Unified Adversarial Augmentation for Improving Palmprint Recognition

Supplementary Material

1. More Details of Proposed Method

1.1. Process of Spatial Transformer

In the spatial transformer network, we first apply an affine transformation matrix for point-to-point transformations. Let the transformation matrix be denoted as A_t, which is defined in Equation 6. The coordinates of the pixels in the input and output images are denoted as (x_s, y_s) and (x_t, y_t), respectively. Then, we define a simple normalization operation N that normalizes these coordinates to the range between -1 and 1 and normalized coordinates $(u, v) \coloneqq N(x, y)$. Following the general convention [8], the point-to-point transformation is represented as follows,

$$\begin{pmatrix} u_t \\ v_t \\ 1 \end{pmatrix} = A_t^{-1} \begin{pmatrix} u_s \\ v_s \\ 1 \end{pmatrix}. \tag{1}$$

To obtain a differentiable output image, a bilinear sampler is applied, which is defined as follows,

$$\begin{aligned} P(x_t, y_t) = \sum_n^H \sum_m^W P(x_s, y_s) \cdot max(0, 1 - \mid x_s - m \mid) \\ \cdot max(0, 1 - \mid y_s - n \mid), \end{aligned} \tag{2}$$

where $P(x, y)$ denotes the pixel value at the coordinates (x, y). This process ensures that the output image is smoothly transformed and differentiable, allowing for effective training of the network.

1.2. Loss Functions in Palmprint Generation Network

During the training of the palmprint generation network, we employ a composite loss function that encompasses several distinct components: L_{KL}, L_1, L_{ID}, and L_{GAN}. These losses are defined as follows:

$$\begin{aligned} L_{KL} &= -\frac{1}{2}\left(1 + \log \sigma_s^2 - \sigma_s^2 - \mu_s^2\right), \\ L_1 &= \|x_1 - x'\|_1, \\ L_{ID} &= 1 - COS(F^*(x_2), F^*(x')), \\ L_{GAN} &= \mathbb{E}_x\left[\log D(x_1, x')\right] + \mathbb{E}_{x,s}\left[\log\left(1 - D(x_1, G(x_2, s))\right)\right]. \end{aligned} \tag{3}$$

Here, μ_s and σ_s^2 represent the mean and variance of the style code $\mathbf{s}$, respectively. The function $COS(\cdot, \cdot)$ computes the cosine similarity between two vectors, F^* refers to a pretrained palmprint recognition network, and D and G denote the discriminative and generative networks, respectively.

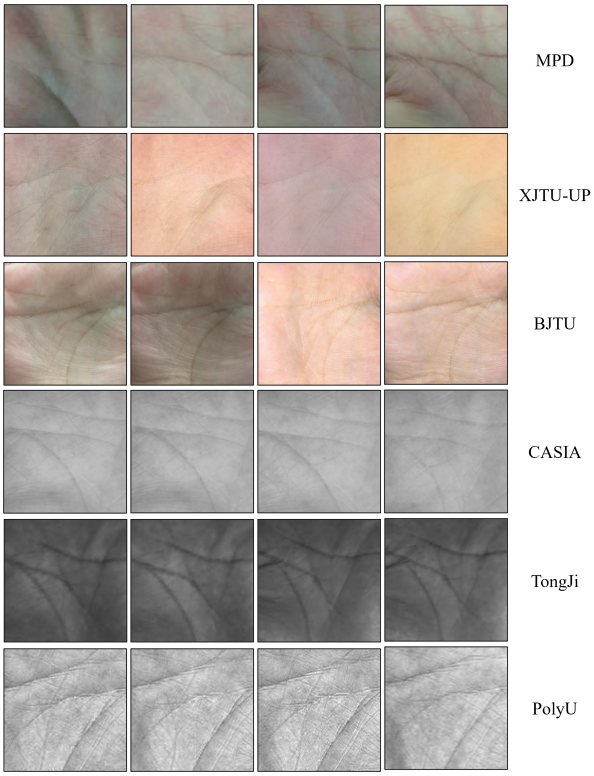

Figure 1. Example images of public palmprint datasets.

2. Public Palmprint Datasets

Figure 1 illustrates a selection of palmprint images from publicly available datasets. Each row of images corresponds to a single identity under one dataset. It is evident that the MPD [19], XJTU-UP [14], and BJTU [2] datasets exhibit significant intra-class variation, while the CASIA [15], TongJi [18], and PolyU [17] datasets demonstrate relatively lower intra-class variation.

3. Additional Experimental Results

3.1. Results with Different Metrics

We present performance on challenging datasets under more metrics, e.g., TAR@FAR=1e-3, TAR@FAR=1e-4, and EER, in Table 1.

3.2. Results with Different Baseline

We conduct additional comparative experiments based on ElasticFace [1] as a recognition baseline. The Resnet-50 [6] is chosen as the backbone. Experimental results are shown in Table 2. We keep the same conclusion as in the main text.

