Towards More Accurate Iris Recognition using Dilated Residual Features

Kuo Wang, Ajay Kumar

***Abstract*—Iris recognition has emerged as the more accurate, convenient and low-cost biometric approach to authenticate human subjects. However, the accuracy offered from currently popular iris recognition algorithms is below the expectations from the community and therefore researchers have recently focused their attention on deep learning based methods. This paper investigates a new deep learning based approach for the iris recognition and attempts to improve the accuracy using a more simplified framework to more accurately recover the representative features. We consider residual network learning with dilated convolutional kernels to optimize the training process and aggregate contextual information from the iris images. Such an approach also alleviates the need for the down-sampling and up-sampling layers, which not only results in a simplified network but also results in outperforming matching accuracy over several classical and state-of-art algorithms for iris recognition, *i.e.*, further improvement in equal error rates by 7.14%, 10.7% and 27.4% on three test databases. Our reproducible experimental results presented in this paper on three publicly available datasets illustrate outperforming results and validate the usefulness of our approach.**

***Index Terms*—Iris recognition, personal identification, biometrics, deep learning.**

I. Introduction

IRIS recognition has emerged as a reliable personal identification method with multifaceted applications in border security control, banking, law enforcement, public welfare distribution and accounting [1][2]. Iris patterns are known to be unique among humans, even among the identical twins, and recovered using low-cost imaging that uses near-infrared illumination to capture textured like patterns between the pupil and sclera. The success of iris recognition largely depends on the accurate extraction of features from the segmented iris patterns. Quite a lot of effort in the past decades has been devoted to the recovery of robust and more accurate iris features [3]. In addition to the Gabor filter based *IrisCode* [4] representation of features, researchers have investigated alternative methods for iris recognition using discrete cosine transform (DCT) [5], discrete Fourier transform (DFT) [6], Radon transform [7] which competitive alternative for iris recognition.

In recent years, deep convolutional neural network based learning methods have gained tremendous success in a range of computer vision and pattern recognition applications. Deep learning based strategies have shown outperforming results for a range of biometrics identification problems, e.g., footstep recognition [8], signature verification[9], face recognition [10], periocular recognition [11], etc. This has also motivated researchers to investigate the strengths of deep convolutional neural networks for iris recognition with interesting insights [12]. Despite encouraging results and promises, the matching accuracy offered from such iris recognition methods is still below the expectations or the potential from this biometric. Therefore further research efforts are required to not only improve the matching accuracy but also to reduce the complexity of network and learning. Deep learning based methods can also offer performance improvement for the cross-spectral iris recognition problem and is related area requiring further research. This paper attempts to further advance iris recognition research using deep learning based approach and provides extensive evaluation of the proposal using publicly available databases.

A. Related Work

One of the most classical and widely adopted approaches for the automated iris recognition is proposed by Daugman [4]. This approach uses band-pass spatial filters, i.e., Gabor filters, to generate the spatial features from the segmented iris images representing the texture. These features are then binarized to generate bar-code like representation which is referred to as the *IrisCode*. The Hamming distance between two *IrisCodes* templates is used as the dissimilarity score to establish the identity of an individual. Such an approach can also be implemented using one-dimensional log-Gabor filters, as in [3], and results in more efficient iris texture feature extraction as compared to those using two-dimensional Gabor filters in [4]. Another frequency-domain approach introduced in [5] and incorporates discrete cosine transform (DCT) coefficients for analyzing spectral contents in image block regions. Such DCT-based feature representation is also binarized for efficient matching and generates promising results. A more promising method in [13] incorporates multi-lobe differential filters (MLDFs) to encode multi-scale and multi-orientation feature representation for the normalized iris templates. The MLDFs are based on ordinal measurements and such measurements can be more efficiently performed using the differences of normal vectors detailed in [15]. Matching iris data acquired from near-infrared images with those from the visible illumination data is widely considered a challenging problem with a range of applications in e-security and surveillance. Among several attempts in the literature, the method in [14] has shown to illustrate the comparatively superior performance and serves a judicious baseline for evaluating the success of other methods for this problem.

Unlike the popularity of deep learning for various computer vision tasks, especially for face recognition, the literature so

TABLE I: Summary of earlier work and performance on same public iris image databases.

Ref.	Method Summary	Employed Databases	Cross-Dataset Evaluation	Recognition Rate (@FAR=0.001)	EER
Masek [3]	Gabor filter based *IrisCodes*.	1) ND-IRIS-0405 Iris Image Dataset (ICE2006) 2) CASIA Iris Image Database V4-distance 3) WVU Non-ideal	No	1) 96.7% 2) 79.3% 3) 88.4%	1) 1.88% 2) 7.71% 3) 6.82%
Sun *et al*. [13]	Ordinal filters based feature encoding.	1) ND-IRIS-0405 Iris Image Dataset (ICE2006) 2) CASIA Iris Image Database V4-distance 3) WVU Non-ideal	No	1) 96.8% 2) 83.0% 3) 90.1%	1) 1.74% 2) 7.89% 3) 5.19%
Zhao *et al*. [14]	Fully convolutional network to encode iris features.	1) ND-IRIS-0405 Iris Image Dataset (ICE2006) 2) CASIA Iris Image Database V4-distance 3) WVU Non-ideal	Yes	1) 97.1% 2) 84.1% 3) 94.3%	1) 1.40% 2) 5.50% 3) 2.63%
Ours	Dilated and residual learning using deep convolutional neural network.	1) ND-IRIS-0405 Iris Image Dataset (ICE2006) 2) CASIA Iris Image Database V4-distance 3) WVU Non-ideal	Yes	1) 97.7% 2) 87.5% 3) 96.1%	1) 1.30% 2) 4.91% 3) 1.91%

far has not yet fully explored its potential for iris recognition. There has been very little attention to exploring the iris recognition solutions using deep learning. An interesting attempt appears with DeepIrisNet in [16], which represents preliminary investigation using a deep-learning based approach for generalized iris recognition problem. This work is essentially a direct application of typical convolutional neural networks (CNN) and inception CNN for iris texture patterns. Minaee et al. [17] use the *VGGNet* to extract deep convolutional features and a two-class support vector machine (SVM) classifier for iris recognition. Another interesting work in this direction appears in [18] and has attempted to investigate the deep belief network (DBN) for iris recognition. Its core component is the optimal Gabor filter selection, while the DBN is a relatively simplified application on the *IrisCode* without iris-specific optimization. Tang et al. [19] have also investigated a lightweight CNN to extract feature maps from iris images. This work also uses ordinal measurements [13] along with the iris mask details, for the iris matching. In [20], Nguyen et al. have investigated several pre-trained CNN models, including AlexNet, VGGNet, InceptionNet, ResNet, and DenseNet, to extract off-the-shelf CNN features for more accurate iris recognition. Reference [21] also investigates performance for the iris recognition using pre-trained *VGGNet* and *ResNet*, and provides promising results. Reference [14] details a new deep learning based framework, referred to as the *UniNet*, which generates feature templates from the fully convolutional network (FCN) for more accurate iris recognition. This network incorporates the bits shifting and masks during the generation of match scores and achieves the state-of-the-art accuracy on several publicly available iris images datasets.

B. Our work

Our work is motivated to further advance the iris recognition capabilities using deep learning based approaches. We specifically focus on generating a robust representation of iris features by incorporating superior feature extraction network that uses dilated convolution kernels to address frequently observed deformations between the matched iris patterns. Our network benefits from the residual learning while the key reason for its simplicity lies in the usage of dilated kernels. In addition to the cross-database performance evaluation, we also broaden the scope of our investigation and ascertain the cross-spectral iris matching capabilities from the network introduced for the generalized iris recognition capabilities.

