A. Brömme, N. Damer, M. Gomez-Barrero, K. Raja, C. Rathgeb,
A. Sequeira, M. Todisco, and A. Uhl (Eds.): BIOSIG 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1

Low-resolution Iris Recognition via Knowledge Transfer

Fadi Boutros[1 2], Olga Kaehm[1], Meiling Fang [1 2], Florian Kirchbuchner[1], Naser Damer[1 2], Arjan Kuijper[1 2]

Abstract: This work introduces a novel approach for extremely low-resolution iris recognition based on deep knowledge transfer. This work starts by adapting the penalty margin loss to the iris recognition problem. This included novel analyses on the appropriate penalty margin for iris recognition. Additionally, this work presents analyses toward finding the optimal deeply learned representation dimension for the identity information embedded in the iris capture. Most importantly, this work proposes a training framework that aims at producing iris deep representations from extremely low-resolution that are similar to those of high resolution. This was realized by the controllable knowledge transfer of an iris recognition model trained for high-resolution images into a model that is specifically trained for extremely low-resolution irises. The presented approach leads to the reduction of the verification errors by more than 3 folds, in comparison to the traditionally trained model for low-resolution iris recognition.

Keywords: Iris recognition, knowledge transfer, deep learning.

1 Introduction

Iris recognition technology is considered one of the most accurate and robust recognition methods, especially when images are captured with relatively high resolution under appropriate camera settings and user collaboration [Da09]. In certain real-world scenarios such as iris recognition from mobile devices [Zh18] or images from the head-mounted display [Bo20b, Bo20c], iris images are of low resolution and usually captured without user cooperation. Such sub-optimal iris captures and low-resolution iris images lead to degradation in recognition performance [Bo22a, Al19, Bo20c]. This raises the need for biometric solutions that can achieve highly accurate recognition performance from sub-optimal and low-resolution iris captures. A common solution to deal with this challenge is to apply the image super resolution method [DP17, Gu19, Al19]. Deshpande et al. [DP17] proposed a local patch-based framework for iris reconstruction. Guo et al. [Gu19] proposed an iris super-resolution method based on adversarial learning. The image reconstruction method improves the visual appearance of the reconstructed images. However, they might not lead to significant improvement in the recognition performance, as they do not enhance identity information in the reconstructed images [Al19]. For details on iris super resolution methods, we refer to [Al19]. Another direction to improve iris recognition performance under challenging capturing settings is to apply biometric fusion [Bo20a]. However, such methods are usually computational costly (several feature extraction models), and they require processing additional biometric modalities to achieve high recognition performance [Bo20a].

Knowledge transfer (KT), also known as knowledge distillation, is a technique in which a deep knowledge learned by one model is transferred to another model. Ge et al. [Ge19]

[1] Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
[2] Mathematical and Applied Visual Computing, TU Darmstadt, Darmstadt, Germany

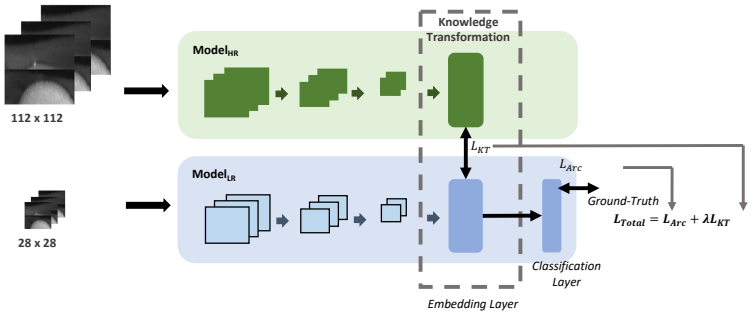

Fig. 1: An overview of the proposed training paradigm for extremely low resolution iris recognition. $Model_{HR}$ is trained with relatively high resolution images of the size 112×112. $Model_{LR}$ is trained with extremely low resolution images of size 28×28. We propose to transfer the knowledge learned by $Model_{HR}$ to $Model_{LR}$. During the training phase of $Model_{LR}$, the model is guided to learn to generate feature representation that behaves similar to the features produced by $Model_{HR}$.

proposed a framework to improve low resolution face recognition performance based on selective knowledge distillation. Moreover, knowledge distillation is commonly adopted to improve compact model accuracy by transferring the knowledge from deeper and more powerful models to shallow models [CH19, Bo22b].

