Few-Shot Learning for Palmprint Recognition via Meta-Siamese Network

Huikai Shao, *Graduate Student Member, IEEE*, Dexing Zhong, *Member, IEEE*, Xuefeng Du, Shaoyi Du, *Member, IEEE*, and Raymond N. J. Veldhuis, *Senior Member, IEEE*

***Abstract*— Palmprint is one of the discriminant biometric modalities of humans. Recently, deep learning-based palmprint recognition algorithms have improved the accuracy and robustness of recognition results to a new level. Most of them require a large amount of labeled training samples to guarantee satisfactory performance. However, getting enough labeled data is difficult due to time consumption and privacy issues. Therefore, in this article, a novel meta-Siamese network (MSN) is proposed to exploit few-shot learning for small-sample palmprint recognition. During each episode-based training iteration, a few images are selected as sample and query sets to simulate the support and testing sets in the test set. Specifically, the model is trained episodically with a flexible framework to learn both the feature embedding and deep similarity metric function. In addition, two distance-based losses are introduced to assist the optimization. After training, the model can learn the ability to get similarity scores between two images for few-shot testing. Adequate experiments conducted on several constrained and unconstrained benchmark palmprint databases show that MSN can obtain competitive improvements compared with baseline methods, where the best accuracy can be up to 100%.**

***Index Terms*— Biometrics, few-shot learning, information security, meta-learning, palmprint recognition.**

I. Introduction

BIOMETRICS is an effective technology using human's physiological or behavioral characteristics for authentication [1]. Recently, several biometric modalities have been widely applied in daily life, such as face recognition [2] and fingerprint recognition [3]. As one of the unique technologies of biometrics, palmprint recognition has received much research attention recently [4], [5]. Generally, researchers applied signal processing methods to analyze the patterns of palmprint for personal authentication [6]. So far, local texture [7] and principal lines [8] have been exploited for feature representation. They are time-invariant with large interclass variance and low intraclass variance. Therefore, promising recognition results have been achieved with high universality, stability, and uniqueness.

Typical procedure of palmprint recognition consists of image acquisition, preprocessing, feature extraction, and matching [9], [10]. Palmprint image acquisitions are usually employed by optical cameras. Preprocessing is mainly adopted to implement noise reduction and region of interest (ROI) extraction. Then, several categories of feature extraction and matching methods are proposed to separate different identities, e.g. encoding-based methods, structure-based methods, statistics methods, and subspace methods [9]. So far, deep learning techniques have emerged as effective tools for automatic visual understanding and obtained the state of the arts in many computer vision tasks. There have been several deep learning-based palmprint recognition models, which outperform other traditional algorithms [5], [11], [12].

However, current palmprint recognition methods based on deep learning indeed have some application disadvantages. One of the major problems is the requirements of large amounts of training samples and labels [13], [14]. Generally, collecting enough data is laborious, and labeling them accurately is also a heavier work. In addition, although palmprint recognition may be more private than face recognition in practical application, sometimes users are still reluctant to provide many images for training and registration due to privacy considerations. Furthermore, even if we have enough data, training on large amounts of data would be computationally expensive with poor generalization ability. Therefore, constructing an accurate and effective palmprint recognition system, which requires less training data, is a significant task.

The problem reflected above is the so-called small-sample-size (SSS) problem [15], which is also called few-shot recognition. It is formulated as follows: given a dataset S consisting of palmprint images, and for each category, we have k samples labeled (k is a small number). The task is how can we recognize the rest images in each category using a little

Manuscript received March 16, 2021; revised April 19, 2021; accepted April 24, 2021. Date of publication April 30, 2021; date of current version May 14, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61105021, in part by the Natural Science Foundation of Zhejiang Province under Grant LGF19F030002, in part by the Natural Science Foundation of Shaanxi Province under Grant 2020JM-073, in part by the Fundamental Research Funds for the Central Universities under Grant xzy022020051, and in part by the China Scholarship Council. The Associate Editor coordinating the review process was Hongrui Wang. *(Corresponding author: Dexing Zhong.)*

Huikai Shao is with the School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: shaohuikai@stu.xjtu.edu.cn).

Dexing Zhong is with the School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China, also with the Pazhou Laboratory, Guangzhou 510335, China, and also with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China (e-mail: bell@xjtu.edu.cn).

Xuefeng Du is with the Department of Computer Science, University of Wisconsin–Madison, Madison, WI 53201 USA (e-mail: dxfsxl@163.com).

Shaoyi Du is with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: dushaoyi@gmail.com).

