Distribution Alignment for Cross-device Palmprint Recognition

Lei Shen[a], Yingyi Zhang[a], Kai Zhao[a,b,*], Ruixin Zhang[a], Wei Shen[c]

[a] *Tencent Youtu Lab*
[b] *University of California, Los Angeles*
[c] *Shanghai Jiaotong University*

Abstract

With the development of IoT and mobile devices, cross-device palmprint recognition is becoming an emerging research topic in multimedia for its great application potential. Due to the diverse characteristics of different devices, *e.g.*resolution or artifacts caused by post-processing, cross-device palmprint recognition remains a challenging problem. In this paper, we make efforts to improve cross-device palmprint recognition in two aspects: 1) we put forward a novel distribution-based loss to narrow the representation gap across devices, and 2) we establish a new cross-device benchmark based on existing palmprint recognition datasets. Different from many recent studies that only utilize instance-level or pairwise-level information between devices, the proposed progressive target distribution loss (PTD loss) uses the distributional information. Moreover, we establish a *progressive target* mechanism that will be dynamically updated during training, making the optimization easier and smoother. The newly established benchmark contains more samples and more types of IoT devices than previous benchmarks, which can facilitate cross-device palmprint research. Extensive comparisons on several benchmarks reveal that: 1) our method outperforms other cross-device biometric recognition approaches significantly; 2) our method presents superior performance compared to SOTA competitors on several general palmprint recognition benchmarks; Code and data are openly available at

*Kai Zhao is the Corresponding author (kz@kaizhao.net).
Lei Shen and Yingyi Zhang contribute equally to this paper.

https://kaizhao.net/palmprint.

Keywords: Palmprint recognition, Deep learning, Loss function, Biometric recognition

1. Introduction

Recently, the palm payment device being developed by Amazon [1] has attracted people's attention. Compared with card payments, palm payments will significantly reduce the waiting time and bring great convenience to people. And the safety of the palmprint recognition system has been confirmed by numerous previous studies [2]. A successful palm payment system will revolutionize the way we do our shopping and significantly impact our daily life in the coming years.

During the last decade, palmprint recognition has switched from early contact and restricted images to unrestricted contactless images. And recently, some studies deal with images taken by mobile phones, such as MPD [3] and XJTU-UP [4]. However, besides smartphones, the devices used in payment application scenarios are diverse and include various IoT devices, such as the aforementioned AmazonOne [1]. Cross-device registration and identification improve payment convenience more effectively. For instance, customers can use their smartphones to register at home and use IoT devices to complete identification and payment at stores. Yet, the heterogeneous visual characteristics between images across devices pose a huge challenge. Existing methods, such as ArcPalm [3], do not take into account the situation of domain shift, which leads to their poor performance in cross-device recognition scenarios.

As shown in Fig. 1 (a) and (b), pictures taken by different devices vary greatly in appearance. Consequently, as illustrated in Fig. 1 (c), though belonging to the same identity, images from different devices present large intra-class variance in the embedding space. This will largely affect the recognition performance on unseen data. To alleviate this problem, we propose a novel loss function, namely Progressive Target Distribution loss (PTD loss), that mini-

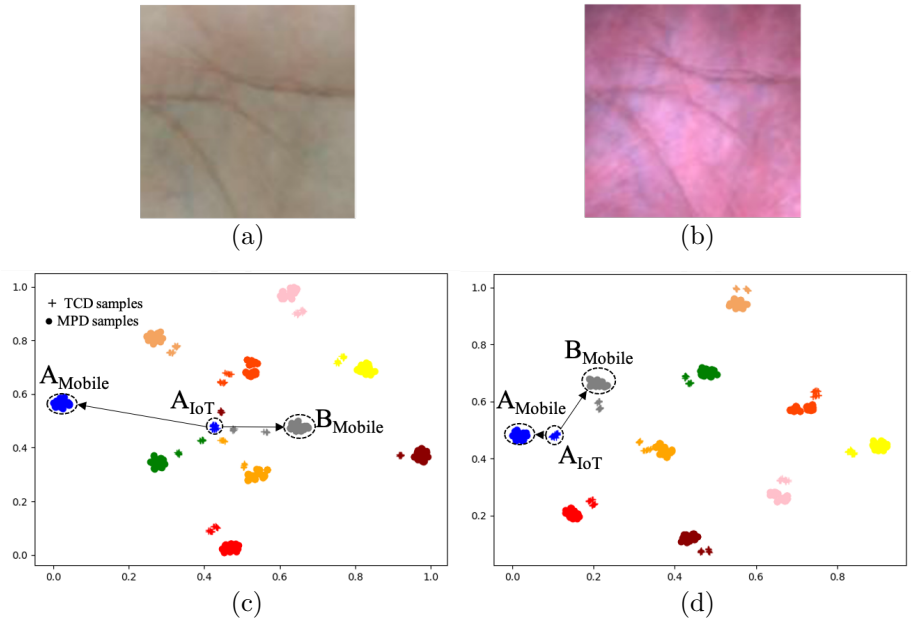

Figure 1: Top: Images taken by mobile phones (a) and IoT devices(b) are diverse in appearance. Bottom: TSNE of palm features extracted by ArcPalm [3] (c) and our proposed method (d). Our method can shorten the distance between samples of the same identity but different devices, *e.g.*A_{Mobile} and A_{IoT}

mizes the distributional gap between cross-device and within-device similarities. Specifically, we estimate the histogram of similarity scores and minimize the distributional distance between the estimated histogram and target distribution. Our innovation is reflected in our proposed 'progress target'. Instead of setting up a fixed target, we propose the 'progressive target' that adjusts to each individual mini-batch.

Additionally, there are not sufficient datasets for cross-device RGB palmprint recognition. We collect a new cross-device palmprint recognition benchmark based on existing palmprint recognition datasets. The newly collected dataset consists of RGB images from different devices, *e.g.*IoT devices, and mobile phones, and it is larger than existing datasets. The main contributions of this work are summarized below.

- We propose a novel and simple loss term to narrow the cross-device gap at the pairwise level. we propose to use a 'progressive target' to guide the

estimated histogram.

- We established a new cross-device palmprint benchmark to improve the study of cross-device RGB palmprint recognition based on existing datasets.
- Extensive experiments on the newly collected dataset and several existing palmprint recognition datasets verified the effectiveness of the proposed method.

We also tested the proposed method on cross-device person re-identification and the results are also in favor of our method.

