

Geometric Synthesis: A Free lunch for Large-scale Palmprint Recognition Model Pretraining

Kai Zhao^{1,2*}, Lei Shen¹, Yingyi Zhang¹, Chuhan Zhou¹, Tao Wang¹,
Ruixin Zhang¹, Shouhong Ding¹, Wei Jia³, and Wei Shen⁴

¹ Tencent Youtu Lab

² University of California, Los Angeles

³ Hefei University of Technology

⁴ Shanghai Jiaotong University

Abstract. Palmprints are private and stable information for biometric recognition. In the deep learning era, the development of palmprint recognition is limited by the lack of sufficient training data. In this paper, by observing that palmar creases are the key information to deep-learning-based palmprint recognition, we propose to synthesize training data by manipulating palmar creases. Concretely, we introduce an intuitive geometric model which represents palmar creases with parameterized Bézier curves. By randomly sampling Bézier parameters, we can synthesize massive training samples of diverse identities, which enables us to pretrain large-scale palmprint recognition models. Experimental results demonstrate that such synthetically pretrained models have a very strong generalization ability: they can be efficiently transferred to real datasets, leading to significant performance improvements on palmprint recognition. For example, under the open-set protocol, our method improves the strong ArcFace baseline by more than 10% in terms of TAR@1e-6. And under the closed-set protocol, our method reduces the equal error rate (EER) by an order of magnitude. Codes are available at <http://kaizhao.net/palmpoint>.

Keywords: Palmprint recognition, Deep Learning, Sample Synthesis, Bézier curve.

1 Introduction

Palm information is privacy-by-design because the palm pattern is concealed inside your hand, and it is almost impossible to be tracked by public cameras without your consent. For its security and privacy, palmprint recognition is being adopted by AmazonOne for identification and payment [1]. In contrast, the

* KZ is the corresponding author: kz@kaizhao.net.

widely used face recognition system can easily track people through public cameras without any consent. As a result, face recognition has received widespread criticism in the last few years due to the privacy concerns it creates [55,17].

It is precisely because of its privacy that palmprint recognition lacks a large-scale public dataset. To the best of our knowledge, the largest open dataset for palmprint recognition contains thousands of identities and tens of thousands of images [33]. In contrast, there are a number of million-scale face recognition datasets either based on webly collected faces [29,4,19] or surveillance cameras [41,58]. The lack of sufficient data has become the main bottleneck for palmprint recognition. In this paper, we propose to synthesize images to augment the training set for palmprint recognition.

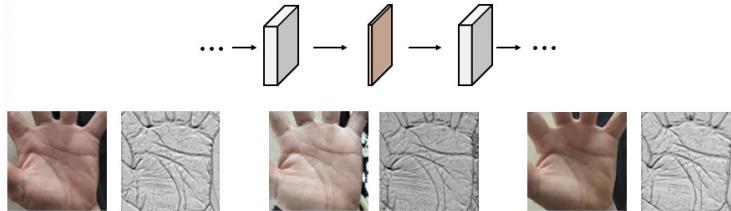


Fig. 1: Top: during training, we squeeze the intermediate features to 1-dimension for visualization. Middle and bottom: the input palm images and intermediate features. The intermediate features suggest that palmar creases are the key information to palmprint recognition.

By visualizing the intermediate features from a CNN-based palmprint recognition model, we observe that the palmar creases play a critical role. Specifically, we insert a ‘squeeze-and-excite’ operation into an existing CNN architecture. The ‘squeeze-and-excite’ first squeezes an intermediate feature map into 1-dimension and excites it back. Then we train this model on palmprint datasets. Finally, we visualize the 1-dimensional feature map on test images. As shown in Fig. 1, most of the texture and color information are ignored by the model, and the palmar creases are largely reserved. This reminds us that we may synthesize palmar creases to enrich the training data for palmprint recognition.

With the above observation, we propose to synthesize training data for palmprint recognition by manipulating palmar creases. An intuitive and simple geometric model is proposed to synthesize palm images by representing palmar creases with several parameterized Bézier curves. The identity of each synthesized data is controlled by the parameters of Bézier curves, *e.g.* number of curves, positions of endpoints, and control points. Our method is able to synthesize massive samples of diverse identities, which enables us to perform large-scale pretraining on such synthetic data. The synthetically pretrained models present promising generalization ability and can be efficiently transferred to real datasets.

Our method holds essential differences from other data generation methods such as generative adversarial networks (GANs) and data augmentation. First

of all, both GANs and data augmentation rely on existing data: either train GANs with or add modifications to existing data. While our method creates new samples without any existing data. Second, neither GANs nor data augmentation can create samples of novel categories, while our method can control the category (identity) of synthesized samples. In addition, GANs require large amount of training data and thus cannot substantially improve the recognition performance. For example, [45] uses GAN-synthesized samples to train face recognition models. However, the synthetically trained models perform worse than models that are directly trained on dataset that is used to train GANs.

Extensive experiments under both open-set and closed-set evaluation protocols demonstrate that our method significantly improve the performance of strong baselines. Additionally, we evaluate on a private million-scale data to further confirm the scalability of our method. The contributions of this paper are summarized as below:

- We visualize the intermediate features of CNN-based palmprint recognition models and observe that the palmar creases play an important role.
- We propose a simple yet effective model to synthesize training data by manipulating creases with parameterized curves. We pretrain deep palmprint recognition models with the synthetic data and then finetune them on real palmprint datasets.
- Extensive evaluation on 13 public datasets demonstrates that the synthetically pretrained models significantly outperform their ‘train-from-scratch’ counterparts and achieve state-of-the-art recognition accuracy.
- We test our method on a million-scale dataset, which is, to the best of our knowledge, the largest evaluation in palmprint recognition. The results verify the scalability of our method, showing its strong potential in the industry-level palmprint recognition.

2 Related Work

2.1 Palmprint Recognition

Traditional palmprint recognition. Traditional palmprint recognition methods in the literature can be roughly classified into two categories: holistic-based and local-based. In holistic-based methods, features are extracted from the whole image and then projected to a space of lower-dimensional to make it more compact and discriminative. PCA [38] and its 2D variant [46] are commonly used in this category. Besides, independent component analysis (ICA) is also used [7]. PCA seeks to find uncorrelated features while ICA attempts to find statistically independent features. Supervised projection methods including Linear Discriminant Analysis (LDA) and 2D-LDA [57] have also been explored. Another interesting method Locality Preserving Projection (LPP) [24] attempts to preserve the local structure of images. Hu *et al.* [26] extend LPP to 2D and Feng *et al.* [13] introduce non-linear kernel to LPP. The holistic-based methods often suffer from degradation caused by distortion, illumination, and noise. To overcome these issues, the palm images are firstly transformed to another domain.

Frequency [25,36], Cosine [35,34] and Radon [53] transforms are commonly used to overcome these degradations.

