DualSANet: Dual Spatial Attention Network for Iris Recognition

Kai Yang
SenseTime Research
Shanghai, China
yangkai@sensetime.com

Zihao Xu
SenseTime Research
Shanghai, China
xuzihao@sensetime.com

Jingjing Fei
SenseTime Research
Shanghai, China
feijingjing1@sensetime.com

Abstract

Compared with other human biosignatures, iris has more advantages on accuracy, invariability and robustness. However, the performance of existing common iris recognition algorithms is still far from expectations of the community. Although some researchers have attempted to utilize deep learning methods which are superior to traditional methods, it is worth exploring better CNN network architecture. In this paper, we propose a novel network architecture based on the dual spatial attention mechanism for iris recognition, called DualSANet. Specifically, the proposed architecture can generate multi-level spatially corresponding feature representations via an encoder-decoder structure. In the meantime, we also propose a new spatial attention feature fusion module, so as to ensemble these features more effectively. Based on these, our architecture can generate dual feature representations which have complementary discriminative information. Extensive experiments are conducted on CASIA-IrisV4-Thousand, CASIA-IrisV4-Distance, and IITD datasets. The experimental results show that our method achieves superior performance compared with the state-of-the-arts.

1. Introduction

Iris recognition [7, 6] is one of the most accurate approaches for person identification, because of the unique, complex, and stable texture patterns in iris. In the last two decades, there are many methods that have been proposed for iris recognition [7, 6, 27, 28, 2]. Here we briefly survey some classical algorithms (For a complete review, please refer to [3, 30]). Gabor filter is a quite popular method which extracts the iris feature representation (IrisCode) based on the segmented and normalized iris image. The Hamming distance between two masked IrisCodes is utilized as the dissimilarity score for identity verification. Other competitive methods for iris recognition generate different feature representation using Radon Transform [50], Discrete Cosine Transform (DCT) [29], Discrete Fourier Transform (DFT) [28].

In recent years, deep learning has achieved state-of-the-art performance in many computer vision tasks, such as image classification [19, 42], object detection [13, 12, 35], semantic segmentation [22, 47, 46, 45], and face recognition [37, 34]. Motivated by the success of deep neural networks in other applications, several approaches based on deep learning are recently explored for iris recognition [21, 10, 31, 11, 49]. However, the methods in [21, 10, 31] did not take non-iris region into account and utilized the global feature which is inappropriate for iris pattern. Gangwar and Joshi [11] utilized DeepIrisNet to extract global feature on detected images without normalization. Zhao and Kumar [49] proposed an UniNet which consists of two subnetwork: FeatNet and MaskNet. FeatNet is utilized to extract spatially corresponding features based on an encoder-decoder structure, and MaskNet is set to perform non-iris region masking. UniNet is a highly competitive benchmark, but the FeatNet only consists of simple convolutional layers and does not explore the effective fusion of features from different levels.

In this paper, we propose a novel network architecture for iris recognition, which is based on the dual spatial attention mechanism, called DualSANet. Figure 1 illustrates the encoder-decoder architecture of DualSANet. The encoder utilizes a pre-trained ResNet-18 [15] model to extract features from different levels, and the decoder contains dual branches to generate dual feature representations. We propose an advanced spatial attention feature fusion module to fuse the features from different levels. Spatial attention [9, 48] is a simple mechanism to encode different importance of every position in a feature map. It is quite appropriate for iris recognition because iris feature is a kind of local features, which means that iris features have different spatial importance in different local areas. The spatial attention feature fusion module in DualSANet can learn the weights of every position and fuse the features from different levels effectively.

In summary, our main contributions can be summarized as follows. 1) We introduce a novel encoder-decoder net-

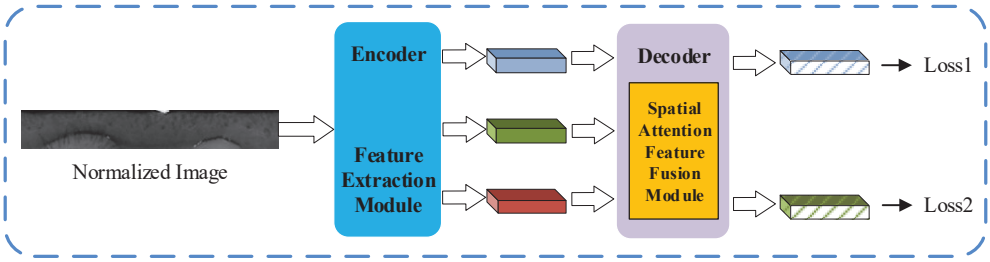


Figure 1. A brief illustration of our encoder-decoder architecture.

work to generate multi-level iris features. 2) We propose a new spatial attention feature fusion module to ensemble features from different levels in DualSANet. 3) We design an effective pipeline for iris recognition, which leverages proposed DualSANet to generate dual feature representations which have complementary discriminative information. Experimental results show the improvement over the state-of-the-art algorithms for iris recognition under fair comparison.

