Cross-sensor iris recognition using adversarial strategy and sensor-specific information

Jianze Wei, Yunlong Wang, Xiang Wu, Zhaofeng He§, Ran He, Zhenan Sun*
CRIPAC & NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China
University of Chinese Academy of Sciences, Beijing, China
§ Beijing IrisKing Co., Ltd
§ Yingu Building, No.9, Bei SI Huan West Road, HaiDian District, Beijing, China
{jianze.wei, yunlong.wang, xiang.wu}@cripac.ia.ac.cn,{zfhe, rhe, znsun}@nlpr.ia.ac.cn

Abstract

Due to the growing demand of iris biometrics, lots of new sensors are being developed for high-quality image acquisition. However, upgrading the sensor and re-enrolling for users is expensive and time-consuming. This leads to a dilemma where enrolling on one type of sensor but recognizing on the others. For this cross-sensor matching, the large gap between distributions of enrolling and recognizing images usually results in degradation in recognition performance. To alleviate this degradation, we propose Cross-sensor iris network (CSIN) by applying the adversarial strategy and weakening interference of sensor-specific information. Specifically, there are three valuable efforts towards learning discriminative iris features. Firstly, the proposed CSIN adds extra feature extractors to generate residual components containing sensor-specific information and then utilizes these components to narrow the distribution gap. Secondly, an adversarial strategy is borrowed from Generative Adversarial Networks to align feature distributions and further reduce the discrepancy of images caused by sensors. Finally, we extend triplet loss and propose instance-anchor loss to pull the instances of the same class together and push away from others. It is worth mentioning that the proposed method doesn't need pair-same data or triplet, which reduced the cost of data preparation. Experiments on two real-world datasets validate the effectiveness of the proposed method in cross-sensor iris recognition.

1. Introduction

Due to uniqueness and long-term stability of iris, iris recognition has been regarded as one of the most reliable biometrics. However, iris recognition heavily relies on the parameters of sensors, including the optical lens, illumination wavelength, and the diameter of iris [14]. For better recognition performance, over the past decades, many advanced sensors which can capture high-quality iris images were launched. Even though new sensors show amazing verification and recognition accuracies, expensive cost of upgrading sensors and re-enrolling lead us to a dilemma where iris images for enrollment and recognizing are acquired by different types of sensors.

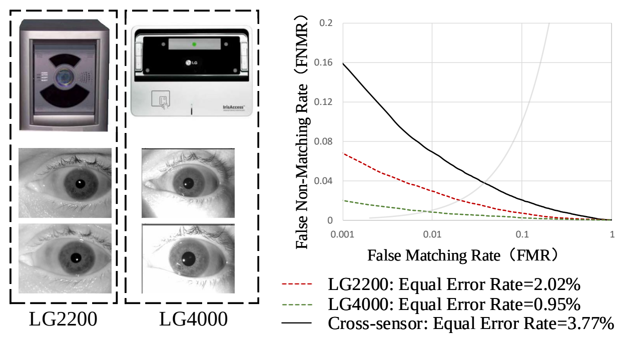


Figure 1. Degradation in performance of cross-sensor matching between LG2200 and LG4000. Left: for the images from LG2200 and LG4000, there exists significant variation in illumination. Right: the value of EER in cross-sensor matching (3.77%) is much larger than that in same-sensor matching (0.95% for LG4000, 2.02% for LG2200).

The reports from the recent publications [7, 3, 5] demonstrate that matching images from different types of sensors, a.k.a. cross-sensor matching, usually degrade performance compared with matching images from the same type of sensors, known as same-sensor matching. Taking cross-sensor matching between LG2200 and LG4000 as an example, the location of illumination, the field of view and camera types are three prominent differences [3] between these two types. These differences make the Equal Error Rate (EER) of cross-sensor matching much higher to that

*Corresponding author

of same-sensor matching (as Figure1). This degradation results from the distribution discrepancy between images acquired by different types of sensors. More specifically, when mapping images from different sensors to a common space, this variation in distribution would increase the intra-class distance and reduce the inter-class distance simultaneously. Thus, narrowing the gaps between distributions is the key to alleviate performance degradation.

To address the distribution discrepancy in cross-sensor matching, Llano et al. [14, 15] explore the possibility of solving problems in pre-processing and propose robust fused segmentation algorithms. However, the final recognition performance is heavily affected by the post-processing method. Thus, many feature-wise methods are proposed, including sparse representation-based method, kernel learning-based method, and Markov random field-based(MRF-based) method. Sparse representation-based methods [24, 30, 31] learn a common sparse dictionary representation to reduce the influence of distribution discrepancy. While kernel learning-based methods [25] improve the metrics and learn kernel matrix to measure the similarity of cross-sensor image pairs. Unlike the previous types of methods, MFR-based methods [13, 21] are built upon a Markov random field to map the recognizing iris coding into the enrolling coding space nonlinearly. However, these methods need good professional knowledge and lots of time to tune their optimal parameters.