Local-based methods extract local features on the image and then fuse these features globally for recognition. Competitive coding (CompCode) [31] uses 2-D Gabor filters to extract orientation information from palm lines. FastCompCode [66] proposes a binary code for effective representation and matching. Other coding-based methods include SMCC [69], RLOC [27], ContourCode [30], double orientation code [11], *et al.* Wu *et al.* [59] extract local SIFT features and match palm images with RANSAC. Qian *et al.* [44] extract histogram of orientations.

Deep learning based palmprint recognition. Inspired by the success of deep learning in other recognition tasks such as person re-identification and face recognition [9,64], many researchers attempt to use deep learning technologies for palmprint recognition. Dian *et al.* [10] use the AlexNet as the feature extractor and match palm images with Hausdorff distance. Svoboda *et al.* [52] train CNNs with a novel loss function related to the d-prime index. Recently, margin-based loss functions have been proven to be effective for face recognition. The large margin loss [68] and additive angular margin loss [63] are introduced to palmprint recognition and impressive performance has been achieved. Graph neural networks are also used for palmprint recognition to model the geometric structure of palmprints [47]. Shao *et al.* [49] combine deep learning with hash coding to build efficient palmprint recognition models. Different from these studies that introduce new architectures or loss functions, our proposed method focuses on synthesizing training data for deep palmprint recognition.

2.2 Data Synthesis for Deep Models

Data synthesis aims at synthesizing training data to reduce the cost of data acquisition and labeling. Gaidon *et al.* [15] render the street views to pretrain deep models for object tracking. Tremblay *et al.* [54] render similar scenes for object detection. Yao *et al.* [60] use a graphic engine to simulate a large amount of training data for autonomous driving. Varol *et al.* [56] synthesize images from 3D sequences of human motion capture data for human pose estimation. Sharingan [43] Combines synthetic and real data for unsupervised geometry estimation. Baek *et al.* [3] synthesize depth maps with generative adversarial networks [18] for depth-based human pose estimation. To reduce the gap between synthetic and natural images, Shrivastava *et al.* [50] proposed the Simulated+Unsupervised learning paradigm and Chen *et al.* [6] propose a layer-wise learning rate selection method to improve the synthetic-to-real generalization performance. All these methods synthesize samples of *existing* and *known* categories, while our proposed method aims at generating samples for *novel* categories and augmenting the training identities for palmprint recognition.

3 Methodology

As illustrated in Fig. 2a, the palmprints are roughly composed of several (usually 3~5) principal lines and a number of thin wrinkles. To imitate the geometric

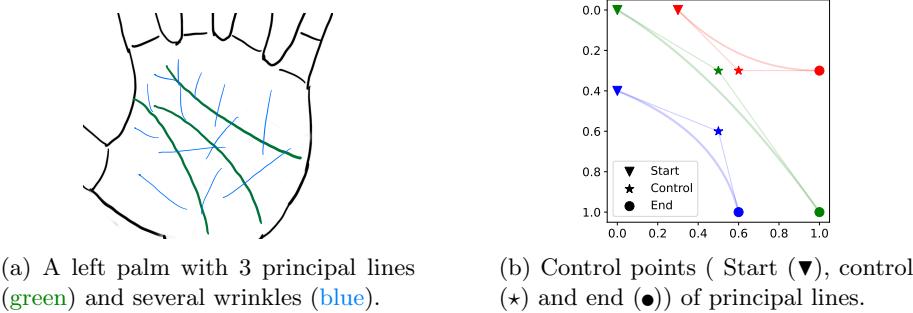
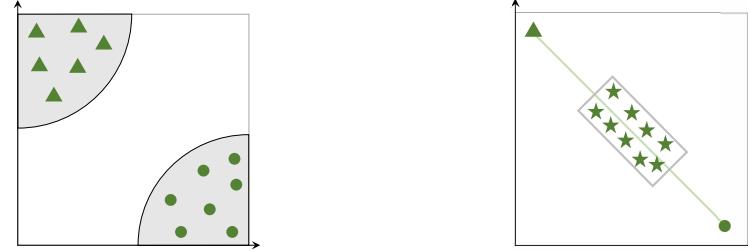


Fig. 2: An example hand (left) and control points of principal lines (right).

appearance of palmprints, we use the Bézier curves to parameterize the palmar creases. Specifically, we use several Bézier curves to represent the principal lines and the wrinkles. For simplicity, we use second-order Bézier curves with three parametric points in a 2D plane, a control point, a start point, and an end point. Fig. 2b gives an example of the parametric points 3 principal lines of a left hand. Next, we will take the left hand as an example to detailedly illustrate how we determine the parameters of Bézier curves, and the case for the right hand can be regarded as the mirror of the left hand.

3.1 Palmar creases with Bézier Curves

Let N and S be the number of total identities and number of samples for each identity, we will generate $N \times S$ samples in total. For each identity, we synthesize



(a) Start and end points of principal lines are sampled from top-right and bottom-left corners.
(b) The control point is sampled from a rectangle that is parallel to the line connecting starting and end points.

Fig. 3: Start (▼), end (●) and control (★) points.

m principal lines and n wrinkles, where m and n are sampled from uniform distributions: $m \sim U(2, 5)$ and $n \sim U(5, 20)$. Take left hand as an example, the starting and end points of principal lines are randomly sampled from top-left and bottom-right corner of the plane, as shown in Fig. 3a. While the start and the end points of wrinkles are randomly sampled from the whole plane. Then, given the starting and end points, the control point is sampled from a rectangle that is parallel to the line connecting two points, as shown in Fig. 3b.

3.2 Parameters of Bèziers

We first determine the number of principal lines and wrinkles for each identity, and then randomly sample start, end, and control points for each crease.

principal lines. For each identity, we sample m principal lines starting from the top-left corner to the bottom-right corner. For left palms, the start/end points for each principal line are sampled from the top-left and bottom-right corner. Given the start/end points of a Bèzier, its control point is randomly sampled from a rectangle area in the middle of the line connecting start and end points. The details about the synthesis of principle lines are illustrated in Appendix A.

wrinkles. We generate $n = 5 \sim 15$ wrinkles for each identity. We do not restrict the directions of wrinkles and their start, end, and control points are randomly sampled from the whole plane:

$$Q = \text{random}(0, 1, \text{size} = (n, 3, 2)).$$

3.3 Within-identity Diversity

We enhance the within-identity diversity of synthesized samples in two aspect: 1) we add small random noises to the parameters so that each sample is a little different from others; 2) we use a randomly selected natural image as the background of the synthesized sample.

Random noise. Given parameters of a specific identity, we add small noises to P and Q to synthesize diverse samples. Formally, the parameters for the j -th sample of identity i are:

$$\begin{aligned} P_j^i &= P^i + N_p \\ Q_j^i &= Q^i + N_q, \end{aligned} \tag{1}$$

where $N_p \sim \mathcal{N}(\mu, 0.04)$ and $N_q \sim \mathcal{N}(\mu, 0.01)$ are small gaussian noises. Each crease is rendered with a random color c and stroke width w .