2. Related Work

The first automated iris recognition system was proposed by Daugman [7, 6] in 1993. In his pioneering work, he proposed generation of Gabor phase-quadrant feature descriptor (IrisCode) to extract on segmented and normalized iris images. The Hamming distance between two masked IrisCodes is utilized for verification. IrisCode is an effective feature descriptor, achieving very low false accept rates. This approach can be implemented by log-Gabor filters [27] which extracts iris texture feature more efficiently than two-dimensional Gabor filters [6]. Monro *et al*. [29] leveraged Discrete cosine transform (DCT) coefficients to analyze spectral contents in image block regions, and generated promising results by matching binary feature. Sun and Tan [39] utilized multi-lobe differential filters (MLDFs) to encode multi-orientation and multi-scale feature representation for normalized templates. In addition, researchers have also proposed varieties of feature descriptors for iris recognition, which could be grouped into four main categories, phased-based method [28] , zero-crossings representation [2, 36, 38], texture analysis [51, 26, 23], and intensity variation analysis [24, 25].

Learning the feature representation from iris data can get a more robust and optimal representation for iris recognition task. However, unlike the popularity of deep learning for various computer vision tasks, especially for face recognition, the literature so far has not yet fully explored its potential for iris recognition. There has been few breakthroughs in iris recognition and very little attention to exploring the iris recognition solutions using deep learning. Liu *et al*. [21] proposed a deep learning based framework DeepIris to learn relational features to measure the similarity between pairs of normalized iris images. Gangwar and Joshi [10] treated iris recognition as a classical classification task and proposed DeepIrisNet, including convolutional layers, pooling layers, and fully connected layers. He *et al*. [14] attempted to explore the deep belief network (DBN) for iris recognition. An optimal Gabor filter selection is the key component, while the DBN is only a simplified application on the IrisCode without iris-specific optimization. Tang *et al*. [43] proposed a lightweight CNN architecture suitable for iris datasets with small-scale labeled images. Nguyen *et al*. [31] utilized off-the-shelf CNNs pretrained on the ImageNet [19] to extract iris features from the segmented and normalized iris images. The extracted features are classified by a support vector machine (SVM).

However, these works treat iris recognition as a direct application of convolutional neural networks without consideration of the adaptation for iris pattern: 1) These methods do not take non-iris region into account. 2) The most discriminative information in iris pattern comes from local features rather than global features based on the early promising works [51, 26, 23, 24]. Therefore, convolutional layers are more effective than fully connected layers in iris recognition. To process the non-iris region of an input image, DeepIrisNet2 in [10] presented a simple effective supervised learning framework to obtain iris representation on the detected and segmented images. Zhao and Kumar [49] proposed a framework called UniNet, which consists of two sub-networks: FeatNet and MaskNet. FeatNet is designed for extracting discriminative iris features, and MaskNet is set to perform non-iris region masking. They also introduce a extended triplet loss to train their network. This method is a highly competitive benchmark and confirms the effectiveness of the spatially corresponding features. However, it does not explore the structure of the network to obtain better features and the impact of features from different levels. Based on this work, Wang and Kumar [44] utilized residual network to learning with dilated convolutional kernels to optimize the training process and aggregate contex-

tual information from the iris images. In addition, some researches explored the effect of data on training iris recognition model. Liu *et al*. [20] used Gaussian, triangular fuzzy median smoothing filters to preprocess the images by fuzzifying the region beyond the boundary to improve the signal-to-noise ratios.

3. Proposed Algorithm

The proposed algorithm consists of three steps: image preprocessing, network architecture, and loss function. The image preprocessing module includes location, segmentation, and normalization of the iris images. The network architecture is based on an encoder-decoder structure where encoder is part of a pre-trained ResNet-18 which extracts features from different levels and decoder consists of bilinear upsampling and spatial attention feature fusion module. The extended triplet loss with batch hard policy is utilized to train the network.