The remaining of this paper is organized as follows. In Sect. 2, related studies are reviewed. Sect. 3 states the details of the proposed PTD loss. Experimental results and dataset details are presented in Sect. 4. Finally, Sect. 5 concludes the paper.

2. Related Work

2.1. Traditional Palmprint Recognition Methods.

Traditional palmprint recognition methods can be roughly divided into two categories: 1) holistic-based and 2) local descriptor based. In holistic-based methods, features are extracted from the whole image and then projected to a lattent space of lower-dimensional to make it more compact and discriminative. Supervised and unsupervised projection methods such as Principle Component Analysis (PCA) [5], Independent Component Analysis (ICA) [6], Mix Factor Analysis (MFA) [7], and Linear Discriminant Analysis (LDA) [8] are used for dimension reduction. Gui *et al.* [9] use Locality Preserving Projection (LPP) to preserve local structures of palmprints. Hu *et al.* [10] extend LPP to 2D-LPP. Holistic-based methods often suffer from degradation caused by distortion, illumination, and noise. To overcome these issues, some studies try to transform the data from image domain to another domain. Frequency [11] and Cosine [12] transforms are commonly used to overcome these degradations.

Local descriptor based methods extract local features of the image and then construct a global description by feature fusion. The coding-based methods have the advantages of small storage space and fast processing speed, but the accuracy is slightly inferior. Earlier in [13], Zhang *et al.* designed the classic competitive code (CompCode) by using multiple 2-D Gabor filters to extract orientation information. FastCompCode [14] proposes a binary code for effective representation and matching. Li *et al.* [15] extract the local micro-structure tetra patterns. Jia *et al.* [16] propose the robust line orientation code (RLOC) [16]. Wu *et al.* [17] extract local SIFT features and match palm images with RANSAC. Qian *et al.* [18] extract histogram of orientations.

2.2. Deep Learning for Palmprint Recognition.

Many recent studies use deep learning as feature extractor for palmprint recognition. In general, these deep learning based methods belong to holistic-based palmprint recognition. Dian *et al.* [19] first use the AlexNet as the feature extractor and match palm images with Hausdorff distance. Svoboda *et al.* [20] propose a new loss function related to the d-prime index. Shao *et al.* [21] use deep metric learning to obtain discriminative features. Zhao *et al.* [22] design a novel CNN architecture for generic palmprint recognition in numerous scenarios. Recently, margin-based loss functions have been proven to be effective for face recognition. The large margin loss [23] and additive angular margin loss [3] are introduced to palmprint recognition. Fei *et al.* [24] propose a compact CNN-based surface representation for 3D palmprint recognition. Different from these studies that introduce new architectures or loss functions, our proposed method focuses on synthesizing training data for deep palmprint recognition.

3. Proposed Method

As mentioned before, the heterogeneous visual characteristics existing between cross-device images pose a huge challenge. According to Fig. 1, the selected baseline ArcPalm [3] performs poorly in cross-device recognition sce-

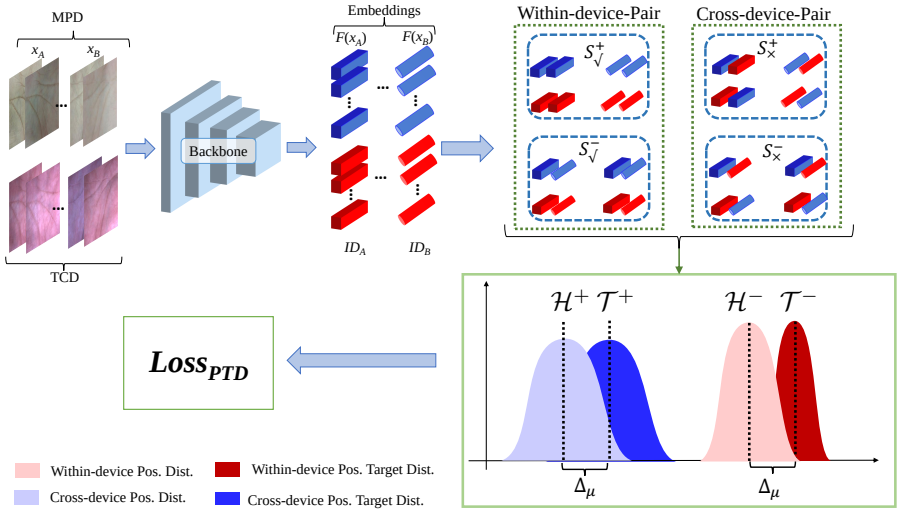

Figure 2: Overall framework of the target distribution loss. During training, each mini-batch contains M IDs, and each ID contains N samples. Colors (red and blue) represent embeddings of different devices and shapes represent different identities.

narios. To solve the domain shift problem caused by heterogeneous image characteristics, we propose a PTD loss based on the baseline ArcPalm, which improves the cross-device recognition performance of the model by expanding the difference between the intra-class and inter-class similarity. The overall framework of the PTD loss is shown in Fig. 2. In this section, we will expand the principle and details of the PTD loss.

3.1. From Pairwise Similarity to Histogram

Pairwise Similarity. Let I be an arbitrary input image and $\mathbf{F}$ be the feature extractor, a convolutional neural network in our case. We denote the extracted features as $X = \mathbf{F}(I)$, for simplicity we use X to represent a sample. The similarity between a pair of samples X_i, X_j is defined as their cosine distance in the embedding space:

$$s_{i,j} = \cos(X_i, X_j) = \frac{X_i \cdot X_j}{||X_i|| \cdot ||X_j||}.$$

Positive and Negative Pairs. Given an arbitrary pair of samples $\langle X_i, X_j \rangle$

and we denote their identity label as y_i, y_j. A pair is referred to as the *positive pair* if $y_i = y_j$ and otherwise the *negative pair*. Given a batch of samples $B = \{X_i, X_2, ..., X_N\}$ where N is the batch size, we construct positive pairs and negative pairs according to their identity labels:

$$\begin{aligned} \mathcal{S}^+ &= \big\{\langle X_i, X_j\rangle \;\mid\; \forall X_i, X_{j,i\neq j} \in B, y_i = y_j\big\} \\ \mathcal{S}^- &= \big\{\langle X_i, X_j\rangle \;\mid\; \forall X_i, X_{j,i\neq j} \in B, y_i \neq y_j\big\}, \end{aligned} \tag{1}$$

where $\mathcal{S}^+$ and $\mathcal{S}^-$ are the collections of positive/negative pairs. Note that the construction of positive pairs requires the sophisticated design of the batch sampler because if randomly sampled, there may be no positive pairs. More information about the batch sampler will be detailed in the experiments.