This work proposes improving the low-resolution iris recognition using KT. Specifically, we propose to transfer the knowledge in the embedding space from a model learned from high-resolution iris images to the model learned from extremely low-resolution iris images. We achieved that by guiding the low-resolution iris model during the training to learn to produce feature representation that behaves similarly to the one produced model learned from high-resolution iris images. We also provide three ablation studies on the penalty margin selection of penalty margin softmax loss, the deep feature representation dimensions, and the effectiveness of using data augmentation in the iris recognition model training. The achieved results in this work pointed out the effectiveness of our proposed approach in improving the low-resolution iris recognition performance. For example, the Equal Error Rate (EER) is reduced by our approach on a testing subset (resolution of 28×28) of CASIA-IRIS-M1-S3 [Zh18] from 0.3272 to 0.0907.

2 Approach

This work presents a novel approach for low resolution iris recognition using deep KT. We propose to guide the model trained on low resolution iris images ($Model_{LR}$) to learn to generate feature representation that behaves similarly to the feature representation of model trained on high resolution iris images ($Model_{HR}$). Figure 1 presents an overview of the proposed training paradigm. The training objective of the $Model_{LR}$ is to learn multi-class classification, i.e., Cross-entropy loss, and to learn to produce feature representations f_{LR} that behaves similar to ones f_{HR} produced by $Model_{HR}$. The second learning objective is achieved by minimizing the L_2 distance between f_{LR} and f_{HR}, as it details later in this section. This section presents iris processing, including eye detection, segmentation, and iris unrolling, as they are initial steps prior to iris feature extraction [Zh18]. Then, this section presents deep iris feature extraction and our proposed KT.

Iris processing This work uses CASIA-IRIS-M1-S3 [Zh18] to train the proposed approach (dataset details provided in Section 3). Images in CASIA-IRIS-M1-S3 [Zh18] were acquired using mobile phone equipped with near-infrared (NIR) iris-scanning technology and provided as full face images, as shown in Figure 2. The iris processing includes the following operations: 1) Eye detection and cropping from face images, 2) Iris segmentation, and 3) Iris unrolling.

Given a face image, we first crop the left and right eye area from the input face image. To achieve that, for each input image, we calculate the eye centers (left and right), the width and the height of the eye area. The width and the height of the eye area are calculated based on standard proportion of different face components in human anthropometry [RW10, BSM13] and they are given by:

$$Width_{eye} = 0.67 * \frac{Height_{face}}{2} \tag{1}$$

and

$$Height_{eye} = 0.49 * \frac{Width_{face}}{2}, \tag{2}$$

where $Height_{face}$ and $Width_{face}$ are the height and the width of the face bounding box extracted from multi-task cascaded convolutional neural network (MTCNN) [Zh16]. We consider the eye landmark points obtained from MTCNN as the center of the eye. Once the eye images are cropped, we utilize the Multi-scale Segmentation Network (Eye-MMS) approach [Bo19] to segment the eye images and extract the inner and outer boundary of the iris. Finally, the iris is normalized using the rubber sheet model by unrolling it to a rectangular image [Da04].

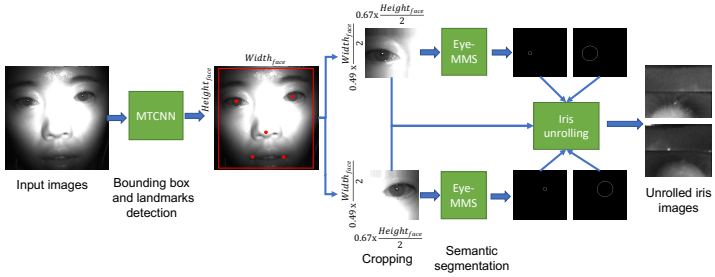

Fig. 2: An overview of data processing. The bounding box and landmarks are first extracted from face images using MTCNN. Then, the left and the right eye areas are cropped. Eye images are further processed by Eye-MMS to obtain the inner- and outer-boundary of the iris. Finally, the iris is unrolled using rubber sheet.