3.1. Image Preprocessing

For a fair comparison, our experiments utilize a freely available system Osiris v4.1 [32] for iris location, segmentation, and normalization. Iris Segmentation consists in isolating the iris texture from other elements of image such as eyelids, eyeslashes, spotlights and/or shadows. In addition, the segmentation module generates a binary mask, which indicates which pixels of the image belong to iris texture. The contours of the iris correspond to an optimal path retrieved by the Viterbi algorithm for joining in an optimal way, the points of high gradients under the constraint that the resulting curve has to be closed [41]. The iris texture is mapped into a size-invariant band called the normalized iris image. This transformation is carried out by exploiting a parameterization of the iris boundaries obtained by the segmentation module. In our image preprocessing, normalization is based on Daugman's rubber sheet model [7] to unwrap the iris texture. We also set the resolution after normalization to 64×512 uniformly. The normalization process allows the alignment of any two iris images to be compared. Figure 2 illustrates the key steps for iris image preprocessing which contains location, segmentation, and normalization. The normalization process allows the alignment of any two iris images to be compared.After image preprocessing, we can get the normalized iris texture image and iris/noise mask image. The texture image is used for extracting FCN feature maps and mask is actually a hard attention used in the extended triplet loss.

3.2. Network Architecture

Detailed structure of proposed DualSANet is shown in Figure 3. The proposed architecture can generate multi-level spatially corresponding feature representations via an encoder-decoder structure. A number of studies [33, 18] have shown that $attention$ plays an important role in many vision tasks, we also propose a new spatial attention feature fusion module, so as to ensemble these features more effectively. Earlier promising works on iris recognition [7, 6, 27, 24, 25] indicated that local features matter more than global features. We do not employ very deep network which has large receptive field and consider more to exploit low-level features and mid-level features, here we adopt ResNet-18 as the backbone network.

3.2.1 Encoder Module

Encoder module: we use the part of the standard pre-trained ResNet-18 model as the backbone to extract features from different levels. The block1 presents the 'c1' layer in ResNet-18, and the block2 presents the 'c2' layer in ResNet-18. Low-level features are necessary to preserve spatial details and textural features, high-level features have the ability to capture more context information. We utilize a spatial attention feature fusion module to combine features from different levels.

3.2.2 Spatial Attention Feature Fusion Module

The detailed structure of SAFFM is illustrated in Figure 3 and Table 1. For the given multi-level input feature maps, we first concatenate them and get an input $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$. Then, a convolutional layer and batch normalization is utilized to balance the scales of features from different levels. We halve the numbers of output channels in consideration of reducing computation and get an feature map $\mathbf{F} \in \mathbb{R}^{C/2\times H\times W}$. Next, we propose a specific spatial attention module to refine the fused features. SAFFM infers a 3D attention map $\mathbf{Tr}(\mathbf{F}) \in \mathbb{R}^{C/2\times H\times W}$. The refined feature map $\mathbf{F}'$ is computed as:

$$\mathbf{F}' = \mathbf{F} + \mathbf{F} \otimes \mathbf{Tr}(\mathbf{F}) \tag{1}$$

where $\otimes$ denotes broadcast element-wise multiplication along the channel dimension. We adopt a residual learning scheme along with the attention mechanism to facilitate the gradient flow.

Usually, iris images have some non-iris region, including eyelid, eyelash and reflection. Therefore, each point in the spatially corresponding iris feature matters differently. The spatial attention module utilizes a small network to compute a weighted coefficient for each position. Spatial attention plays a feature selection role in the network and pays more attention to the features which have more discriminative information.

Considering the completeness of input feature $\mathbf{X}$, we do not reduce resolution in SAFFM forward computation. In concrete, we utilize two simple conv-bn-relu structures and one conv-sigmoid structure to generate the spatial attention

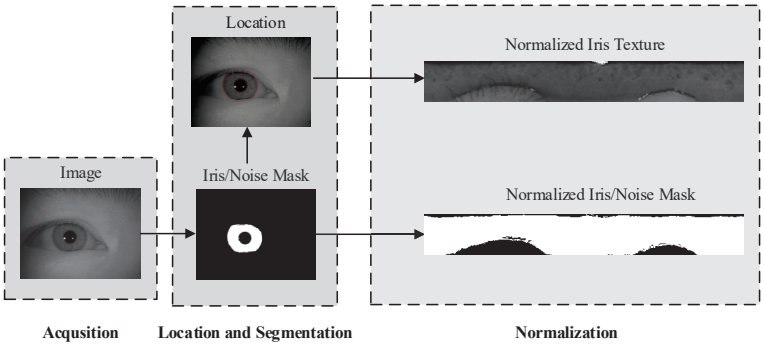


Figure 2. Illustration of key steps for iris image preprocessing.

