning, the training samples are down-sampled to a fixed size of 70×70 after data augmentation. The batches are fed into the a convolutional layer, which is followed by a residual layer, and both of them contain 8 feature maps whose kernel size is 5×5. Such block with one convolutional layer and one residual layer is considered as a MiCoRe layer. The output of the residual layer is fed into a second convolutional layer with the same kernel size as the above ones, but the number of feature maps is extended to 16. A max-pooling layer, whose size is 2×2 is then implemented with a stride of 2 pixels. As a result, the size of the 16 feature maps becomes 35×35 before the second residual network comes in. The output of the second residual network is sent to another set of MiCoRe block with the same parameters as the first two layers except that the kernel size is 3×3. Another max-pooling layer is implemented with the same configuration as the first one before the output is fed to the flatten layer. Finally, the flatten layer is followed by two fully connected layers, each of which has 512 neurons. The softmax technique is utilized after the second fully connected layer. The ReLU activation function is implemented after all the convolutional layers, residual layers and the first fully connected layer.

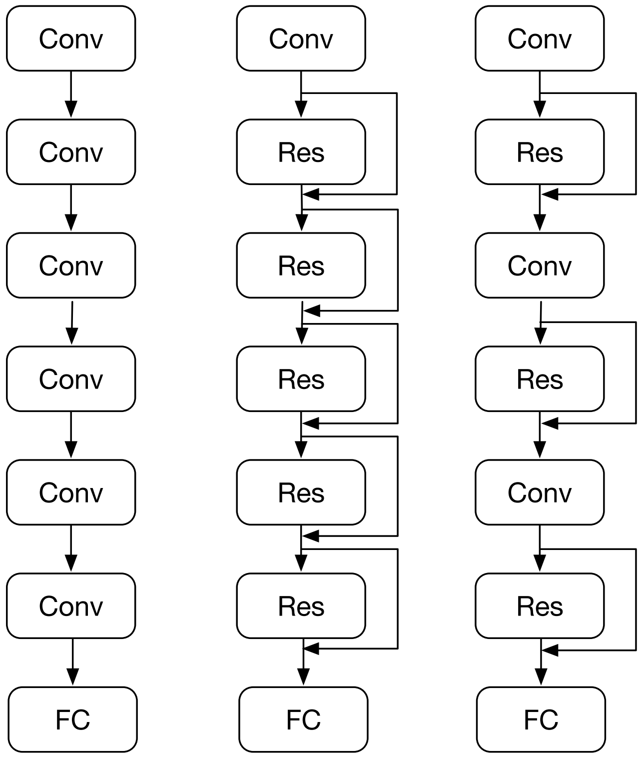


FIGURE 4. **An illustration of the pure convolutional neural network (Left), residual network (Middle), and proposed MiCoRe-Net (Right).**

1) MIXED CONVOLUTIONAL AND RESIDUAL NETWORK (MiCoRe-Net)

In the past few years, convolutional neural network as shown in Fig. 4 has achieved great success in the field of computer vision, but it is difficult to further optimize it as the performance reaches a bottleneck [23]. Deep residual network as shown in Fig. 4 is a successful improvement which breaks through the limitation of saturation when training a plain convolutional neural network. However, there exists the evidence indicating that when the architecture is relatively shallow, which (comparing with those containing hundreds of layers that works on extremely large datasets with thousands of classes and millions of samples, such as ImageNet [24]), residual network is not easy to train comparing with CNN (See Section 4.5).

Based on the above concerns, we propose a novel architecture called MiCoRe-Net, in which convolutional layers and residual layers are placed one after another to form a mixed convolutional and residual network as shown in Fig. 4. The MiCoRe-Net with one convolutional layer followed by a residual layer can obtain the advantages of both kinds of architechtures. It can not only learn faster as CNN, but overcome the saturation issue caused by the plain CNN, which is the superiority of residual network. In this study, we implement a three-stacked-layer MiCoRe-Net, which contains 3 convolutional layers and 3 residual layers, to deal with the eye recognition task. Details of the performance of the MiCoRe-Net are provided in Section 4.5.