Iris feature extraction ResNet50 [Du20] architecture is used as the backbone of $Model_{HR}$ and $Model_{LR}$ to extract iris feature representation. $Model_{HR}$ and $Model_{LR}$ model are trained with the penalty margin softmax loss. Specifically, we utilize ArcFace loss [De19] as training loss which is given by:

$$L_{Arc} = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{s(cos(\theta_{y_i}+m))}}{e^{s(cos(\theta_{y_i}+m))} + \sum\limits_{j,j\neq y_i}^{c} e^{scos(\theta_j)}}, \tag{3}$$

where θ is the angle between the feature representation and the weights of the last fully connected layer, i.e. θ_{y_i} is the angle between the feature f_{y_i} of a sample i (belongs to class y) and y_i-th class center. N is the batch size, c is the number of classes, s is the scaling parameter and m is the penalty margin.

Template-driven Knowledge transfer This work aims to enhance the performance of the $Model_{LR}$, i.e model with low-resolution images, by driving it to learn to produce feature representations similar to the ones produced by the $Model_{HR}$, i.e model with high-resolution images. We achieve that by adding an additional loss (mean square error loss L_{MSE}) to the main training loss (L_{Arc}). Specially, we propose to minimize the L_2 distance between the feature representations f_{LR} produced by $Model_{LR}$ and the feature representations f_{HR} produced by $Model_{HR}$. The KT loss is given by:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{D}\sum_{h=1}^{D}(f_{HR}^{i}[h] - f_{LR}^{i}[h])^2, \tag{4}$$

where f_{HR}^{i} and f_{LR}^{i} are the feature representation obtained from sample i using $Model_{HR}$ and $Model_{LR}$, respectively. D is the feature representation dimension, i.e. size of the feature extraction layer. N is the batch size. The final combined loss of L_{MSE} and L_{Arc} is given by:

$$L = L_{Arc} + \lambda L_{MSE}, \tag{5}$$

where λ is a weight parameter to balance the two losses during the training. We followed the setting in [Bo22b] and set λ to 100.

3 Experimental setups

Datasets We use CASIA-IRIS-M1-S3 [Zh18] to train and evaluate the presented models in this work. CASIA-IRIS-M1-S3 [Zh18] contains images of 360 identities with 10 images per identity and it is split equally for training and testing i.e. 180 identity for testing and 180 identity for training. The testing set is identity disjoint form the training set. The images were acquired with a mobile phone equipped with NIR iris-scanning technology and provided as face images. The high-resolution ($CASIA-IRIS_{HR}$) iris images are obtained from CASIA-IRIS-M1-S3 as described in Section 2. Each face image results in two iris images (left and right). All images in $CASIA-IRIS_{HR}$ are of size 112×112. The extreme low-resolution ($CASIA-IRIS_{LR}$) are obtained by resizing $CASIA-IRIS_{HR}$ images to 28×28 pixels using the bilinear interpolation. The left and right iris images of each image in $CASIA-IRIS_{HR}$ and $CASIA-IRIS_{LR}$ are considered as of different identities. MTCNN failed to detect faces in 42 images. These images are removed from the dataset.

Evaluation Metrics The presented models are evaluated by using the following metrics [Ma06]: Equal Error Rate (EER) as well as FMR10, FMR100 and FMR1000 which are the lowest false none match rate (FNMR) for a false match rate (FMR) FMR$\leq$ 10.0%, $\leq$ 1.0% and $\leq$ 0.1%, respectively. We also plotted the Receiver Operating Characteristics (ROC) curve and reported the area under the curve (AUC). For each experimental result, we reported the verification performance separately of the left (L) and right (R) iris as well as the overall performance of left and right iris (LR).

Model training setup The presented feature extraction models are implemented using Pytorch. All models are trained with Stochastic Gradient Descent (SGD) optimizer with an initial learning rate of $1e^{-1}$ and a batch size of 256. The learning rate is reduced by a factor of 10 after 200, 280 and 320 epochs. The training is stopped after 400 epochs. The scale parameter of ArcFace is set to 64 following [De19]. For penalty margin of ArcFace, we conduct an ablation study (Section 4) to select optimal penalty margin by training three models with $m = 0.40$, $m = 0.45$ or $m = 0.5$, respectively. These models are trained and tested on $CASIA - IRIS_{HR}$ i.e image size of 112×112 and noted as $Model_{HR} - M4$, $Model_{HR} - 45$, and $Model_{HR} - 5$, respectively. After selecting the optimal penalty margin of ArcFace, we study the optimal deeply learned representation by setting the embedding size to 128, 256 and 512. The trained three models are noted as $Model_{HR} - Em128$, $Model_{HR} - Em256$ and $Model_{HR} - Em512$, respectively. Additionally, as the size of the training dataset is relatively small, we propose to increase the variation of the training data by using RandAugment technique [Cu19]. Finally, the $Model_{LR}$ is trained first on $CASIA - IRIS_{LR}$ without KT, noted as $Model_{LR}$ and then, we repeat the training with the proposed training paradigm (noted as $Model_{LR} - KT$).