Cross-device and Within-device Pairs. Our proposed loss function can be used to improve cross-device palmprint recognition by pulling cross-device similarity of positive pairs. To this end, we have to build within-device and cross-device positive pairs. Given arbitrary sample X with identity label y, we denote its device as d. Then within-device and cross-device positive pairs are defined as:

$$\begin{aligned} \mathcal{S}^+_{\checkmark} &= \big\{\forall\langle X_i, X_j\rangle \in \mathcal{S}^+, \quad d_i = d_j\big\} \\ \mathcal{S}^+_{\text{✗}} &= \big\{\forall\langle X_i, X_j\rangle \in \mathcal{S}^+, \quad d_i \neq d_j\big\} \\ \mathcal{S}^-_{\checkmark} &= \big\{\forall\langle X_i, X_j\rangle \in \mathcal{S}^-, \quad d_i = d_j\big\} \\ \mathcal{S}^-_{\text{✗}} &= \big\{\forall\langle X_i, X_j\rangle \in \mathcal{S}^-, \quad d_i \neq d_j\big\}, \end{aligned} \tag{2}$$

where $\mathcal{S}_{\checkmark}$ represents the collection of within-device pairs and $\mathcal{S}_{\text{✗}}$ is the collection of cross-device pairs.

Construct Histograms. Next, we construct the histogram of positive and negative pairs according to their similarities. The ordinary discrete histogram is not decomposable and thus cannot be used in the CNN architecture for end-to-end optimization. Ustinova *et al.* [25] propose the histogram loss that builds decomposable histograms by interpolating between discrete values. Take positive pairs as an example, let $\mathcal{H}^+ \in \mathbb{R}^R$ be the histogram with nodes: $t_1 = -1, ..., t_R = 1$.

Where R is the dimension of the resulting histogram and the interval between two nodes is $\frac{2}{R-1}$. The r-th node of the histogram is:

$$h_r^+ = \frac{1}{|\mathcal{S}^+|} \sum_{\langle X_i, X_j \rangle \in \mathcal{S}^+} \delta_{i,j,r}. \tag{3}$$

where

$$\delta_{i,j,r} = \begin{cases} (s_{ij} - t_{r-1})/\Delta; & \text{if} \quad s_{ij} \in [t_{r-1}, t_r] \\ (t_{r+1} - s_{ij})/\Delta; & \text{if} \quad s_{ij} \in [t_{r-1}, t_r] \\ 0; & otherwise. \end{cases} \tag{4}$$

Eq. (4) linearly interpolates between two nodes.

The histogram of negative similarities can be calculated analogously. We denote the positive/negative histogram as $\mathcal{H}^+, \mathcal{H}^- \in \mathbb{R}^R$.

Algorithm 1: Algorithmic outline of the proposed method.

Input: Model F, initial parameters Θ, learning rate α.

while *not converged* **do**

- Random select a mini-batch B;
- Collect sample pairs according to Eq. (1) or Eq. (2);
- Forward pass $\mathbf{F}(B)$ to get features;
- Construct histogram $\mathcal{H}$ according to Eq. (3);
- Get target distribution according to Eq. (6).
- Construct target distribution $\mathcal{T}$ using $\hat{\mu}, \hat{\sigma}$;
- Compute loss according to Eq. (7)
- Backward to get gradient $\Delta\Theta$;
- Update Θ by: $\Theta := \Theta - \alpha \cdot \Delta\Theta$

end

Output: Θ

Though the construction of histograms is inspired by [25], our proposed loss function is quite different from [25]: [25] tries to shrink the intersection between positive and negative histograms, our proposed method explicitly setup the target distribution for estimated histograms. Moreover, we propose *progressive targets* to smooth the ease the optimization.

3.2. Distributional loss with Progressive Targets

Distributional Losses. After the positive/negative histograms are constructed, we set target distributions for them and then minimize the distance between histograms and their respective targets under specific metrics, *e.g.*Kullback–Leibler divergence.

Let $\mathcal{T}$ be the target distribution and $\mathcal{H}$ be the estimated histogram. Our method minimizes the divergence between them:

$$\begin{aligned} \mathcal{L}_{hist}^{+} &= \mathrm{D}(\mathcal{H}^{+}, \mathcal{T}^{+}) \\ \mathcal{L}_{hist}^{-} &= \mathrm{D}(\mathcal{H}^{-}, \mathcal{T}^{-}), \end{aligned} \tag{5}$$

where $\mathrm{D}(\cdot, \cdot)$ is a distance metric on discrete distributions. In our experiments, we use Kullback–Leibler divergence to measure the distributional distances for its simplicity.

Progressive Targets. Since the batch statistics vary very significantly during training, using a fixed target distribution may lead to difficulty in optimization and unstable results. As illustrated in Fig. 3 (a), when the estimated histogram is far away from the fixed target, the intersection of their support sets is nearly empty. Computing KL divergence under such conditions is an ill-posed problem. To avoid such a situation and make the training procedure smooth, we propose the *progressive target* in the distribution-based loss.

Fig. 6 presents the similarity distributions of a single device dataset, PolyU, and the CrossDevice-A dataset. The similarities roughly follow a gaussian distribution. Therefore, we define the target histogram as a gaussian distribution with μ and σ differ a little bit from the estimated histograms. Specifically, let μ, σ be the mean and variance of the estimated histogram, the mean and the variance of its target distribution are given by:

$$\begin{aligned} \hat{\mu} &= \begin{cases} \mu + \Delta_{\mu}, & \text{if } \mathcal{H}^{+} \\ \mu - \Delta_{\mu}, & \text{if } \mathcal{H}^{-} \end{cases} \\ \hat{\sigma} &= \mu - \Delta_{\sigma}, \end{aligned} \tag{6}$$

where $\Delta_\mu, \Delta_\sigma$ are small delta values.

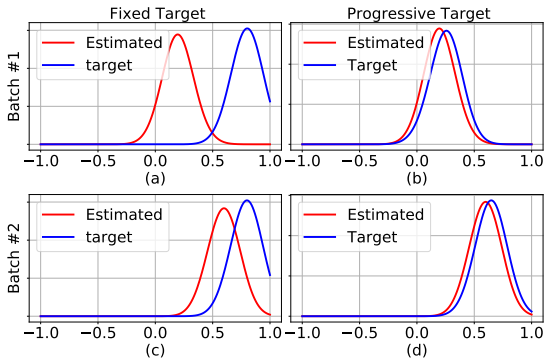

Figure 3: Left: fixed distributional targets; Right: our proposed progressive targets. Top: Batch 1, Bottom: Batch 2.