Recently, the breakthrough of deep learning in computer vision indicated that feature extraction based on deep learning methods is more competitive than handcrafted feature extraction in exploiting the potential for iris recognition. Applying the deep neural network in iris recognition has become a new way to improve recognition performance, as well as cross-sensor iris matching. Gangwar et al. [8] design a deep neural network and train this model with fine-tuning tricks to solve the cross-sensor matching. However, the weight-shared network in [8, 22] does not consider the variations in textures of images caused by the different types of sensors.

This motivates us to design a two-path network for various textures from different sensors. However, our experiment shows that this two-path network could not provide a satisfactory improvement in cross-sensor matching. The experimental result forces us to consider a new structure combining the shared network and two-path network.

In this paper, we put forward Cross-sensor iris network (CSIN) to address distribution discrepancy problem in cross-sensor matching. For the proposed CSIN, there are three effective ways to learn more discriminative features. Firstly, sensor-specific information is noises rather than discriminative clues in cross-sensor matching. To weaken the influence of the noise, extra convolutional neural networks (CNNs) are employed to extract sensor-specific information as residual components and further decrease the impact of sensor variation on cross-sensor matching. Secondly, due to the success of adversarial strategy in image generation and domain adaptation, the adversarial strategy has become an important and popular solution to the distribution gap. We build sensor adversarial network(SAN) upon this strategy to narrow this gap. Thirdly, for better generalization on unseen data, the instance-anchor loss is developed by introducing metric learning. The developed loss could drag the instances to the corresponding center and push away from other centers.

The main contributions are summarized as follows: 1) We propose CSIN by considering sensor-specific information. In CSIN, sensor-specific information is represented by residual components, and we narrow the distribution gap in cross-sensor matching by removing residual components. 2) Based on the metric learning, instance-anchor loss is proposed to reduce the intra-class gap. Compared with triplet loss, the proposed instance-anchor loss alleviates overfitting problems. 3) The experimental results conducted on two real-world datasets demonstrate that the proposed method shows obvious improvement in cross-sensor matching.

The rest of this paper is organized as follows. In Section 2, we provide a brief review of related works, especially methods for specific domain or tasks. Section 3 presents our proposed method in detail. In Section 4, we give the introduction of datasets and the details of experimental evaluation. Finally, the conclusion is given in Section 5.

2. Related work

2.1. Domain adaptation

Recently, domain adaptation has continuously developed rapidly and drawn widespread attention of researchers. In domain adaptation, it leverages the prior knowledge from one distribution on the similar task of the other distribution.

Compared with cross-sensor matching, both two fields aim to narrow the gap between distributions, while there are still two differences: 1) The popular tasks in domain adaptation are classification and semantic segmentation, which is much easier than the recognition task. 2) The distribution discrepancy in domain adaptation is more difficult to be narrowed compared with cross-sensor matching.

Non-deep-learning methods in domain adaptation can be roughly divided into two categories, instance-based adaptation methods and feature-based adaptation methods. Instance-based adaptation methods weight the data from known distribution to train the classifier, like TrAdaBoost [6] and Transfer Joint Matching (TJM) [18]. Feature-based adaptation methods map the data from different distributions to a common space, such as Transfer Component Analysis (TCA) [23] Joint Distribution Analysis (JDA) [17] etc.