4 Ablation study

This section presents three ablation studies on the selection of the optimal penalty margin, the need for data augmentation and the optimal size of deep feature representation.

Penalty margin ArcFace [De19] investigated several penalty margin values and suggested that penalty margin of 0.5 is the optimal for face recognition. However, the identity nature in iris might differ from the well-studied penalty margins in face recognition. Thus, we investigate in this ablation study the selection of optimal penalty margin by training three instances of $Model_{HR}$ with ArcFace loss and margin of $m = 0.40$, $m = 0.45$ and $m = 0.5$, respectively. The achieved verification performances of these models are presented in Table 1. It can be observed that the model trained with margin of $m = 0.45$ achieved the best verification performance in term of FMR10, FMR100 and FMR1000 and obtained a very close EER value to the model trained with $m = 0.5$. Therefore, we fixed the margin to $m = 0.45$ in the rest of the experiments.

Margin	**ERR**			**FMR10**			**FMR100**			**FMR1000**		
	L	R	LR	L	R	LR	L	R	LR	L	R	LR
0.40	0.0781	0.0825	0.0810	0.0712	0.0759	0.0753	0.1481	0.1579	0.1597	0.2607	0.2538	0.2642
0.45	0.0769	0.0802	0.0796	0.0695	0.0746	**0.0733**	0.1443	0.1421	**0.1471**	0.2497	0.2365	**0.2457**
0.50	0.0787	0.0794	**0.0794**	0.0719	0.0726	0.0734	0.1510	0.1520	0.1547	0.2498	0.2550	0.2557

Tab. 1: Evaluation results of different penalty margins.

Data augmentation In this experiment, the impact of data augmentation is analysed. In our work, we adapt RandAugment [Cu19] technique to increase the variety of the training data, as the used CASIA-IRIS-M1-S3 [Zh18] is a relatively small-scale dataset. RandAugment [Cu19] uses a simple parameterization for targeting augmentation to the particular model and dataset sizes. The results showed that RandAugment obtained a notable performance improvement with minimal computational cost, i.e., only two hyperparameters. In our work, we set the number of operation of $N = 2$ and the magnitude of $M = 9$ as suggested in [Cu19]. Then, we trained two instances of $Model_{HR}$ with and without data augmentation, respectively. The achieved results of these models are presented in Table 2.

It can be clearly noticed that including data augmentation is beneficial for improving the iris recognition performance.

Data augmentation	**EER**			**FMR10**			**FMR100**			**FMR1000**		
	L	R	LR	L	R	LR	L	R	LR	L	R	LR
w/o	0.0769	0.0802	0.0796	0.0695	0.0746	0.0733	0.1444	0.1421	0.1471	0.2497	0.2365	**0.2457**
w.	0.0705	0.0653	**0.0686**	0.0632	0.0545	**0.0588**	0.1322	0.1312	**0.1312**	0.2431	0.2516	0.2512

Tab. 2: The achieved verification performances of models trained without (first row) and with data augmentation (second row).

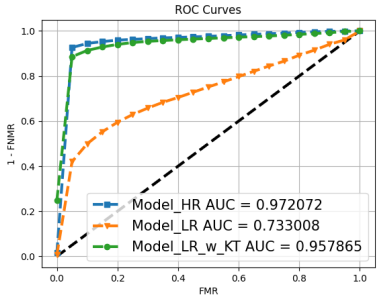

Fig. 3: ROCs achieved by $Model_{HR}$, $Model_{LR}$ and $Model_{LR}$ trained with KT.

Deep feature representation dimension Given that the optimal penalty margin is 0.45 and data augmentation is beneficial for iris recognition, we further conduct in this experiment an evaluation on the dimensions of deep feature representation. We trained three instances of $Model_{HR}$ with feature dimension of 128, 256 and 512. The achieved results of these three models are presented in Table 3. The best verification performance is achieved by the model trained with feature dimension of 256.