Examples of the estimated histogram and target distribution can be found in Fig. 3 (b)&(d).

Total loss. In general, the final loss function consists of the *positive histogram loss* pull positive pairs, the *negative histogram loss* to push negative pairs, and the *ArcFace* for classification. In addition to the aforementioned terms, we add the *mean loss* to explicitly enlarge the average a similar gap between positive/negative pairs:

$$\mathcal{L}_{mean} = \text{mean}(\mathcal{H}^+) - \text{mean}(\mathcal{H}^-)$$

The total loss in our method can be formulated as:

$$\mathcal{L} = \alpha(\mathcal{L}^+_{hist} + \mathcal{L}^-_{hist}) + \beta\mathcal{L}_{mean} + \mathcal{L}_{ArcFace}, \tag{7}$$

where α, β are balance factors.

The overall pipeline of the proposed method is summarized in line 1.

4. Cross Device Palm Datasets

To the best of our knowledge, no palm dataset is dedicated to cross-device RGB palmprint recognition. Here 'cross-device' refers to RGB images that are

Figure 4: Example images of the four source datasets. Each column represents images of the same identity taken by diverse devices.

taken by sensors, *e.g.*digital cameras, mobile phone, IoT devices *et al.*. To fulfill the blank, we collect two cross-device palmprint recognition datasets, namely CrossDevice-A and CrossDevice-B, for improving the research of cross-device palm recognition. Example images of the newly collected dataset and their respective original datasets are shown in Fig. 4.

4.1. Data collection and annotation

CrossDevice-A. Images of CrossDevice-A are from MPD [3] and TCD [26] datasets where the images are taken by mobile and IoT devices, respectively. MPD and TCD datasets contain 400/600 identities and 16,000/12,000 images, respectively. As part of the identities present in both TCD and MPD, we construct CrossDevice-A by selecting intersect identities from TCD and MPD.

For each identity in MPD, we select top-5 candidate identities from TCD by computing the cosine similarity between identities. Formally, we define identity-level similarity as the average of instance-level similarities:

$$\mathcal{S}(\mathrm{ID}^m, \mathrm{ID}^n) = \frac{1}{||\mathrm{ID}^m|| \cdot ||\mathrm{ID}^n||} \sum_i \sum_j \cos(I_i^m, I_j^n),$$

where ID^m and ID^n are two identities and I_i^m (I_j^n) is the i-th (j-th) image in identity ID^m (ID^n). The instance-level similarity is computed based on features that are extracted using a pretrained palmprint recognition model.

After selecting top-5 most-similar identities, we manually verify and select the matched identity. The verification and selection is performed by two distinct human annotators for the sake of accuracy.

CrossDevice-B. Images of CrossDevice-B are from MOHI [27] and WEHI [27] datasets. Since the original MOHI and WEHI datasets are designated for hand shape recognition, therefore, the palmprint is not very clear. Consequently, the CrossDevice-B dataset is more challenging than CrossDevice-A. The construction of CrossDevice-B is simpler than that of CrossDevice-B since the identities are strictly matched in the two source datasets.

The statistics of the two new datasets and their respective source datasets are summarized in Tab. 1.

New Dataset	#IDs	#samples	Source Dataset	#IDs	#samples	Device
CrossDevice-A	310	18,600	MPD	400	16,000	Mobile
			TCD	600	12,000	IoT
CrossDevice-B	200	6,000	MOHI	200	3,000	Mobile
			WEHI	200	3,000	WebCam

Table 1: Statistics of two newly collected datasets and their respective source datasets.

5. Experiments

5.1. Implementation details

Training settings. Our method is implemented using the PyTorch framework. We first train a base model using the ArcFace loss and then finetune with ArcFace + PTD loss. The loss weights in PTD loss are set to: $\alpha = 2.0$, $\beta = 0.05$. The deltas in Eq. (6) are set to $\Delta_\mu = 0.07$ and $\Delta_\sigma = 0.05$. We train the model for 26 epochs in both first-round training and finetune. The initial learning rate is set to $5e^{-3}$ and the learning rate decays with a factor of 0.1 at 14, 18, and 24 epochs. We use the stochastic gradient descent algorithm to optimize the

Datasets	Capture type	#images	#IDs
CASIA [28]	Contactless	5,502	624
IITD [29]	Contactless	2,601	460
PolyU [30]	Contactless	1,140	114
TCD [26]	Contactless	12,000	600
MPD [3]	Smartphones	16,000	400

Table 2: Datasets used in our experiments for general palmprint recognition.

model, the movement is set to 0.9 and weight-decay is set to $5e^{-4}$. For more implementation details, please visit `https://kaizhao.net/palmprint`.

Datasets. We test our proposed method as well as competitors on CASIA [28], IITD [29], PolyU [30], TCD [26], MPD [3] and our newly collected datasets. Unless otherwise stated, we use the 5-fold cross-validation where 4/5 of images are used for training and others for testing. The statistics of these datasets are summarized in Tab. 1 and Sec. 5.1.

ROI Extraction. Given a palm image, we first detect two landmarks and then crop the center area of the palm according to the landmarks. Fig. 5 illustrates the landmarks (A and B) and ROI of the left hand. As shown in Fig. 5, we use the intersection of the index finger and little finger as the first landmark (A), and the intersection of the ring finger and middle finger as the second landmark (B). The landmarks are detected with a YOLOv3-based detector.

Network Architecture. Following the common practice of many previous studies [31], we use the Inception-ResNet (also known as Inception-v3) network as our backbone. Specifically, we use its 50 layer variant for the compromise between efficiency and performance, we will refer to the backbone network as 'IR50' for short.

Besides IR50, we took into account the limitations of computing power and speed requirements in some scenarios and selected MobileFaceNet [32] as Back-

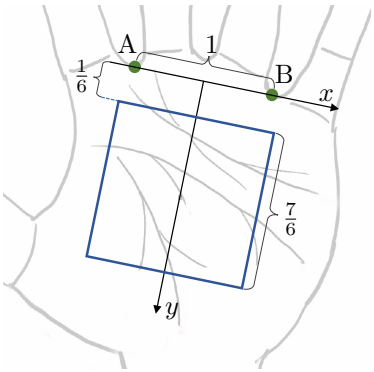

Figure 5: ROI extraction of a left hand. We first detect two key points (green dots) A and B and then set up the coordinates where the x-axis is the line across A and B, and the y-axis is perpendicular. The ROI is a square on the upper half of y-axis.

bone to complete a series of experiments. In our experiments, MobileFaceNet was abbreviated as 'MobileNet'.

