

# Alignment Free and Distortion Robust Iris Recognition

Min Ren<sup>1,2</sup>, Caiyong Wang<sup>1,2</sup>, Yunlong Wang<sup>2</sup>, Zhenan Sun<sup>2</sup>, Tieniu Tan<sup>2</sup>

<sup>1</sup>University of Chinese Academy of Sciences <sup>2</sup>CRIPAC, CASIA, Beijing, P.R. China

{min.ren, caiyong.wang, yunlong.wang} @cripac.ia.ac.cn

{znsun, tnt} @nlpr.ia.ac.cn

## Abstract

*Iris recognition is a reliable personal identification method but there is still much room to improve its accuracy especially in less-constrained situations. For example, free movement of head pose may cause large rotation difference between iris images. And illumination variations may cause irregular distortion of iris texture. To match intra-class iris images with head rotation robustly, the existing solutions usually need a precise alignment operation by exhaustive search within a determined range in iris image preprocessing or brute-force searching the minimum Hamming distance in iris feature matching. In the wild environments, iris rotation is of much greater uncertainty than that in constrained situations and exhaustive search within a determined range is impracticable. This paper presents a unified feature-level solution to both alignment free and distortion robust iris recognition in the wild. A new deep learning based method named Alignment Free Iris Network (AFINet) is proposed, which utilizes a trainable VLAD (Vector of Locally Aggregated Descriptors) encoder called NetVLAD [18] to decouple the correlations between local representations and their spatial positions. And deformable convolution [5] is leveraged to overcome iris texture distortion by dense adaptive sampling. The results of extensive experiments on three public iris image databases and the simulated degradation databases show that AFINet significantly outperforms state-of-art iris recognition methods.*

## 1. Introduction

Iris recognition technology is an attractive topic of biometrics and has been widely applied in person identification because of the rich and unique texture information human iris can provide. Since the first successful algorithm for iris recognition [7], great improvements have been made so far to solve high-quality iris image matching under constrained conditions.

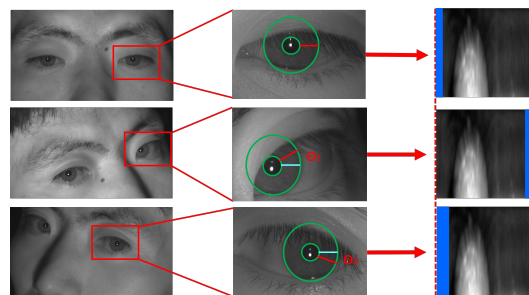


Figure 1. The first row shows the normal iris sample under constrained situations. The rest two rows show the situation without constrain: a glimpse while walking at a distance. The rotation shift of iris may be much greater than constrained situations. The rotation shift of iris becomes horizontally shift in the normalized iris images as shown in the right column.

However, iris recognition in the wild environments, especially where subjects are standing at a distance and not cooperative, is still challenging. Standing distance, gaze direction and shooting angle exhibit significant variations as shown in Figure 1. Two severe obstacles for accurate iris recognition arise: (i) The rotation of iris is of much greater uncertainty than that in constrained situations, which can occasionally reach  $45^\circ$ . (ii) Distortion of iris texture is more serious than constrained situations because of the complicated illumination in the wild.

Currently, almost all existing methods for iris recognition, including handcraft-designed methods and deep learning based methods, need to align iris samples by a cumbersome process during both training and testing [7] [21] [17] [24]: shift the normalized iris image horizontally within a small and determined range to exhaustively search the best position for matching. However, this process only works well in constrained situations. While in the wild environments, where the range of iris rotation is not determined, the method of exhaustion is impracticable to align iris samples beyond the searching range. Moreover, iris texture may have irregular distortions under different illumination conditions. And these distortions are difficult to

be precisely modeled and corrected in preprocessing step. Existing solutions are to design robust features against iris pattern distortions. However considerable high rates of false rejection are still reported in recent iris evaluations because the difference between intra-class iris patterns can not be perfectly resolved in practical scenarios. So misalignment and distortion are important factors affecting iris recognition accuracy in the wild and they are open problems in iris biometrics needing more in-depth investigation.