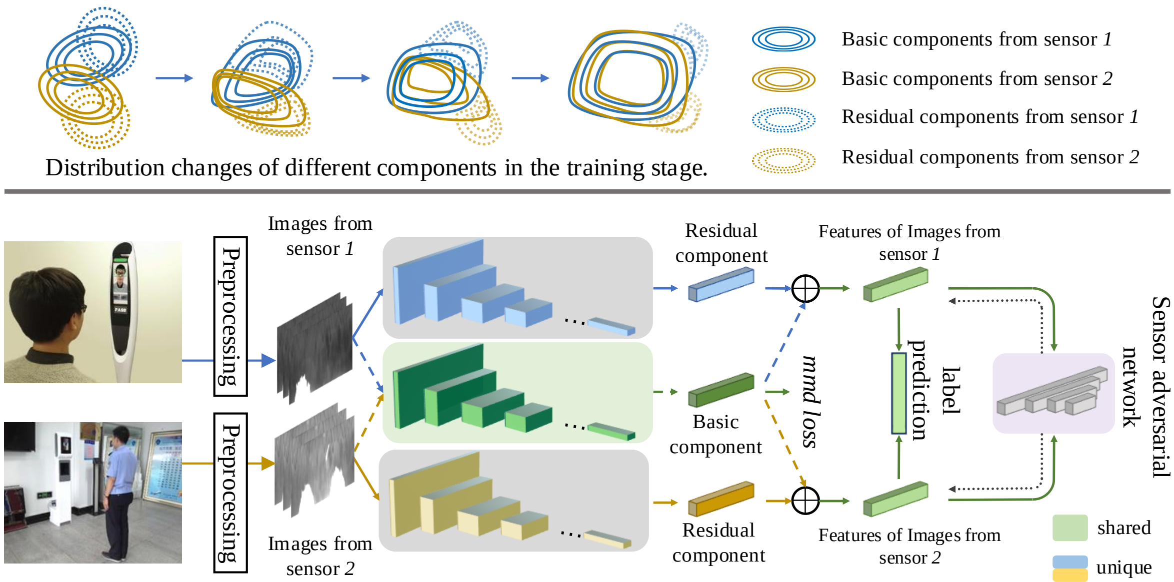


Figure 2. An illustration of our proposed CSIN architecture. For the weight-shared network for cross-sensor matching, there usually exists a large gap between the distributions of extracted features (basic components) from different sensors. And the gap is caused by sensor-specific information and difficult to eliminate. In order to decrease the influence of this gap, we introduce residual components to approximate the distribution gap. Within iterations, minimizing Maximum Mean Discrepancy loss narrows the distribution gap between basic components from different sensors gradually. The adversarial strategy makes it difficult to identify sources of features with less sensor-specific information.

Nowadays, applying deep learning for domain adaptation has become mainstream. In order to reduce the impact of distribution discrepancy, there are two ways to solve it. The first way is to design the loss function to measure the distribution discrepancy, such as the distance between distribution centers or the distance between distribution covariance [27]. The other way borrows the idea from the adversarial strategy of Generative Adversarial Networks (GAN) [9], the approaches attributed to this category narrow the gap between distributions by fooling the domain classifier which predicts the source of data [26, 16].

2.2. Sensor identification

Contrary to cross-sensor matching, sensor identification aims to identify sensor according to the acquired image.

Based on the prior work's conclusion that the noise pattern of images is highly related to sensor, many approaches in the literature are proposed based on noise analysis [11]. Lukas et al. [19] consider the principle of digital imaging and propose an identification method based on Fixed Pattern Noise (FPN) and Photo-Response Non-Uniformity Noise (PRNU). Chen et al. [4] improve the computation of PRNU using maximum likelihood estimate. Bartlow et al. [2] propose a wavelet-based Wiener filtering approach to approximate PRNU of images. Lawgaly et al. [12] observe that bright images and dark images could provide different noise patterns, and further propose weighted averaging-based Sensor Pattern Noise (SPN) estimation. In addition, it is also effective using texture analysis and quality assessment for sensor identification [1].

Last few years, the development of deep learning provides a new direction for sensor identification. Marra et al. [20] present a deep-learning method based on convolutional neural networks (CNN) for sensor identification. The success of this attempt proves that it's feasible to employ neural networks in sensor identification.

3. Proposed approach

Due to the fact that it is difficult for the weight-shared network to reduce the influence of distribution discrepancy, we proposed CSIN with a trident structure, as shown in Figure 2. CSIN employs a shared network and extra networks to extract basic components and residual components of normalized images respectively. Then basic components plus residual components to generate features with less sensor-specific information. For smaller distribution gap, we introduce the adversarial strategy from GAN in SAN.

The detail of the proposed model will be introduced in this section.

3.1. Feature extraction network

Since sensor-specific information frustrates the alignment of distributions from different sensors. To obtain the feature with less sensor-specific information, extra CNNs are added to extract residual components.

Here, for normalized images from sensor i ($i = 1, 2$), I_i, there are two CNN feature extractors to generate feature representation. One for extraction of basic components Θ^b, the other for residual component extraction, Θ_i^r,($i = 1, 2$). The CNN feature extraction process can be denoted as

$$\begin{aligned} f_i^b &= \Theta^b(I_i, \theta) \\ f_i^r &= \Theta_i^r(I_i, \theta_i), \end{aligned} \tag{1}$$

where θ and θ_i denote CNN parameters for basic component extractor and residual component extractors. Then the feature with less sensor-specific information is

$$f_i = f_i^b + \alpha f_i^r, \tag{2}$$

where α is the trade-off parameter.

For smaller distribution gap of basic components from different sensors, MMD (Maximum Mean Discrepancy) loss, a popular loss in domain adaptation, is employed here to narrow the distribution gap. And MMD loss can be written as:

$$\mathcal{L}_{mmd} = \frac{1}{n_1} \sum_{x_1 \in f_1} x_1 + \frac{1}{n_2} \sum_{x_2 \in f_2} x_2, \tag{3}$$

where n_1 and n_2 are the number of instances from sensor 1 and sensor 2 respectively.

In addition, to ensure that the residual components only contain sensor-specific information, some orthogonal losses are necessary. One for less redundancy between basic and residual components, i.e.,

$$\mathcal{L}_{o\_i} = \left\{f_i^b\right\}^T \times f_i^r, (i = 1, 2). \tag{4}$$

The other orthogonal loss for smaller overlap between residual components from different sensors, and it can be formulated as

$$\mathcal{L}_{o\_c} = \{f_1^r\}^T \times f_2^r. \tag{5}$$

For CNN extractors of the proposed model, they can be replaced by arbitrary models, which guarantees the extendibility of the proposed model.