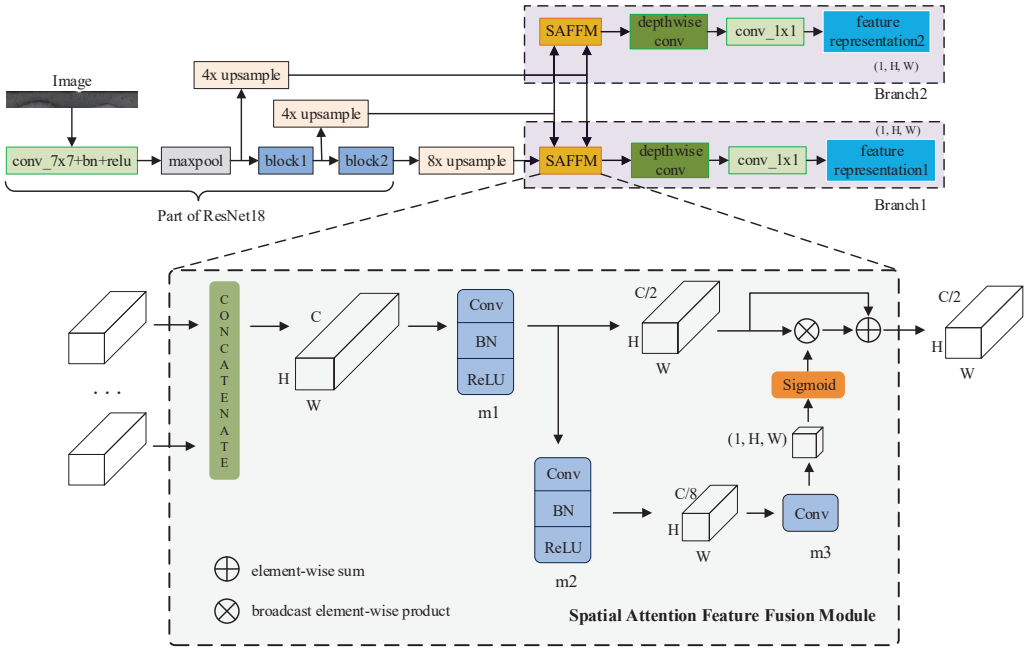


Figure 3. Detailed structure of DualSANet.

feature. The first 3×3 convolution eliminates the aliasing effect of direct concatenation and reduces the channel dimension from C to $C/2$. Such feature after first conv-bn-relu is regarded as $\mathbf{F}$. The second 1×1 convolution reduces the channel dimension to $C/8$ much further. Finally, the last 1×1 convolution generates spatial weighted response map $\mathbf{Tr}(\mathbf{F})$, and the operation of sigmoid compresses the output between 0 and 1 which represents weighted coefficient of each position.

3.2.3 Decoder Module

Decoder module: the decoder contains two branches. Each branch consists of bilinear upsample, spatial attention feature fusion module, depthwise convolution layer, and last convolutional layer. Bilinear upsample is utilized to upsample all features into the same resolution as input image. Spatial attention feature fusion module with attention branch is utilized to fuse features from different levels and give a spatial weight to fused features. Then, we utilize depthwise convolution [5, 17] to reduce the parameters. The last convolutional layer generates a 1×H×W feature representation.

Module Name	Layer Type	Output Channel	Kernel Size	Padding	Input Size	Output Size
m1	Convolution	C/2	3x3	1	C×H×W	C/2×H×W
	BN	C/2	-	-	C/2×H×W	C/2×H×W
	ReLU	-	-	-	C/2×H×W	C/2×H×W
m2	Convolution	C/8	1x1	0	C/2×H×W	C/8×H×W
	BN	C/8	-	-	C/8×H×W	C/8×H×W
	ReLU	-	-	-	C/8×H×W	C/8×H×W
m3	Convolution	1	1x1	0	C/8×H×W	1×H×W

Table 1. Details of m1, m2, m3.

3.3. Loss Function

The basic Triplet Loss is proposed by FaceNet [37]. It is defined as:

$$L = \sum_{i}^{N}[\|\boldsymbol{f}(x_i^a) - \boldsymbol{f}(x_i^p)\|_2^2 - \|\boldsymbol{f}(x_i^a) - \boldsymbol{f}(x_i^n)\|_2^2 + \alpha]_+ \quad (2)$$

While the symbol $[\bullet]_+$ represents $max(\bullet, 0)$. This loss makes sure that, the anchor-positive distance is closer than the anchor-negative distance by at least a margin α which is the tuning hyper-parameter. While Batch Hard is an improved triplet selection method [16], it is defined as:

$$L_{BH}(\theta; X) = \sum_{i=1}^{P}\sum_{a=1}^{K}[m + \max_{p=1..K} D(f_\theta(x_a^i), f_\theta(x_p^i)) - \min_{\substack{j=1..P,\\ n=1..K,\\ j\neq i}} D(f_\theta(x_a^i), f_\theta(x_n^j))]_+ \quad (3)$$

Within a mini-batch, it randomly sample P classes and then randomly sample K images of each class. For each sample in the batch, this method select the hardest positive and the hardest negative samples within the batch to form a triplet for computing the loss.