We use the state-of-the-art loss function for face recognition, ArcFace [33], as the classification loss.

Evaluation protocol. We evaluate palmprint recognition results in terms of EER, TPR@FAR, and top-1 accuracy. When evaluating the cross-device setting, we use images of a device as a query and images of other devices as a gallery.

5.2. Results on General Palmprint Recognition.

In this section, we evaluated the performance of the baseline models on five public within-device palmprint datasets, and report it in Table 3 in detail. For all methods, we conduct 5-fold validation and record the average performance. Since the evaluation protocol of original papers vary significantly and almost each paper has its evaluation protocol, we put the results which strictly follow the original protocols in the supplementary material.

According to Table 3, the Top-1 Accuracy and EER of the baseline on each within-device dataset outperform the state-of-the-art methods. It can also be seen that based on the baseline ArcPalm, PTD can significantly improve performance. In particular, the Top-1 accuracy of the ArcPalm-IR+PTD achieved

Methods	CASIA	IITD	PolyU	TCD	MPD
PalmNet [34]	97.17 / 3.21	97.31 / 3.83	99.95 / 0.39	99.89 / 0.40	91.88 / 6.22
FERNet [35]	97.65 / 0.73	99.61 / 0.76	99.77 / 0.15	98.63 / -	- / -
DDBC [36]	96.41 / -	96.44 / -	-	98.73 / -	- / -
CompCode [13]	79.27 / 1.08	77.79 / 1.39	99.21 / 0.68	- / -	- / -
OLOF [37]	73.32 / 1.75	73.26 / 2.09	99.55 / 0.23	- / -	- / -
DoN [38]	99.30 / 0.53	99.15 / 0.68	100.00 / 0.22	- / -	- / -
C-LMCL [23]	- / -	- / -	100.00 / 0.13	99.93 / 0.26	- / -
JCLSR [39]	98.94 / -	98.17 / -	- / -	- / -	- / -
ArcPalm-IR50 [3]	98.91 / 0.59	99.85 / 0.47	100.00 / 0.08	99.90 / 0.21	99.18 / 0.81
ArcPalm-IR50 + PTD	**99.85 / 0.37**	**100.00 / 0.20**	**100.00 / 0.05**	**100.00 / 0.04**	**99.78 / 0.43**

Table 3: Top-1 Accuracy/EER (%) of the proposed method compared to other methods on several general palmprint recognition datasets.

100% on IITD, PolyU, and TCD, and close to 100% on CASIA and MPD.

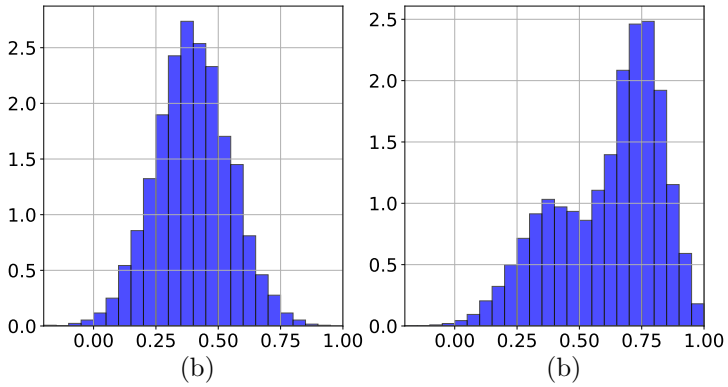

Figure 6: (a) Histogram of positive similarities on single device dataset PolyU (b) Histogram of positive similarities on dual device dataset CrossDevice-A.

5.3. Results on Cross-device Palmprint Recognition.

Here we test our proposed method on the newly collected cross-device datasets: CrossDevice-A and CrossDevice-B.

Experimental settings. In the cross-device setting, we collect both cross-device positive pairs $\mathcal{S}_{\text{✗}}^{+}$ and within-device positive pairs $\mathcal{S}_{\checkmark}^{+}$, as depicted in Eq. (2). As illustrated in Fig. 6 (b), the similarity histogram of positive similarities on cross-device dataset present dual-peak distribution, this reveals that we can model it with a mixture of Gaussians.

Based on the above intuition, we set up a target for cross-device and within-device pairs individually. Let $\mathcal{S}_{\text{✗}}^{+}$ and $\mathcal{S}_{\checkmark}^{+}$ be collection of cross-detice positive similarities and within-device positive similarities. We setup two target distributions $\mathcal{T}_{\checkmark}^{+}, \mathcal{T}_{\checkmark}^{+}$ for them. The negative similarities $\mathcal{S}_{\text{✗}}^{-}, \mathcal{S}_{\checkmark}^{-}$ and their targets $\mathcal{T}_{\checkmark}^{-}, \mathcal{T}_{\checkmark}^{-}$ are defined accordingly. Finally, the $\mathcal{L}_{hist}^{+}$ and $\mathcal{L}_{hist}^{-}$ in Eq. (5) are broken into two parts:

$$\begin{aligned} \mathcal{L}_{hist}^{+} &= \mathrm{D}(\mathcal{H}_{\checkmark}^{+}, \mathcal{T}_{\checkmark}^{+}) + \mathrm{D}(\mathcal{H}_{\text{✗}}^{+}, \mathcal{T}_{\text{✗}}^{+}) \\ \mathcal{L}_{hist}^{-} &= \mathrm{D}(\mathcal{H}_{\checkmark}^{-}, \mathcal{T}_{\checkmark}^{-}) + \mathrm{D}(\mathcal{H}_{\text{✗}}^{-}, \mathcal{T}_{\text{✗}}^{-}). \end{aligned} \tag{8}$$

	CrossDevice-A		CrossDevice-B	
Methods	Top-1	EER	Top-1	EER
PalmNet [34]	84.51	11.60	55.45	34.25
ArcPalm-MB	96.31	2.17	67.96	14.88
ArcPalm-MB+PTD	**98.46**	**1.25**	**69.72**	**13.39**
ArcPalm-IR50	98.79	1.17	67.63	14.62
ArcPalm-IR50+PTD	**99.19**	**0.95**	**71.20**	**13.50**

Table 4: Top-1 accuracy and EER (%) of the ArcPalm-based methods under the cross-device evaluation protocol on CrossDevice-A and CrossDevice-B datasets. IR50 and MB indicate the IR50 and MobileFaceNet backbones, respectively.