Raymond N. J. Veldhuis is with the Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, 7500 AE Enschede, The Netherlands (e-mail: r.n.j.veldhuis@utwente.nl).

Digital Object Identifier 10.1109/TIM.2021.3076850

labeled training data. It is similar to the practical palmprint recognition scenarios, where a few images are registered in the database and the query images need to be matched with the registration images to determine the tester's identity. One of the effective solutions is meta-learning [16], which aims to train the deep neural networks (DNNs) for generalizing on different tasks. Inspired by it, in this article, we propose a few-shot palmprint recognition method called meta-Siamese network (SN) (MSN).

In order to help DNN to generalize to new palmprint images, our MSN follows the structure of SN [17], which firstly extracts feature from image pairs using weight-shared convolutional neural network (CNN). Then, the feature vectors extracted are concatenated together and input to follow-up decision network to obtain their similarity. Finally, the mean square error (MSE)-based similarity losses are adopted and backpropagated so that it can verify whether palmprint image pairs are from the same individual.

However, different from SN, meta-learning is introduced to improve the generalization ability and the model is trained through episode-based iteration. Specifically, images of N categories are firstly randomly sampled from train set, and for each category, k images are selected, which are matched with the images in query set (denoted as a N-way, k-shot episode task). In the testing stage, a similar strategy is adopted to select support set and testing set from test set, where the former is labeled and the latter is unlabeled, and N-way, k-shot tasks are also randomly sampled. The model is trained on a large number of N-way, k-shot tasks in the training set, and finally it can adapt the new N-way, k-shot tasks to obtain the similarity scores of different palmprint image pairs. In order to help optimize the model and improve the performance, two distance-based losses, contrastive loss [18] or binomial deviance (BD) [19], are introduced to constrain the distance between features in the feature space. Specially, to increase the flexibility, the convolutional blocks are incorporated in the decision network instead of pure stacked fully connected (FC) layers so that the entire neural networks can adapt quickly. Experiments on several popular benchmark palmprint datasets reveal the outperforming accuracy and generalization ability of our model. The details can be found in Second III. The overview of MSN is shown in Fig. 1.

The contributions can be summarized as follows.

1) MSN is proposed for efficient few-shot palmprint recognition. Its core is to directly imitate the identification task of the test in the training phase to improve the accuracy. After the episode-based training on the tasks in training set, the model can be applied to the test set for new few-shot palmprint recognition tasks.
2) MSE-based similarity loss is applied to measure the similarity scores of palmprint image pairs to determine their categories. Distance-based losses are further constructed to assist in training the model and improve the accuracy.
3) Adequate experiments are conducted on several constrained and unconstrained benchmark palmprint databases. From the results, MSN can obtain promising performance and the best accuracy can be up to 100%. Furthermore, compared with the previous models, MSN can outperform others to obtain the state-of-the-art palmprint recognition.

Compared with our previous work in [20], we have made some significant improvements. First, in addition to the previous similarity loss, two other losses are constructed to constrain the distance of image pairs in the feature space directly, i.e. contrastive loss and BD. Though obtaining the category relations between two images matched through neural networks can reduce the impact of manual intervention, the distance constraints on them will be beneficial and improve the performance, which is shown in the results. Here, distance-based losses can make positive matching features closer while negative matching features farther in the feature space. Second, eight new unconstrained palmprint databases and four benchmark palmprint databases are introduced in the experiments to verify the effectiveness of our modified algorithms. Third, more adequate analyses and comparisons are conducted with the state-of-the-art algorithms to demonstrate the superiority of our algorithms, especially the recent few-shot palmprint recognition methods.

The remainder of this article is structured as follows. Section II reviews some related works. Our methods are described in Section III in detail. Section IV presents our experiments and results on several databases. Analysis of results is in Section V. Section VI gives a conclusion for this article.

II. Related Work

A. Palmprint Recognition

Traditional palmprint recognition algorithms mainly extract its rich main line, texture, and wrinkle features. One kind of the commonly used methods is based on orientation code. They convolve palmprint images with a list of Gabor filters with several orientations and convert them into codes as features, such as competitive code [21], binary orientation co-occurrence vector (BOCV) [22], extend BOCV (E-BOCV) [23], double-orientation code (DOC) [24], discriminative and robust competitive code (DRCC) [25], and so on. Using the multiplication and addition schemes, Fei *et al.* [26] fused the apparent and latent direction features of palmprint and proposed a unique double-layer direction extraction method, called apparent and latent direction code (ALDC). Luo *et al.* [27] proposed local line directional patterns (LLDP) which operated in local line-geometry space for palmprint recognition. Zhang *et al.* [6] established a contactless palmprint database and proposed CR_CompCode for palmprint identification with low computational complexity. Fei *et al.* [26] extracted six discriminant direction binary codes (DDBCs) for each pixel of palmprint image and concatenated them as the global feature vector, called discriminant direction binary palmprint descriptor (DDBPD). Toward more accurate direction representations, Jia *et al.* [28] extracted the direction features of palmprint on more levels such as multiscale, multidirection level, and multiregion. Zhang *et al*. [6] proposed a unique local descriptor to extract both direction descriptors and thickness features, called local microstructure tetra pattern (LMTrP).