For one normalized iris image, our network extracts dual spatial feature representations defined as $\boldsymbol{f}^1$ and $\boldsymbol{f}^2$. $\boldsymbol{f}^1$ and $\boldsymbol{f}^2$ fuse different level features and have the complementary discriminative information. The feature representations have the same resolution with the input. Euclidean distance is to metric the corresponding representations between the compare pair. Our loss is defined as:

$$L_s = \sum_{i=1}^{P}\sum_{a=1}^{K}[m + \max_{p=1..K} MMSD(\boldsymbol{f}_{ia}^s, \boldsymbol{f}_{ip}^s) - \min_{\substack{j=1..P,\\ n=1..K,\\ j\neq i}} MMSD(\boldsymbol{f}_{ia}^s, \boldsymbol{f}_{jn}^s)]_+ \quad (4)$$

$$L_{total} = \lambda_1 L_1 + \lambda_2 L_2 \quad (5)$$

While s=1, 2, $MMSD(f^{1,s}, f^{2,s})$ is the *Minimum Shifted and Masked Distance* proposed in [49]. In our case, it is defined as:

$$MMSD(\boldsymbol{f}^{1,s}, \boldsymbol{f}^{2,s}) = \min_{-B\leq b\leq B} SD(\boldsymbol{f}_b^{1,s}, \boldsymbol{f}^{2,s}) \quad (6)$$

While B is the Shift size, SD is the distance between the features which take masks into consideration:

$$SD = \frac{\sum((\boldsymbol{f}^{1,s} - \boldsymbol{f}^{2,s}) \odot (\boldsymbol{f}^{1,s} - \boldsymbol{f}^{2,s}) \odot (\boldsymbol{m}^1 \odot \boldsymbol{m}^2))}{\sum(\boldsymbol{m}^1 \odot \boldsymbol{m}^2)} \quad (7)$$

Where $\odot$ represents the element-wise product of two matrices, $\sum$ represents the sum of matrix here. $\boldsymbol{m}^1$ and $\boldsymbol{m}^2$ are the binary masks for two feature maps, in which zero means the current position is non-iris. In (6), the subscript b means the feature map has been shifted horizontally by b pixels, i.e., a shifted feature map has the following spatial correspondence with the original one:

$$\begin{aligned} \boldsymbol{f}_b[x_b, y] &= \boldsymbol{f}[x, y] \\ x_b &= (x - b + w) \bmod w \end{aligned} \quad (8)$$

3.4. Feature extracting and matching

Our network architecture generates dual feature representations and we obtain the final feature representation by weighted average of the feature representations which is computed as follows:

$$\boldsymbol{f}_{final} = \lambda_1 \boldsymbol{f}^1 + \lambda_2 \boldsymbol{f}^2 \quad (9)$$

While λ_1 and λ_2 is as same as defined in loss function.

The $Minimum\ shifted\ and\ Masked\ Distance(MMSD)$ is the dissimilarity metric between two final compare feature representations. The false reject rate (FRR) at different false accept rate (FAR) and equal error rate (EER) are the main evaluation criterion.

4. Experiments

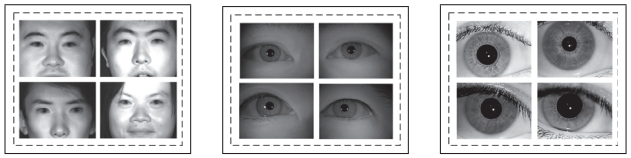
4.1. Datasets

We employed the following three publicly available datasets in our experiments, as shown in Figure 4:

CASIA-IrisV4-Thousand: the dataset(subset) [4]. The thousand subset includes 20,000 iris images from 2000 eyes of 1000 persons. Thus, each subject has both left and right eye. We just use the left eye for training and test. We split the first 900 person as the training and the last 100 person as the test. The test set includes 1000 samples, containing 4,500 intra pairs and 495,000 inter pairs.

CASIA-IrisV4-Distance: this dataset (subset) [4] includes 2,246 samples from 142 subjects. Each sample captures the upper part of face and therefore contains both left and right irises. We train an easy eye detector based on [35] to crop the eye regions from the original images. All of the right eye iris images are utilized as training set, and all of the left eye iris images were used as test set. The test set contains 20,702 intra pairs and 2,969,533 inter pairs.