Random Background. For each sample we select a random image from the imangenet [8] dataset as the background of the synthesized sample.

Finally, sample S_j^i is synthesized with:

$$S_j^i = \text{synthesize}(P_j^i, Q_j^i, c, w, I)$$

A overall algorithmic pipeline is illustrated in Appendix B and some synthetical samples can be found in Appendix C.

4 Experimental Settings

In this section, we introduce the detailed experimental settings including data preparation and evaluation protocols. Our experiments are mainly based on the ArcFace [9], a strong baseline for palm recognition [63]. During training, we use the ArcFace loss as supervision. During testing, we extract 512 dimensional features for each sample and the cosine similarity is used as the distance measurement.



Fig. 4: Example ROIs of different datasets.

4.1 Datasets and Data Preprocessing

Datasets. We use 13 public datasets in our experiments. The statistical information of these datasets is summarized in Tab. 1 and example ROIs of these datasets are shown in Fig. 4.

Name	#IDs	#Images	Device	Name	#IDs	#Images	Device
MPD [63]	400	16,000	Phone	COEP [2]	167	1,344	Digital camera
XJTU_UP [48]	200	30,000	Phone	TCD [63]	600	12,000	Contactless
MOHI [22]	200	3,000	Phone	IITD [32]	460	2,601	Contactless
GPDS [14]	100	2,000	Web cam	CASIA [51]	620	5,502	Contactless
WEHI [22]	200	3,000	Web cam	PolyU-MS [61]	500	24,000	Contactless
PolyU(2d+3d) [28]	400	8,000	Web cam	CASIA-MS [21]	200	7,200	Contactless
XJTU_A [48]	114	1,130	CMOS camera				

Table 1: Statistics of the 13 public palmprint datasets.

The images in CASIA-MS [61] dataset are captured with multi-spectral devices and we only use visible spectra images. We remove the overlapped identities in MPD [63] and TCD [63] datasets. Finally, there are 3,268 identities and 59,162 images used in our experiments.

ROI extraction. We follow the protocol of [62] for 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).

Then we set up a coordinate where \vec{AB} is the x -axis and its perpendicular is the y -axis.

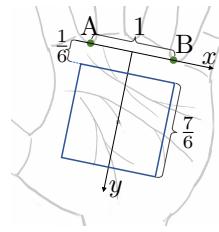


Fig. 5: ROI extraction of a left hand.

Suppose $|\vec{AB}| = 1$ is the unit length, we crop a square with side length $7/6$ as the ROI. The ROI of the right hand can be extracted similarly. Some example ROIs used in our experiments are shown in Fig. 4.

4.2 Open-set Protocol

Dataset split. For the open-set protocol, we select part of identities from each dataset and combine them as a large training set, and the other identities are merged as a large test set. We test two different split settings. In the first setting, half of the identities are used for training, and half are used for testing. In the second setting, $1/4$ of the identities are used for training and others for testing. The number of samples and identities in the two splits are summarized in Tab. 2.

Evaluation. The performance under the open-set protocol is evaluated in terms of TAR@FAR, where TAR and FAR stand for ‘true accept rate’ and ‘false accept rate’, respectively. Specifically, given several test images, we randomly sample several positive pairs where the two samples share the same identity, and negative pairs whose samples are from distinct identities. Let p^+, p^- be the positive/negative pairs and $sim(p)$ be the similarity between a pair of samples. We first fix the FAR and then calculate a proper threshold τ from negative pairs, finally we compute TAR using that threshold on the positive pairs.

Take FAR=1e-3 as an example, we can search for a threshold τ on the negative pairs satisfying:

$$FAR = 10^{-3} = \frac{|\{p^- \mid sim(p^-) > \tau\}|}{|\{p^-\}|}.$$

With the threshold, we then calculate the TAR on the positive pairs:

$$TAR = \frac{|\{p^+ \mid sim(p^+) > \tau\}|}{|\{p^+\}|}.$$

Accordingly, we can calculate TAR under various FARs. In our experiments, we report the performance under FAR=1e-3, 1e-4, 1e-5, 1e-6.

4.3 Closed-set Protocol

The closed-set experiments are conducted on five datasets: CASIA, IITD, PolyU, TCD, and MPD. We perform 5-fold cross-validation on each dataset and report the average performance. The experiments are conducted individually on the five datasets. We use top-1 accuracy, EER to evaluate the performance of closed-set palmprint recognition. To compute the top-1 accuracy, we randomly select one sample from each identity as the *registry* and other samples are *queries*. Let

Split	mode	#IDs	#Images
train:test = 1:1	train	1,634	29,347
	test	1,632	29,815
train:test = 1:3	train	818	14,765
	test	2,448	44,397

Table 2: Training/test splits of the open-set protocol.

$\mathcal{R} = \{r_i\}$ be the set of registries and $\mathcal{Q} = \{q_j\}$ the query set, and $sim(q_i, r_j)$ is the similarity between two samples. $y(\cdot)$ tells the identity label of a sample. The successfully matched queries are these queries that are of the same identity with their nearest registries:

$$\mathcal{Q}^+ = \left\{ q_i \mid y\left(\operatorname{argmax}_{r_j \in \mathcal{R}} sim(q_i, r_j)\right) = y(q_i) \right\}$$

Finally, the top-1 accuracy is the number of successfully matched queries divided by the total number of queries: $acc = |\mathcal{Q}^+|/|\mathcal{Q}|$. The EER is a point where FAR (False Acceptance Rate) and FRR (False Rejection Rate) intersect.

5 Experimental Results

In this section, we first compare our method with other traditional and deep learning based palmprint recognition methods. Then we conduct ablation studies to show the contribution of each component in our method.

5.1 Implementation Details

We implement our method with the PyTorch [42] framework. Two backbone networks, ResNet50 [23] and MobileFaceNet [5], are used in our experiments. The Bézier curves are generated with an opensource package⁵.

Data Synthesizing. By default, the synthetic dataset contains 4,000 identities, and each identity contains 100 samples. The size of the synthetic image is 224×224 . The stroke width w for principal lines and wrinkles are randomly selected from $1.5 \sim 3$ and $0.5 \sim 1.5$, respectively. We randomly blur the synthetic images using a gaussian kernel to improve generalization.

Model Training. For our proposed method, we first pretrain models on synthesized data for 20 epochs and then finetune on real palmprint datasets for 50 epochs. For the baseline, we directly train the models on real datasets for 50 epochs. We use the cosine annealing learning rate scheduler with a warmup start. The maximal learning rate for pretraining is 0.1 and 0.01 for finetune. All models are trained with mini-batch SGD algorithm. The momentum is 0.9 and weight decay is set to 1e-4. We use the *additive angular margin loss* (ArcFace [9]) with margin $m = 0.5$ and scale factor $s = 48$. Besides, we linearly warm up the margin from 0 in the first epoch to improve stability. We use 4 GPUs to run all training experiments and each GPU process 32 images in a batch, in total the effective batchsize is 128.