3.2. Sensor adversarial network

The successful application of adversarial strategy in domain adaptation suggests that it is feasible to measure the distribution gap using neural network [26]. This inspires us to apply adversarial network to cross-sensor matching.

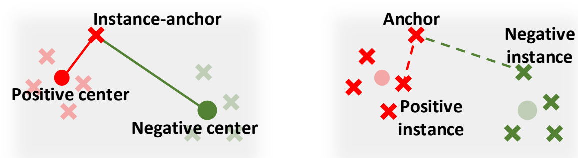
In this paper, we present Sensor adversarial network (SAN) built upon Conditional Domain Adversarial Networks(CDAN) [16] which identifies sensor not only according to features but also according to label predictions. Since both labels from different sensors are available, SAN uses labels of features instead of predictions. And corresponding adversarial loss can be rewritten as

$$\begin{aligned} \mathcal{L}_{ad} = &- \frac{1}{n_1} \sum_{x_1, l_1 \in f_1, y_1} \log\left[\Lambda\left(x_1, l_1, \lambda\right)\right] \\ &- \frac{1}{n_2} \sum_{x_2, l_2 \in f_2, y_2} \log\left[1 - \Lambda\left(x_2, l_2, \lambda\right)\right] \end{aligned} \tag{6}$$

where Λ is SAN and λ is its parameters. n_1 and n_2 are the number of instances from sensor 1 and sensor 2 respectively. The SAN improves the discriminability of features in cross-sensor matching.

3.3. Instance-anchor loss

In cross-sensor matching, distribution discrepancy increases the intra-class distance greatly, resulting in an obvious degradation on final performance. Aiming to reduce the effect of distribution discrepancy, we proposed instance-anchor loss by borrowing the idea from metric learning.



(a) Instance-anchor loss (b) Triplet loss

Figure 3. An example for instance-anchor loss.

Compared with the traditional metric loss, such as triplet loss, tripHard loss, the proposed instance-anchor loss uses the class centers instead of the positive and negative instances of triplet. As shown in Figure 3, instances (plotted as "x") with the same color belong to the same class and their centers are plotted as the corresponding color dots. Triplet loss computes the distances from anchor to positive and negative instances of a triplet. While our loss computes the distances from instance to corresponding center (positive center) and from instance to other centers (negative centers). This change avoids trivial generation of triples and fully exploits the potentiality of mini-batch track. The instance-anchor loss can be written as,

$$\mathcal{L}_{ia} = M + d(f_i, c) - \min(d(f_i, \overline{c})) + \sum_{i \in \{1,2\}} \mathcal{L}_{cl}(p_i, y_i) \tag{7}$$

where $\mathcal{L}_{cl}$ is cross-entropy loss function, p_i is prediction of f_i and y_i is the label of f_i. M denote margin between different classes, and $d(f, c)$ computes the distance between feature f and corresponding center c ($\overline{c}$ denotes the other centers). For better integration with cross-entropy loss, we use cos-distance to measure the distance,

$$d_{cos}(t_1, t_2) = 1 - \frac{t_1^T t_2 / (\|t_1\|_2 \times \|t_2\|_2)}{2}, \tag{8}$$

where t_1 and t_2 are two arbitrary vectors, and set $M = 1$.

3.4. Optimization

With the above definitions, we can derive the following objective function for the proposed CSIN, i.e.

$$\mathcal{L}_{tol} = \mathcal{L}_{ia} + \beta_1 \mathcal{L}_{mmd} + \beta_2 \sum_{i \in \{1,2\}} \mathcal{L}_{o\_i} + \beta_3 \mathcal{L}_{o\_c} - \mathcal{L}_{ad}, \quad (9)$$

Then the minimax game of our method is

$$\max_{\lambda} \mathcal{L}_{ad} \quad (10)$$

$$\min_{\theta, \theta_i (i=1,2)} \mathcal{L}_{tol} \quad (11)$$

This minimax problem can be optimized by the alternative optimization method which is widely used in GAN.

4. Experiment

4.1. Data

In order to evaluate our method, we conduct experiments on two public real-world datasets.

ND cross-sensor dataset

To support the development of cross-sensor iris recognition, Notre Dame University constructs the first publically available dataset. The images from this dataset are collected by two iris sensors, LG2200 and LG4000. And these two sensors are different in the location of illumination, the field of view and camera type. This dataset contains 29,986 images from the LG4000 and 116,564 images from the LG2200, and both sensors acquired eyes of 676 unique subjects.