B. DATA AUGMENTATION

Unlike feature extraction approaches with local descriptors, deep neural networks rely on the diversity of the representation of the features. Compared with traditional classifiers, deep neural networks take the advantage of its ability on learning various features from training samples, rather than obtaining a decision boundary simply. Data augmentation plays a very important role for a deep learning based eye recognition system, since the raw images are captured in a variety of situations, such as different distances, angles etc. We set the data augmentation strategies, including rotations, cropping, rotation after cropping, and the combination of them to investigate their influence on the system performance. We select a number of parameters for each strategy and compare the results in every situation.

C. ReLU ACTIVATION FUNCTION

Traditional activation functions such as sigmoid as shown in Eq. 1 and tanh as shown in Eq. 2 are widely used in neural networks. However, Rectified Linear Units (ReLU) becomes more and more popular on large datasets since convolutional neural network has been created in recent years [25]–[27].

$$f(x) = \frac{1}{1+e^{-x}} \tag{1}$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2}$$

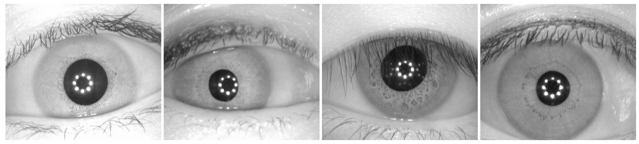
The ReLU function is suitable for forward/back propagation because of its concise form as shown in Eq. 3, compared with the complex operations of sigmoid and tanh. Moreover, the ReLU function, which holds the feature of non-saturating nonlinearity, is reported to run several times faster during the training process than the sigmoid and tanh with saturating nonlinearities. Although ReLU has certain disadvantages, for example, the units might be fragile if learning rate is high, or the gradient will lose on negative values, and its faster learning speed is beneficial for the acquisition of better performance, which is what we concern the most. Therefore, we follow the most state-of-art deep learning algorithms to adopt ReLU as the activation function.

$$f(x) = \max(0, x) \tag{3}$$

IV. EXPERIMENTAL RESULTS

A. SETUP

Since very few studies focus on the eye recognition, the potential data sources come from the following two kinds of database: (1) face database, and (2) iris database. The former one usually contains face images, from which the eye parts can be segmented. But because the eyes only occupy a very small region of the whole face, the resolution of the samples is a problem. Therefore, we use the iris database that contains the complete eye part as the dataset for eye recognition. In this study, the CASIA-Iris-IntervalV4 [28] and UBIRIS.v2 [29] iris datasets are used for the verification of the proposed algorithm.



(a)

(b)

FIGURE 5. Samples from the two databases used in our experiments. (a) The CASIA-Iris-IntervalV4 database. (b) The UBIRIS.v2 database.

CASIA-Iris-IntervalV4 is one of the most widely used standard database for iris recognition, and the samples from it contain sufficient regions around the eye. The database is with 249 classes, each of which has samples from the left and right eyes, and both contain 1 to 10 samples. To assure the validity and scalability of the research, we pick out the classes that contain at least 5 samples from either left or right eyes. If both of them meet the requirement, we randomly choose one to form the dataset we use in this experiment. We randomly choose one sample from each class as the testing set, and the rest as the training set. With this strategy, our subset contains 218 classes, with 1,346 training samples and 218 testing ones.