Two key benefits from our iris recognition approach can be summarized as the following. Firstly, the overall matching performance is enhanced using the within dataset and cross-dataset matching scenario. The dilated convolution kernels employed in the network can support nonlinearly expanding receptive fields without degrading the resolution or coverage. Improvement in the matching accuracy can also be attributed to the usage of residual learning blocks, which can learn the residual information by increasing the depth and enrich the learning capability of the model. The experimental results presented in Section III of this paper, using three publicly available databases, indicate outperforming results and validate the effectiveness of the proposed approach. Table I summarizes the comparative performance from our approach in this paper with other competing methods on three public iris images databases. The performance evaluation presented in Section IV-C also indicate the effectiveness of the approach for cross-spectral iris matching problem. Another benefit from our model, over the earlier work in [14], lies in its simplicity. The dilated convolutional kernel can learn the information from different scale without down-sampling. The *UniNet* in [14] requires parameters from up-sampling layers which are not trainable in CNN. This increases the complexity of the model and the potential for introducing more errors from the trained network. Although the new architecture in this paper uses more layers, as compared with the one in [14], the parameters trained are not increased. This is because the element-wise combination and instance normalization layers from residual network do not include any trained parameters.

Rest of this paper is organized as follows. Section II presents the methodology of our approach and includes the details of the network and training considered in this work. It provides details on the dilated convolutional kernels, feature learning and triplet selection employed to train the network. Section III provides details on experiments, protocols and publicly available databases employed for the within and cross-database performance evaluation. The results in this section also include ablation test results. Section IV provides discussion which also includes the cross-spectral iris matching results, failure cases and other promising deep learning architectures. Key conclusions from this paper are summarized in the last section of this paper.

II. Iris Recognition using Dilated Kernels and Residual Feature Learning

The framework for accurate iris recognition investigated in this work is shown in Fig. 1. Our model refers to the *UniNet* proposed in [14] which is also an important baseline compared to this work. The whole framework includes the *MaskNet* and *FeatNet*, and we first learn the parameters in *MaskNet*. After that, we freeze all those params in *MaskNet* and fine tune the weights in our feature extraction architecture, referred to as Dilated Residual Feature Net (DRFNet), based on the extensive triplet loss [14]. In our testing phase, all the segmented iris images are fed into the network, and it will produce iris templates and respective masks automatically. With the binarized iris templates and corresponding masks, we calculate the hamming distance and use it as the matching score to distinguish the genuine and the imposter.

This work focuses on the improvement in the architecture of DRFNet in order to generate more accurate binary feature maps compared with *FeatNet*. The new architecture of this branch incorporates the dilated convolutional neural networks [22] and residual learning kernels [23], and all the detail settings are shown in Table II. We also optimize the training process with offline triplets selection [24]. The details of all those techniques are introduced in the following sections.

A. Dilated Convolutional Neural Network

Dilated convolutional neural network was recently introduced in [22] to address the semantic segmentation problem. In CNN, the output from the previous layer, say j , is always used as the input for the next layer $(j+1)$. Let us represent $I_j(x,y)$ as the 2-D input for our network and $f_j(x,y)$ as a discrete convolutional kernel. The output feature map from $(j+1)$th layer using conventional convolution operations can be described as:

$$I_{j+1}(x,y) = (I_j * f_j)(x,y) = \sum_m \sum_n I_j(m,n) f_j(x-m, y-n) \tag{1}$$

TABLE II: The details of Dilated Residual Feature Net (DRFNet).

Layer Name	Layer Type	Kernel Size	Output Channel	Dilated Factor
Conv1	Convolution	3×3	16	1
Tanh1	*tanh*	-	16	-
Pre_Conv1	Convolution	1×1	32	1
Imnorm1	Instance normalization	-	32	-
Conv2	Convolution	3×3	32	2
Tanh2	*tanh*	-	32	-
Res2	Elementwise sum	-	32	-
Tanh3	*tanh*	-	32	-
Pre_Conv2	Convolution	1×1	64	1
Imnorm2	Instance normalization	-	64	-
Conv3	Convolution	3×3	64	4
Tanh4	*tanh*	-	64	-
Res3	Elementwise sum	-	64	-
Tanh5	*tanh*	-	64	-
Res	Concatenate	-	112	-
Conv4	Convolution	3×3	1	1
Tanh6	*tanh*	-	1	-

where $*$ is the convolution operator. If the filter f is a dilated kernel and let k be a dilation factor, the dilated convolutional operator $*_k$ can be defined as:

$$I_{j+1}(x,y) = (I_j *_k f_j)(x,y) = \sum_m \sum_n I_j(m,n) f_j(x-km, y-kn) \tag{2}$$

We can observe that the dilation factor k is the key to control receptive field of the convolutional kernel. Without losing the resolution and convergence, this controlling factor can be exponentially increased. In our model, the dilated convolutional kernel is incorporated as:

$$I_{j+1}(x,y) = (I_j *_{2^j} f_j)(x,y) \quad for \quad j = 0,1,...n \tag{3}$$

Our network model uses three scales of dilated convolutional kernels whose receptive fields are illustrated in Fig. 2. It is straightforward to observe that the use of such kernels can process the feature maps from different scales, without increasing the number of training parameters. The up-sampling and down-sampling layers are discarded, which simplifies the network structure. Nonetheless, we can aggregate more information to train the network and introduce less error without the down-sampling layers and up-sampling layers. Incorporating such a network is expected to enhance the matching accuracy as more robust feature maps are generated from the trained network. The dense prediction of features is generated from the feature stack, by aggregating the features from the three different scales, with a concatenation layer as shown in Fig. 1.

B. Residual Learning Blocks

Our discussion in the previous section, using Eq. (2), explains the generation of feature maps using the dilated convolution kernels. The dilated kernel in our iris recognition framework is not learning the feature maps for the next

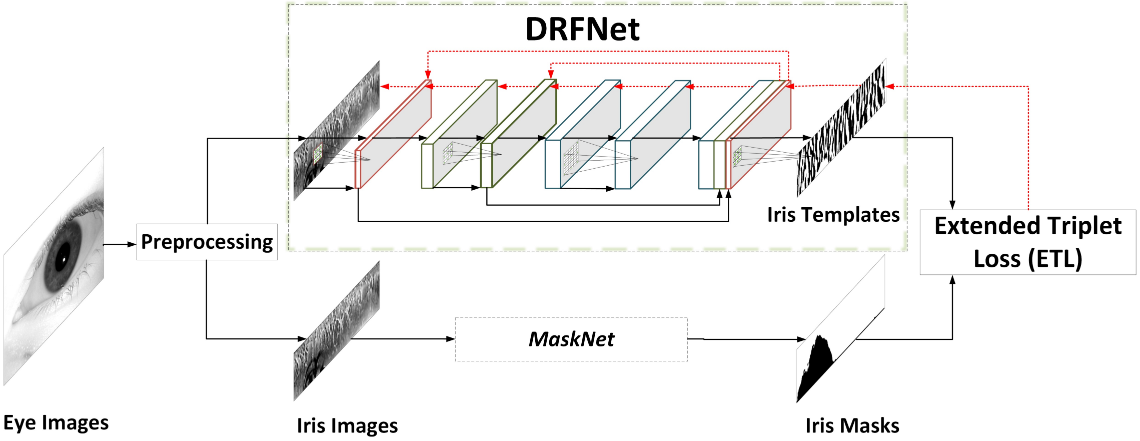


Fig. 1: The framework of accurate iris recognition using the fully convolutional network with dilated residual learning.