MIR dataset

Based on the mobile module produced by IrisKing, Chinese Academy of Sciences' Institute of Automation builds the known largest dataset for mobile iris recognition. The dataset is collected at the same time but varied over three collection distances, 20cm, 25 cm, and 30 cm. There are two databases in MIR, MIR-Train for training and MIR-Test for testing. The MIR-Train consists of 4500 images from 150 subjects, while the MIR-Test consists of 12,000 images from 400 subjects. The difficulty of MIR lies in distance changes, eyeglasses, specular reflections, defocus and so on.

4.2. Experiment setting

In order to get a good initialization model to avoid overfitting, we use the following databases for training an initial model: Bath, CASIA-IrisV4 (Thousand, Lamp and Interval), CASIA-CSIR2015. For iris image preprocessing, [13]'s preprocessing method is used and we resize the normalized images to "128×128".

Referring to the previous work, two evaluation protocols are employed in our experiment. The first is 'open set' protocol, and it is adopted for each dataset. In 'open set' protocol, there is no overlap between classes/subjects of the training set and the testing set. The other protocol is 'half-open set', adopted by [13, 8], which means that testing set contains both same and different classes/subjects with the training set. The latter is only used in the experiment on the ND cross-sensor dataset.

Due to the fact that MRI has been divided into two parts and there is no overlap between their classes/subjects, we follow these settings and only process ND cross-sensor (ND) dataset according to above two protocols. For ND dataset, the training set is constructed by selecting the first 100 classes with 10 images for each class. The testing set using 'open set' protocol contains 2686 random selected images from new 223 classes (refer from Table III in [8]) While the testing set using 'half-open set' protocol contains 1343 random selected images from 100 known classes and 1343 random selected images from new 123 classes.

In order to evaluate the proposed method, we compare it with recently proposed methods, including Maxout [32], ResNet [10], LightCNN [28], Deepirisnet [8], Disentangled variational representation(DVR) [29]. Thanks to the implementation friendly provided by the authors[1], the scores of Maxout, LightCNN9, DVR are gained by running their algorithms and we assume that these results are their best performance. For the rest performance scores of compared methods, we directly collect from the authors' publication.

For subsequent experiments, we do not tune these parameters carefully. We set $\alpha = 2$, $\beta_1 = 0.1$, $\beta_2 = 0.1$, $\beta_3 = 0.1$. The length of features that we adopt is 256, which is much small than Deepirisnet (4096). The learning rate is set to 0.001 for all networks. Feature extractors of our model are LightCNN9 [28] for its high-level performance in biometrics.

4.3. Cross-sensor matching

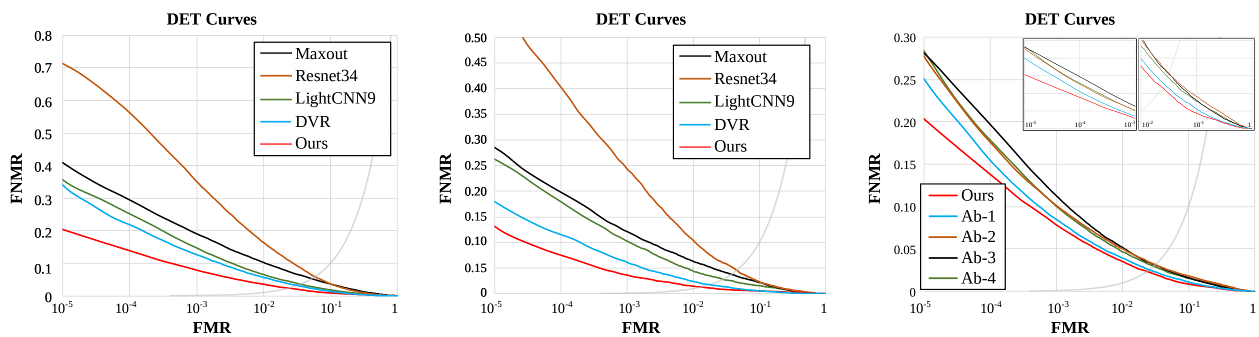
In order to evaluate the recognition performance of the proposed method for cross-sensor matching, we conduct this experiment on a real-world cross-sensor dataset, ND cross-sensor dataset. Table 1 reports the particular experimental results under two different evaluation protocols, and the highest results are highlighted in bold. Furthermore, the DET curves are shown in Figure 4(a) and Figure 4(b).

According to Table 1, Figure 4(a) and Figure 4(b), the proposed CSIN achieves better and comparable results than compared methods. Specifically, our proposed method has the lowest EER, 2.35% using 'open set' protocol and 1.31% using 'half-open set' protocol. More concretely, through three effective efforts including residual component, adversarial strategy and instance-anchor loss, the value of EER drops by 33.80% ('open set' protocol) and 53.55% ('half-open set' protocol) compared with that of LightCNN9, respectively; the value of FNMR@FMR = 10^{-5} drops by 50.21% ('open set' protocol) and 50.92% ('half-open set'

[1] https://github.com/AlfredXiangWu/LightCNN

Table 1. Comparing the performance of the proposed method with some existing approaches on the ND cross-sensor dataset (%).