The UBIRIS.v2 is one of the most famous noisy color iris datasets, which is captured in non-cooperative situations. It contains 2 sessions and 261 subjects, with both left and right eyes which is 522 in total, and the number of samples is 11,102. The images are collected in the visible wavelength with different distances, i.e., in the range of four to eight meters, and on-the-move. The dataset contains not only frontal perspective samples, but left and right off-angle ones as well. Therefore, it covers a variety of samples and can test robust biometric recognition systems with efficiency. We extract the left eye samples from session one as the subset used in our experiment. Therefore, there are 258 classes which contain 3,866 samples that meet the requirements. We pick out the most extreme sample, namely, the one captured at a distance of four meters in a right off-angle view from each class as the testing set, and the rest as the training set. The images from the UBIRIS.v2 dataset are converted to 8-bit gray-scale ones. Images from both datasets are resized to a resolution of 70×70 which is smaller than the resolution of the iris (from CASIA-Iris-IntervalV4: 320×240, and UBIRIS.v2: 800×600) before feeding to the deep neural network.

The MiCoRe-Net are implemented with Tensorflow (Version 0.9.7) on a workstation with a Titan X (Pascal) GPU. For the purpose of performance comparison, the SVM and kNN approaches are implemented in Matlab (Version 2015b) with the provided toolbox.

TABLE 1. The performances with different data augmentation strategies.

Index	Rotation	Cropping°	Cropping interval†	Rotation after Cropping	No. of samples	Accuracy
Database: CASIA-Iris-IntervalV4						
1	None	None	NA	None	1,346	33.03%
2	$\pm15°$	None	NA	None	4,038	56.88%
3	$\pm5°$, $\pm15°$, $\pm30°$	None	NA	None	9,422	64.68%
4	None	10	5	None	13,460	65.14%
5	None	5,10,20,30	5	None	118,448	91.28%
6	$\pm5°$, $\pm15°$, $\pm30°$	5,10,20,30	5	None	126,524	93.58%
7	None	5,10,20,30	5	$\pm10°$	352,652	97.71%
8	$\pm5°$, $\pm15°$, $\pm30°$	5,10,20,30	5	$\pm10°$	360,728	**99.08%**
9	None	5,10,20,30,40,50	5	None	390,340	97.25%
10	$\pm5°$,$\pm10°$,$\pm15°$,$\pm30°$	5,10,20,30,40	5	$\pm5°$,$\pm10°$,$\pm15°$‡	940,854	**99.08%**
Database: UBIRIS.v2						
11	None	None	NA	None	3,608	9.30%
12	$\pm15°$	None	NA	None	10,824	22.48%
13	$\pm5°$, $\pm15°$, $\pm30°$	None	NA	None	25,265	50.78%
14	None	15	15	None	18,040	55.81%
15	None	15,30,60,90	15	None	317,504	91.47%
16	None	15,30,60,90	10	None	613,360	96.12%
17	$\pm5°$, $\pm15°$, $\pm30°$	15,30,60,90	15	None	339,152	91.47%
18	None	15,30,60,90	15	$\pm10°$	411,312	94.57%
19	$\pm5°$, $\pm15°$, $\pm30°$	15,30,60,90	15	$\pm10°$*	432,960	95.74%

° This value listed in this column is used for cropping the image with a filter window whose width and height is that smaller than the training sample. For example, suppose the resolution of the training sample is 320×240. A cropping parameter of 20 means that we use a filter window, whose size is 300×200 to crop the training sample.

† This value indicates how many pixels we move along the width and height between two cropping operations.

‡ The rotation operation is not implemented on the sub-images with a cropping parameter of 40.

* The rotation operation xxis not implemented on the sub-images with cropping parameters of 60 and 90.

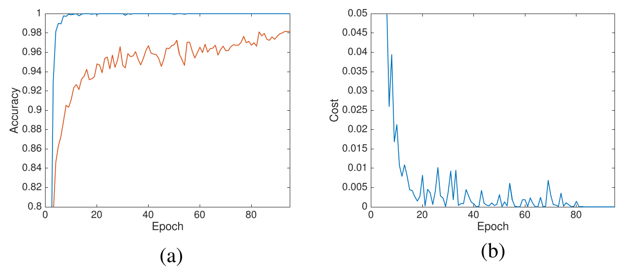


FIGURE 6. Overall training and testing accuracy, and the cross entropy cost. (a) Training and testing accuracy. (b) The cost.