IITD Iris Database: the IITD dataset [1] contains 2,240 image samples from 224 subjects. We use the same split as [49]. All of the right eye iris images are utilized as the training set while the first five eye images are utilized as test set. The test set contains 2,240 intra pairs and 624,400 inter pairs.



(a) (b) (c)

Figure 4. (a) examples of CASIA-IrisV4-Distance (b) examples of CASIA-IrisV4-Thousand (c) examples of IITD

4.2. Implemtation Details

Our network is implemented with PyTorch. We use the same optimizer and learning policy for all experiments. The optimizer is SGD. The initial learning rate is set to 0.001, which is down-scaled by 10 after epoch 50, 100, 150, and the max epoch is 200. We train all the networks with 8 Nvidia TITAN Xp. On each GPU, to form a mini-batch, we sample P persons and K images for each person. For CASIA dataset, P is set to 6, and K is set to 6. For IITD dataset, P is set to 10, and K is set to 3 because each person only have 5 images to train. The margin m is set to 20, the coefficients of two losses are set as λ_1=0.5, λ_2=0.5, and the Shift size is set to 5.

4.3. Ablation Study

In this subsection, we investigate the effect of each compoment in our proposed DualSANet in detail. In the following experiments, We use the ResNet18 network pre-trained on ImageNet dataset [8] as the backbone and conduct the experiments on the CASIA-IrisV4-Thousand dataset.

Baseline: our baseline just have single feature representation from the Branch1 without Spatial Attention Feature Fusion Module.

Ablation Study for the number of feature representation: we propose the dual feature representations as outputs. We also evaluate single representation from the Branch1 and triple feature representations that the third branch is from ahead of the block1. Results are shown in Table2 which indicates the scheme of dual feature representations is the best choice. The reason is that one feature representation from Branch1 missing some low-level information and the third representation from ahead of the block1 is too shallow to learn discriminative information.

Method	FRR@FAR			EER
	1e-5	1e-4	1e-3	
Base	4.78%	2.58%	1.09%	0.40%
Base+Dual	**3.67%**	**1.58%**	**0.69%**	**0.36%**
Base+Trip	8.93%	3.29%	1.22%	0.58%

Table 2. Setting an appropriate number of feature representations is important. 'Base' means the baseline network with just one feature representation from Branch1, 'Dual' means dual feature representations and 'Trip' means triple feature representations which the third branch is from ahead of the block1.

Ablation for SAFFM: we propose the spatial attention feature fusion module to combine the different level features. We evaluate a simple concatenation of these features, channel attention feature fusion module (CAFFM) and spatial attention feature fusion module (SAFFM). The channel attention operation we used is the same as SENet [18]. Results are shown in Table 3, which indicate that 'SAFFM' is the best fusion module. 'CAFFM' may harm the performance because of the global operation in 'CAFFM' is inappropriate for iris pattern.

Method	FRR@FAR			EER
	1e-5	1e-4	1e-3	
Base+Dual	3.67%	1.58%	0.69%	0.36%
Base+Dual+CA	4.82%	1.93%	0.64%	0.34%
Base+Dual+SA	**3.16%**	**1.53%**	**0.58%**	**0.31%**

Table 3. Performance comparison between different feature fusion modules. 'Base' represents the baseline model, 'SA' represents the spatial attention feature fusion module, 'CA' represents the channel attention feature fusion module, and the other is just simple concatenation.

Ablation Study for the coefficients of two feature representations: we evaluate different coefficients for the two feature representations, results are shown in Table 4.

Ablation Study for the Shift: we evaluate the effect of Shift. The size of Shift is set to 5. Results are shown in Table 5.

Method	FRR@FAR			EER
	1e-5	1e-4	1e-3	
Base+Dual+SA(0.1, 0.9)	4.98%	1.96%	1.07%	0.48%
Base+Dual+SA(0.3, 0.7)	4.92%	2.58%	0.84%	0.39%
Base+Dual+SA(0.5, 0.5)	**3.16%**	**1.53%**	**0.58%**	**0.31%**
Base+Dual+SA(0.7, 0.3)	5.42%	1.73%	0.62%	0.37%
Base+Dual+SA(0.9, 0.1)	16.22%	9.09%	4.8%	1.84%

Table 4. Setting appropriate coefficients is important to learn discriminative feature representations.