5.2 Open-set Palmprint Recognition

We first test our method under the "open-set" protocol. Details about the "open-set" protocol can be found in Sec. 4.2. We test our method under two different training test ratios: 1:1 and 1:3, quantitative results are in Tab. 3. The TAR *v.s.* FAR curves of the 1:1 setting are in Fig. 6.

⁵ <https://bezier.readthedocs.io/>

Method	Backbone	train : test = 1 : 1				train : test = 1 : 3			
		TAR@ 1e-3	TAR@ 1e-4	TAR@ 1e-5	TAR@ 1e-6	TAR@ 1e-3	TAR@ 1e-4	TAR@ 1e-5	TAR@ 1e-6
CompCode [31]	N/A	0.4800	0.4292	0.3625	0.2103	0.4501	0.3932	0.3494	0.2648
FastCompCode [66]	N/A	0.4243	0.3649	0.1678	0.2103	0.4188	0.3568	0.3100	0.2748
LLDP [39]	N/A	0.7382	0.6762	0.5222	0.1247	0.7372	0.6785	0.6171	0.2108
Ordinal Code [51]	N/A	0.4628	0.4074	0.3462	0.1993	0.4527	0.3975	0.3527	0.2422
BOCV [20]	N/A	0.4930	0.4515	0.3956	0.2103	0.4527	0.3975	0.3527	0.2422
RLOC [27]	N/A	0.6490	0.5884	0.4475	0.1443	0.6482	0.5840	0.5224	0.3366
DOC [11]	N/A	0.4975	0.4409	0.3712	0.1667	0.4886	0.4329	0.3889	0.2007
PalmNet [16]	N/A	0.7174	0.6661	0.5992	0.1069	0.7217	0.6699	0.6155	0.2877
C-LMCL [68]	MB	0.9290	0.8554	0.7732	0.6239	0.8509	0.7554	0.7435	0.5932
ArcFace [9]	MB	0.9292	0.8568	0.7812	0.7049	0.8516	0.7531	0.6608	0.5825
ArcFace+Ours [9]	MB	0.9640	0.9438	0.9102	0.8437	0.9407	0.8861	0.7934	0.7012
C-LMCL [68]	R50	0.9545	0.9027	0.8317	0.7534	0.8601	0.7701	0.6821	0.6254
ArcFace [9]	R50	0.9467	0.8925	0.8252	0.7462	0.8709	0.7884	0.7156	0.6580
ArcFace+Ours [9]	R50	0.9671	0.9521	0.9274	0.8956	0.9424	0.8950	0.8217	0.7649

Table 3: Quantitative performance under the open-set protocol where the performance are evaluated in terms of TAR@FAR. ‘MB’ represents the MobileFaceNets [5] backbone and ‘R50’ is resnet50.

As shown in Tab. 3, since traditional methods do not rely on training data, they behave similar performance under 1:1 and 1:3 settings, and deep learning based methods perform much better under the 1:1 setting than under the 1:3 setting. Among all traditional methods, LLDP [39] performs the best. Deep Learning based methods [16,68,9] significantly outperform traditional methods, and margin-based methods, *e.g.* C-LMCL [68] and ArcFace [9], present superior performance. Our proposed method remarkably improves the ArcFace baseline and achieves state-of-the-art performance under both 1:1 and 1:3 settings. Under the 1:3 setting, our performance even exceeds the performance of ArcFace under the 1:1 setting.

5.3 Closed-set Palmprint Recognition

Here we report quantitative results of our method as well as other methods under the closed-set protocol. Our experiments are conducted on five datasets, and the performance is evaluated in terms of top-1 accuracy and EER. Detailed setting about the experiments was described in Sec. 4.3.

As shown in Tab. 4, though the results on the closed-set protocol are nearly saturated, our method still improves the baseline with a clear margin, advancing the top-1 accuracies to nearly 100%. Besides, our method significantly decreases the EER to an unprecedented level of 1e-3, surpassing all existing methods.

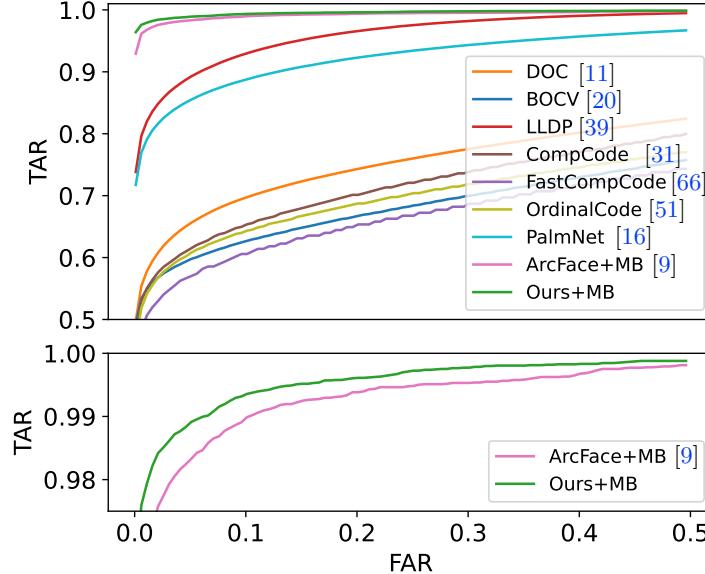


Fig. 6: FAR *v.s.* TAR curves of various methods under the open-set 1:1 settings. The ArcFace and our method are based on the MobileFaceNet backbone.

Method	CASIA	IITD	PolyU	TCD	MPD
CompCode [31]	79.27 / 1.08	77.79 / 1.39	99.21 / 0.68	- / -	- / -
Ordinal Code [51]	73.32 / 1.75	73.26 / 2.09	99.55 / 0.23	- / -	- / -
DoN [67]	99.30 / 0.53	99.15 / 0.68	100.0 / 0.22	- / -	- / -
PalmNet [16]	97.17 / 3.21	97.31 / 3.83	99.95 / 0.39	99.89 / 0.40	91.88 / 6.22
FERNet [40]	97.65 / 0.73	99.61 / 0.76	99.77 / 0.15	98.63 / -	- / -
DDBC [12]	96.41 / -	96.44 / -	-	98.73 / -	- / -
RFN [37]	- / -	99.20 / 0.60	- / -	- / -	- / -
C-LMCL [68]	- / -	- / -	100.0 / 0.13	99.93 / 0.26	- / -
JCLSR [65]	98.94 / -	98.17 / -	- / -	- / -	- / -
ArcFace [9] + MB	97.92 / 0.009	98.73 / 0.012	98.58 / 0.014	98.83 / 0.008	96.12 / 0.022
ArcFace [9] + MB + Ours	99.75 / 0.004	100.0 / 0.000	100.0 / 0.000	100.0 / 0.000	99.96 / 0.001

Table 4: Top-1 accuracy and EER under the ‘closed-set’ protocol. Our method significantly improves the top-1 accuracy and EER with a clear margin.