During testing, we use images of one device as the gallery and images of the other device as registry to evaluate the cross-device recognition performance.

Quantitative evaluations The quantitative evaluation results are reported in Tab. 4. We compare the proposed method with several recent palmprint recognition methods such as PalmNet [34] and ArcPalm [3] using MobileNet [40] and Inception network [31] as backbone.

For CrossDevice-A, the Top-1 Accuracy of ArcPalm-IR50+PTD and ArcPalm-MobileNet+PTD are 0.4% and 2.15% higher than ArcPalm-IR50 and ArcPalm-MobileNet, respectively. While IR50+PTD's EER and MobileNet's EER are 0.22% and 0.92% lower.

As for CrossDevice-B, it can be seen that IR50+PTD's Top-1 Accuracy is increased by 3.57%, and its EER is decreased by 1.12%. For MobileNet+PTD, its Top-1 Accuracy is increased by 1.76%, and its EER is decreased by 1.49%.

We drew the ROC curve of the SOTA method PalmNet [34], ArcPalm [3]-MobileNet, ArcPalm-MobileNet+PTD, ArcPalm-IR50, and ArcPalm-IR50+PTD on the CrossDevice-A dataset. According to Fig. 7, we can see that ArcPalm-X+PTD outperforms SOTA and its corresponding baseline model, which proves the effectiveness of our method.

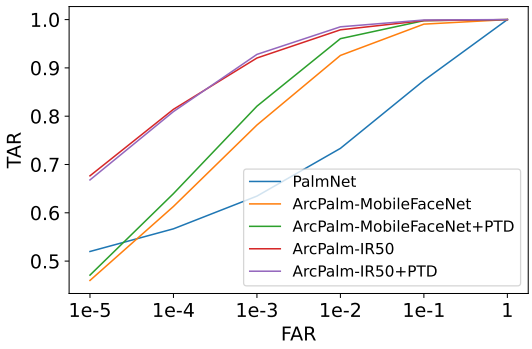

Figure 7: ROC curv on crossPalm-A

5.4. Results on Cross-dataset Palmprint Recognition

In this experiment, we test the cross-dataset generalization of our method. The model is trained on one dataset and evaluated on another dataset. Results

Method	Training	Test	Top-1	EER
C-LMCL [23]	TCD	PolyU	99.93	0.58
ArcPalm-IR50	TCD	PolyU	98.63	0.83
ArcPalm-IR50+PTD	TCD	PolyU	**99.93**	**0.56**
C-LMCL [23]	PolyU	TCD	98.72	1.46
ArcPalm-IR50	PolyU	TCD	97.09	1.74
ArcPalm-IR50+PTD	PolyU	TCD	**98.74**	**1.43**

Table 5: Top-1 Accuracy (%) and EER (%) of the proposed method on cross-dataset recognition

in Sec. 5.4 imply that our method show substantial advantages against the ArcPalm baseline under the challenging cross-dataset setting.

5.5. Ablation Study

We conduct several ablation experiments to verify the effectiveness of the proposed method and test the robustness of our method against hyper-parameters. All experiments in this section are conducted on the CrossDevice-A dataset.

Effectiveness of each component. We first ablate the effectiveness of each component in Eq. (7). According to Table Tab. 6, each of our modules can im-

prove the performance of the baseline model. Not only that, the Top-1 accuracy and EER including the PTD loss of all modules are the best, indicating that each module is an indispensable part of the PTD loss.

$\mathcal{L}_{hist}^{+}$	$\mathcal{L}_{hist}^{-}$	$\mathcal{L}_{mean}$	EER	Top-1
✗	✗	✗	96.31	2.17
✗	✗	✓	96.42	1.99
✓	✓	✗	97.79	1.30
✓	✓	✓	98.46	1.25

Table 6: Ablation study of each components in Eq. (7) on the CrossDevice-A dataset.

PTD *v.s.* fixed targets. In the first experiments, we verify the effectiveness of the proposed method by comparing the performance of PTD and a counterpart with a fixed target. According to Tab. 7, PTD outperforms fixed target in terms of both Top-1 accuracy and EER. When using within-device distribution as the target, its performance is not as good as the proposed PTD loss. Therefore, we believe that the PTD loss target setting is more reasonable and effective. For the fixed target counterpart, we set up a fixed target for positive/negative pairs, respectively. We set $\mu = 0.7, \sigma = 0.01$ for positive pairs and $\mu = 0, \sigma = 0.01$ for negative pairs. Results are in Tab. 7.

Methods	Top-1	EER
ArcPalm-MB	96.31	2.17
ArcPalm-MB+Within-device Target *	96.75	1.94
ArcPalm-MB+Fixed Target	97.61	1.79
ArcPalm-MB+PTD	**98.46**	**1.25**

Table 7: Aablation study of different target settings on CrossDevice-A (%). MB means using the MobileFaceNet as backbone.

PTD under different hyper-parameters. Here we report the performance of PTD under different hyper-parameters, these hyper-parameters are: loss

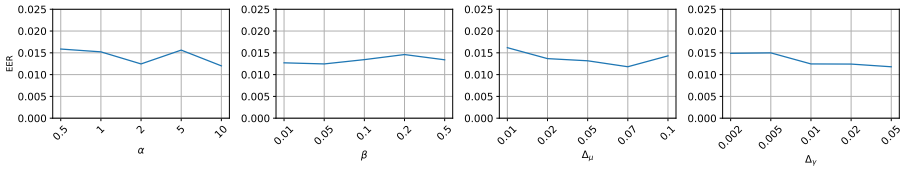

Figure 8: Performance (EER) of our method with different hyper-parameters. When tuning one hyper-parameter, we fix others to the default value.

weights α, β in Eq. (7) and deltas $\Delta_\mu, \Delta_\sigma$ in Eq. (6). When tuning one hyper-parameter, we fix others as the default values. The results in Fig. 8 reveal that our method is robust against the choice of loss weights α, β and deltas $\Delta_\mu, \Delta_\sigma$.