A novel deep learning based framework named Alignment Free Iris Network (AFINet) is proposed in this paper for both alignment free and distortion robust iris recognition. Our model consists of three components: local feature extractor, spatial-invariant encoder and classifier. Deformable convolution [5] is incorporated into convolutional neural network for dense adaptive sampling to overcome the distortion of iris texture. Inspired by NetVLAD [18], which is a trainable VLAD encoder, we decouple the correlations between local representations and their spatial positions. Full connected layers are adopted as classifier.

The main contributions of this paper are summarized as follows: (i) a new deep learning based framework is proposed for alignment free and distortion robust iris recognition, (ii) results of extensive experiments show that our method remarkably outperforms several state-of-art methods in various settings.

## 2. Related Works

Numerous of iris recognition methods were proposed since the successful introduction of Gabor filters in 1993 [7]. One of the challenge problems for iris recognition is iris feature extraction. 1D log-Gabor filters was proposed [14] for efficient iris feature extraction. Discrete cosine transforms (DCT) was exploited [13] to produce binary iris features by analyzing frequency information of the iris image. 2D discrete Fourier transforms (DFT) proposed in [15] also attempted to match iris images by frequency analysis. An universal framework of iris feature extraction named Ordinal Measures (OMs) was proposed by Sun *et al.* [21], which extract spatial texture features for matching. Adaptive bloom filters were applied for alignment-free cancelable iris biometric [19]. However, this method was a framework tailored for feature transformation rather than feature extraction, so it was limited in general iris feature extraction methods.

Deep learning based methods for iris recognition have been developed in recent years. Liu *et al.* proposed DeepIris in [16] for heterogeneous iris matching by pairwise filter learning. FCN based framework was investigated in [24] for iris recognition and a method named UniNet was proposed. A CNN based method named MaxoutCNNs was proposed in [17] for iris and periocular recognition on mobile devices.

Overall these methods need cumbersome alignment op-

Table 1. Layer configurations of feature extractor.

Layer	Kernel size	Stride	Output channels
Conv1	$9 \times 9$	1	48
Maxpooling1	$2 \times 2$	2	-
DeformConv1	$3 \times 3$	1	-
Conv2	$5 \times 5$	1	96
Maxpooling2	$2 \times 2$	2	-
DeformConv2	$3 \times 3$	1	-
Conv3	$5 \times 5$	1	128
Maxpooling3	$2 \times 2$	2	-
DeformConv3	$3 \times 3$	1	-
Conv4	$4 \times 4$	1	192
Maxpooling4	$2 \times 2$	2	-
DeformConv4	$3 \times 3$	1	-

eration before iris feature extraction and can not handle the distortion of iris texture well.

## 3. Alignment Free Iris Networks

This section describes the proposed deep learning framework for alignment free and distortion robust iris recognition. Iris rotation on 2D image planes, which presents as horizontal shifting in the normalized iris image, is much more serious in the wild environments. To solve this problem, the proposed framework decouples the correlations between local representations and their spatial positions to generate rotation-insensitive representation. Iris texture distortions, usually caused by varying illuminations, are resolved by dense adaptive sampling. AFINet can be trained in an end-to-end fashion for alignment free and distortion robust iris recognition.

### 3.1. Iris Image Preprocessing

For all of the iris images this paper, we apply the same preprocessing procedure which contains four steps: 1) Eye detection using Haar-like Adaboost detectors [22]. 2) Pupilary and iris boundaries localization by method of [8]. 3) Annular iris image normalized to a rectangle image with  $128 \times 128$  resolution by rubber sheet model [7]. 4) Pixel intensity normalization by the mean and standard deviation of each database.

### 3.2. Network Architecture

The proposed network contains three components as shown in Figure 2: local feature extractor, aggregation encoder and classifier.