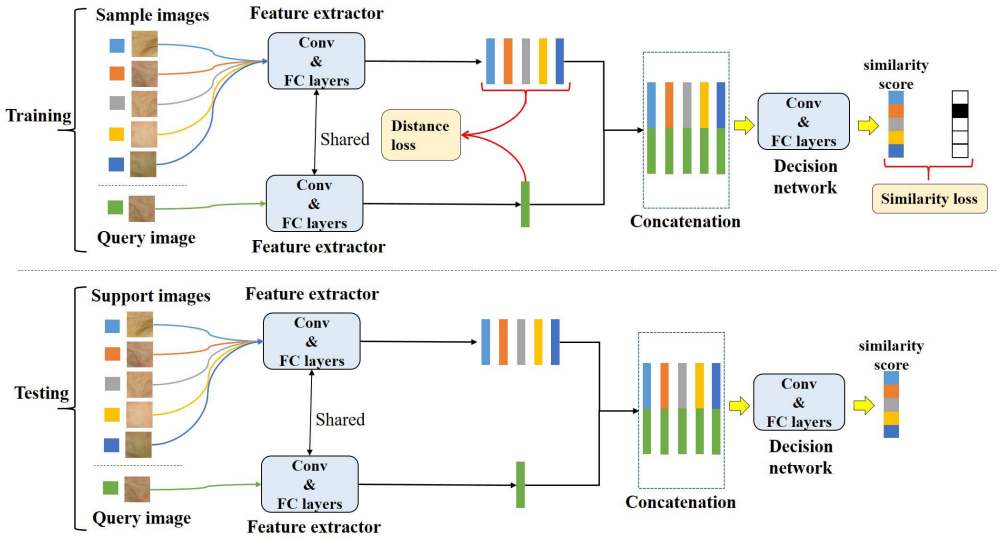

Fig. 1. Overview of our MSN (5-way, 1-shot). Sample and query images are selected from training set randomly to imitate the N-way, k-shot recognition tasks in test set, where support and testing images are selected randomly during testing. Sample and query images are input into weight-shared feature extractors to get feature vectors. Then, they are concatenated and input into decision network to obtain similarity score. Similarity loss and distance loss are constructed to optimize the model. After training, the network can be used for new N-way, k-shot tasks to distinguish the identity of other images in the test set.

However, the methods abovementioned need to design and extract feature manually, which are easily affected by subjective factors, especially for unconstrained palmprint images. Recently, CNN exhibits powerful feature extraction capability, and deep learning-based palmprint recognition methods obtain the state-of-the-art performance. Genovese *et al.* [11] proposed a novel CNN model called PalmNet using principal component analysis (PCA) and Gabor responses. Shao and Zhong [29] adopted deep hashing network (DHN) for palmprint recognition and fused it with dorsal hand vein to perform multibiometrics, which transferred palmprint images into binary codes. Meraoumia *et al.* [30] proposed a novel deep learning architecture to extract texture information for palmprint identification, called PCANet. Shao and Zhong [29] proposed graph neural network (GNN) for few-shot palmprint recognition, where the features extracted by CNN were processed into nodes and the edges were used to represent similarities between images. Matkowski *et al.* [5] established a new palmprint database collected from uncontrolled and uncooperative environment and proposed an end-to-end deep learning algorithm for ROI extraction and feature matching. Zhao and Zhang [31] proposed deep discriminative representation (DDR) to extract high-level discriminative features for palmprint recognition. Some researchers also proposed algorithms for cross-dataset recognition, such as PalmGAN [32] and transfer convolutional autoencoder [29].

These deep learning-based palmprint recognition methods can obtain promising accuracy. However, they require many labeled palmprint images, which consumes a lot of manpower and material resources. Different from them, in this article, we focus on few-shot palmprint recognition and propose a novel MSN method, which can achieve good performance based on a few labeled data.