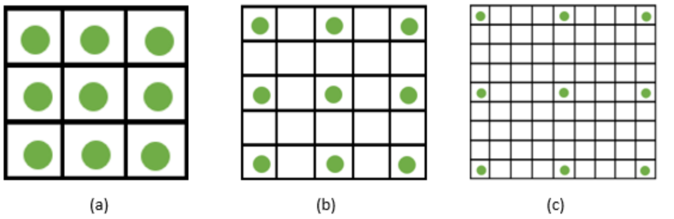


Fig. 2: The dilated convolutional kernels with (a) dilation factor of one and receptive field of 3×3, (b) dilation factor of two and receptive field 5 × 5, (c) dilation factor of four and receptive field of 9 × 9.

layer but the residual information. Deep residual network in [23] was introduced to learn the residual information from input and ease the network training because the classical CNN structure poses difficulties in approximating the identity mapping due to the multiplications from few nonlinear layers. Given a vector input $\mathbf{x}$ for the network layer and its desired output vector $O(\mathbf{x})$, the residual learning aims to learn the residue $R(\mathbf{x})$ for generating the desired output:

$$O(\mathbf{x}) = R(\mathbf{x}) + \mathbf{x} \tag{4}$$

From the Eq. (4), we can find that the input is processed in two different branches. One is for identity mapping, whereas the other one is for the residual information learning. If the identity mapping is an optimal solution, the network could simply train the residue $R(\mathbf{x})$ towards 0. If the residual features are minimizing the loss, we will get a new feature map $O(\mathbf{x})$ from the combination. Compared with the plain network, it does not need more parameters or produce more computational complexity. In our model, we use the dilated convolutional kernel to learn the residual information as introduced in Section II-A, so the output from the dilated residual learning kernels can be represented with Eq. (5).

$$I_{j+1}(x, y) = (I_j *_{2^j} f_j)(x, y) + I_j(x, y) \tag{5}$$

The channel number of outputs $O(\mathbf{x})$ can be different from input $\mathbf{x}$ because we need more feature maps with the increase of network depth. We use a pre-convolutional layer with kernel size 1 to control the number of channels of different layers. We also use instance normalization instead of batch normalization which performs better in the dense prediction. The dilated residual feature learning blocks used in our network is shown in Fig. 3.

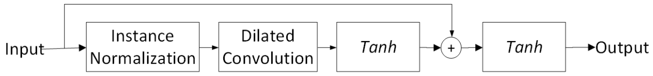


Fig. 3: The dilated residual feature learning block.

C. Triplets Selection

The triplets selection aims to optimize the training process for the triplet network as shown in Fig. 4. Triplet pairs are generated from the combination of an anchor sample, with a positive sample and a negative sample. These pairs are respectively fed into the three network branches, each with same parameters. Each of these network branches correspondingly generates feature maps F_A, F_P and F_N, which are used to compute the extended triplet loss (ETL). The ETL l is defined as follows:

$$l = \frac{1}{M}\sum_{i=1}^{M} max(||F_P - F_A||^2 - ||F_N - F_A||^2 + \gamma, 0) \tag{6}$$

where M is the batch size and γ is a hyperparameter controlling the margin between the anchor-positive distance and anchor-negative distance. It is important to select triplet samples/pairs that can generate non-zero training loss to ensure the effective and efficient training of the network. This means that given F_A we prefer to select corresponding F_P (hard positive) with $argmax_{F_P}||F_P - F_A||^2$ and its similarly F_N (hard negative) such that $argmin_{F_N}||F_N - F_A||^2$. However, it is not feasible to generate the desired triplets by computing the $argmin$ and $argmax$ across the whole training set. In addition, this can also lead to poor training since the mislabelled and/or noisy iris samples (outliers) would dominate the selection of such hard pairs. There are two choices to alleviate such limitations. First is to generate the triplets offline after a few thousands of iteration, using the most recent network model to determine the hard pairs in every training subsets. Second choice is to generate triplets online by selecting the

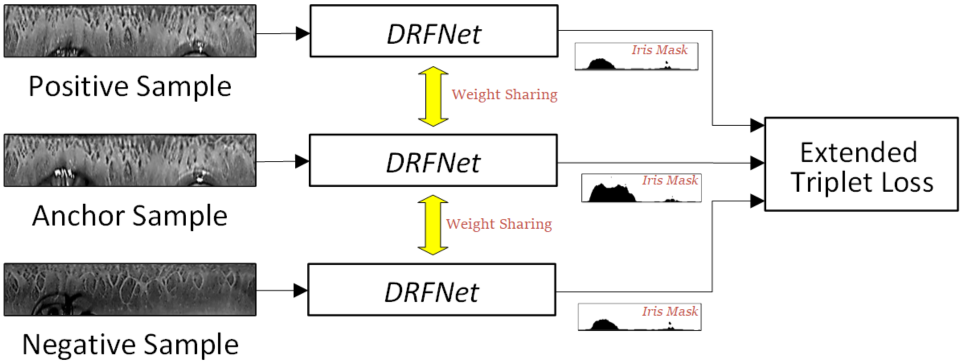


Fig. 4: Training *DRFNet* using triplet architecture.

hard positive and negative samples from a mini-batch, but it can result in a significant amount of computing overhead. In our work, we employ the offline generation by dividing the whole training sets into several parts and compute $argmin$ and $argmax$ within each training subset. All the subjects in each subset are considered to have a meaningful representation of the anchor positive distances. In our experiments, we also find that all anchor positive approach instead of only hard positive is more stable and results in faster convergence of the loss during the training process. Selection of hardest negatives for network training can often lead to a poor local minimum during the training process. To mitigate such limitations, we preferred to select iris image triplets with the following constraints:

$$||F_N - F_A||^2 + \alpha > ||F_P - F_A||^2 > ||F_N - F_A||^2 \qquad (7)$$

Such negative exemplars can be referred to as *semi-hard*, as they are not expected to be outliers because they lie inside the margin α, but still represent challenging samples because their squared distance is less than the anchor positive distance. Since the impostor training samples or the negative matching pairs are much more than the genuine or the positive pairs, we generate triplets from all the genuine pairs, with respective negative samples, which can meet constraints for the optimization during the training process.

III. Experiments and Results

We perform a series of reproducible experiments [25] on three publicly available datasets to evaluate the effectiveness of our proposed framework. We firstly introduce three publicly available iris images datasets employed in this work and our experimental protocols. The comparative results from our approach and other state of the art methods are also detailed in the following section.

A. Databases and Protocols

We employ the following three datasets in our experiments to ascertain our performance improvement. The sample iris images from different datasets are illustrated in Fig. 5.

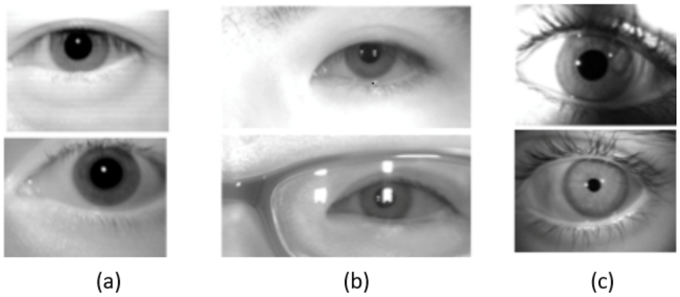


Fig. 5: Sample iris images from (a) ND-IRIS-0405 dataset (b) CASIA.v4-distance dataset (c) WVU Non-ideal dataset.

1) ND-IRIS-0405 Iris Image Dataset (ICE2006): This database [26] includes 64,980 iris samples acquired from 356 subjects using LG 2200 iris biometrics sensor. We employ the first 25 left eye images from all the subjects to train our model and test it with the first 10 right eye images from all the subjects. For the 68 left eye subjects with less than 25 samples, we use all the available images during the training phase. The test set for the performance evaluation therefore generates 14,791 number of genuine match scores and 5,743,130 number of imposter match scores during the performance evaluation.