5.4 Cross-dataset Validation

We perform cross-dataset validation to test the generalization of the proposed method. We train our method, as well as the baseline (ArcFace), on one dataset and test the performance on the other dataset. We test 5 different cross-dataset settings using the MobileFaceNet backbone, results are summarized in Tab. 5. The performance is evaluated in terms of both TAR@FAR, EER.

As shown in Tab. 5, our method consistently improves the performance of ArcFace on all the 5 settings, suggesting strong cross-dataset generalization ability.

Datasets	Method	TAR@FAR=			Top-1	EER
		1e-3	1e-4	1e-5		
M→P	AF	0.9759	0.9499	0.9210	99.93	0.007
	Ours	0.9935	0.9766	0.9622	100.0	0.002
T→P	AF	0.9347	0.8981	0.8509	98.22	0.018
	Ours	0.9918	0.9748	0.9591	100.0	0.003
I→P	AF	0.9364	0.9001	0.8020	97.67	0.019
	Ours	0.9688	0.9224	0.8728	99.04	0.009
T→I	AF	0.8533	0.7872	0.7306	97.47	0.033
	Ours	0.9896	0.9864	0.9745	98.85	0.007
M→I	AF	0.9927	0.9846	0.9717	99.76	0.004
	Ours	1.0000	1.0000	1.0000	100.0	0.000

Table 5: Cross-dataset validation. ‘M’, ‘P’, ‘T’ and ‘I’ represent MPD, PolyU, TCD, and IITD datasets, respectively. M→P indicates the model is trained on M and evaluated on P.

5.5 Palmprint Recognition at Million Scale

To verify the scalability of our proposed method, we test our method on our internal dataset with million samples. The training set contains 19,286 identities and 2.87 million samples, while the test set has 1,000 identities and 0.18 million samples. The images of the dataset are collected parallelly in three places by 19 difference mobile phones (different brands and modes) and 2 IoT cameras. Images of each identity was collected in one session by 4 devices (2 IoT and 2 random mobile phones) and 4 different man-made light conditions. More detailed information and example images of this dataset can be found in Appendix D.

We synthesize 20,000 identities and totally 2 million samples to pretrain the models in this experiment. The performance is evaluated under open-set protocol and we report both TAR@FAR and TAR *v.s.* FAR curves in Tab. 6 and Fig. 7, respectively. The results show that our method consistently improves the performance of the baseline ArcFace method, showing great potential in large-scale palmprint recognition.

Method	Backbone	TAR@				
		1e-5	1e-6	1e-7	1e-8	1e-9
AF [9]	MB	0.9911	0.9770	0.9550	0.9251	0.8833
		0.9934	0.9803	0.9605	0.9301	0.9015
AF [9]	R50	0.9997	0.9986	0.9964	0.9931	0.9879
		0.9999	0.9996	0.9975	0.9943	0.9911

Table 6: Palmprint recognition performance on million scale dataset.

5.6 Palmprint Recognition with Limited Identities

The model performance under a limited number of training identities is critical to privacy-sensitive conditions where collecting training set with large-scale

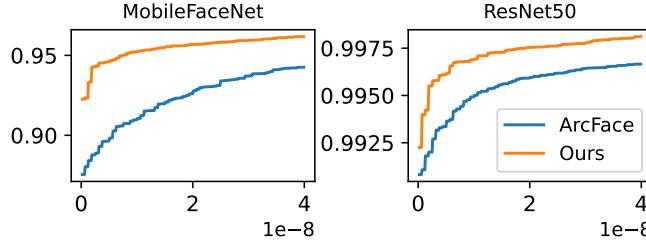


Fig. 7: FAR v.s. TAR curves of ArcFace (AF) and our method on the million-scale dataset.

identities is infeasible. Here we test our method with various training identities. Specifically, under the open-set protocol (train:test = 1:1), we fix the test set and train models with 400, 800 and 1,600 identities. As demonstrated in Tab. 7,

Method	#ID	TAR@FAR=			
		1e-3	1e-4	1e-5	1e-6
ArcFace [9]	1,600	0.9292	0.8568	0.7812	0.7049
ArcFace [9]+Ours		0.9640	0.9438	0.9102	0.8437
ArcFace [9]	800	0.8934	0.7432	0.7104	0.6437
ArcFace [9]+Ours		0.9534	0.9390	0.9025	0.8164
ArcFace [9]	400	0.8102	0.7050	0.6668	0.3320
ArcFace [9]+Ours		0.9189	0.8497	0.7542	0.6899

Table 7: Performance under various training identities. The models are based on the MobileFaceNet backbone.

our method maintains high performance while the ArcFace baseline degrades quickly as the drop of training identities. Even trained with 400 identities, our method still performs on par with the ArcFace counterpart that is trained with 1,600 identities, showing its superiority in identity-constrained scenarios.

5.7 Ablation Study

In this section, we ablate the components and design choices of our method. All the experiments in this ablation study are conducted using the MobileFaceNet [5] and evaluated under the open-set protocol.

Creases synthesis. The main components in our synthesized samples are the principal lines, the wrinkles, and the background images. Tab. 8 presents the results of models with and without these components. ‘P’, ‘W’ and ‘B’ represent the principal lines, wrinkles and image background in the synthesized samples, respectively.

Synthesizing principal lines significantly improves the performance over

P W B	TAR@FAR=			
	1e-3	1e-4	1e-5	1e-6
Baseline	0.9102	0.8259	0.7458	0.7217
✓	0.9514	0.9003	0.7613	0.7513
✓ ✓	0.9597	0.9307	0.8949	0.8061
✓ ✓ ✓	0.9640	0.9438	0.9102	0.8437

Table 8: Ablation of design choices in our method.

the baseline at higher FARs, and the improvements at lower FARs are marginal. With wrinkles, the performance can be further improved especially at lower FARs. Finally, using natural images as the background helps achieve higher performance.

Compared to imagenet pretrain. Many down-stream vision tasks, *e.g.* detection, and segmentation, strongly rely on the imagenet [8] pretrained models. In this experiment, we compare the performance of our synthetically pretrained models to the imagenet pretrained models. We pretrain the MobileFaceNets with the imagenet dataset and our synthesized samples and compare their performance under the open-set protocol (train:test = 1:1). For imagenet pretraining, we follow the training configuration of [23]. It is worth noting that there are 1.2 million images in the imagenet training set and our synthesized dataset consists of only 0.4 million samples (4,000 identities with 100 samples per identity). As demonstrated in Tab. 9, even pretrained with one-third of samples, our proposed method still outperforms the imagenet pretrained model with a clear margin, especially under lower FARs. The experimental result tells that our synthesized dataset is specifically more suitable for palm-print recognition than general vision datasets, *e.g.* imagenet.