FNMR	'Open set' protocol				'Half-open set' protocol			
	@FMR=10^{-3}	@FMR=10^{-4}	@FMR=10^{-5}	EER	@FMR=10^{-3}	@FMR=10^{-4}	@FMR=10^{-5}	EER
Maxout [32]	19.06	29.25	40.88	5.24	12.12	19.75	28.69	3.78
ResNet34 [10]	35.11	56.15	71.26	5.86	24.46	40.18	56.56	4.58
LightCNN9 [28]	14.68	25.21	40.87	3.55	10.26	17.85	26.02	2.82
DeepirisNet [8]	–	–	–	–	–	–	–	1.91
DVR [29]	12.64	21.78	34.18	3.19	6.18	11.62	17.75	1.78
Ours	**7.87**	**13.81**	**20.35**	**2.35**	**3.64**	**7.71**	**12.77**	**1.31**



(a) The ND dataset('open set' protocol). (b) The ND dataset ('half-open set' protocol). (c) Ablation study for feature extraction.

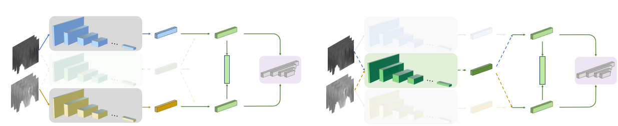
Figure 4. DET curves of different methods and ablation models on the ND cross-sensor datasets. The red solid lines are the results of the proposed method.

protocol) compared with that of LightCNN9, respectively. The more valuable thing to note is that the recognition results using 'open test' are also competitive to the results using 'half-open set' protocol.

In addition, we note that Resnet34 shows the weaker performance of cross-sensor matching compared with other basic networks, i.e. Maxout and LightCNN9. And the gap of performance is more obvious when using 'half-open set' protocol. This result is contrary to our intuition because ResNet34 without pooling layers should be better at feature extraction of texture. We think it is caused by the following. According to the review of Section 2.2, the texture in iris images contains not only iris information but also sensor-specific information. During feature extraction, Resnet34 extracts much sensor-specific information, and the sensor-specific information enlarges the distribution discrepancy, resulting in disturbing the performance in cross-sensor matching. Meanwhile, for CNN models with maxout units, the network with deeper layers shows more robust performance in cross-sensor matching. And we believe that the network with deeper layers could extract higher-level features, which are helpful to resist the influence of sensor-specific information.

4.4. Ablation study for feature extraction

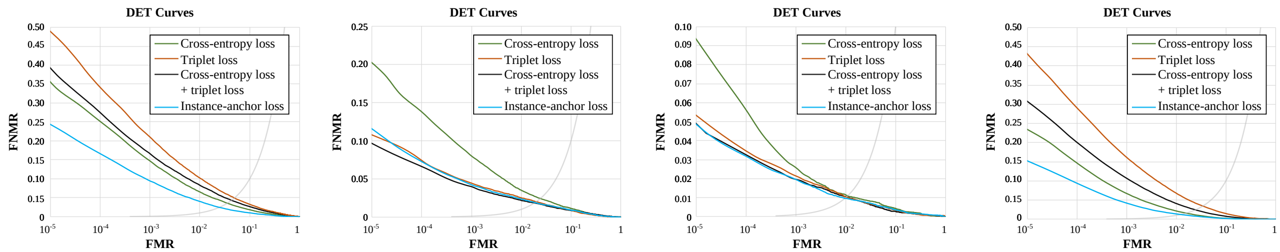
To demonstrate the effectiveness of feature extraction with trident structure in cross-sensor matching, the following ablations are conducted. As shown in Figure 5, we obtain 'Ab-1' and 'Ab-2' networks by removing and only retraining the shared network. 'Ab-3' and 'Ab-4' networks are designed by removing the branch for different sensors. With the same experimental setting as Section 4.3 and 'open set' protocol, we list quantitative results in Table 2, and these results are also shown in Figure 4(c).



(a) Ab-1. (b) Ab-2. (c) Ab-3. (d) Ab-4.

Figure 5. Models for Ablation study. 'Ab-1' removes weight-shared network. 'Ab-2' only retains the weight-shared network. 'Ab-3' discards the weight-speicific network for sensor 1, while 'Ab-4' discards the other weight-speicific network.

From Table 2 and Figure 4(c), we can obtain the following conductions. Firstly, no matter which part is removed, the recognition performance of ablation models would decrease due to the increase of distribution discrepancy. Secondly, our proposed model and ablation models improve



(a) The ND dataset. (b) The ND-2200 dataset. (c) The ND-4000 dataset. (d) The MIR dataset.

Figure 6. DET curves of different losses on the ND cross-sensor, ND-2200, ND-4000 and MIR datasets. The blue solid lines are the results of the proposed loss function.