Superposition of convolution layers is able to extract discriminative representation from raw data [4][23]. This kind of architecture is adopted as the first component which contains four convolution layers. The configurations of the lo-

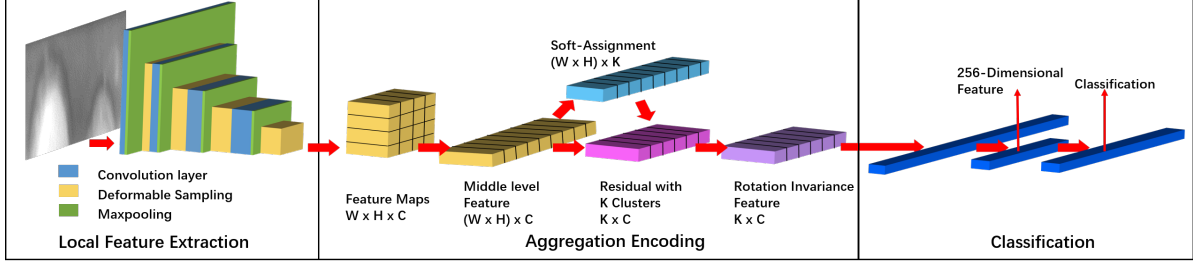


Figure 2. Framework of Alignment Free Iris Network consist of three stages: local feature extractor which overcomes iris texture distortion by dense adaptive sampling while extracts local features from iris images, aggregation encoder which aggregates local features to rotation insensitive representations and classifier which reduces the feature dimension before classification.

cal feature extractor are shown in Table 1. The structure of convolutional layers is based on MaxoutCNNs [17]. Deformable convolution [5] is adopted to overcome iris distortion which will be discussed in next section. All of the convolution layers adopt Maxout activation function and all of the deformable convolution layers use linear activation function.

The local features need to be aggregated to yield rotation-insensitive representations. Actually, feature aggregation is a significant research topic, and some classical methods have been proposed, including bag-of-visual-words [10], VLAD [9][20] and Fisher vector [11][6]. We make use of a trainable aggregation strategy, NetVLAD [18], a variant of the VLAD to decouple the correlations between local representations and their spatial positions, and more details will be shown later.

Two fully connected layers play the parts of dimension reduction and classification. The first fully connected layer reduces the feature dimension to 256 and its output serves as feature for matching. The output of second fully connected layers is used for loss computation.

### 3.3. Deformable Sampling

Iris texture may have irregular distortions under complicated illumination. Although iris images are normalized to a rectangle image with a fixed resolution, the distortions are difficult to be corrected and this problem is still a important factor affecting iris recognition accuracy in the wild.

Deformable convolution provides an effective approach to tackle this problem by dense adaptive sampling. The outputs of deformable sampling layers are adaptive sampled from the inputs. The adaptive sampling operation is trained according to supervised signals. This operation can be viewed as a nonlinear function which adaptively corrects the iris texture distortion and maps the intra-class iris samples to close positions of feature space.

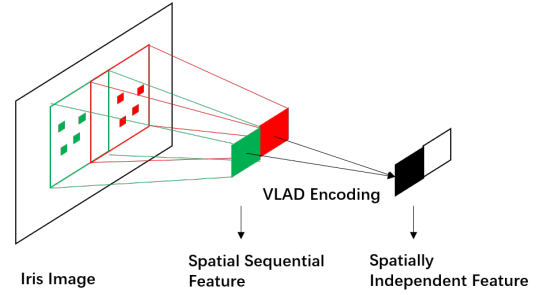


Figure 3. Spatially independent encoding. The similar local patterns of iris image at different positions will activate different nodes of the feature map. After spatially independent encoding, the similar local patterns activate the same nodes and the representation will be rotation insensitive.

### 3.4. Aggregation Encoding

The key component of feature aggregation is the trainable VLAD encoder. Feature maps of layer DeformConv4 with dimension  $W \times H \times C$  are re-organized as  $W \times H$   $C$ -dimension vectors. Each of these vector represents a special patch of the iris image, and they are spatially correlated. The similar local patterns of iris image at different positions are expected to be decoupled with their positions and aggregated into the same representation which means activate the same nodes of network as shown in Figure 3. So, these vectors are expected to be encoded in a spatially independent way for rotation insensitive representation. NetVLAD is a differentiable spatially independent method, and it has been shown to outperform average and max pooling for vector aggregation [18], which makes it perfectly suited for our task.