Number of synthesized samples and identities. By default, we synthesize 4,000 identities and each of them has 100 images. In this ablation, we fix one

Pretrain	TAR@FAR=			
	1e-3	1e-4	1e-5	1e-6
Imagenet	0.9608	0.9135	0.8294	0.7256
Ours	0.9640	0.9438	0.9102	0.8437

Table 9: Comparison of imagenet and our synthetically pretrained models.

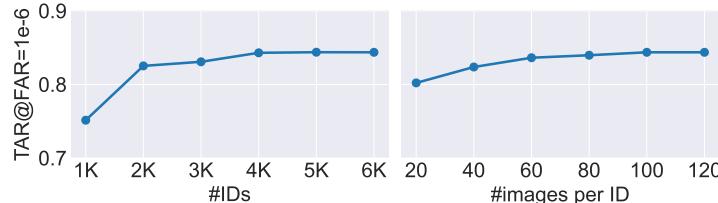


Fig. 8: TAR@FAR=1e-6 of models pretrained with different synthetic samples and identities.

number as the default and vary the other, and evaluate the finetuned performance in terms of FAR@1e-6. The results in Fig. 8 reveal that increasing both the number of samples and identities improves the performance. The number of identities has a greater impact on the fine-tuned performance and the number of samples has less impact.

6 Conclusion

We proposed a simple yet effective geometric model to synthesize palmar creases by manipulating parameterized Bézier curves. The synthetic samples are used to pretrain deep palmprint recognition models and improve model performance. Different from other data synthesizing methods, our method synthesizes samples of novel categories to augment both the identities and samples of the training set. Competitive results on several public benchmarks demonstrate the superiority and great potentials of our approach. Besides, experiments on a million-scale dataset verify the scalability of our method. We also believe our method could benefit some other tasks, *e.g.*fingerprint recognition.

Acknowledgment

We would like to acknowledge Haitao Wang and Huikai Shao for their assistance in processing experimental results.

References

1. Amazon one. <https://one.amazon.com/> 1
2. The coep dataset. <https://www.coep.org.in/resources/coeppalmpintdatabase> 7
3. Baek, S., Kim, K.I., Kim, T.K.: Augmented skeleton space transfer for depth-based hand pose estimation. In: CVPR. pp. 8330–8339 (2018) 4
4. Cao, Q., Shen, L., Xie, W., Parkhi, O.M., Zisserman, A.: Vggface2: A dataset for recognising faces across pose and age. In: FG 2018. pp. 67–74. IEEE (2018) 2
5. Chen, S., Liu, Y., Gao, X., Han, Z.: Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In: Chinese Conference on Biometric Recognition. pp. 428–438. Springer (2018) 9, 10, 13
6. Chen, W., Yu, Z., Wang, Z., Anandkumar, A.: Automated synthetic-to-real generalization. pp. 1746–1756. PMLR (2020) 4
7. Connie, T., Jin, A.T.B., Ong, M.G.K., Ling, D.N.C.: An automated palmprint recognition system. Image and Vision computing **23**(5), 501–515 (2005) 3
8. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: CVPR. pp. 248–255. Ieee (2009) 6, 14
9. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: CVPR. pp. 4690–4699 (2019) 4, 6, 9, 10, 11, 12, 13
10. Dian, L., Dongmei, S.: Contactless palmprint recognition based on convolutional neural network. In: IEEE ICSP. pp. 1363–1367. IEEE (2016) 4
11. Fei, L., Xu, Y., Tang, W., Zhang, D.: Double-orientation code and nonlinear matching scheme for palmprint recognition. PR **49**, 89–101 (2016) 4, 10, 11
12. Fei, L., Zhang, B., Xu, Y., Guo, Z., Wen, J., Jia, W.: Learning discriminant direction binary palmprint descriptor. IEEE TIP **28**(8), 3808–3820 (2019) 11
13. Feng, G., Hu, D., Zhang, D., Zhou, Z.: An alternative formulation of kernel lpp with application to image recognition. Neurocomputing **69**(13–15), 1733–1738 (2006) 3
14. Ferrer, M.A., Vargas, F., Morales, A.: Bispectral contactless hand based biometric system. In: CONATEL 2011. pp. 1–6. IEEE (2011) 7
15. Gaidon, A., Wang, Q., Cabon, Y., Vig, E.: Virtual worlds as proxy for multi-object tracking analysis. In: CVPR. pp. 4340–4349 (2016) 4

16. Genovese, A., Piuri, V., Plataniotis, K.N., Scotti, F.: Palmnet: Gabor-pca convolutional networks for touchless palmprint recognition. *IEEE TIFS* **14**(12), 3160–3174 (2019) [10](#), [11](#)
17. Gibney, E.: The battle for ethical ai at the world’s biggest machine-learning conference. *Nature* **577**(7791), 609–610 (2020) [2](#)
18. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (eds.) *NeurIPS*. vol. 27. Curran Associates, Inc. (2014), <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf> [4](#)
19. Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J.: Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In: *ECCV*. pp. 87–102. Springer (2016) [2](#)
20. Guo, Z., Zhang, D., Zhang, L., Zuo, W.: Palmprint verification using binary orientation co-occurrence vector. *PRL* **30**(13), 1219–1227 (2009) [10](#), [11](#)
21. Hao, Y., Sun, Z., Tan, T., Ren, C.: Multispectral palm image fusion for accurate contact-free palmprint recognition. In: *ICIP*. pp. 281–284. IEEE (2008) [7](#)
22. Hassanat, A., Al-Awadi, M., Btoush, E., Al-Btoush, A., Alhasanat, E., Altarawneh, G.: New mobile phone and webcam hand images databases for personal authentication and identification. *Procedia Manufacturing* **3**, 4060–4067 (2015). <https://doi.org/10.1016/j.promfg.2015.07.977> [7](#)
23. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR*. pp. 770–778 (2016) [9](#), [14](#)
24. He, X., Niyogi, P.: Locality preserving projections. *NeurIPS* **16**, 153–160 (2003) [3](#)
25. Hennings-Yeomans, P.H., Kumar, B.V., Savvides, M.: Palmprint classification using multiple advanced correlation filters and palm-specific segmentation. *IEEE TIFS* **2**(3), 613–622 (2007) [4](#)
26. Hu, D., Feng, G., Zhou, Z.: Two-dimensional locality preserving projections (2dlpp) with its application to palmprint recognition. *PR* **40**(1), 339–342 (2007) [3](#)
27. Jia, W., Huang, D.S., Zhang, D.: Palmprint verification based on robust line orientation code. *PR* **41**(5), 1504–1513 (2008) [4](#), [10](#)
28. Kanhangad, V., Kumar, A., Zhang, D.: Contactless and pose invariant biometric identification using hand surface. *IEEE TIP* **20**(5), 1415–1424 (2010). <https://doi.org/10.1109/TIP.2010.2090888> [7](#)
29. Kemelmacher-Shlizerman, I., Seitz, S.M., Miller, D., Brossard, E.: The megaface benchmark: 1 million faces for recognition at scale. In: *CVPR*. pp. 4873–4882 (2016) [2](#)
30. Khan, Z., Mian, A., Hu, Y.: Contour code: Robust and efficient multispectral palmprint encoding for human recognition. In: *ICCV*. pp. 1935–1942. IEEE (2011) [4](#)
31. Kong, A.K., Zhang, D.: Competitive coding scheme for palmprint verification. In: *ICPR*. vol. 1, pp. 520–523. IEEE (2004) [4](#), [10](#), [11](#)
32. Kumar, A.: Incorporating cohort information for reliable palmprint authentication. In: Indian Conference on Computer Vision, Graphics and Image Processing. pp. 583–590. ICVGIP ’08, IEEE, Bhubaneswar, India (2008). <https://doi.org/10.1109/ICVGIP.2008.73> [7](#)
33. Kumar, A.: Toward more accurate matching of contactless palmprint images under less constrained environments. *IEEE TIFS* **14**(1), 34–47 (2018) [2](#)
34. Laadjel, M., Al-Maadeed, S., Bouridane, A.: Combining fisher locality preserving projections and passband dct for efficient palmprint recognition. *Neurocomputing* **152**, 179–189 (2015) [4](#)