B. DETAILS OF THE TRAINING PROCEDURE

We first investigate the relationship between the training accuracy, testing accuracy, and cost on the CASIA-Iris-IntervalV4 dataset. The learning rate is set to 10^{-4}, and the cross entropy is used for calculating the cost. From the result shown in Fig. 6, we discover that the training accuracy gets saturated much faster than the testing accuracy (100% vs. around 95%), and overfitting problems do exist to some extent. However, the cost remains at a relative high level for a long period of time. Cost reduction makes a contribution to the system performance, as the testing accuracy increases gradually. From Fig. 6(b) we find that after 80 epochs of training, the cost becomes to be stable, which is less than 10^{-8} in our case. Based on the above results, we train the neural network for 100 epochs, which is a little more than the starting point from which the cost becomes stable, in the following experiments.

C. EVALUATION ON DATA AUGMENTATION STRATEGIES

In this sub-section, we investigate the influences of the system performance on both datasets via different data augmentation strategies, including rotation, cropping, rotation after cropping, and the combination of them whose results are shown in Table 1.

We can see that without data augmentation, the performance is quite unsatisfied which are 33.03% and 9.30%, respectively. As different kinds of data augmentation strategies are implemented, the accuracies gradually increase. We find out that the performance and the number of training samples approximately have a logarithmic relationship after data augmentation, unless one or more strategies are largely abused, for example, in Index 9 in Table 1, or the performance has reached a bottleneck, for example, in Index 10 in Table 1. The best performance acquired on the CASIA-Iris-IntervalV4 and UBIRIS.v2 datasets are 99.08% and 96.12%, respectively.

Further analysis indicate that rotation operations make much less contribution than cropping operations to the performance as shown in Index 5 and 6 in CASIA-Iric-IntervalV4, or even no contribution at all as shown in Index 15 and 17 in UBIRIS.v2. By looking at the samples in the datasets we discover that the images are seldom captured with various degrees of rotation, but on the other hand, a lot of them are centered at different locations, which more or less explains why cropping works much better than rotation. Another thing that worths noting is that the cropping size and interval also has an influence to the performance. It is obvious that different cropping sizes and intervals perform differently on different datasets. A small cropping window with a large interval leads to higher accuracy in UBIRIS.v2 than CASIA-Iris-IntervalV4. This is reasonable since the resolution of the samples in UBIRIS.v2 is 800×600, which is nearly three times of the width and height of those in CASIA-Iris-IntervalV4 which are 320×240. Cropping a large image with a window is to retain most of the useful information, but is useless for increasing the performance. Correspondingly, the cropping parameters in UBIRIS.v2 are also three times of those in CASIA-Iris-IntervalV4, which perfectly explains the simulation result.

D. PERFORMANCE COMPARISON WITH CLASSIC APPROACHES

To compare the performance between our proposed approach with classic ones, we implement several traditional classifiers, namely, the kNN and SVM classifier, and plain convolutional neural network (CNN) on the CASIA-Iris-IntervalV4 dataset. To control the training time in a tolerable range, we take a data augmentation strategy that limits the number of training samples less than 200,000. Rotations of $\pm 5°$, $\pm 15°$, $\pm 30°$, and cropping with windows whose sizes are 5,10,20,30 smaller than the width and height of the raw image (with a five-pixel move along the width and height each time) are implemented before training the classifiers/neural networks. The number of images in the training set for this experiment is 126,524.

For the SVM classifier, we use a Gaussian kernel function. The number of neighbors is set from 2 to 20 when implementing the kNN classifier, and the best performance is presented (k=5). The architecture of the CNN approach is set to be the same as the MiCoRe-Net, which is described in Section 3, except that the residual layers are substituted with convolutional ones, remaining other parameters, such as number of layers, number of feature maps, kernel sizes, etc. unchanged.