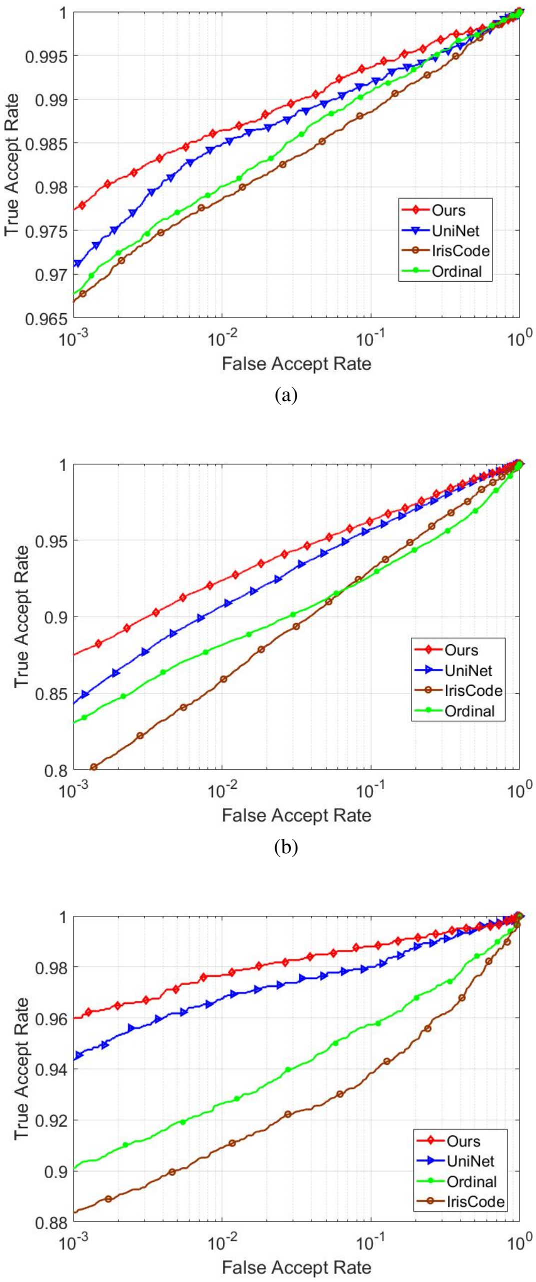
2) CASIA Iris Image Database V4-distance: This subset of database [27] contains 2,446 instances from 142 different subjects. The origin image samples are the upper part of faces including the iris information. We automatically segment the iris region with a publicly available eye detector [28]. The training set comprises all the right eye image samples in this database. The test set for the performance evaluation consists of all the left eye samples and therefore generated 20,702 number of genuine match scores and 2,969,533 number of imposter match scores.

3) WVU Non-ideal Iris Database - Release 1: The WVU non-ideal iris dataset [29] consists of 3,042 iris image samples from 231 subjects. This database includes iris images acquired in real imaging environments, in which samples are also acquired under off-angles, with blur, sensor noise, and the occlusions. The training set in our experiments consists of all the right eye image instances while the test set consists of first 5 left eye instances from all the subjects and this protocol is same as in earlier references [14]. The test set data therefore generates 2,251 number of genuine match scores and

643,565 number of imposter match scores which is used for the performance evaluation.

B. Within-Database Performance Evaluation

In this configuration, we train our network with the ND-Iris-0405 dataset as the initial model. The trained model is further fine-tuned with the training images from the target dataset to generate other models. The model of ND-IRIS-0405 dataset is trained from scratch. Our initial learning rate is set to be 0.01. The total number of iterations is set to 60,000 while our learning policy chose multiple steps with the step size 15,000. The parameter update scheme is stochastic gradient descent (SGD) with momentum while the momentum is 0.9. The ratio of positive pairs and negative pairs in triplets generation is 1:5. During the fine-tuning for the other two datasets, we change the initial learning rate to 0.001.



(a)

(b)

(c)

Fig. 6: Comparative receiver characteristic curve (ROC) results from within dataset matching using (a) ND-IRIS-0405 dataset (b) CASIA.v4-distance dataset (c) WVU Non-ideal dataset.

TABLE III: Summary of equal error rates (EER) from the within dataset performance evaluation.

	ND-IRIS-0405	**CASIA.v4-distance**	**WVU Non-ideal**
IrisCode	1.88%	7.71%	6.82%
Ordinal Filters	1.74%	7.89%	5.19%
UniNet	1.40%	5.50%	2.63%
Ours	1.30%	4.91%	1.91%

TABLE IV: Summary of area under curve (AUC) from the within dataset performance evaluation.

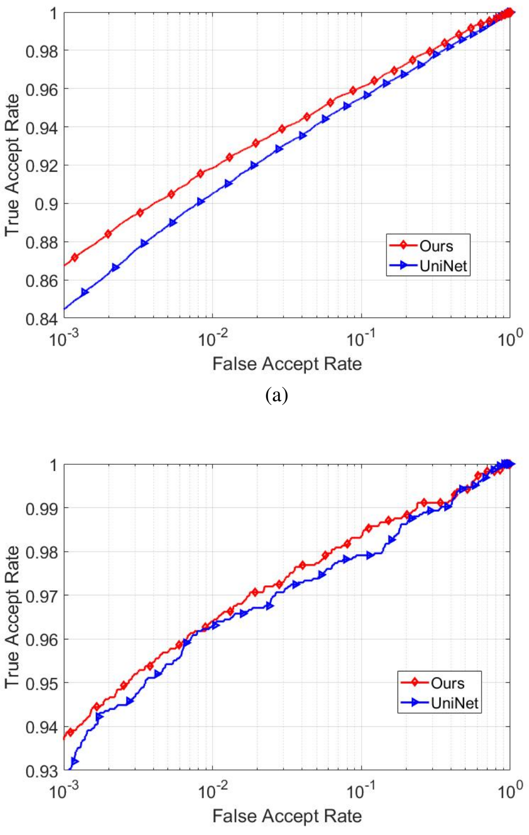
	ND-IRIS-0405	**CASIA.v4-distance**	**WVU Non-ideal**
UniNet	0.9963	0.9819	0.9921
Ours	0.9970	0.9843	0.9941

TABLE V: Comparative evaluation of the network complexity during the performance evaluation.

	Feature extraction time per sample (ms)	**Parameters Number**	**FLOPs**
UniNet	8.93	129,872	883.88M
Ours	8.12	125,264	828.67M

All the models are evaluated by their respective test set samples. We also compare the performance from our approach with three other competing benchmark methods: *IrisCode* [3] is a popular and widely used benchmark for the iris recognition performance evaluation. Instead of Gabor employing the filter, ordinal filters in [13] have also shown a superior performance when employed as an iris features extractor. *UniNet*[14] employs the fully convolutional network to generate binarized feature templates. The comparative experimental results from our approach and the respective benchmark methods are presented in Fig. 6. The equal error rates (EERs) from the respective methods are summarized in Table III.

The receiver characteristic curves (ROC) in Fig. 6 indicates that our approach can offer consistently outperforming results on three different public databases. The extent of the performance improvement however varies for the different databases and the observed improvement is also supported by EER in Table III. Some references have also illustrated area under curve to ascertain the comparative performance. Therefore, we also provide the area under the curve (AUC) in Table IV, using the results from our framework and the results from *UniNet*[14] which has shown to offer promising results in the literature. In addition to the consistent improvement in the matching accuracy, our model also offers a simplified network architecture as it discards the pooling layer and the up-sampling layer. Such simplification can also be observed from Table V which provides average feature extraction time for the iris image samples, the number of parameters and float



(a)

(b)

Fig. 7: Comparative ROC results from cross-dataset matching using (a) CASIA.v4-distance dataset (b) WVU Non-ideal dataset.

point operations (FLOPs) in the trained model. Our machine configuration is Intel i9-7900x with 32 GB memory, and all the experiments use one NVIDIA GTX 1080Ti card with 11GB memory.