Method	FRR@FAR			EER
	1e-5	1e-4	1e-3	
Base+Dual+SA	3.16%	1.53%	0.58%	0.31%
Base+Dual+SA+Shift	**1.38%**	**0.6%**	**0.31%**	**0.27%**

Table 5. Performance comparison with Shift and No-Shift

4.4. Results and Comparison

We present comparative experimental results with other methods. 2D Gabor filters based IrisCode [7, 6] have been the most widely deployed iris feature descriptor. IrisCode has a number of advanced versions. From the publicly available ones, we selected OSIRIS [32], which is an open source tool for iris recognition. Our image preprocessing also utilizes the OSIRIS for a fair comparison. Another widely accepted method is based on 1D log-Gabor filters [27]. Ordinal filters based method proposed in [40] is also a powerful method of iris feature extracting. Zhao *et al*. proposed UniNet [49] to achieve a highly competitive benchmark, and they also tuned other benchmarks as good performance as possible. We take these tuned methods as benchmarks. The comparison results are shown in Table 6 which show our methods outperform other methods.

5. Conclusion

In this paper we propose a dual spatial attention network, namely DualSANet, to extract dual spatially corresponding iris feature representations for iris recognition. We use pre-trained ResNet-18 as the encoder backbone to extract multi-level features. In decoder architecture, we propose a new spatial attention feature fusion module(SAFFM) to fuse multi-level features. The architecture generates dual discriminative feature representaitons that fuse different level features. An extended triplet loss is utilized to train our network. The experimental results show our DualSANet significantly outperforms state-of-the-art methods on CASIA-IrisV4-Thousand, CASIA-IrisV4-Distance, and IITD.

References

[1] Iitd iris dataset, 2000.

[2] Wageeh W Boles and Boualem Boashash. A human identification technique using images of the iris and wavelet transform. *IEEE transactions on signal processing*, 46(4):1185–1188, 1998.

[3] Kevin W Bowyer and Mark J Burge. *Handbook of iris recognition*. Springer, 2016.

[4] CASIA. Download the separated subsets below, 2000.

[5] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[6] John Daugman. How iris recognition works. In *The essential guide to image processing*, pages 715–739. Elsevier, 2009.

[7] John G Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE transactions on pattern analysis and machine intelligence*, 15(11):1148–1161, 1993.

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[9] Jun Fu, Jing Liu, Haijie Tian, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. *arXiv preprint arXiv:1809.02983*, 2018.

[10] Abhishek Gangwar and Akanksha Joshi. Deepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2301–2305. IEEE, 2016.

[11] Abhishek Gangwar, Akanksha Joshi, Padmaja Joshi, and R Raghavendra. Deepirisnet2: Learning deep-iriscodes from scratch for segmentation-robust visible wavelength and near infrared iris recognition. *arXiv preprint arXiv:1902.05390*, 2019.

	CASIA-IrisV4 -Thousand		CASIA-IrisV4 -Distance		IITD	
	FRR	EER	FRR	EER	FRR	EER
IrisCode(OSIRIS) [32]	9.34%	3.2%	19.93%	6.39%	1.61%	1.11%
IrisCode(log-Gabor) [27]	-	-	20.72%	7.71%	1.81%	1.38%
Oridinal [40]	-	-	16.93%	7.89%	1.70%	1.25%
UniNet-WithinDB [49]	-	-	11.15%	3.85%	1.19%	0.73%
Our-baseline	1.09%	0.40%	13.33%	3.91%	0.89%	0.72%
Our-best	**0.31%**	**0.27%**	**10.67%**	**3.23%**	**0.54%**	**0.45%**

Table 6. Summary of false reject rates(FRR)at 0.1% false accept rate(FAR) and equal error rates(EER) for comparison

[12] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[13] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[14] Fei He, Ye Han, Han Wang, Jinchao Ji, Yuanning Liu, and Zhiqiang Ma. Deep learning architecture for iris recognition based on optimal gabor filters and deep belief network. *Journal of Electronic Imaging*, 26(2):023005, 2017.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[16] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.

[17] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[18] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[19] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[20] Ming Liu, Zhiqian Zhou, Penghui Shang, and Dong Xu. Fuzzified image enhancement for deep learning in iris recognition. *IEEE Transactions on Fuzzy Systems*, 2019.

[21] Nianfeng Liu, Man Zhang, Haiqing Li, Zhenan Sun, and Tieniu Tan. Deepiris: Learning pairwise filter bank for heterogeneous iris verification. *Pattern Recognition Letters*, 82:154–161, 2016.

[22] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[23] Li Ma, Tieniu Tan, Yunhong Wang, and Dexin Zhang. Personal identification based on iris texture analysis. *IEEE transactions on pattern analysis and machine intelligence*, 25(12):1519–1533, 2003.