TABLE 2. Performance comparison with different methods.

Approach	Accuracy
Support Vector Machine	71.10%
k-Nearest Neighbor	85.78%
Convolutional Neural Network (CNN)	91.74%
MiCoRe-Net (Proposed)	**93.58%**

Table 2 presents the performance of each method. We can see that both deep learning approaches perform much better than traditional classifiers. Specifically, the proposed method with MiCoRe-Net overcomes the one with convolutional neural network (93.58% vs. 91.74%).

E. PERFORMANCE EVALUATION WITH DIFFERENT ARCHITECTURES

We verify the effectiveness of our proposed MiCoRe-Net by comparing the performance with other architectures. The candidates include 3-layer convolutional network and MiCoRe-Net, and 6-layer convolutional network, residual network, and MiCoRe-Net. The results shown in Table 3 indicate that our proposed MiCoRe-Net achieves the highest accuracy on the CASIA-Iris-IntervalV4 dataset (with the same data augmentation strategy as Index 8 in Table 1).

To further dig out the reason behind the success of MiCoRe-Net, we further analyze the training accuracy as shown in Fig. 7 and the cross entropy cost as shown in Fig. 8 over iterations on every architecture. With considerations on both of these results and the highest accuracies as shown in Table 3, we discover that pure convolutional networks converge faster than other architectures, but their accuracies remain saturated as the neural networks get deeper, since they are faced with the problem of saturation. Moreover, with the analysis of the cost at the very end of the training procedure, we find out that the cost of pure residual network is still oscillating, whereas the MiCoRe-Net has converged for a long period of time as shown in Fig. 9. Such result indicates that pure residual network is not easy to train when the framework is relatively shallow. Based on the above experimental results, we can conclude with confidence that the MiCoRe-Net takes the advantages of fast convergence

TABLE 3. Performance comparison with different architectures.

Index	Approach	Accuracy
1	Conv → Conv → Conv	98.17%
2	Conv → Res → Conv	97.71%
3	Conv → Conv → Conv → Conv → Conv → Conv	98.17%
4	Conv → Res → Conv → Res → Conv → Res	**99.08%**
5	Conv → Res → Res → Res → Res → Res	97.25%

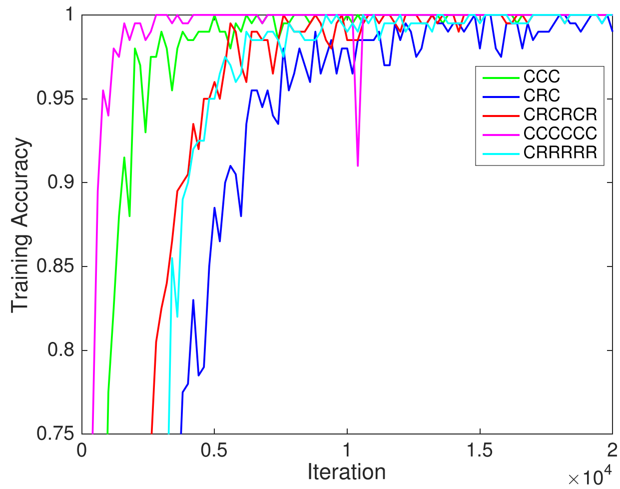


FIGURE 7. Training accuracies on different architectures.

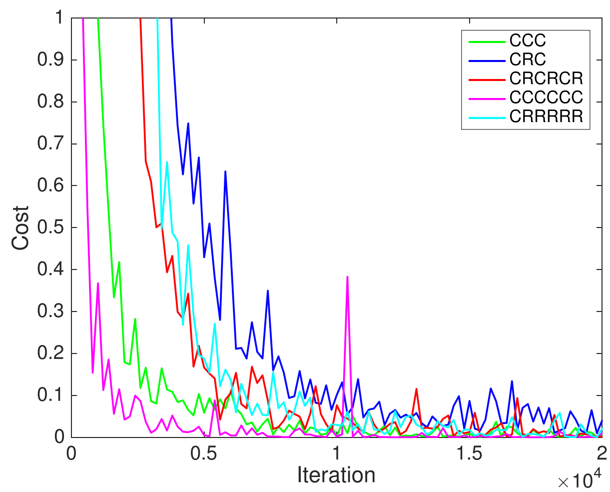


FIGURE 8. Costs on different architectures.