	All Search		Indoor Search	
Method	Rank-1	mAP	Rank-1	mAP
PIG [41]	45.10	25.50	52.70	42.70
SIM [42]	57.47	53.75	-	-
DDAG [42]	54.75	53.02	61.02	67.98
AGW [43]	47.50	47.65	54.17	59.81
AGW+PTD	**60.13**	**57.42**	**62.51**	**70.64**

Table 8: Evaluations (%) of PTD loss on SYSU-MM01 [44] dataset.

	Visible to Infrared		Infrared to Visible	
Method	Rank-1	mAP	Rank-1	mAP
PIG [41]	48.50	49.30	48.10	48.90
SIM [42]	75.29	74.47	78.30	75.24
DDAG [45]	68.06	61.80	69.34	63.46
AGW [43]	70.05	66.37	-	-
AGW+PTD	**78.73**	**78.12**	**79.63**	**77.84**

Table 9: Evaluations (%) of PTD loss on RegDB [46] dataset

5.6. Experiments on Person Re-identification

Person re-identification (ReID) is a similar task which aims to learning discriminative features to identity pedestrians. Recently, many deep learning based

ReID methods have been proposed [43, 47, 48]. To further verify the generalization of our proposed method, we also test our method on cross-device person re-identification. task. We use the AGW [43] as the baseline to verify the effectiveness of our method. The experiments on ReID are conducted on SYSU-MM01 [44] and RegDB [46] datasets. The images of SYSU-MM01 dataset are captured by RGB cameras under two scenes: indoor and ourdoor. We evaluate both the all-to-all and the indoor-to-outdoor performance. The RegDB is a cross-device and cross-modal dataset that the images are captured by RGB (visible) and infrared cameras. We evaluate all-to-all and the infrared-to-visible performance on RegDB dataset. The results in Tab. 8 and Tab. 9 clearly demonstrate that our method consistently improves the performance of a strong baseline with considerable margins.

6. Conclusion and Future Work

In this paper, we propsoe a new loss function for cross-device palmprint recognition. Our contributions are summarized as follows. We first reorganized a large-scale palm benchmark dataset consists of two subsets. Second, we propose the progressive target loss (PTD loss) which progressively narrows the gap between representations of cross-device samples. Extensive experiments have demonstrated the superior of our method. The proposed dataset will benefit the research of cross-device palmprint recognition, and the proposed method may be also helpful to other biometric recognition tasks, *e.g.*face recognition. Though it is effective, our method works in a supervised manner which means the device labels are required during training, which limits its application to unsupervised conditions where device labels are unavailable. Besides, our method brings extra computation to estimate the distribution of sample similarities.

References

[1] Amazon one, https://one.amazon.com/.

[2] L. Fei, G. Lu, W. Jia, S. Teng, D. Zhang, Feature extraction methods for palmprint recognition: A survey and evaluation, IEEE Trans. Syst. Man Cybern. Syst. 49 (2) (2018) 346–363. doi:10.1109/TSMC.2018.2795609.

[3] Y. Zhang, L. Zhang, R. Zhang, S. Li, J. Li, F. Huang, Towards palmprint verification on smartphones, arXiv preprint arXiv:2003.13266 (2020).

[4] H. Shao, D. Zhong, X. Du, Efficient deep palmprint recognition via distilled hashing coding, in: The IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR '19, IEEE, Long Beach, CA, USA, 2019, pp. 0–0. doi:10.1109/CVPRW.2019.00098.

[5] G. Lu, D. Zhang, K. Wang, Palmprint recognition using eigenpalms features 24 (9-10) (2003) 1463–1467.

[6] T. Connie, A. T. B. Jin, M. G. K. Ong, D. N. C. Ling, An automated palmprint recognition system, Image and Vision computing 23 (5) (2005) 501–515.

[7] J. Chen, L. Liao, W. Zhang, L. Du, Mixture factor analysis with distance metric constraint for dimensionality reduction, Pattern Recognition 121 (2022) 108156.

[8] M. Wang, Q. Ruan, Palmprint recognition based on two-dimensional methods, in: ICSP, Vol. 4, IEEE, 2006.

[9] J. Gui, W. Jia, L. Zhu, S.-L. Wang, D.-S. Huang, Locality preserving discriminant projections for face and palmprint recognition, Neurocomputing 73 (13-15) (2010) 2696–2707.

[10] D. Hu, G. Feng, Z. Zhou, Two-dimensional locality preserving projections (2dlpp) with its application to palmprint recognition, Pattern Recognition 40 (1) (2007) 339–342.

[11] H. Li, L. Wang, Palmprint recognition using dual-tree complex wavelet transform and compressed sensing, in: Proceedings of 2012 International

Conference on Measurement, Information and Control, Vol. 2, IEEE, 2012, pp. 563–567.

[12] L. Leng, M. Li, C. Kim, X. Bi, Dual-source discrimination power analysis for multi-instance contactless palmprint recognition, Multimedia Tools and Applications 76 (1) (2017) 333–354.

[13] A.-K. Kong, D. Zhang, Competitive coding scheme for palmprint verification, in: Proceedings of the 17th International Conference on Pattern Recognition, Vol. 1 of ICPR '17, IEEE, Cambridge, UK, 2004, pp. 520–523. doi:10.1109/ICPR.2004.1334184.

[14] Q. Zheng, A. Kumar, G. Pan, Suspecting less and doing better: New insights on palmprint identification for faster and more accurate matching, IEEE TIFS 11 (3) (2015) 633–641.

[15] G. Li, J. Kim, Palmprint recognition with local micro-structure tetra pattern, Pattern Recognition 61 (2017) 29–46.

[16] W. Jia, D.-S. Huang, D. Zhang, Palmprint verification based on robust line orientation code, Pattern Recognition 41 (5) (2008) 1504–1513.

[17] X. Wu, Q. Zhao, W. Bu, A sift-based contactless palmprint verification approach using iterative ransac and local palmprint descriptors, Pattern Recognition 47 (10) (2014) 3314–3326.

[18] J. Qian, J. Yang, G. Gao, Discriminative histograms of local dominant orientation (d-hldo) for biometric image feature extraction, Pattern Recognition 46 (10) (2013) 2724–2739.

[19] L. Dian, S. Dongmei, Contactless palmprint recognition based on convolutional neural network, in: IEEE ICSP, IEEE, 2016, pp. 1363–1367.

[20] J. Svoboda, J. Masci, M. M. Bronstein, Palmprint recognition via discriminative index learning, in: 23rd International Conference on Pattern

Recognition, ICPR '16, IEEE, Cancun, Mexico, 2016, pp. 4232–4237. doi:10.1109/ICPR.2016.7900298.