B. Few-Shot Learning

CNN uses spatial filters to compute the weighted sum in image patches, which may cause huge computation complexity and weak generalization ability. As an application of meta-learning, few-shot learning is introduced to solve the problem of overfitting when CNN is faced with sparse data, which samples auxiliary tasks to help DNN generalize [33]. There exist three types of few-shot learning methods, namely recurrent neural network (RNN)-based, initialization-based, and metric-based models. RNN-based methods take the advantage of memory and train RNN models to remember the seen tasks [34]. When facing tasks with a different distribution, it compares them with its memory [35]. The second models aim to help DNN learn to update its parameters within a few gradient steps; therefore, the sampled tasks in training resemble the task setting in testing [33]. The metric-based models learn the image embedding functions so that the subspace features after embedding are easier to classify [36].

Specifically, Snell *et al.* [37] proposed prototypical networks for few-shot learning and computed distances to prototype features of every category to perform the classification. Then, Ren *et al.* [38] modified it and proposed a novel few-shot learning algorithm when unlabeled examples were available for producing prototypes. The model was trained in an end-to-end way on episodes and learned to leverage the unlabeled examples successfully. Sung *et al.* [33] proposed relation network (RN) to learn a deep distance metric to compare images and classify new classes by computing their relation scores. Finn *et al.* [39] proposed a model-agnostic meta-learning algorithm, and the parameters could be explicitly trained using a few data for a new task.

Kang *et al.* [40] proposed a few-shot object detector using a meta-feature extractor and a reweighting learner within a one-stage detection structure. Recently, researchers also proposed GNN and knowledge graph-based algorithms for few-shot learning and achieved promising results, such as graph few-shot learning (GFL) [41], GNN [42], hybrid knowledge routed modules (HKRM) [43], and so on. Liu *et al.* [44] proposed a novel transductive propagation network (TPN) for few-shot classification task. They proposed to learn a graph construction module to exploit the manifold structure in the data and classify the entire test set at once. Wang *et al.* [45] proposed a novel dubbed instance credibility inference (ICI) method to perform the distribution support of unlabeled instances for few-shot learning. Ravi and Beatson [46] proposed a novel meta-learning method to amortize hierarchical variational inference across tasks so that a few steps of Bayes by backprop can produce useful task-specific approximate posteriors. Li *et al.* [47] proposed TargetNet and MetaNet to learn transferable knowledge across different tasks, which can construct parameters for similar unseen tasks.

The improvement of our work compared with [17] and [48] is that we add meta-episode training to improve the generalization ability. Besides, we add CNN blocks in the decision network which enables more flexible training. Compared with RN [33], we introduced two distance-based losses to help optimize the feature distributions, which can achieve promising improvement on the performance. Compared with [35], our model does not require designing complex memory encoding and decoding structure.

III. Methods

A. Task Description

Few-shot palmprint recognition can be formulated as training a classifier to recognize the remaining images given a few labeled images for each category in the test set. Due to lack of enough labeled data, if the classifier is directly trained using traditional optimization algorithms (e.g. *softmax* loss), it will suffer from overfitting and may not obtain satisfactory performance. So the meta-learning is proposed [16]. Suppose there is an image dataset, $\{(x_0, y_0), \ldots, (x_i, y_i), \ldots, (x_n, y_n)\}$, and y_i is the label of x_i. The dataset is split into training set and test set. The test set is further divided into support set and testing set, and we aim to match the images of testing set to the support set accurately. It is similar to palmprint identification scenario, where the test images are matched to the registration samples in the database to determine their identity. During training, the training set is separated to sample set and query set to simulate the few-shot recognition setting in the test set. Note that the label space of the sample/query sets is different from that of support/testing sets. In addition, the images and labels are not used at all during training.

Similar to [49], in each episode-based training iteration, M images of N categories are randomly selected from training set, and $N \times k < M$ images are further randomly selected to denote the small sample set as $\{(x_i, y_i), i = 1, \ldots, N \times k\}$ for training, where each class has k samples. Then, the remaining images are denoted as query set, $\{(x_i, y_i), i = N \times k + 1, \ldots, M\}$. Finally, N-way k-shot tasks are formed to train the model in an episode manner. In testing, the samples are randomly selected from test set to form the support and testing sets, which is similar to training set but the support set has labels and the testing set has no labels. The losses obtained on the sample and query sets are backpropagated to adapt the model to new tasks. Therefore, after the episode-based training, the model can be applied to the test set for new few-shot palmprint recognition tasks.