TABLE 4. The performances with different data augmentation strategies on CASIA-Iris-Thousand.

Database: CASIA-Iris-Thousand						
Index	Rotation	Cropping	Cropping interval	Rotation after cropping	No. of samples	Accuracy
1	None	None	NA	None	9,000	32.3%
2	$\pm 15°$	None	NA	None	27,000	55.2%
3	$\pm 5°$, $\pm 15°$, $\pm 30°$	None	NA	None	63,000	65.8%
4	None	10	5	None	90,000	60.7%
5	None	20	10	None	90,000	65.7%
6	None	10,20,40,60	10	None	792,000	88.0%
7	$\pm 5°$, $\pm 15°$, $\pm 30°$	20,20,40,60	10	None	846,000	**88.7%**
8	$\pm 5°$, $\pm 15°$, $\pm 30°$	5,10,20,30	5	$\pm 10°$	2,412,000	83.9%
9	$\pm 5°$, $\pm 15°$, $\pm 30°$	20,20,40,60	10	$\pm 10°$	2,412,000	88.2%

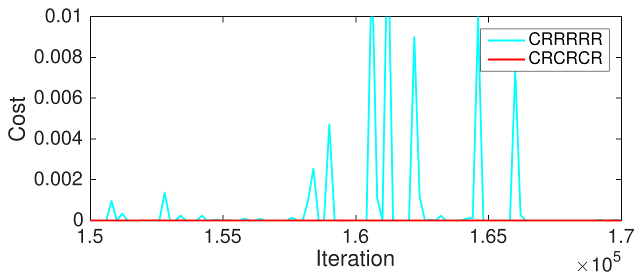


FIGURE 9. Costs on the ResNet and MiCoRe-Net architectures at the very end.

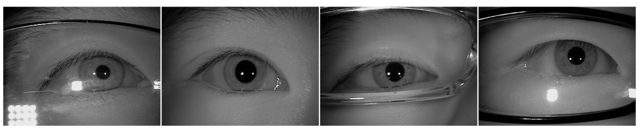


FIGURE 10. Samples from the CASIA-Iris-Thousand database.

from convolutional network, and of the non-saturation feature from residual network. The MiCoRe-Net is the best choice for the eye recognition.

F. EXAMINATION IN HARSH ENVIRONMENT

In order to test the robustness of the our MiCoRe-Net architecture on the eye recognition task, we utilize the most challenging database, namely, CASIA-Iris-Thousand as shown in Fig. 10 for the verification purpose. CASIA-Iris-Thousand contains 1,000 subjects, and each subject has 10 images for both left and right eyes, so the total number of images is 20,000. The samples are collected under a variety of external and internal conditions, such as different illuminations, or interferences from eyeglasses and specular reflections [30].

Following the idea in Section 4.3, we test the performance of the proposed framework with different data augmentation strategies as shown in Table 4. We find that the performance is much better when implementing a cropping move interval of 10, rather than other values. This is reasonable and consistent with our analysis in Section 4.3, since the resolution of the samples in CASIA-Iris-Thousand is 640×480, which is roughly twice of the width and height as those in CASIA-Iris-Thousand, and 2/3 as those in UBIRIS.v2. The highest accuracy acquired on this database is 88.7%, which the rotation and cropping parameters are set to $\pm 5°$, $\pm 15°$, $\pm 30°$, and 10, 20, 30, 60 (with a move of 10 pixels), respectively.