Table 1. Comparative performance on challenging datasets, under the open-set protocol, with ArcFace [5] as baseline and ResNet-50 [6] as the backbone for all methods. Results are presented in TAR@FAR and EER

Methods	MPD			XJTU-UP			BJTU		
	EER↓	@1e-3↑	@1e-4↑	EER↓	@1e-3↑	@1e-4↑	EER↓	@1e-3↑	@1e-4↑
Baseline [5]	0.0940	0.6898	0.5915	0.0480	0.8482	0.7716	0.0953	0.6849	0.5963
RandAugment [4]	0.0896	0.6851	0.5890	0.0330	0.8839	0.8104	0.0664	0.7466	0.6252
TrivialAugment [12]	0.0715	0.7332	0.6262	0.0253	0.9172	0.8588	0.0572	0.7876	0.6888
AdaAug [3]	0.0667	0.7469	0.6516	0.0262	0.9106	0.8517	0.0514	0.7903	0.6910
MADAug [7]	0.0427	0.8145	0.7073	0.0174	0.9416	0.8839	0.0406	0.7825	0.7026
AdvProp [16]	0.0939	0.6895	0.5932	0.0472	0.8490	0.7706	0.0955	0.6867	0.5997
Fast AdvProp [11]	0.0932	0.6916	0.5939	0.0468	0.8516	0.7748	0.0949	0.6901	0.6013
CFSM [9]	0.0856	0.7207	0.6313	0.0324	0.9035	0.8462	0.0634	0.7630	0.6776
ARoFace [13]	0.0465	0.8226	0.7291	0.0204	0.9196	0.8536	0.0580	0.7990	0.7111
UAA-p	0.0254	0.9022	0.8254	0.0087	0.9759	0.9517	0.0257	0.9044	0.8550
UAA-s	**0.0183**	**0.9301**	**0.8675**	**0.0069**	**0.9820**	**0.9597**	**0.0210**	**0.9258**	**0.8716**

Table 2. Comparative performance on both challenging and controlled datasets under the open-set protocol, with ElasticFace [1] as baseline and ResNet-50 [6] as the backbone for all methods. Results are presented in TAR@FAR.

Methods	MPD		XJTU-UP		BJTU		TongJi		PolyU		CASIA	
	1e-5	1e-6	1e-5	1e-6	1e-5	1e-6	1e-5	1e-6	1e-5	1e-6	1e-5	1e-6
Baseline [1]	0.4589	0.4004	0.6660	0.5924	0.5034	0.4802	0.9369	0.9180	0.9642	0.9422	0.8574	0.8229
RandAugment [4]	0.5117	0.4252	0.7675	0.7031	0.5608	0.5300	0.9770	0.9686	0.9538	0.9329	0.9276	0.8894
TrivialAugment [12]	0.5643	0.4839	0.7993	0.7384	0.6546	0.5701	0.9767	0.9633	0.9580	0.9221	0.9553	0.9341
AdaAug [3]	0.5354	0.4655	0.7409	0.6628	0.5738	0.5537	0.9680	0.9554	0.9622	0.9462	0.9298	0.9057
MADAug [7]	0.6083	0.5091	0.8349	0.7877	0.6585	0.6412	0.9744	0.9565	0.9697	0.9535	0.9545	0.9411
AdvProp [16]	0.4601	0.3988	0.6670	0.5913	0.5102	0.4822	0.9375	0.9172	0.9639	0.9432	0.8593	0.8246
Fast AdvProp [11]	0.4602	0.4013	0.6664	0.6062	0.5052	0.4806	0.9386	0.9190	0.9635	0.9429	0.8607	0.8291
CFSM [9]	0.5461	0.4554	0.8108	0.7624	0.6254	0.5904	0.9549	0.9349	0.9668	0.9406	0.9011	0.8722
ARoFace [13]	0.6354	0.5568	0.8006	0.7115	0.6489	0.6179	0.9771	0.9553	0.9583	0.9116	0.9593	0.9460
UAA-p	0.7513	0.6738	0.9262	0.8869	0.7967	0.7468	**0.9868**	**0.9757**	**0.9847**	**0.9746**	0.9802	0.9738
UAA-s	**0.7957**	**0.7024**	**0.9319**	**0.9000**	**0.7887**	**0.7641**	0.9846	0.9665	0.9792	0.9663	**0.9830**	**0.9753**

Table 3. More results of ablation experiments. 'K-T' and 'K-G' denote the optimization steps for the spatial transformation module and the palmprint generation network, respectively. 'γ' represents the augmentation ratio. The UAA is implemented in a sequential manner.

K-T	K-G	γ	MPD		XJTU-UP		BJTU	
			1e-5	1e-6	1e-5	1e-6	1e-5	1e-6
1	1	0.5	0.7779	0.6811	0.9245	0.8928	0.8142	0.7850
1	2	0.5	0.8075	0.7289	0.9332	0.8917	0.8192	0.7892
2	1	0.5	0.7550	0.6678	0.9085	0.8591	0.8170	0.7887
2	2	0.5	0.7681	0.6656	0.9105	0.8636	0.8311	0.8079
1	1	0.25	0.7474	0.6631	0.9068	0.8736	0.7313	0.7045
1	1	0.50	0.7779	0.6811	0.9245	0.8928	0.8142	0.7850
1	1	0.75	0.8140	0.7243	0.9293	0.8822	0.7956	0.7626

3.3. More Ablation Study

Different Numbers of Optimization Steps In the adversarial training, we employ PGD [10], which utilizes a K-step iterative process to optimize the control vector $\mathbf{z}$. We try a different number of optimization steps K for both the spatial transformation module and the palmprint generation network, with the experimental results presented in Table 3.