35. Leng, L., Li, M., Kim, C., Bi, X.: Dual-source discrimination power analysis for multi-instance contactless palmprint recognition. *Multimedia Tools and Applications* **76**(1), 333–354 (2017) 4
36. Li, H., Wang, L.: Palmprint recognition using dual-tree complex wavelet transform and compressed sensing. In: Proceedings of 2012 International Conference on Measurement, Information and Control. vol. 2, pp. 563–567. IEEE (2012) 4
37. Liu, Y., Kumar, A.: Contactless palmprint identification using deeply learned residual features. *IEEE TBBIS* **2**(2), 172–181 (2020) 11
38. Lu, G., Zhang, D., Wang, K.: Palmprint recognition using eigenpalms features. *PRL* **24**(9-10), 1463–1467 (2003) 3
39. Luo, Y.T., Zhao, L.Y., Zhang, B., Jia, W., Xue, F., Lu, J.T., Zhu, Y.H., Xu, B.Q.: Local line directional pattern for palmprint recognition. *PR* **50**, 26–44 (2016) 10, 11
40. Matkowski, W.M., Chai, T., Kong, A.W.K.: Palmprint recognition in uncontrolled and uncooperative environment. *IEEE TIFS* (2019). <https://doi.org/10.1109/TIFS.2019.2945183> 11
41. Maze, B., Adams, J., Duncan, J.A., Kalka, N., Miller, T., Otto, C., Jain, A.K., Niggel, W.T., Anderson, J., Cheney, J., et al.: Iarpa janus benchmark-c: Face dataset and protocol. In: ICB. pp. 158–165. IEEE (2018) 2
42. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. *NeurIPS* **32**, 8026–8037 (2019) 9
43. PNVR, K., Zhou, H., Jacobs, D.: Sharingan: Combining synthetic and real data for unsupervised geometry estimation. In: CVPR. pp. 13974–13983 (2020) 4
44. Qian, J., Yang, J., Gao, G.: Discriminative histograms of local dominant orientation (d-hldo) for biometric image feature extraction. *PR* **46**(10), 2724–2739 (2013) 4
45. Qiu, H., Yu, B., Gong, D., Li, Z., Liu, W., Tao, D.: Synface: Face recognition with synthetic data. In: CVPR. pp. 10880–10890 (2021) 3
46. Sang, H., Yuan, W., Zhang, Z.: Research of palmprint recognition based on 2dpca. In: International Symposium on Neural Networks. pp. 831–838. Springer (2009) 3
47. Shao, H., Zhong, D.: Few-shot palmprint recognition via graph neural networks. *Electronics Letters* **55**(16), 890–892 (2019) 4
48. Shao, H., Zhong, D., Du, X.: Effective deep ensemble hashing for open-set palmprint recognition. *Journal of Electronic Imaging* **29**(1), 013018 (2020) 7
49. Shao, H., Zhong, D., Du, X.: Deep distillation hashing for unconstrained palmprint recognition. *IEEE TIM* **70**, 1–13 (2021) 4
50. Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., Webb, R.: Learning from simulated and unsupervised images through adversarial training. In: CVPR. pp. 2107–2116 (2017) 4
51. Sun, Z., Tan, T., Wang, Y., Li, S.Z.: Ordinal palmprint representation for personal identification [representation read representation]. In: CVPR. vol. 1, pp. 279–284. IEEE (2005) 7, 10, 11
52. Svoboda, J., Masci, J., Bronstein, M.M.: Palmprint recognition via discriminative index learning. In: ICPR. pp. 4232–4237. IEEE (2016) 4
53. Tamrakar, D., Khanna, P.: Noise and rotation invariant rdf descriptor for palmprint identification. *Multimedia Tools and Applications* **75**(10), 5777–5794 (2016) 4
54. Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., Birchfield, S.: Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In: CVPRW. pp. 969–977 (2018) 4

55. Van Noorden, R.: The ethical questions that haunt facial-recognition research (2020) [2](#)
56. Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M.J., Laptev, I., Schmid, C.: Learning from synthetic humans. In: CVPR. pp. 109–117 (2017) [4](#)
57. Wang, M., Ruan, Q.: Palmprint recognition based on two-dimensional methods. In: ICSP. vol. 4. IEEE (2006) [3](#)
58. Whitelam, C., Taborsky, E., Blanton, A., Maze, B., Adams, J., Miller, T., Kalka, N., Jain, A.K., Duncan, J.A., Allen, K., et al.: Iarpa janus benchmark-b face dataset. In: CVPRW. pp. 90–98 (2017) [2](#)
59. Wu, X., Zhao, Q., Bu, W.: A sift-based contactless palmprint verification approach using iterative ransac and local palmprint descriptors. PR **47**(10), 3314–3326 (2014) [4](#)
60. Yao, Y., Zheng, L., Yang, X., Naphade, M., Gedeon, T.: Simulating content consistent vehicle datasets with attribute descent. In: ECCV. pp. 775–791. Springer (2020) [4](#)
61. Zhang, D., Guo, Z., Lu, G., Zhang, L., Zuo, W.: An online system of multispectral palmprint verification. IEEE TIM **59**(2), 480–490 (2009) [7](#)
62. Zhang, Y., Zhang, L., Liu, X., Zhao, S., Shen, Y., Yang, Y.: Pay by showing your palm: A study of palmprint verification on mobile platforms. In: 2019 IEEE International Conference on Multimedia and Expo (ICME). pp. 862–867. IEEE (2019) [7](#)
63. Zhang, Y., Zhang, L., Zhang, R., Li, S., Li, J., Huang, F.: Towards palmprint verification on smartphones. arXiv preprint arXiv:2003.13266 (2020) [4, 6, 7](#)
64. Zhao, K., Xu, J., Cheng, M.M.: Regularface: Deep face recognition via exclusive regularization. In: CVPR. pp. 1136–1144 (2019) [4](#)
65. Zhao, S., Zhang, B.: Joint constrained least-square regression with deep convolutional feature for palmprint recognition. IEEE TSMC (2020) [11](#)
66. Zheng, Q., Kumar, A., Pan, G.: Suspecting less and doing better: New insights on palmprint identification for faster and more accurate matching. IEEE TIFS **11**(3), 633–641 (2015) [4, 10, 11](#)
67. Zheng, Q., Kumar, A., Pan, G.: A 3d feature descriptor recovered from a single 2d palmprint image. IEEE TPAMI **38**(6), 1272–1279 (2016) [11](#)
68. Zhong, D., Zhu, J.: Centralized large margin cosine loss for open-set deep palmprint recognition. IEEE TCSVT (2019). <https://doi.org/10.1109/TCSVT.2019.2904283> [4, 10, 11](#)
69. Zuo, W., Lin, Z., Guo, Z., Zhang, D.: The multiscale competitive code via sparse representation for palmprint verification. In: CVPR. pp. 2265–2272. IEEE (2010) [4](#)