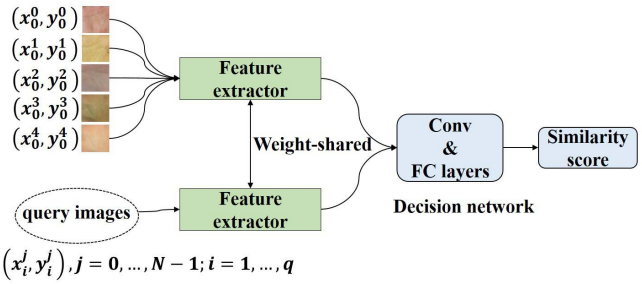

Fig. 2. Pipeline of 5-way, 1-shot palmprint recognition. Five images of five categories are randomly selected to be matched with the query images, and then their similarity scores are obtained.

B. One-Shot Palmprint Recognition

For one-shot learning, there is only one labeled image for each category in the sample set and support set. A schematic of 5-way, 1-shot palmprint recognition is shown in Fig. 2. During training, firstly, N images from N categories (one image in each category) in training set are randomly sampled to form the sample set $\{(x_0^0, y_0^0), (x_0^1, y_0^1), \ldots, (x_0^{N-1}, y_0^{N-1})\}$, and the remaining images of the same N classes are denoted as the query set $\{(x_i^i, y_i^j), j = 0, \ldots, N-1; i = 1, \ldots, q\}$. Here, we suppose each class consists of q images in total. During every episode-based iteration, one image, x_0^s, in the sample set and one image, x_i^j, in the query set are input to two CNN feature extractor modules $f(\cdot)$, respectively, which have the same weights, to obtain two feature vectors, $f(x_0^s)$ and $f(x_i^j)$. Then, they are concatenated together in the depth dimension to obtain the concatenated feature map, $C(f(x_0^s), f(x_i^j))$. Finally, by down-sampling and dimension transformation in the decision network $D(\cdot)$, a similarity score with respect to the current image pair is calculated. When the score is higher, it means the two images are more likely to belong to the same individual. Therefore, for one episode training iteration, the similarities between different image pairs are compared for $N \times N \times (q-1)$ times.

For parameter optimization, an MSE-based similarity loss is adopted. A one-hot vector, Y, is constructed to denote the ground-truth value, which is a $N \times N \times (q-1)$ dimensional vector. The t-*th* value of Y is formulated as follows:

$$Y_t = \begin{cases} 1, & \text{if } t\text{-thimage pair is positive} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The predicted similarity score is set as

$$\hat{Y}_t = D\big[C\big(f\big(x_0^s\big), f\big(x_i^j\big)\big)\big]. \tag{2}$$

Then, the similarity loss is adopted

$$L_s = \sum_{t=1}^{N\times N\times(q-1)} (Y_t - \hat{Y}_t)^2. \tag{3}$$

Though the loss above can be optimized to constrain the similarity score of positive matching approaching 1 and the score of negative matching approaching 0, the score is obtained through decision network, which needs to be carefully designed and takes some time to train. Therefore, inspired by metric learning, another distance-based loss is constructed for auxiliary optimization, which can be contrastive loss [18] or BD [19].

For images x_0^s and x_i^j, the contrastive loss is

$$L_c\left(x_0^s, x_i^j\right) = S_{0i}\, d\left(f\left(x_0^s\right), f\left(x_i^j\right)\right) + (1 - S_{0i}) \max\left(m - d\left(f\left(x_0^s\right), f\left(x_i^j\right)\right), 0\right) \tag{4}$$

where $d(f(x_0^s), f(x_i^j))$ is the Euclidean distance between $f(x_0^s)$ and $f(x_i^j)$, and m is a distance margin. S_{0i} is the image pair label and set to 1 for positive image pairs and 0 for negative image pairs.

For BD, it is

$$L_{\mathrm{BD}}\left(x_0^s, x_i^j\right) = \log\left(1 + e^{-(2c_{0i}-1)\beta_1 c_{0i} - \beta_2)\alpha}\right) \tag{5}$$

$$c_{0i} = \frac{f\left(x_0^s\right)^T f\left(x_i^j\right)}{\left\| f\left(x_0^s\right)\right\| \left\| f\left(x_i^j\right)\right\|} \tag{6}$$

where c_{0i} is their cosine similarity score. β_1 and β_2 are scaling and translation parameters and set to 2 and 0.5, like [50]; α is a balance weight and set to 1 for positive image pairs and 25 for negative image pairs.

Through these two losses, the features of positive matching can be as close as possible and the features of negative matching can be far away.

Therefore, the overall loss is

$$L = L_s + wL_d \tag{7}$$

where w is a parameter to balance the weights of two losses and L_d can be L_c or L_{BD}.