From the results we also find that, as the accuracy achieves around 88%, the performance on longer benefits from augmentation operations, even though the strategies are not abused. The straightforward reason might be that we limit the augmentation strategies only to rotation, cropping, and rotation after cropping, which cannot cover the diversity of the samples in the CASIA-Iris-Thousand database. Moreover, since the CASIA-Iris-Thousand database contains much more samples than CASIA-Iris-IntervalV4 and the UBIRIS.v2, a six-layer MiCoRe-Net may not be deep enough to train a satisfactory network. Due to the space limitation of this paper, we leave the potential improvements against the above two issues as a future work.

V. CONCLUSION

In this paper, we propose a novel eye recognition framework that overcomes the defects in the traditional iris recognition system. Such framework removes the feature extraction and pro-processing procedures, which increase the learning complexity, and are influenced by a number of issues such as the low resolution of the sample, and the interferences from eyelid, eyeglasses, etc. To apply the eye recognition problem properly, we propose a new deep learning architecture called mixed convolutional and residual network (MiCoRe-Net), which inserts a plain convolutional layer between every two residual layers. The MiCoRe-Net takes both advantages of fast learning from convolutional neural network, and non-saturation features from residual network. Also, we study how different data augmentation strategies have influences on the system performance, and it turns out that such eye recognition framework relies on a proper augmentation strategy to work well.

Experimental results show that our approach achieves highest accuracy within a proper number of training epoch/iteration. With the MiCoRe-Net, the best accuracy rates on the CASIA-Iris-IntervalV4 and the UBIRIS.v2 datasets are 99.08% and 96.12%, respectively, which not only exceeds the performances of traditional methods and other previously proposed deep learning based architectures, but is competitive with the most state-of-art iris recognition algorithms as well.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Tieniu Tan and Dr. Hugo Proenca for providing permissions of the access of the datasets CASIA-Iris-IntervalV4, CASIA-Iris-Thousand and UBIRIS.v2, respectively. And we gratefully acknowledge the support of NVIDIA Corporation with the donation of the TITAN X GPU used for this research.

REFERENCES

[1] J. Daugman, "How iris recognition works," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, Jan. 2004.

[2] C. Li, W. Zhou, and S. Yuan, "Iris recognition based on a novel variation of local binary pattern," *Vis. Comput.*, vol. 31, no. 10, pp. 1419–1429, 2015.

[3] K. Roy, P. Bhattacharya, and R. C. Debnath, "Multi-class svm based iris recognition," in *Proc. Int. Conf. Comput. Inf. Technol.*, 2008, pp. 1–6.

[4] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast iris segmentation for iris biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1670–1684, Sep. 2009.

[5] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148–1161, Nov. 1993.

[6] A. D. Rahulkar, L. M. Waghmare, and R. S. Holambe, "A new approach to the design of hybrid finer directional wavelet filter bank for iris feature extraction and classification using *k-out-of-n*: A post-classifier," *Pattern Anal. Appl.*, vol. 17, no. 3, pp. 529–547, 2014.

[7] B. Patel, R. P. Maheshwari, and B. Raman, *Compass Local Binary Patterns for Gender Recognition of Facial Photographs and Sketches*. Amsterdam, The Netherlands: Elsevier, 2016.

[8] L. Liu, P. Fieguth, M. Pietikäinen, and S. Lao, "Median robust extended local binary pattern for texture classification," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 2319–2323.

[9] P. Karczmarek, A. Kiersztyn, W. Pedrycz, and M. Dolecki, "An application of chain code-based local descriptor and its extension to face recognition," *Pattern Recognit.*, vol. 65, pp. 26–34, May 2017.

[10] C. Ding, J. Choi, D. Tao, and L. S. Davis, "Multi-directional multi-level dual-cross patterns for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 518–531, Mar. 2016.