We provide a brief overview of NetVLAD in this paper (full details are available in [18]). Formally,  $W \times H$   $C$ -dimensional vectors  $\{v_i\}$  are represented by their residuals with  $K$  cluster centers. Each  $v_i$  are assignmented to cluster centers in a soft way. Spatially independent feature, which means rotation insensitive feature in our task, are produced

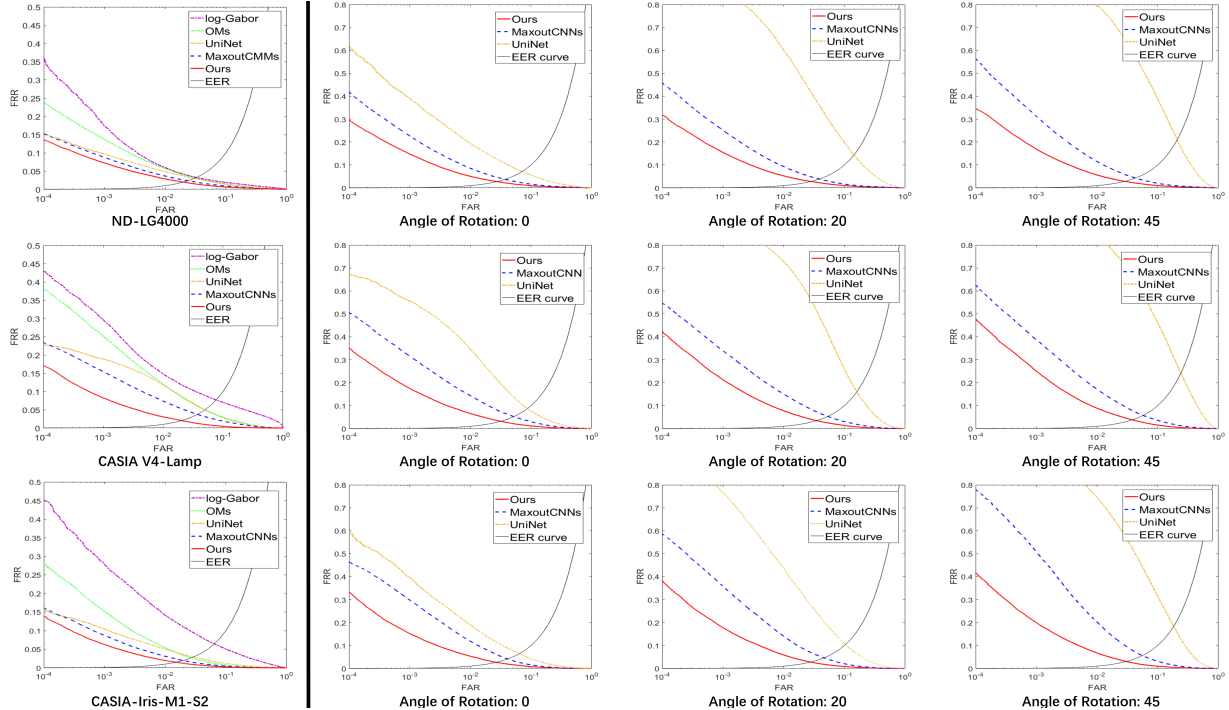


Figure 4. DET curves of the experiments. The first row is DET curves on ND-CrossSensor-Iris-2013 Dataset - LG4000, the second row is DET curves on CASIA Iris Image Database V4 - Lamp, the third row is DET curves on CASIA-Iris-M1-S2. The first column from left shows the results of experiments without iris rotation. The rest three columns shows the results of experiments with iris rotation. The second column from left shows the results under  $0^\circ$  iris rotation, the third column from left shows the results under  $20^\circ$  iris rotation and the right-most column shows the results under  $45^\circ$  iris rotation. Note that the settings of experiments without iris rotation is different from the experiments under  $0^\circ$  iris rotation. The models in experiments without iris rotation were trained on original databases, and the models in experiments under  $0^\circ$  iris rotation were trained on the degraded databases.

according to equation 1 (the dimension is  $K \times C$ ):

$$V(k, j) = \sum_{i=1}^{W \times H} \frac{e^{W_k^T v_i + b_k}}{\sum_{k'} e^{W_{k'}^T v_i + b_{k'}}} (v_i(j) - c_k(j)) \quad (1)$$

where  $\{W_k\}$ ,  $\{b_k\}$ ,  $\{c_k\}$  are trainable parameters, with  $k \in \{1, 2, \dots, K\}$ ,  $j \in \{1, 2, \dots, C\}$ .  $\{c_k\}$  is the cluster centers and the second term of equation 1 represents residuals between middle level feature  $\{v_i\}$  and  $\{c_k\}$ . The first term of equation 1 represents the soft-assignment weight of  $\{v_i\}$  and it can be implemented by  $1 \times 1$  convolution followed by softmax. The final rotation insensitive feature is obtained by performing L2 normalization.