At the testing stage, the support set and testing set are constructed in a similar way to the sample set and query set. For each of the $N \times (q-1)$ testing images, it is compared with every image in the support set, and the image in the support set which generates the highest similarity score is selected. Then, the label of support image selected is compared with the label of testing image to evaluate whether our model generates the right recognition result, and the accuracy can be calculated.

C. k-Shot Palmprint Recognition

For k-shot palmprint recognition, similar to one-shot recognition, k images are sampled instead of one image for each category in sample set and the remaining images are regard as query images. The difference is that we input k sample images along with query images to MSN at each episode training. In order to help the neural networks to see more information of a certain class, all k feature maps of sample images are summed up and then their averaged feature map is obtained. Afterward, the averaged feature vector is concatenated with the features from the query images and input to decision network to get similarity score. Similarly, the distance between the averaged feature vector and query feature is also constrained through distance-based losses. After that, the following matching and comparison procedures are the same as the one-shot palmprint recognition. The diagram of few-shot palmprint recognition scenario is shown in Fig. 3.

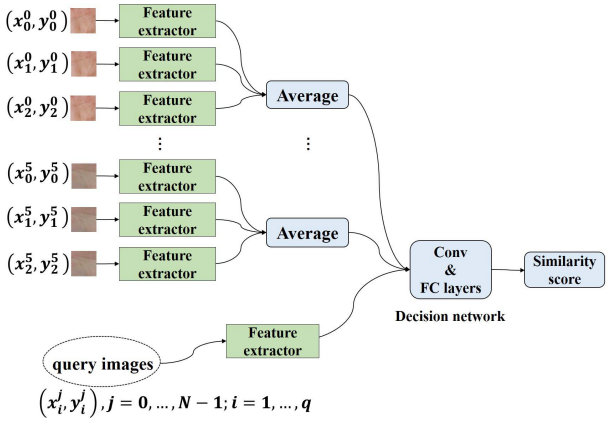

Fig. 3. Pipeline of few-shot recognition (5-way, 3-shot). For every category, three images are selected as sample images to get the features, which are averaged to a feature map. Then, the averaged feature maps are matched with query features to obtain the similarity scores.

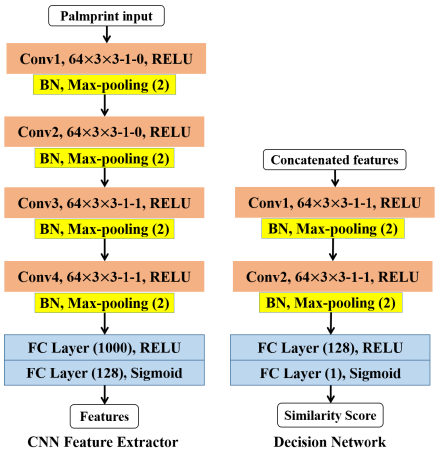

Fig. 4. Structure of two sub-modules, namely CNN feature extractor and decision network. The configuration of each layer is shown in the box. For instance, the "64 × 3 × 3.1-0" means there are 64 convolutional filters with 3 × 3 kernel size, 1 stride, and 0 padding. "FC layer (128)" means we have 128 hidden units in the FC layer. Activation functions and batch normalization are also demonstrated in each box.

In this article, CNN feature extractor is adopted to obtain convolutional features and decision network is adopted to get similarity score. Their parameters are shown in Fig. 4.

IV. Experiments and Results

A. Databases

PolyU multispectral palmprint database consists of four spectral bands, i.e. blue, red, green, and near-infrared (NIR) [51]. Under each spectral band, there are 6000 images

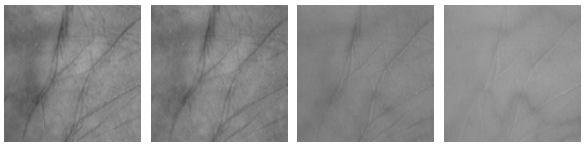

Fig. 5. Some typical ROI samples of PolyU multispectral palmprint database. (a) Blue, (b) green, (c) red, and (d) NIR.

Fig. 6. Some typical samples of Tongji palmprint database. (a) and (b) Original images. (c) ROI.

collected from 500 different palms of 250 individuals, including 195 males and 55 females. For each palm, 12 images were collected in two sessions, and six images in each session. The palmprint images are oriented and cropped to form ROIs with the size of 128 × 128 pixels. According to different spectral bands, this database can be used as four sub-databases, recorded as blue, red, green, and NIR. Fig. 5 shows some typical samples.