C. Cross-Database Performance Evaluation

One of the key benefits expected from the deep learning based iris recognition lies in the generalization, i.e. capability of offer high matching performance using the trained model which is trained using completely different or independent iris database. Therefore, such cross-database performance evaluation was also performed to ascertain the effectiveness of our iris recognition approach. In this cross-database performance evaluation, we employ the model that is directly trained using the ND-IRIS-0405 iris image database [26] and use it to ascertain the matching performance for the CASIA.v4-distance database [27] and WVU non-ideal iris image dataset [29], directly without any fine-tuning. The number of test images are same as described for respective databases in previous section. This evaluation aims to validate generalization capability of the framework when there are limited or no training samples accessible from the target iris database. The comparative performance from the respective databases is shown in Fig. 7. Table VI summarized respective EER values from this cross-database performance evaluation.

TABLE VI: Summary of EER from the cross dataset performance evaluation.

	CASIA.v4-distance	**WVU Non-ideal**
UniNet	5.61%	3.67%
Ours	5.13%	2.31%

TABLE VII: Summary of AUC from the cross dataset performance evaluation.

	CASIA.v4-distance	**WVU Non-ideal**
UniNet	0.9804	0.9915
Ours	0.9837	0.9927

TABLE VIII: Summary of EER from the ablation study.

	ND-IRIS-0405	**CASIA.v4-distance**	**WVU Non-ideal**
ResNet	1.37%	5.54%	2.71%
Dilated Net	1.32%	5.58%	2.33%
UniNet	1.40%	5.50%	2.63%
Ours	1.30%	4.91%	1.91%

These results for the cross-database matching also indicate the improvement from our framework and reveal the generalization capability of our framework. We also computed the AUC between the *UniNet* and our framework to ascertain the significance of performance improvement and is presented in Table VII.

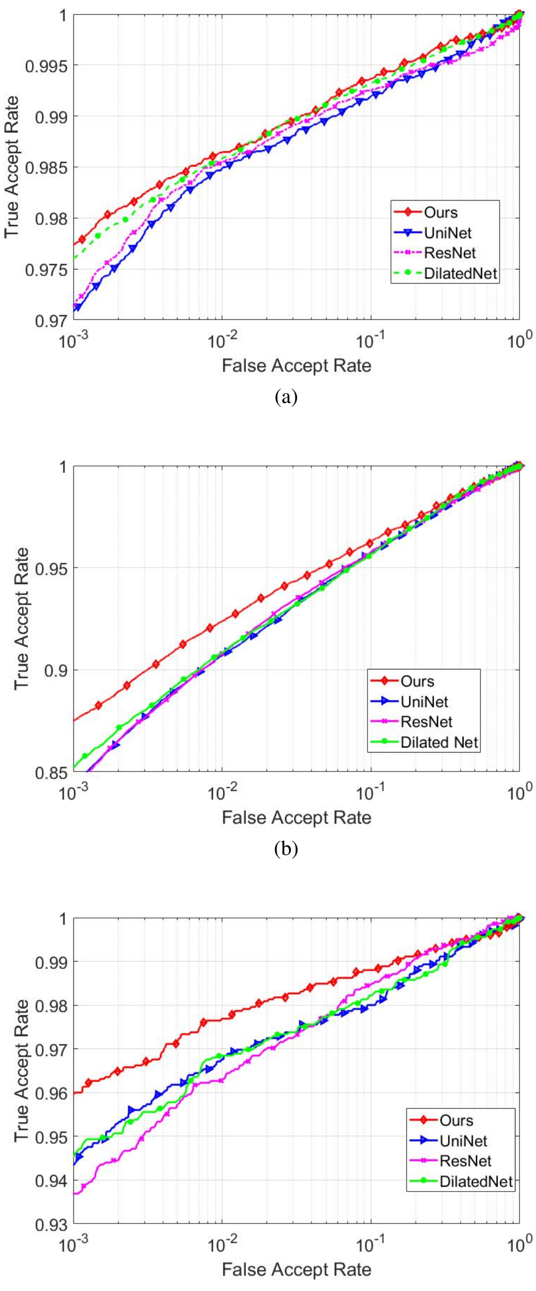
D. Ablation Study

We performed two ablation experiments, including the dilated net and residual net, to further investigate the extent of performance enhancement. This ablation study helps us to ascertain the effectiveness of our approach or network, over other well-established networks (like *ResNet*, *UniNet*, *DilatedNet*), for the iris recognition problem in this work. The EER results are shown in Table VIII, and the corresponding ROCs are illustrated in Fig. 8.

From the results shown above, we can observe the usage of both the residual learning blocks and dilated kernel contribute to the increase in iris matching accuracy. The residual block learns the residual information and ensemble across layers and makes the learning pattern more representative. The dilated kernels can also learn more accurate features by discarding the down-sampling (maximum pooling) layer and the up-sampling (bilinear) layer whose parameters are not tunable in *UniNet*[14]. The dilated convolutional kernel can maintain the resolution of input data and help to preserve the contributions from thin and small texture features that are important in correctly matching iris images.

IV. Discussion

In this section, we provide additional experimental results to further evaluate the effectiveness and robustness of iris recognition framework introduced in this paper. These additional

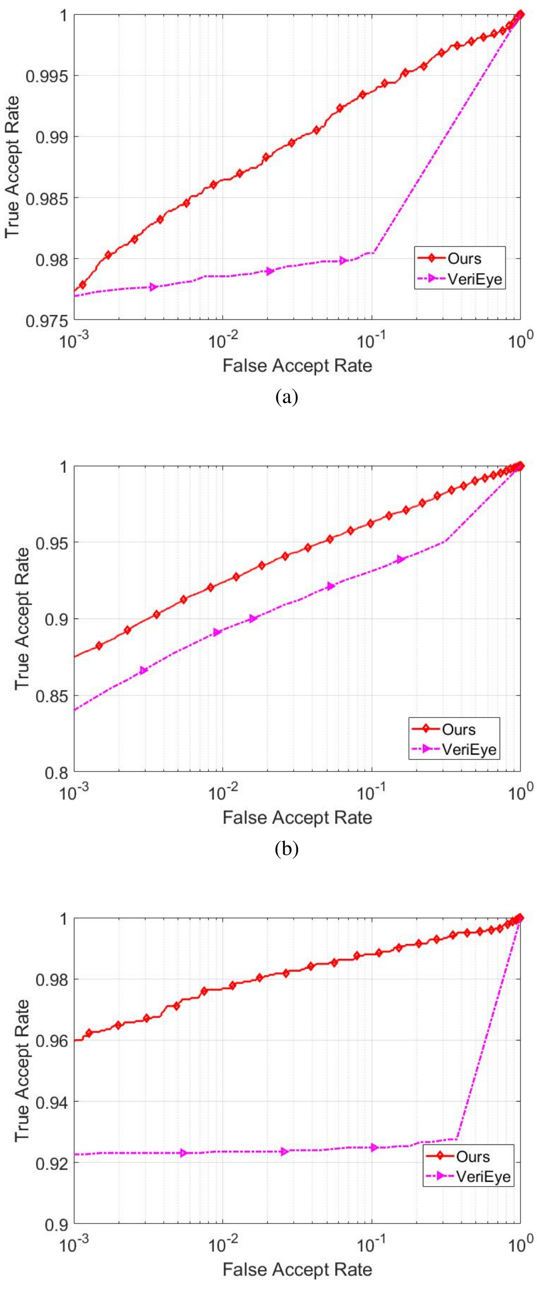


(a)

(b)

(c)

Fig. 8: Comparative ROC results from ablation test using (a) ND-IRIS-0405 dataset (b) CASIA.v4-distance dataset (c) WVU Non-ideal dataset.