### 3.5. Implementation Details

Network initialization and training details are provided in this section.

The cluster centers of NetVLAD are initialized in the same way of [18]: clustering the local vectors with k-means into  $K$  clusters. The number of clusters is 25. The other parameters of the network are initialized by sampling from a standardized normal distribution.

The first component of the network used as a local feature extractor is pre-trained on each iris database. After that, the VLAD encoder is adopted for training. Cross entropy loss is used for measuring loss of classification. The initial learning rate of local feature extractor is  $10^{-4}$ . And the initial learning rate of VLAD encoder and fully connected layers is  $10^{-2}$ . The learning rates of whole network are divided by 10 until validation error stagnates. The weight decay and momentum, which are  $10^{-4}$  and 0.9 respectively, are fixed during training.

## 4. Experiments and Results

Extensive experiments are conducted and results are shown in this section to evaluate the performance of proposed method.

### 4.1. Databases

**ND CrossSensor Iris 2013 Dataset-LG4000.** This database [1] contains 29,986 iris samples from 1,352 classes and is one of the most popular iris database for iris recognition research. The training set of this database consists of 676 classes which are selected randomly. The testing set

Table 2. False reject rates (FRR) at  $10^{-4}$  false accept rates (FAR) and equal error rates (EER) of experiments without rotation.

	ND-LG4000		CASIA V4-Lamp		CASIA-Iris-M1-S2	
	FRR	EER	FRR	EER	FRR	EER
log-Gabor	35.90%	3.25%	42.93%	7.69%	45.03%	6.57%
OMs	23.88%	3.12%	38.07%	4.99%	28.02%	2.69%
UniNet	15.23%	3.01%	22.76%	4.96%	15.60%	2.95%
MaxouCNNs	15.27%	2.36%	23.36%	3.73%	16.08%	1.97%
Ours	<b>13.55%</b>	<b>1.98%</b>	<b>17.13%</b>	<b>2.04%</b>	<b>13.97%</b>	<b>1.51%</b>

Table 3. False reject rates (FRR) at  $10^{-4}$  false accept rates (FAR) and equal error rates (EER) of experiments with different rotation.

	ND-LG4000		CASIA V4-Lamp		CASIA-Iris-M1-S2	
	FRR	EER	FRR	EER	FRR	EER
<b>Angle difference between sample pairs: 0°</b>						
UniNet	61.71%	7.13%	67.13%	8.74%	59.26%	6.39%
MaxouCNNs	41.81%	3.75%	50.56%	5.19%	46.31%	4.12%
Ours	<b>29.65%</b>	<b>2.76%</b>	<b>35.13%</b>	<b>3.18%</b>	<b>33.19%</b>	<b>2.67%</b>
<b>Angle difference between sample pairs: 20°</b>						
UniNet	96.03%	14.34%	97.84%	16.02%	93.20%	10.42%
MaxouCNNs	45.88%	3.87%	54.65%	5.30%	58.65%	4.60%
Ours	<b>31.82%</b>	<b>2.81%</b>	<b>42.19%</b>	<b>3.47%</b>	<b>38.22%</b>	<b>2.77%</b>
<b>Angle difference between sample pairs: 45°</b>						
UniNet	98.62%	20.96%	98.94%	24.14%	98.70%	18.30%
MaxouCNNs	56.36%	4.32%	62.59%	5.75%	77.91%	5.57%
Ours	<b>34.80%</b>	<b>2.85%</b>	<b>47.87%</b>	<b>3.80%</b>	<b>41.64%</b>	<b>3.04%</b>

is the other 676 classes which are not overlapped with the training set. The training set and test set contain 15,012 and 14,974 samples respectively.

**CASIA Iris Image Database V4-Lamp.** This database [2] contains 16,212 iris samples from 819 classes. A lamp was turned on/off close to the subject to introduce elastic deformation of iris texture due to pupil expansion and contraction under different illumination conditions. So this database is suitable for evaluating the ability of the proposed method to deal with iris distortion. The training set of this database contains 409 classes which are selected randomly. The testing set is the rest 410 classes which are not overlapped with the training set. The training set and test set contain 8,115 and 8,097 samples respectively.