Tongji contactless palmprint database contains 12 000 images captured from 300 individuals [6]. For each category, there are 20 palm images. During acquisition, the volunteers can move their hands up and down freely within the enclosed space. A ring white LED light source was used and the background was still pure black, which reduces the difficulty of recognition. Some examples are shown in Fig. 6, and ROIs with the size of 128 × 128 pixels are extracted based on [6].

Xi'an Jiaotong University Unconstrained Palmprint (XJTU-UP) database was established using five mobile phones, i.e., Samsung Galaxy Note5, iPhone 6S, LG G4, HUAWEI Mate8, and MI8 [52], [53]. The volunteers can freely move their hands during acquisition and the backgrounds are complex natural scenes. Two kinds of illuminations were adopted, indoor natural illumination and the flash lighting of mobile phone. There are ten sub-databases in XJTU-UP database, named HUAWEI Mate8 under Natural illumination (HN), iPhone 6s under Natural illumination (IN), LG G4 under Natural illumination (LN), MI8 under Natural illumination (MN), Samsung Galaxy Note5 under Natural illumination (SN), HUAWEI Mate8 under Flash illumination (HF), iPhone 6s under Flash illumination (IF), LG G4 under Flash illumination (LF), MI8 under Flash illumination (MF), and Samsung Galaxy Note5 under Flash illumination (SF). Each sub-database contains RGB hand images from 100 individuals and each category has ten images. The ROIs are extracted based on the key points and methods provided by [52]. Fig. 7 shows some exemplary samples of XJTU-UP database.

Furthermore, three other palmprint databases are also adopted to evaluate the model, NUIG_Palm1 [54], CASIA [55], and IITD [56]. NUIG_Palm1 database is also collected by mobile phones under different conditions,

TABLE I
SOME DETAILS OF DIFFERENT PALMPRINT DATABASES

Database	Time	Type	Main characteristics	Images	Hands
PolyU multispectral	2010	Contacted	Constrained	24,000	500
Tongji	2017	Contactless	Constrained	12,000	600
XJTU-UP	2020	Mobile phones	Unconstrained	20,000	200
NUIG_Palm1	2017	Mobile phones	Unconstrained	1616	81
CASIA	2005	Contactless	Constrained	5502	624
IITD	2006	Contactless	Constrained	2601	460

which consists of 1620 image from 81 hands. CASIA palmprint database contains 5502 palm images collected from 312 individuals. There are 2601 images collected from 230 individuals in IITD palmprint database. All of them are cropped to form the ROIs using official methods. After preprocessing, five images are selected for each palm. The details of these palmprint databases are summarized in Table I.

B. Implementation Details

In this article, the image dataset is randomly split into the training set and test set with the ratio of 1:1. The model is evaluated in 15 different constrained and unconstrained palmprint sub-datasets on different task settings, which are 5-way/1-shot, 5-way/3-shot, 5-way/5-shot, 15-way/1-shot, 15-way/3-shot, and 15-way/5-shot recognition tasks. Images are resized to 128 × 128 and input to the networks. The entire experiments are implemented using PyTorch framework on NVIDIA GPU GTX 1080 and i7-3.30 GHz processors. The base learning rate is set to 0.001, and the Adam Optimizer and Stochastic Gradient Descent (SGD) are adopted.

C. Results

1) Performance on PolyU Multispectral Palmprint Database: For each sub-database, 3000 images from 250 categories are used as training set and the remaining images are used as test set. For 5-way, 1-shot recognition scenario, the size of sample set is 5 and the size of query set is 55. For 5-way, 3-shot recognition scenario, the size of sample set is 15 and the size of the query set is 45. For 5-way, 5-shot recognition scenario, the size of sample set is 25 and the size of query set is 35. Similarly, for 15-way, 1-shot recognition scenario, the size of sample set is 15 and the size of query set is 165. The results on different experiment settings are shown in Table II. The top 1 accuracy is marked in bold (same below). "MSE" means the overall loss $L = L_s$, "Con" means $L = L_s + wL_c$, and "BD" means $L = L_s + wL_{\mathrm{BD}}$, and the description below is similar, unless otherwise specified. From the results, for Blue, the optimal accuracy is 100% on 5-way, 5-shot recognition using similarity loss and contrastive loss. For Green, the optimal accuracy is 100% both on 5-way, 3-shot and 5-way, 5-shot recognition using similarity loss and BD loss. For NIR, the optimal accuracy is 100% on 5-way, 5-shot recognition using similarity loss and BD loss. For Red, the accuracies on