Table 2. Quantitative results for ablation study on the ND dataset using 'open set' protocol (%).

FNMR	@FMR=10^{-3}	@FMR=10^{-4}	@FMR=10^{-5}	EER
Ours	**7.87**	**13.81**	**20.35**	**2.35**
Ab-1	9.71	17.59	26.66	2.87
Ab-2	10.08	17.99	28.25	2.96
Ab-3	11.32	19.71	28.14	3.12
Ab-4	10.14	17.61	27.81	3.17

the value of EER compared with LightCNN9. These improvements come from the application of adversarial strategy which sensibly aligns distributions from different sensors. Thirdly, the ablation models with only one residual component extractor do not show more competitive performance than previous ablation models We believe that these asymmetrical ablation models extract residual components with wrong sensor-specific information. Thus, the proposed model uses the feature extraction module with a symmetrical structure.

4.5. Same-sensor matching

In this part, we evaluate instance-anchor loss on the ND cross-sensor dataset and MIR dataset. In addition, two subsets of the ND dataset are also used for evaluation, ND-4000 contains all images from LG4000, and ND-2200 contains all images from LG2200. The loss functions for comparison include 1) cross-entropy loss; 2) triplet loss; 3) cross-entropy loss + triplet loss. For a fair comparison, we employ LightCNN9 as feature extractor uniformly. This results of instance-anchor loss and other compared loss are reported in Table 3. The DET curves for these loss functions are plotted in Figure 6.

From Table 3 and Figure 6, it is significant that instance-anchor loss provides better recognition performance than others on ND, ND-4000, MIR datasets and very competitive performance on ND-2200 dataset. Compared with cross-entropy loss and triplet loss, the improvement of instance-anchor loss is over 5.5% on all the above datasets. This improvement is due to that instance-anchor loss provides not only cross-entropy loss but also the metric loss which could reduce intra-class distance and enlarge the inter-class distance simultaneously.

Table 3. Comparing the performance of the proposed instance-anchor loss with some existing losses on the MIR dataset (%). Notation: l_1: cross-entropy loss; l_2: triplet loss; $l3$: cross-entropy loss + triplet loss; l_4: instance-anchor loss.

Dataset	Loss	@FMR=10^{-3}	@FMR=10^{-5}	EER
ND dataset	l_1	14.68	40.87	3.55
	l_2	20.99	48.54	4.91
	l_3	16.64	39.13	4.31
	l_4(ours)	**9.40**	**24.44**	**2.45**
ND-2200	l_1	7.95	20.22	2.43
	l_2	4.43	11.26	1.99
	l_3	**3.97**	**9.81**	**1.83**
	l_4(ours)	4.23	11.59	1.88
ND-4000	l_1	2.57	9.34	1.09
	l_2	2.16	5.26	1.03
	l_3	1.97	4.64	1.04
	l_4(ours)	**1.95**	**4.62**	**0.95**
MIR	l_1	6.72	23.54	1.59
	l_2	16.14	43.05	3.43
	l_3	10.68	30.91	2.43
	l_4(ours)	**4.19**	**14.99**	**1.22**

In addition, we can also observe from Table 3 that cross-entropy loss shows better performance compared with (cross-entropy loss +) triplet loss on ND and MIR datasets. While we obtain a completely opposite conclusion on the ND-2200 and ND-4000 datasets. This is due to that disturb of eyeglasses in the MIR dataset and cross-sensor images in the ND dataset greatly increase the risk of over-fitting. However, Training with triplet loss is easy to get stuck in trouble of overfitting, resulting in serious degradation of performance on the ND and MIR datasets.

5. Conclusion

In this paper, we proposed a CSIN for the cross-sensor matching task. Compared with the previous cross-sensor iris matching, the proposed CSIN narrows the distribution gap by considering sensor-specific information and employing adversarial strategy. Furthermore, we borrow the idea from metric learning and propose instance-anchor loss which decreases the intra-class distance and increases the inter-class distance simultaneously. In the further, we will try to improve performance from image quality enhancement and sensor-specific information elimination.

6. Acknowledgement

This work is supported by National Natural Science Foundation of China (Grant No. 61622310), National Key Research and Development Plan (Grant No. 2017YFC0821602), Beijing Natural Science Foundation (Grant No. JQ18017), Youth Innovation Promotion Association CAS (Grant No. 2015109)

References

[1] A. Agarwal, R. Keshari, M. Wadhwa, M. Vijh, C. Parmar, R. Singh, and M. Vatsa. Iris sensor identification in multi-camera environment. *Information Fusion*, 45:333–345, 2019.

[2] N. Bartlow, N. Kalka, B. Cukic, and A. Ross. Identifying sensors from fingerprint images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 78–84, 2009.

[3] K. W. Bowyer, S. E. Baker, A. Hentz, K. Hollingsworth, T. Peters, and P. J. Flynn. Factors that degrade the match distribution in iris biometrics. *Identity in the Information Society*, 2(3):327–343, 2009.