[11] P. Zhou, Y. Peng, J. Shen, B. Zhang, and W. Yang, "Local dual-cross ternary pattern for feature representation," in *Proc. Chin. Conf. Biometric Recognit.*, 2016, pp. 600–608.

[12] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using local phase quantization," in *Proc. Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.

[13] S. Bhattacharya, A. Dasgupta, and A. Routray, "Robust face recognition of inferior quality images using local Gabor phase quantization," in *Proc. Technol. Symp.*, 2017, pp. 294–298.

[14] F. Jafari and H. R. Kanan, "Disguised face recognition by using local phase quantization and singular value decomposition," *J. Comput. Robot.*, vol. 9, no. 1, pp. 51–60, 2016.

[15] K. V. Arya and G. K. Gupta, "Robust iris identification system using local descriptors," in *Proc. Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2014, pp. 744–748.

[16] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2049–2058, Nov. 2015.

[17] S. Zhan, Q.-Q. Tao, and X.-H. Li, "Face detection using representation learning," *Neurocomputing*, vol. 187, pp. 19–26, Sep. 2016.

[18] N. Liu, M. Zhang, H. Li, Z. Sun, and T. Tan, "Deepiris: Learning pairwise filter bank for heterogeneous iris verification," *Pattern Recognit. Lett.*, vol. 82, no. 2, pp. 154–161, 2015.

[19] Z. Zhou, Y. Du, N. L. Thomas, and E. J. Delp, "Multimodal eye recognition," *Proc. SPIE*, vol. 7708, p. 770806, Apr. 2010.

[20] Z. Zhou, E. Y. Du, N. L. Thomas, and E. J. Delp, "A comprehensive multimodal eye recognition," *Signal Image Video Process.*, vol. 7, no. 4, pp. 619–631, 2013.

[21] Z. Zhou, E. Y. Du, C. Belcher, N. L. Thomas, and E. J. Delp, "Quality fusion based multimodal eye recognition," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2012, pp. 1297–1302.

[22] B. Cyganek and S. Gruszczyíski, "Hybrid computer vision system for drivers' eye recognition and fatigue monitoring," *Neurocomputing*, vol. 126, pp. 78–94, Feb. 2014.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 770–778.

[24] J. Deng, W. Dong, R. Socher, and L. J. Li, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[26] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

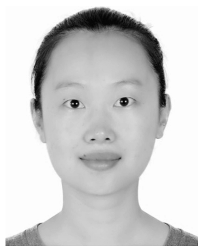
[27] P. Qin, W. Xu, and J. Guo, "An empirical convolutional neural network approach for semantic relation classification," *Neurocomputing*, vol. 190, pp. 1–9, May 2016.

[28] *The CASIA-Iris-IntervalV4 Database*. Accessed: Jan. 2010. [Online]. Available: http://biometrics.idealtest.org/

[29] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1529–1535, Aug. 2010.

[30] *The CASIA-Iris-Thousand database*. Accessed: Jan. 2010. [Online]. Available: http://biometrics.idealtest.org/dbDetailForUser.do?id=4

ZI WANG is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville. His research interests include deep learning, reinforcement learning, and computational biology.



CHENGCHENG LI is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville. Her research interests include deep learning, pattern recognition, and generative models.

HUIRU SHAO is currently pursuing the master's degree with the School of Information Science and Engineering, Shandong Normal University, China. Her research interests include gaze tracking and gaze information analysis.

JIANDE SUN received the Ph.D. degree in communication and information system from Shandong University, Jinan, China, in 2005. He has been the Visiting Researcher with the Technical University of Berlin, University of Konstanz, and Carnegie Mellon University, and a Post-Doctoral Researcher with the Institute of Digital Media, Peking University, Beijing, China, and with the State Key Laboratory of Digital-Media Technology, Hisense Group. He has published over 60 journal and conference papers. He has co-authored two books. His current research interests include multimedia signal processing, analysis and applications, multimedia security, gaze-based HCI, and image/video.