(a)

(b)

(c)

Fig. 9: Comparative receiver characteristic curve (ROC) results from our algorithms and commercial products VeriEye SDK 9.0 on (a) ND-IRIS-0405 dataset (b) CASIA.v4-distance dataset (c) WVU Non-ideal dataset.

experimental results are organized into different subsections; including comparative evaluation with a commercial system, comparative performance with other deep learning architecture and effectiveness of our approach for cross-spectral iris matching problem.

A. Comparison with Commercial tools

We also performed experiments to comparatively evaluate the effectiveness of our iris recognition approach with popular commercial iris recognition system VeriEye [30]. Fig. 9 and Table IX presents such comparative results from our approach and those using such commercial iris recognition system on three different public iris database employed in our work.

In order to ensure a fair comparison, the size of the test samples for these comparative experiments is exactly same as employed for the experiments in Section III-A, i.e. 14,791 genuine and 5,743,130 impostor match scores for the tests using ND-IRIS-0405 iris database, 20,702 genuine and 2,969,533 impostor match scores for the CASIA.v4-distance iris database, and 2,251 genuine and 643,565 impostor scores for the WVU non-ideal iris database. Our comparative results

TABLE IX: Summary of EER for the comparison with commercial tools.

	ND-IRIS-0405	CASIA.v4-distance	WVU Non-ideal
VeriEye	1.33%	7.05%	7.20%
Ours	1.30%	4.91%	1.91%

TABLE X: Summary of EER from the cross-spectral iris recognition performance evaluation.

	EER
IrisCode	19.48%
MRF	18.40%
Ours	17.03%

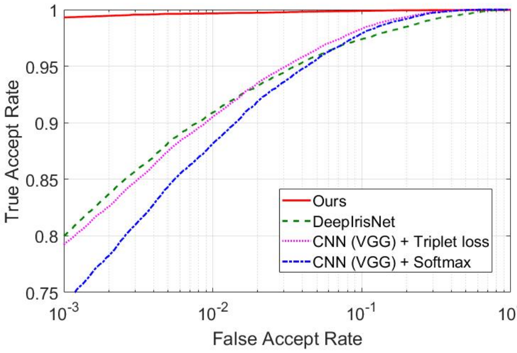


Fig. 10: Comparative ROC results from other deep learning based algorithms using the ND-IRIS-0405 dataset.

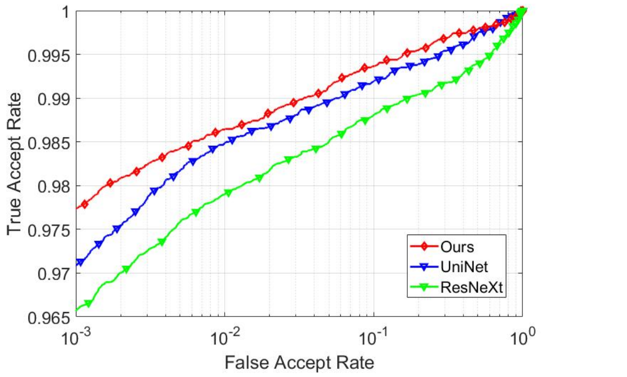


Fig. 12: Comparative ROC results from *ResNeXt* on the ND-Iris-0405 dataset.

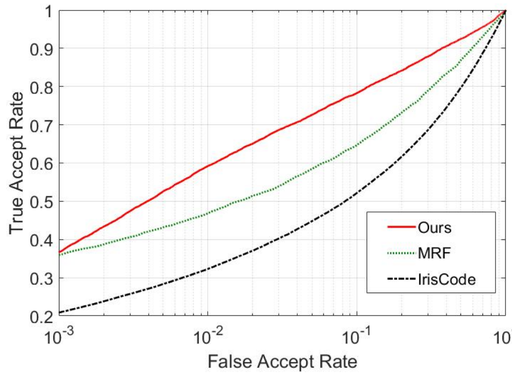


Fig. 11: Comparative ROC results from cross-spectral iris recognition using the PolyU-cross-spectral iris dataset.

in Fig. 9 consistently indicate outperforming results from such performance evaluation.

B. Comparison with Other Methods

We also performed experiments to comparatively ascertain the performance from other deep learning based methods ascertain the effectiveness of our framework. These experimental results are reported using ND-IRIS-0405 database. The first model is the CNN with softmax cross-entropy loss which has been widely employed in the literature [31][32]. The second model considered for the performance evaluation uses triplet loss introduced in FaceNet [24] which offers great potential for the recognition problems. The CNN employed in the two models mentioned above is *VGG-16* [33] which has emerged as an effective feature extractor. The last model, DeepIrisNet [16], is implemented based on inception CNN with softmax cross-entropy loss, and it is designed to address the iris recognition problem. However, the original model is not publicly available, so we implemented this model from the details described in this paper and performed the comparative experiments. The comparative ROCs from this set of experiments are shown in Fig. 10 and indicate that our other approach can significantly outperform other deep learning based methods considered for the evaluation.

C. Cross-spectral Iris Recognition

In order to further ascertain the robustness of the presented iris recognition approach, we performed some experiments for the cross-spectral iris matching problem. We employed publicly available PolyU-cross-spectral iris image dataset [34] and used same train-test protocols as detailed in [34]. The first baseline results are reported using the *IrisCode* generated from 1D-log Gabor filter response [3]. We also employed the state-of-the-art approach in that paper [34] which uses Markov random field (MRF) to synthesize the iris images for the accurate cross-spectral matching. The comparative experimental results are presented in Fig. 11 and respective EER are summarized in Table X. These results, although for a different problem, are quite encouraging and indicate the robustness of our iris-recognition approach for the cross-spectral iris matching problem.

D. Comparison with Aggregated Residual Network

A variant of *ResNet*, codenamed as *ResNeXt*, was recently proposed in [35] which includes several parallel ResNet branches with the same topology, and introduces a hyper-parameter called cardinality-the number of independent paths, to provide a new way of adjusting the model capacity. In this experiment, we attempted to expand our model with cardinality 8 to ascertain possible improvement of wide architecture. Such comparative experimental results using ND-IRIS-0405

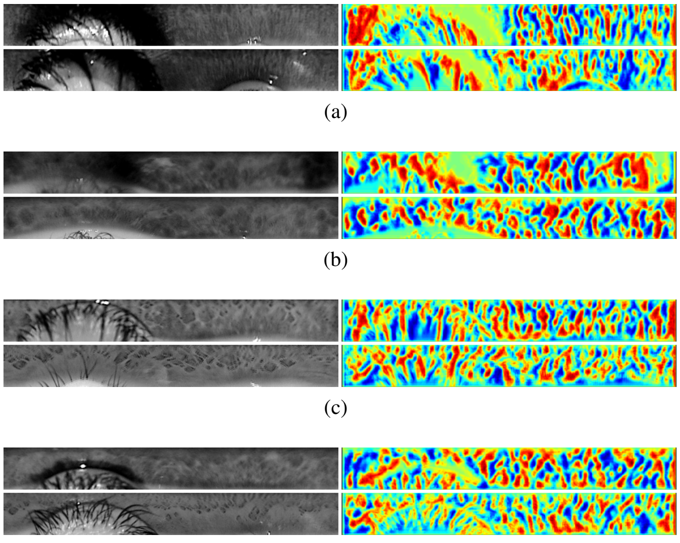


Fig. 13: Sample images from ND-IRIS-0405 Dataset that failed to correctly match; (a)-(b) genuine image sample pairs that failed to match and (c)-(d) imposter image pairs that incorrectly matched.

iris dataset are illustrated in Fig. 12. These results also suggest outperforming results from our approach.