F. dimension	**EER**			**FMR10**			**FMR100**			**FMR1000**		
	L	R	LR	L	R	LR	L	R	LR	L	R	LR
128	0.0705	0.0653	0.0686	0.0632	0.0545	0.0588	0.1322	0.1312	0.1312	0.2431	0.2516	0.2512
256	0.0665	0.0637	**0.0654**	0.0579	0.0529	**0.0556**	0.1247	0.1219	**0.1249**	0.2244	0.2395	**0.2325**
512	0.0682	0.0651	0.0673	0.0610	0.0561	0.0576	0.1297	0.1242	0.1300	0.2348	0.2410	0.2412

Tab. 3: Evaluation results of models with different features dimensions. The model with a feature dimension of 256-D leads to the best verification performance based on all considered metrics.

5 Results

We evaluated and reported the verification performance of the model trained and tested on extreme low resolution iris images ($model_{LR}$) i.e. 28×28. The achieved results are presented as part of Table 4. It can be observed that the verification performances are significantly degraded using extremely low resolution iris images ($model_{LR}$) in comparison to the case when the model is trained and tested on relatively high resolution iris images ($model_{HR}$). This degradation in the verification performance can be seen in all reported evaluation metrics. For example, EER value is increased from 0.0654 (obtained by $Model_{HR}$) to 0.3272 (obtained by $Model_{LR}$).

When the $Model_{LR}$ model is trained and evaluated with our proposed KT approach, it can be clearly observed that our proposed approach significantly enhanced the verification performance of low-resolution iris, as shown in Table 4 and Figure 3. $Model_{LR}$ with KT outperformed $Model_{LR}$ in terms of all evaluation metrics. For example, the EER is

significantly reduced from 0.3272 to 0.0907, corresponding to an improvement of around 72.28%. It can also be noticed that the $Model_{LR}$ with KT achieved competitive results to the $Model_{HR}$. For example, the achieved EERs by $Model_{HR}$ and $Model_{LR}$ with KT are 0.0654 and 0.0907, respectively. These achieved results proved the effectiveness of our proposed approach in improving the low resolution iris recognition.

Model	**EER**			**FMR10**			**FMR100**			**FMR1000**		
	L	R	LR	L	R	LR	L	R	LR	L	R	LR
$Model_{HR}$	0.0665	0.0637	0.0654	0.0579	0.0529	0.0556	0.1247	0.1219	0.1249	0.2244	0.2395	0.2325
$Model_{LR}$	0.3157	0.3324	0.3272	0.4811	0.5142	0.5012	0.7069	0.7310	0.7192	0.8424	0.8499	0.8486
$Model_{LR}$ w. KT	0.0945	0.0868	**0.0907**	0.0927	0.0814	**0.0867**	0.1826	0.1783	**0.1796**	0.2938	0.3104	**0.3046**

Tab. 4: The achieved verification performance by our proposed KT training paradigm. The results in the first row are achieved by $Model_{HR}$. The results in the second row are achieved by $Model_{LR}$ trained only with ArcFace loss using extremely low resolution images of 28×28. The results in the third row are achieved by $Model_{LR}$ trained using extremely low resolution images with our proposed knowledge transformation from $Model_{HR}$. The results showed that our proposed approach significantly enhanced the iris recognition accuracy on extremely low resolution images.

6 Conclusion

We presented in this paper a novel approach to improving low-resolution iris recognition by transferring the knowledge from the model learned on high-resolution iris images to the model learned on extremely low-resolution iris images. Three different ablation studies are conducted in this work to analyze and study the penalty margin selection, the deep feature representation dimensions, and the effectiveness of using random data augmentation in the iris recognition model training. By demonstrating extensive experiment evaluations on the CASIA-IRIS-M1-S3 dataset, we proved the benefits of the proposed approach in improving the verification performance of extremely low-resolution iris recognition. The model trained with our proposed method significantly improved the verification performance compared to the model trained with conventional classification learning.

Acknowledgment: This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE. This work has been partially funded by the German Federal Ministry of Education and Research (BMBF) through the Software Campus Project.

References

[Al19] Alonso-Fernandez, Fernando; Farrugia, Reuben A.; Bigün, Josef; Fiérrez, Julian; Gonzalez-Sosa, Ester: , A Survey of Super-Resolution in Iris Biometrics With Evaluation of Dictionary-Learning, 2019.

[Bo19] Boutros, Fadi; Damer, Naser; Kirchbuchner, Florian; Kuijper, Arjan: Eye-MMS: Miniature Multi-Scale Segmentation Network of Key Eye-Regions in Embedded Applications. In: ICCV Workshops. IEEE, pp. 3665–3670, 2019.