[4] M. Chen, J. Fridrich, and M. Goljan. Digital imaging sensor identification (further study). In *Proceedings of SPIE*, volume 6505, page 65050P. International Society for Optics and Photonics, 2007.

[5] R. Chu, L. Zhen, Y. Han, H. Ran, and S. Z. Li. Learning gabor magnitude features for palmprint recognition. In *Proceedings of Asian Conference on Computer Vision*, pages 22–31, 2007.

[6] W. Dai, Q. Yang, G. R. Xue, and Y. Yu. Boosting for transfer learning. In *International Conference on Machine Learning*, 2007.

[7] J. S. Doyle, K. W. Bowyer, and P. J. Flynn. Variation in accuracy of textured contact lens detection based on sensor and lens pattern. In *Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–7, 2013.

[8] A. Gangwar and A. Joshi. Deepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition. In *Proceedings of IEEE International Conference on Image Processing*, pages 2301–2305, 2016.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Proceedings of Neural information processing systems*, pages 2672–2680, 2014.

[10] 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*, pages 770–778, 2016.

[11] N. Kalka, N. Bartlow, B. Cukic, and A. Ross. A preliminary study on identifying sensors from iris images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 50–56, 2015.

[12] A. Lawgaly, F. Khelifi, and A. Bouridane. Weighted averaging-based sensor pattern noise estimation for source camera identification. In *Proceedings of IEEE International Conference on Image Processing*, pages 5357–5361, 2014.

[13] N. Liu, J. Liu, Z. Sun, and T. Tan. A code-level approach to heterogeneous iris recognition. *IEEE Transactions on Information Forensics and Security*, 12(10):2373–2386, 2017.

[14] E. G. Llano, J. M. C. Vargas, M. S. García-Vázquez, L. M. Z. Fuentes, and A. A. Ramírez-Acosta. Cross-sensor iris verification applying robust fused segmentation algorithms. In *Proceedings of IAPR International Conference on Biometrics*, pages 17–22, 2015.

[15] E. G. Llano, M. S. G. Vázquez, J. M. C. Vargas, L. M. Z. Fuentes, and A. A. R. Acosta. Optimized robust multi-sensor scheme for simultaneous video and image iris recognition. *Pattern Recognition Letters*, 101:44–51, 2018.

[16] M. Long, Z. Cao, J. Wang, and M. I. Jordan. Conditional adversarial domain adaptation. In *Proceedings of the Neural Information Processing Systems*, pages 1645–1655, 2018.

[17] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu. Transfer feature learning with joint distribution adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2200–2207, 2013.

[18] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu. Transfer joint matching for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1410–1417, 2014.

[19] J. Lukas, J. Fridrich, and M. Goljan. Determining digital image origin using sensor imperfections. In *Image and Video Communications and Processing 2005*, volume 5685, pages 249–261, 2005.

[20] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva. A deep learning approach for iris sensor model identification. *Pattern Recognition Letters*, 113:46–53, 2018.

[21] P. R. Nalla and A. Kumar. Toward more accurate iris recognition using cross-spectral matching. *IEEE Transactions on Image processing*, 26(1):208–221, 2017.

[22] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan. Iris recognition with off-the-shelf cnn features: A deep learning perspective. *IEEE Access*, PP(99):1–1, 2017.

[23] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2011.

[24] J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha. Secure and robust iris recognition using random projections and sparse representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9):1877–1893, 2011.

[25] J. K. Pillai, M. Puertas, and R. Chellappa. Cross-sensor iris recognition through kernel learning. *IEEE Transactions*

on Pattern Analysis and Machine Intelligence, 36(1):73–85, 2014.

[26] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7167–7176, 2017.

[27] D. Wang, P. Cui, and W. Zhu. Deep asymmetric transfer network for unbalanced domain adaptation. In *Proceeding of AAAI Conference on Artificial Intelligence*, pages 443–450, 2018.

[28] X. Wu, R. He, Z. Sun, and T. Tan. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13(11):2884–2896, 2018.

[29] X. Wu, H. Huang, V. M. Patel, R. He, and Z. Sun. Disentangled variational representation for heterogeneous face recognition. In *AAAI Conference on Artificial Intelligence*, 2019.

[30] L. Xiao, Z. Sun, R. He, and T. Tan. Margin based feature selection for cross-sensor iris recognition via linear programming. In *Proceedings of IAPR Asian Conference on Pattern Recognition*, pages 246–250, 2013.

[31] L. Xiao, Z. Sun, H. Ran, and T. Tan. Coupled feature selection for cross-sensor iris recognition. In *Proceedings of IEEE International Conference on Biometrics*, pages 1–6, 2013.

[32] Q. Zhang, H. Li, Z. Sun, and T. Tan. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Transactions on Information Forensics and Security*, 13(11):2897–2912, 2018.