**CASIA-Iris-M1-S2.** This database [3] includes 6000 images captured from 200 subjects. Iris samples were captured by mobile devices at different distances. Both eyes of subjects are captured, so that 400 classes are contained in this database. The training set of this database contains 200 classes which are selected randomly. The testing set is the other 200 classes which are not overlapped with the training set. Hence, the numbers of samples of training set and testing set are both 6,000.

## 4.2. Experiments without Iris Rotation

Experiments on the three databases were launched firstly. The proposed method was compared with the state-of-art approaches, including log-Gabor filters [14], Ordinal Measures (OMs) [21], MaxouCNNs [17] and UniNet [24]. Both handcraft-designed methods and deep learning based methods are contained for comparison.

The results are shown in the first column from left of Figure 4 and Table 2. Significant improvements from proposed method can be observed on all three databases especially on the CASIA V4-Lamp. Such results suggest that the proposed method is a better feature extractor for iris data. Two-folds reasons are considered. Firstly, our proposed method handles small range of iris rotation which exists in these databases better than others. Secondly, the distortion of iris texture is well overcome because of the deformable sampling and this can be observed more obviously on CASIA V4-Lamp which introduced elastic deformation of iris texture under different illumination conditions.

## 4.3. Experiments with Iris Rotation

To evaluate the performance of the proposed method, iris samples of training set were degraded by rotating a random angle, and the angle was sampled from a uniform distribution from  $0^\circ$  to  $45^\circ$ . Three configurations were taken into

consideration during testing:  $0^\circ$ ,  $20^\circ$  and  $45^\circ$  difference between sample pairs for matching (horizontal translations of normalized iris images were applied during implementation according to angles). Handcraft-designed methods were not contained because they can not work without the strategy of exhaustive search.

The results are shown in right three columns of Figure 4 and Table 3. The proposed method surpasses other methods by a large margin, and the gain in accuracy grow with the increase of angle of rotation. These results suggest that our network does capture rotation insensitive features from the iris images, and the representation of iris image produced by our network is suitable for alignment free and distortion robust iris recognition. Experimental evidence are shown in next section.

#### 4.4. Experimental Analysis

To demonstrate the capability of capturing the similar local patterns under different rotation angles, saliency maps [12] were adopted to visualize the important areas of iris image by back-propagation. Saliency maps of the same iris sample under different rotation angle are shown in Figure 5.

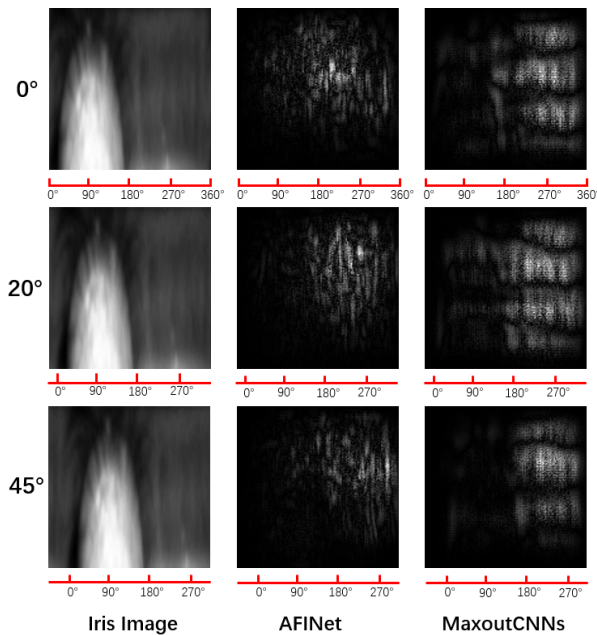


Figure 5. Saliency maps of the same iris sample under different rotation angle. The left column is normalized iris image under three different rotation angles:  $0^\circ$ ,  $20^\circ$ ,  $45^\circ$ . The middle column is the saliency maps of proposed network (the brighter positions represent higher responses which means more important for classification). The saliency maps of MaxoutCNNs are shown in right column as comparison.