[Bo20a] Boutros, Fadi; Damer, Naser; Raja, Kiran B.; Ramachandra, Raghavendra; Kirchbuchner, Florian; Kuijper, Arjan: Fusing Iris and Periocular Region for User Verification in Head Mounted Displays. In: FUSION. IEEE, pp. 1–8, 2020.

[Bo20b] Boutros, Fadi; Damer, Naser; Raja, Kiran B.; Ramachandra, Raghavendra; Kirchbuchner, Florian; Kuijper, Arjan: Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation. Image Vis. Comput., 104:104007, 2020.

[Bo20c] Boutros, Fadi; Damer, Naser; Raja, Kiran B.; Ramachandra, Raghavendra; Kirchbuchner, Florian; Kuijper, Arjan: On Benchmarking Iris Recognition within a Head-mounted Display for AR/VR Applications. In: IJCB. IEEE, pp. 1–10, 2020.

[Bo22a] Boutros, Fadi; Damer, Naser; Raja, Kiran B.; Kirchbuchner, Florian; Kuijper, Arjan: Template-Driven Knowledge Distillation for Compact and Accurate Periocular Biometrics Deep-Learning Models. Sensors, 22(5):1921, 2022.

[Bo22b] Boutros, Fadi; Siebke, Patrick; Klemt, Marcel; Damer, Naser; Kirchbuchner, Florian; Kuijper, Arjan: PocketNet: Extreme Lightweight Face Recognition Network Using Neural Architecture Search and Multistep Knowledge Distillation. IEEE Access, 10:46823–46833, 2022.

[BSM13] Bakshi, Sambit; Sa, Pankaj Kumar; Majhi, Banshidhar: Optimized Periocular Template Selection for Human Recognition. BioMed Research International, 2013, 2013.

[CH19] Cho, Jang Hyun; Hariharan, Bharath: On the Efficacy of Knowledge Distillation. In: ICCV. IEEE, pp. 4793–4801, 2019.

[Cu19] Cubuk, Ekin D.; Zoph, Barret; Shlens, Jonathon; Le, Quoc V.: , RandAugment: Practical automated data augmentation with a reduced search space, 2019.

[Da04] Daugman, John: Iris recognition border-crossing system in the UAE. International Airport Review, 8(2), 2004.

[Da09] Daugman, John: How iris recognition works. In: The essential guide to image processing, pp. 715–739. Elsevier, 2009.

[De19] Deng, Jiankang; Guo, Jia; Xue, Niannan; Zafeiriou, Stefanos: ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: CVPR. Computer Vision Foundation / IEEE, pp. 4690–4699, 2019.

[DP17] Deshpande, Anand; Patavardhan, Prashant: Super resolution and recognition of long range captured multi-frame iris images. IET Biometrics, 6:360 –368, 09 2017.

[Du20] Duta, Ionut Cosmin; Liu, Li; Zhu, Fan; Shao, Ling: Improved Residual Networks for Image and Video Recognition. In: ICPR. IEEE, pp. 9415–9422, 2020.

[Ge19] Ge, Shiming; Zhao, Shengwei; Li, Chenyu; Li, Jia: Low-Resolution Face Recognition in the Wild via Selective Knowledge Distillation. IEEE Trans. Image Process., 28(4):2051–2062, 2019.

[Gu19] Guo, Yanqing; Wang, Qianyu; Huang, Huaibo; Zheng, Xin; He, Zhaofeng: Adversarial Iris Super Resolution. In: ICB. IEEE, pp. 1–8, 2019.

[Ma06] Mansfield, A: Information technology–Biometric performance testing and reporting–Part 1: Principles and framework. ISO/IEC, pp. 19795–1, 2006.

[RW10] Ramanathan, Venkatesh; Wechsler, Harry: Robust human authentication using appearance and holistic anthropometric features. Pattern Recognition Letters, 31(15):2425–2435, 2010.

[Zh16] Zhang, Kaipeng; Zhang, Zhanpeng; Li, Zhifeng; Qiao, Yu: Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. IEEE Signal Process. Lett., 23(10):1499–1503, 2016.

[Zh18] Zhang, Qi; Li, Haiqing; Sun, Zhenan; Tan, Tieniu: Deep Feature Fusion for Iris and Periocular Biometrics on Mobile Devices. IEEE Trans. Inf. Forensics Secur., 13(11):2897–2912, 2018.