Table 4. Performance of the palmprint generation method BézierPalm [20] and the method that combines it with our proposed UAA.

Methods	MPD		XJTU-UP		BJTU	
	1e-5	1e-6	1e-5	1e-6	1e-5	1e-6
BézierPalm [20]	0.6834	0.6020	0.8300	0.7697	0.7750	0.7596
UAA-s	0.8075	0.7289	0.9332	0.8917	0.8192	0.7892
combined	0.8882	0.8212	0.9462	0.9196	0.9148	0.8839

Different Augmentation Ratios During training, we augment the training samples in a batch by 25%, 50%, and 75%, respectively. The results are shown in Table 3.

Evaluation of Identity Consistency We quantitatively evaluate identity consistency using paired original and augmented images (x, x'), and a pre-trained recognition model F, expressed as $C = \frac{1}{N}\Sigma_{i=1}^{N} cos\_sim(F(x_i), F(x_i'))$. For N=10k, we report $C_{geometric} = 0.88$, $C_{textural} = 0.92$ and $C_{combined} = 0.80$.

Palmprint generation methods Table 4 shows that combining with palmprint generation methods, our proposed UAA achieves better performance.

References

[1] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1578–1587, 2022. 1, 2

[2] Tingting Chai, Shitala Prasad, and Shenghui Wang. Boosting palmprint identification with gender information using deepnet. *Future Generation Computer Systems*, 99:41–53, 2019. 1

[3] Tsz-Him Cheung and Dit-Yan Yeung. Adaaug: Learning class and instance-adaptive data augmentation policies. In *International Conference on Learning Representations*, 2022. 2

[4] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020. 2

[5] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, pages 4690–4699, 2019. 2

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 1, 2

[7] Chengkai Hou, Jieyu Zhang, and Tianyi Zhou. When to learn what: Model-adaptive data augmentation curriculum. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1717–1728, 2023. 2

[8] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015. 1

[9] Feng Liu, Minchul Kim, Anil Jain, and Xiaoming Liu. Controllable and guided face synthesis for unconstrained face recognition. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XII*, pages 701–719. Springer, 2022. 2

[10] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017. 2

[11] Jieru Mei, Yucheng Han, Yutong Bai, Yixiao Zhang, Yingwei Li, Xianhang Li, Alan Yuille, and Cihang Xie. Fast advprop. In *International Conference on Learning Representations*, 2022. 2

[12] Samuel G. Müller and Frank Hutter. Trivialaugment: Tuning-free yet state-of-the-art data augmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 774–782, 2021. 2

[13] Mohammad Saeed Ebrahimi Saadabadi, Sahar Rahimi Malakshan, Ali Dabouei, and Nasser M Nasrabadi. Aroface: Alignment robustness to improve low-quality face recognition. In *European Conference on Computer Vision*, pages 308–327. Springer, 2024. 2

[14] Huikai Shao, Dexing Zhong, and Xuefeng Du. Effective deep ensemble hashing for open-set palmprint recognition. *Journal of Electronic Imaging*, 29(1):013018, 2020. 1

[15] Zhenan Sun, Tieniu Tan, Yunhong Wang, and Stan Z Li. Ordinal palmprint represention for personal identification [represention read representation]. In *CVPR*, pages 279–284. IEEE, 2005. 1

[16] Cihang Xie, Mingxing Tan, Boqing Gong, Jiang Wang, Alan L Yuille, and Quoc V Le. Adversarial examples improve image recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 819–828, 2020. 2

[17] David Zhang, Wai-Kin Kong, Jane You, and Michael Wong. Online palmprint identification. *IEEE Transactions on pattern analysis and machine intelligence*, 25(9):1041–1050, 2003. 1

[18] Lin Zhang, Lida Li, Anqi Yang, Ying Shen, and Meng Yang. Towards contactless palmprint recognition: A novel device, a new benchmark, and a collaborative representation based identification approach. *Pattern Recognition*, 69:199–212, 2017. 1

[19] Yingyi Zhang, Lin Zhang, Ruixin Zhang, Shaoxin Li, Jilin Li, and Feiyue Huang. Towards palmprint verification on smartphones. *arXiv preprint arXiv:2003.13266*, 2020. 1

[20] Kai Zhao, Lei Shen, Yingyi Zhang, Chuhan Zhou, Tao Wang, Ruixin Zhang, Shouhong Ding, Wei Jia, and Wei Shen. Bézierpalm: A free lunch for palmprint recognition. In *European Conference on Computer Vision*, pages 19–36. Springer, 2022. 2