[24] Li Ma, Tieniu Tan, Yunhong Wang, and Dexin Zhang. Efficient iris recognition by characterizing key local variations. *IEEE Transactions on Image processing*, 13(6):739–750, 2004.

[25] Li Ma, Tieniu Tan, Yunhong Wang, and Dexin Zhang. Local intensity variation analysis for iris recognition. *Pattern recognition*, 37(6):1287–1298, 2004.

[26] Lia Ma, Yunhong Wang, Tieniu Tan, et al. Iris recognition based on multichannel gabor filtering. In *Proc. Fifth Asian Conf. Computer Vision*, volume 1, pages 279–283, 2002.

[27] Libor Masek et al. *Recognition of human iris patterns for biometric identification*. PhD thesis, Master's thesis, University of Western Australia, 2003.

[28] Kazuyuki Miyazawa, Koichi Ito, Takafumi Aoki, Koji Kobayashi, and Hiroshi Nakajima. An effective approach for iris recognition using phase-based image matching. *IEEE transactions on pattern analysis and machine intelligence*, 30(10):1741–1756, 2008.

[29] Donald M Monro, Soumyadip Rakshit, and Dexin Zhang. Dct-based iris recognition. *IEEE transactions on pattern analysis and machine intelligence*, 29(4):586–595, 2007.

[30] Richard Yew Fatt Ng, Yong Haur Tay, and Kai Ming Mok. A review of iris recognition algorithms. In *2008 International Symposium on Information Technology*, volume 2, pages 1–7. IEEE, 2008.

[31] Kien Nguyen, Clinton Fookes, Arun Ross, and Sridha Sridharan. Iris recognition with off-the-shelf cnn features: A deep learning perspective. *IEEE Access*, 6:18848–18855, 2018.

[32] Nadia Othman, Bernadette Dorizzi, and Sonia Garcia-Salicetti. Osiris: An open source iris recognition software. *Pattern Recognition Letters*, 82:124–131, 2016.

[33] Jongchan Park, Sanghyun Woo, Joon-Young Lee, and In So Kweon. Bam: Bottleneck attention module. *arXiv preprint arXiv:1807.06514*, 2018.

[34] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. Deep face recognition. In *bmvc*, volume 1, page 6, 2015.

[35] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.

[36] C Sanchez-Avila, R Sanchez-Reillo, and D de Martin-Roche. Iris-based biometric recognition using dyadic wavelet transform. *IEEE Aerospace and Electronic Systems Magazine*, 17(10):3–6, 2002.

[37] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[38] Noh Seung-In, Kwanghyuk Bae, Yeunggyu Park, and Jaihie Kim. A novel method to extract features for iris recognition system. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 862–868. Springer, 2003.

[39] Zhenan Sun and Tieniu Tan. Ordinal measures for iris recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 31(12):2211–2226, 2008.

[40] Zhenan Sun and Tieniu Tan. Ordinal measures for iris recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 31(12):2211–2226, 2009.

[41] Guillaume Sutra, Sonia Garcia-Salicetti, and Bernadette Dorizzi. The viterbi algorithm at different resolutions for enhanced iris segmentation. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 310–316. IEEE, 2012.

[42] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[43] Xingqiang Tang, Jiangtao Xie, and Peihua Li. Deep convolutional features for iris recognition. In *Chinese Conference on Biometric Recognition*, pages 391–400. Springer, 2017.

[44] Kuo Wang and Ajay Kumar. Toward more accurate iris recognition using dilated residual features. *IEEE Transactions on Information Forensics and Security*, 14(12):3233–3245, 2019.

[45] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 325–341, 2018.

[46] Hengshuang Zhao, Xiaojuan Qi, Xiaoyong Shen, Jianping Shi, and Jiaya Jia. Icnet for real-time semantic segmentation on high-resolution images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 405–420, 2018.

[47] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[48] Hengshuang Zhao, Yi Zhang, Shu Liu, Jianping Shi, Chen Change Loy, Dahua Lin, and Jiaya Jia. Psanet: Point-wise spatial attention network for scene parsing. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 267–283, 2018.

[49] Zijing Zhao and Ajay Kumar. Towards more accurate iris recognition using deeply learned spatially corresponding features. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3809–3818, 2017.

[50] Yingbo Zhou and Ajay Kumar. Personal identification from iris images using localized radon transform. In *2010 20th International Conference on Pattern Recognition*, pages 2840–2843. IEEE, 2010.

[51] Yong Zhu, Tieniu Tan, and Yunhong Wang. Biometric personal identification based on iris patterns. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 2, pages 801–804. IEEE, 2000.