[21] H. Shao, D. Zhong, Towards open-set touchless palmprint recognition via weight-based meta metric learning, Pattern Recognition 121 (2022) 108247.

[22] S. Zhao, B. Zhang, Deep discriminative representation for generic palmprint recognition, Pattern Recognition 98 (2020) 107071.

[23] D. Zhong, J. Zhu, Centralized large margin cosine loss for open-set deep palmprint recognition, IEEE Trans. Circuits Syst. Video Technol. (2019). doi:10.1109/TCSVT.2019.2904283.

[24] L. Fei, B. Zhang, Y. Xu, W. Jia, J. Wen, J. Wu, Precision direction and compact surface type representation for 3d palmprint identification, Pattern Recognition 87 (2019) 237–247.

[25] E. Ustinova, V. Lempitsky, Learning deep embeddings with histogram loss, in: Advances in Neural Information Processing Systems, NIPS '16, Curran Associates, Inc., Barcelona, SPAIN, 2016, pp. 4170–4178.

[26] L. Zhang, L. Li, A. Yang, Y. Shen, M. Yang, Towards contactless palmprint recognition: A novel device, a new benchmark, and a collaborative representation based identification approach, Pattern Recognit. 69 (2017) 199–212. doi:10.1016/j.patcog.2017.04.016.

[27] A. Hassanat, M. Al-Awadi, E. Btoush, A. Al-Btoush, E. Alhasanat, G. Altarawneh, New mobile phone and webcam hand images databases for personal authentication and identification, Procedia Manufacturing 3 (2015) 4060–4067. doi:10.1016/j.promfg.2015.07.977.

[28] Y. Hao, Z. Sun, T. Tan, C. Ren, Multispectral palm image fusion for accurate contact-free palmprint recognition, in: 15th IEEE International Conference on Image Processing, ICIP '08, IEEE, San Diego, CA, USA, 2008, pp. 281–284. doi:10.1109/ICIP.2008.4711746.

[29] A. Kumar, Incorporating cohort information for reliable palmprint authentication, in: Sixth Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP '08, IEEE, Bhubaneswar, India, 2008, pp. 583–590. doi:10.1109/ICVGIP.2008.73.

[30] V. Kanhangad, A. Kumar, D. Zhang, Contactless and pose invariant biometric identification using hand surface, IEEE Trans. Image Process. 20 (5) (2010) 1415–1424. doi:10.1109/TIP.2010.2090888.

[31] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-first AAAI conference on artificial intelligence, AAAI '17, San Francisco, CA, USA, 2017, pp. 4278–4284.

[32] S. Chen, Y. Liu, X. Gao, Z. Han, Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices, in: Chinese Conference on Biometric Recognition, CVBR '18, Springer, Xinjiang, China, 2018, pp. 428–438. doi:10.1007/978-3-319-97909-0_46.

[33] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, CVPR '19, IEEE, Long Beach, CA, USA, 2019, pp. 4690–4699. doi:10.1109/CVPR.2019.00482.

[34] A. Genovese, V. Piuri, K. N. Plataniotis, F. Scotti, Palmnet: Gabor-pca convolutional networks for touchless palmprint recognition, IEEE Trans. Inform. Forensics Secur. 14 (12) (2019) 3160–3174. doi:10.1109/TIFS.2019.2911165.

[35] W. M. Matkowski, T. Chai, A. W. K. Kong, Palmprint recognition in uncontrolled and uncooperative environment, IEEE Trans. Inform. Forensics Secur. (2019). doi:10.1109/TIFS.2019.2945183.

[36] L. Fei, B. Zhang, Y. Xu, Z. Guo, J. Wen, W. Jia, Learning discriminant direction binary palmprint descriptor, IEEE Trans. Image Process. 28 (8) (2019) 3808–3820. doi:10.1109/TIP.2019.2903307.

[37] Z. Sun, T. Tan, Y. Wang, S. Z. Li, Ordinal palmprint represention for personal identification [represention read representation], in: 2005 IEEE Computer society Conference on Computer Vision and Pattern Recognition, Vol. 1 of CVPR '05, IEEE, San Diego, CA, USA, 2005, pp. 279–284. doi:10.1109/CVPR.2005.267.

[38] Q. Zheng, A. Kumar, G. Pan, A 3d feature descriptor recovered from a single 2d palmprint image, IEEE transactions on pattern analysis and machine intelligence 38 (6) (2016) 1272–1279.

[39] S. Zhao, B. Zhang, Joint constrained least-square regression with deep convolutional feature for palmprint recognition, IEEE Transactions on Systems, Man, and Cybernetics: Systems (2020).

[40] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, 2018, pp. 4510–4520.

[41] M. Jia, Y. Zhai, S. Lu, S. Ma, J. Zhang, A similarity inference metric for rgb-infrared cross-modality person re-identification, in: Proceedings of the 29th International Joint Conference on Artificial Intelligence, IJCAI '20, AAAI Press, 2020, pp. 1026–1032. doi:10.24963/ijcai.2020/143.

[42] G.-A. Wang, T. Zhang, Y. Yang, J. Cheng, J. Chang, X. Liang, Z.-G. Hou, Cross-modality paired-images generation for rgb-infrared person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, AAAI Press, 2020, pp. 12144–12151. doi:10.1609/aaai.v34i07.6894.

[43] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, S. C. Hoi, Deep learning for person re-identification: A survey and outlook, IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).

[44] A. Wu, W.-S. Zheng, H.-X. Yu, S. Gong, J. Lai, Rgb-infrared cross-modality person re-identification, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 5380–5389.

[45] M. Ye, J. Shen, D. J. Crandall, L. Shao, J. Luo, Dynamic dual-attentive aggregation learning for visible-infrared person re-identification, in: European Conference on Computer Vision, ECCV '20, Springer, 2020.

[46] D. T. Nguyen, H. G. Hong, K. W. Kim, K. R. Park, Person recognition system based on a combination of body images from visible light and thermal cameras, Sensors 17 (3) (2017) 605.

[47] Y. Chen, H. Wang, X. Sun, B. Fan, C. Tang, H. Zeng, Deep attention aware feature learning for person re-identification, Pattern Recognition 126 (2022) 108567.

[48] F. Yang, K. Yan, S. Lu, H. Jia, X. Xie, W. Gao, Attention driven person re-identification, Pattern Recognition 86 (2019) 143–155.