A Details synthesis of principal lines

We elaborate the steps of the synthesizing details using a left hand of 3 principal lines.

- Fig. 9 (a): Randomly select starting points (s_1, s_2, s_3) and ending points (e_1, e_2, e_3) along the top-left and bottom-right edge of the coordinate. We set a simple rule to ensure that the lines do not intersect: $S_3 > s_2 > s_1$ and $S_3 > s_2 > s_1$.

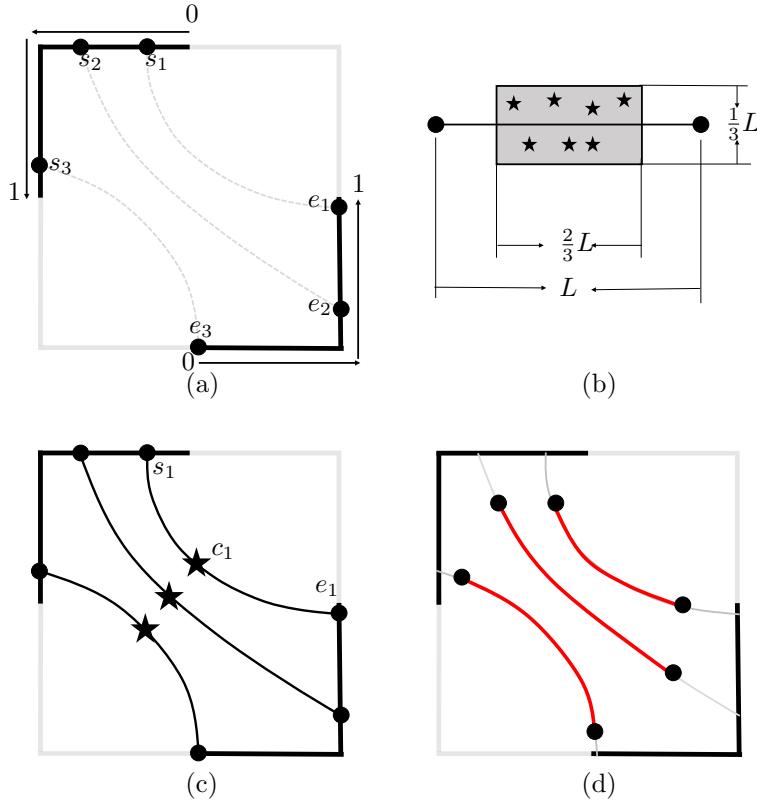


Fig. 9: Detailed steps of principal lines of a left hand.

- Fig. 9 (b): The control points are randomly sampled from a rectangle area. Let L be the length of the line between s and e , the width and height of the rectangle are $\frac{2}{3}L$ and $\frac{1}{3}L$.
- Fig. 9 (c): Bézier curves are determined given the starting, end and control points (s, e, c) . Let $f(t) : \mathbb{R}^{[0,1]} \rightarrow \mathbb{R}^2$ be the parametric function, where $s = f(0)$ and $e = f(1)$.
- Fig. 9 (d): Randomly sample $t_0 \in (0, 0.3)$, $t_1 \in (0.7, 1)$ so that the finally curves (red) are $f(t), t \in [t_0, t_1]$.

B Algorithmic pipeline of the synthesizing process

Algorithm 1 illustrates the pipeline of the synthesizing process. N and S represent the number of identities and number of samples per identity. The between identity randomness and within identity randomness are represented by different colors.

Algorithm 1 Algorithmic pipeline for creases synthesis

```

1: for  $i \in \{1, 2, \dots, N\}$  do
2:    $m = \text{randint}(3, 5)$ 
3:    $n = \text{randint}(5, 15)$ 
4:    $P = \text{random}(0, 1, \text{size}=(m, 3, 2))$ 
5:    $Q = \text{random}(0, 1, \text{size}=(n, 3, 2))$ 
6:   for  $j \in \{1, 2, \dots, S\}$  do
7:      $P_j^i += \text{random}(P, \text{std}=0.04)$ 
8:      $Q_j^i += \text{random}(Q, \text{std}=0.01)$ 
9:      $bg = \text{random\_select(imagenet)}$ 
10:     $S_j^i = \text{synthesize}(P_j^i, Q_j^i, bg)$ 
11:   end for
12: end for

```

C Example of synthesized images

Fig. 10 and Fig. 11 present example synthesized images without and with imangenet images as the background.

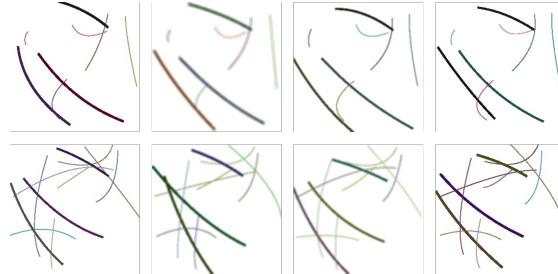


Fig. 10: Example of synthesized images without imangenet images as background. Each row contains sample of the same identity.



Fig. 11: Example of synthesized images with imangenet images as background. Each row contains sample of the same identity.

D Details about the million-scale dataset

The images of the the dataset are collected parallelly in three places by 19 difference mobile phones (different brands and modes) and 2 IoT cameras. Images of each identity was collected in one seesion by 4 devices (2 IoT and 2 random mobile phones) and 4 different man-made light conditions. We provide selected



Fig. 12: Example images of two identities (each row corresponds to an identity) our million-scale dataset.

palms of two identities in the figure below (zoom in for details).