E. Failure Cases in Iris Matching

We analyzed the iris images which failed from our approach and Fig. 13 presents these sample images. These failed cases, i.e. failure of genuine class iris samples to match, can be largely attributed to degradation in the iris image quality, segmentation error, and large off-angle iris images. Fig. 13 provides image samples from the same-class (genuine) which failed to match and also the different-class (impostor) samples which falsely matched from our iris recognition approach. This figure also provides corresponding heat maps, which are generated from the real value output from DRFNet before the binarization step. The cold area towards blue color in these images represents the pixels values close to -1 while the warm area towards red color represents the respective pixel values that are closer to 1. The decision threshold was fixed as 0.3770 for these matching. The matching scores from the genuine pairs are 0.3946 and 0.4063, while the matching scores from the imposter pairs are 0.3569 and 0.3723 respectively.

F. Selection of Network Parameters

In computer vision, the rectified linear unit *ReLU* [36] has been widely used for the DCNN training for its sparsity and reduced likelihood for the vanishing gradient. There are a few reasons for our preference to use *tanh* instead of *ReLU*. Firstly, we expect our final output as a pseudo binary feature since we perform Hamming distance computations during the test phase. The *tanh* whose output ranges from -1 to 1, is therefore more preferable than *ReLU*. Secondly, our network is not very deep and therefore we do not expect serious implications from the problem of vanishing gradients. However, we have also performed experiments using *ReLU*, instead of *tanh* in our model, and these comparative experimental results are shown in Fig. 14. The EER results from our model with *tanh* and the new model with *ReLU* are 1.30% and 1.61% respectively. These experiments were performed on ND-IRIS-0405 database and use same train/test protocol as in Section III-B. These results also help to validate the choice of *tanh* over *ReLU* in our network.

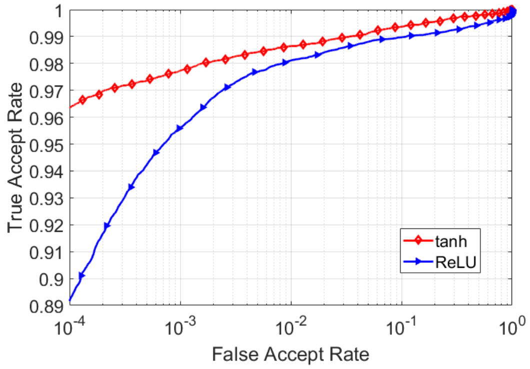


Fig. 14: Comparative ROC results from different nonlinear units.

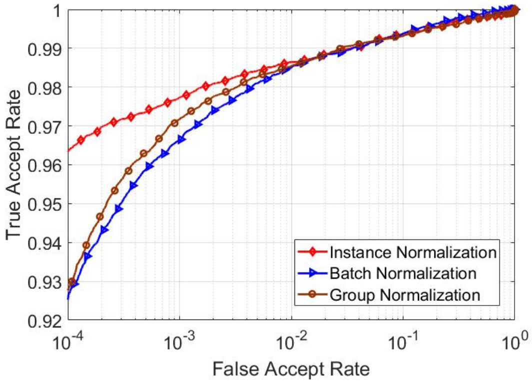


Fig. 15: Comparative ROC results from different normalization schemes.

Normalization is also an important component in the DCNN training, which accelerates the network convergence and also improves the performance. We also performed comparative experiments with other promising normalization methods including the batch normalization [37] and the group normalization [38]. In the batch normalization, we empirically selected batch size of 32 while for the group normalization, the group size of 8 was set for the best performance. These experimental results are shown in Fig. 15. The EER results from instance normalization, batch normalization and group normalization are 1.30%, 1.35% and 1.33% respectively. The experimental results indicate superior performance using the instance normalization for the iris recognition using our network.

The deep bottleneck architecture is another variation of *ResNet*[23] which can further simplify the networks with fewer parameters. Such bottleneck design uses 1×1 convolutional kernels to reduce the number of the channels and then perform the 3×3 convolutions on the less number of layers. After that, 1×1 convolution is performed again to increase the number of channels. We also perform experiments using the bottleneck architecture, with the configuration as shown in

TABLE XI: The details of DRFNet with bottleneck architecture.

Layer Name	Layer Type	Kernel Size	Output Channel	Dilated Factor
Conv1	Convolution	3×3	16	1
Imnorm1	BatchNorm	-	16	-
Scale1	Scale	-	16	-
ReLU1	ReLU	-	16	-
Conv2a_1	Convolution	1×1	8	1
Imnorm2a_1	BatchNorm	-	8	-
Scale2a_1	Scale	-	8	-
ReLU2a_1	ReLU	-	8	-
Conv2a_2	Convolution	3×3	8	2
Imnorm2a_2	BatchNorm	-	8	-
Scale2a_2	Scale	-	8	-
ReLU2a_2	ReLU	-	8	-
Conv2a_3	Convolution	1×1	32	1
Imnorm2a_3	BatchNorm	-	32	-
Scale2a_3	Scale	-	32	-
Conv2b	Convolution	1×1	32	1
Imnorm2b	BatchNorm	-	32	-
Scale2b	Scale	-	32	-
Res2	Elementwise sum	-	32	-
ReLU2	ReLU	-	32	-
Conv3a_1	Convolution	1×1	16	1
Imnorm3a_1	BatchNorm	-	16	-
Scale3a_1	Scale	-	16	-
ReLU3a_1	ReLU	-	16	-
Conv3a_2	Convolution	3×3	16	4
Imnorm3a_2	BatchNorm	-	16	-
Scale3a_2	Scale	-	16	-
ReLU3a_2	ReLU	-	16	-
Conv3a_3	Convolution	1×1	64	1
Imnorm3a_3	BatchNorm	-	64	-
Scale3a_3	Scale	-	64	-
Conv3b	Convolution	1×1	64	1
Imnorm3b	BatchNorm	-	64	-
Scale3b	Scale	-	64	-
Res3	Elementwise sum	-	64	-
ReLU3	ReLU	-	64	-
Res	Concatenate	-	112	-
Conv4	Convolution	1×1	1	1
Imnorm4	BatchNorm	-	1	-
Scale4	Scale	-	1	-
Tanh	*tanh*	-	1	-

Table XI, using the ND-0405 IRIS dataset. The comparative experimental results from DRFNet without bottleneck architecture and DRFNet with bottleneck architecture are shown in Fig. 16. These results indicate noticeable improvement in the true acceptance rate at low false acceptance rates.

G. Comparison with Dilated Residual Network

There have been many promising attempts in the literature to incorporate the dilated kernels or filters to recover the features from multiple scales. Yu et al. in [39] have introduced a dilated residual network for the image classification problem and the presented impressive results for the semantic segmentation problem. This architecture is based on ResNet-18. The authors have divided all the layers into five groups determined by the feature map scale. They also remove the down-sampling in the covolutional layer by changing the convolution stride back to one in the last two groups and change convolutional kernels to dilated convolutional kernels with dilation factor of 2 and 4 respectively. This work also introduce the degridding layers in the model and increase the total number of layers to 26. Authors use bilinear interpolation to up-sample the layer, before the global average pooling, for the semantic segmentation.

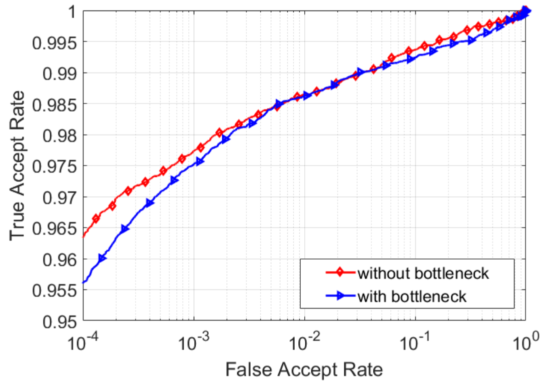


Fig. 16: Comparative ROC results from different *ResNet* blocks.