The similar high response positions of saliency map of

proposed network move right with the rotation of input iris sample as shown in middle column of Figure 5. This phenomenon dose not appear in the right column, which is the saliency maps of MaxoutCNNs. It is demonstrated that our network capture the similar local patterns of iris image under different rotation angles. This ability is not observed in MaxoutCNNs which has similar architecture with our network except the VLAD encoder and deformable sampling.

## 5. Conclusion

In this paper, we analyze the difficulties of iris recognition in the wild and propose a novel framework for alignment free and distortion robust iris recognition. Extensive experiments demonstrate the effectiveness of our model compared to existing methods. Alignment free and distortion robust are open problems for iris recognition deserving further investigation and iris recognition in the wild is still a challenging problem. Our future work would focus on robust representation for iris sample matching.

**Acknowledgment** This work is supported by the National Natural Science Foundation of China (Grant No. U1836217, 61427811, 61573360, 61721004) and the National Key Research and Development Program of China (Grant No. 2017YFC0821602, 2016YFB1001000). Special thanks to Kunbo Zhang and Yuhao Zhu.

## References

- [1] <https://ce.nd.edu/labs/cvrl/data-sets/biometrics-data-sets>. University of Notre Dame, Computer Vision Research Laboratory, Biometrics Data Sets.
- [2] <http://biometrics.idealtest.org/dbDetailForUser.do?id=4>. CASIA Iris V4.
- [3] <http://biometrics.idealtest.org/dbDetailForUser.do?id=13>. CASIA Iris Mobile V1.
- [4] K. Alex, S. Ilya, and H. G. E. Imagenet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems*, 2012.
- [5] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei. Deformable convolutional networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 764–773, 2017.
- [6] P. Florent, L. Yan, S. Jorge, and P. Herve. Large-scale image retrieval with compressed fisher vectors. In *Computer Vision and Pattern Recognition*, pages 3384–3391, 2010.
- [7] D. J. G. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 15(11):1148–1161, 1993.
- [8] Z. He, T. Tan, Z. Sun, and X. Qiu. Toward accurate and fast iris segmentation for iris biometrics. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 31(9):1670–1684, 2009.

- [9] J. Herve, D. Matthijs, S. Cordelia, and P. Patrick. Aggregating local descriptors into a compact image representation. *Proc Cvpr*, 238(6):3304–3311, 2010.
- [10] P. James, C. Ondrej, I. Michael, S. Josef, and Z. Andrew. Object retrieval with large vocabularies and fast spatial matching. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, 2007.
- [11] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid. Aggregating local image descriptors into compact codes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9):1704–1716, 2012.
- [12] S. Karen, V. Andrea, and Z. Andrew. Deep inside convolutional networks: Visualising image classification models and saliency maps. *Computer Science*, 2013.
- [13] M. D. M., S. Rakshit, and D. Zhang. Dct-based iris recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 29(4):586–595, 2007.
- [14] L. Masek. Recognition of human iris patterns for biometric identification. *The University of Western Australia* 2, 2003.
- [15] 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*, 30(10):1741–1756, 2008.
- [16] L. Nianfeng, Z. Man, L. Haiqing, S. Zhenan, and T. Tieniu. Deepiris: Learning pairwise filter bank for heterogeneous iris verification. *Pattern Recognition Letters*, 82(2):154–161, 2016.
- [17] Z. Qi, L. Haiqing, S. Zhenan, and T. Tieniu. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Transactions on Information Forensics & Security*, PP(99):1–1, 2018.
- [18] A. R, G. P, T. A, P. T, and J. Sivic. Netvlad: Cnn architecture for weakly supervised place recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, PP(99):1–1, 2017.
- [19] C. Rathgeb, F. Breiting, and C. Busch. Alignment-free cancelable iris biometric templates based on adaptive bloom filters. In *IAPR International Conference Biometrics*, pages 1–8, 2013.
- [20] A. Relja and Z. Andrew. All about vlad. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1578–1585, 2013.
- [21] Z. Sun and T. Tan. Ordinal measures for iris recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 31(12):2211, 2009.
- [22] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- [23] L. Yann, B. Leon, B. Yoshua, and H. Patrick. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [24] Z. Zijiang and K. Ajay. Towards more accurate iris recognition using deeply learned spatially corresponding features. In *IEEE International Conference on Computer Vision*, pages 3829–3838, 2017.