The differences between our model and the dilated residual network (DRN) [39] are summarized as follows. Firstly, the model in [39] is a modified *ResNet* and they preserve the first three down-sampling layers while our model is shallow network without any down-sampling and up-sampling. Secondly, we concatenate the layers from different scale, but the authors in [39] only up-sample the feature map without concatenating feature maps from different scales. Thirdly, we use the triplet network to calculate the distance between each templates while the authors use the sequential architecture with softmax, using the ground truth, to achieve the semantic segmentation. Fourthly, the loss function is different since we use extended triplet loss while they use softmax cross-entropy loss. Finally, the authors in [39] use degridding layers and our model do not have these layers which can increase the complexity of model (18 layers to 26 layers).

The DRN-C-26 performs well for the semantic segmentation as detailed in [39]. Therefore to ensure a fair comparison, we use it as one branch of triplet networks and compute the extended triplet loss. We follow the same training protocols as detailed in [39], with the initial learning rate 0.1 and the parameter update scheme is SGD with momentum 0.9. The comparative performance using the ROC results from our model and DRN-C-26 is shown in Fig. 17 and the respective EERs are 1.30% and 2.19%.

V. Conclusions and Future Work

This paper has introduced a new approach for more accurate iris recognition. Involuntary pupil dilation and scale changes during the iris imaging constitute the key source for the frequently observed iris deformations. Our approach attempts

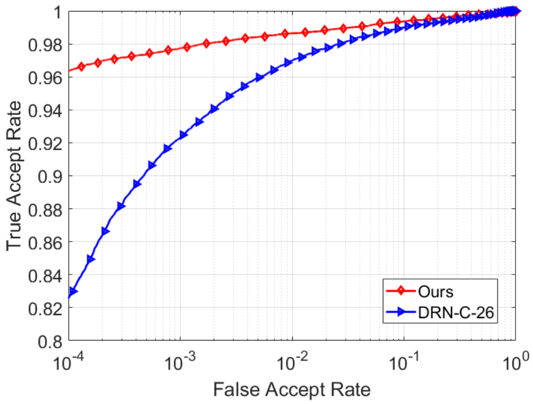


Fig. 17: Comparative ROC results from our model and DRN-C-26.

to address this problem by incorporating the dilated convolutional kernel and residual learning in our framework for more accurate iris matching. Such an approach also simplified the architecture of the deep neural network. The experimental results presented in Section III of this paper, using within-database and cross-database performance evaluation, on three different public iris image databases illustrate outperforming results and validate the effectiveness of our approach. Iris images inherently illustrate ocular information and can be incorporated in the deep neural network model to further improve the iris image matching accuracy and it is part of further research in this area. Our current work uses a *MaskNet* which was separately trained. Development of an end-to-end architecture which can simultaneously ignore masked bits, or incorporate end-to-end *MaskNet* training, is highly desirable and part of further work in this area.

References

[1] J. Daugman, “Iris recognition border-crossing system in the UAE,” *International Airport Review*, vol. 8, no. 2, 2004.

[2] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, “Image understanding for iris biometrics: A survey,” *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 281–307, 2008.

[3] L. Masek, “Recognition of human iris patterns for biometric identification,” *The University of Western Australia*, 2003.

[4] J. Daugman, “How iris recognition works,” in *The essential guide to image processing*. Elsevier, 2009, pp. 715–739.

[5] D. M. Monro, S. Rakshit, and D. Zhang, “Dct-based iris recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 586–595, 2007.

[6] K. Miyazawa, K. Ito, T. Aoki, K. Kobayashi, and H. Nakajima, “An effective approach for iris recognition using phase-based image matching.” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1741–1756, 2008.

[7] Y. Zhou and A. Kumar, “Personal identification from iris images using localized Radon transform,” in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 2840–2843.

[8] O. C. Reyes, R. Vera-Rodriguez, P. Scully, and K. B. Ozanyan, “Analysis of spatio-temporal representations for robust footstep recognition with deep residual neural networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[9] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, “Exploring recurrent neural networks for on-line handwritten signature biometrics,” *IEEE Access*, vol. 6, no. 5128-5138, pp. 1–7, 2018.

[10] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701–1708.

[11] Z. Zhao and A. Kumar, “Improving periocular recognition by explicit attention to critical regions in deep neural network,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 12, pp. 2937–2952, 2018.

[12] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcao, and A. Rocha, “Deep representations for iris, face, and fingerprint spoofing detection,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 864–879, 2015.

[13] Z. Sun and T. Tan, “Ordinal measures for iris recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2211–2226, 2009.

[14] Z. Zhao and A. Kumar, “Towards more accurate iris recognition using deeply learned spatially corresponding features,” in *Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy*, 2017, pp. 22–29.

[15] Q. Zheng, A. Kumar, and G. Pan, “A 3d feature descriptor recovered from a single 2d palmprint image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 1272–1279, 2016.

[16] A. Gangwar and A. Joshi, “Deepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition,” in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 2301–2305.

[17] S. Minaee, A. Abdolrashidiy, and Y. Wang, “An experimental study of deep convolutional features for iris recognition,” in *2016 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2016, pp. 1–6.

[18] F. He, Y. Han, H. Wang, J. Ji, Y. Liu, and Z. Ma, “Deep learning architecture for iris recognition based on optimal Gabor filters and deep belief network,” *Journal of Electronic Imaging*, vol. 26, no. 2, p. 023005, 2017.

[19] X. Tang, J. Xie, and P. Li, “Deep convolutional features for iris recognition,” in *Chinese Conference on Biometric Recognition*. Springer, 2017, pp. 391–400.

[20] 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. 18 848–18 855, 2018.

[21] H. Menon and A. Mukherjee, “Iris biometrics using deep convolutional networks,” in *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2018, pp. 1–5.

[22] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” in *International Conference on Learning Representations (ICLR)*, 2016.

[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[24] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815–823.

[25] Web link to download the source code and executable files for the approach detailed in this paper:. [Online]. Available: http://www.comp.polyu.edu.hk/ csajaykr/drfiris.htm

[26] K. W. Bowyer and P. J. Flynn, “The nd-iris-0405 iris image dataset,” *arXiv preprint arXiv:1606.04853*, 2016.

[27] Casia.v4 iris database. [Online]. Available: http://biometrics.idealtest.org/dbDetailForUser.do?id=4

[28] Opencv based face and eye detector. [Online]. Available: http://docs.opencv.org/trunk/d7/d8b/tutorial_py_face_detection.html

[29] S. Crihalmeanu, A. Ross, S. Schuckers, and L. Hornak, “A protocol for multibiometric data acquisition, storage and dissemination,” *Technical Report, WVU, Lane Department of Computer Science and Electrical Engineering*, 2007.

[30] Verieye sdk 9.0. [Online]. Available: http://www.neurotechnology.com/verieye.html

[31] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[32] Y. Sun, X. Wang, and X. Tang, “Deep learning face representation from predicting 10,000 classes,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1891–1898.

[33] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[34] P. R. Nalla and A. Kumar, “Toward more accurate iris recognition using cross-spectral matching,” *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 208–221, 2017.

[35] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in *Computer Vision and*

Pattern Recognition (CVPR), 2017 IEEE Conference on. IEEE, 2017, pp. 5987–5995.

[36] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807–814.

[37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[38] Y. Wu and K. He, "Group normalization," *arXiv preprint arXiv:1803.08494*, 2018.

[39] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 472–480.