TABLE II
FEW-SHOT RECOGNITION ACCURACIES (%) ON MULTISPECTRAL DATABASE

Data bases	Loss	5-way			15-way		
		1-shot	3-shot	5-shot	1-shot	3-shot	5-shot
Blue	MSE	99.956	99.993	99.997	99.043	99.984	**99.999**
	Con	**99.973**	99.995	**100**	**99.856**	**99.987**	99.996
	BD	**99.973**	**99.997**	99.997	99.143	**99.987**	99.996
Green	MSE	99.917	99.989	99.993	98.747	99.992	99.988
	Con	99.907	99.991	99.996	**99.633**	**99.994**	99.981
	BD	**99.933**	**100**	**100**	99.003	99.989	**99.992**
NIR	MSE	99.987	99.996	99.995	99.113	**99.999**	99.995
	Con	**99.993**	99.996	99.997	**99.149**	99.998	99.995
	BD	99.967	**99.997**	**100**	98.440	99.996	**99.997**
Red	MSE	99.967	99.995	99.996	98.999	99.996	**100**
	Con	**99.986**	99.997	**99.997**	**99.827**	**99.999**	99.998
	BD	99.980	**100**	**99.997**	99.416	99.995	**100**

TABLE III
FEW-SHOT RECOGNITION ACCURACIES (%) ON TONGJI PALMPRINT DATABASE

Loss	5-way			15-way		
	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot
MSE	99.890	99.947	**99.989**	99.718	99.959	99.988
Con	**99.933**	**99.971**	**99.989**	**99.789**	99.962	**99.991**
BD	99.893	99.960	99.985	99.733	**99.982**	99.990

5-way, 3-shot using similarity loss and 15-way, 5-shot recognition using similarity loss and BD loss are 100%. In summary, almost all results of different settings exceed 99%.

2) Performance on Tongji Contactless Palmprint Database: For Tongji palmprint database, 6000 images from 300 categories are selected as training set and the remaining 6000 images are as test set. Each category has 20 images, so for 5-way, 1-shot recognition scenario, the size of sample set is 5 and the size of query set is 95. For 5-way, 3-shot recognition scenario, the size of sample set is 15 and the size of query set is 85. For 5-way, 5-shot recognition scenario, the size of sample set is 25 and the size of query set is 75. For 15-way, 1-shot recognition scenario, the size of sample set is 15 while the size of query set is 285. For 15-way, 3-shot recognition scenario, the size of sample set is 45 and the size of query set is 255. For 15-way, 5-shot recognition scenario, the size of sample set is 75 and the size of query set is 225. The results are presented in Table III and the optimal accuracy is 99.991% on 15-way, 5-shot recognition using similarity loss and contrastive loss. Though the database is collected in a contactless manner, its performance is also relatively good.

3) Performance on XJTU-UP Database: XJTU-UP database consists of ten sub-databases. For each sub-database, the first 100 categories are used as training set and the remaining images are used for testing. Each category contains ten images, and for 5-way, 1-shot recognition scenario, the size of sample set is 5, while the size of query set is 45. For 5-way, 3-shot recognition scenario, the size of sample set is 15 and the size of query set is 35. For 5-way, 5-shot recognition scenario, the size of sample set is 25 and the size of query set is also 25. Similarly, for 15-way, 1-shot recognition scenario, the size of sample set is 15 and the size of query set is 135. For 15-way, 3-shot recognition scenario, the size of sample set is 45 and the size of query set is 105. For 15-way, 5-shot recognition scenario, the size of sample set is 75 and the size of query set is 75. The results are listed in Table IV.

TABLE IV
FEW-SHOT RECOGNITION ACCURACIES (%) ON XJTU-UP DATABASE

Data bases	Loss	5-way			15-way		
		1-shot	3-shot	5-shot	1-shot	3-shot	5-shot
IF	MSE	99.323	99.933	99.941	98.734	99.545	**99.785**
	Con	**99.500**	**99.953**	**99.954**	**98.896**	**99.676**	99.714
	BD	99.380	99.819	99.829	98.258	99.593	99.766
IN	MSE	99.243	99.878	99.867	98.686	99.633	**99.833**
	Con	**99.353**	**99.908**	99.861	98.469	99.335	99.656
	BD	99.273	99.804	**99.909**	**99.151**	**99.718**	99.798
HF	MSE	98.813	99.598	99.725	98.066	99.320	99.991
	Con	98.873	**99.682**	**99.800**	98.218	**99.359**	99.512
	BD	**98.940**	99.582	99.676	**98.280**	99.323	**99.996**
HN	MSE	99.293	99.729	99.891	**99.642**	99.553	99.659
	Con	**99.520**	99.802	**99.915**	98.516	99.537	**99.684**
	BD	99.293	**99.808**	99.801	98.262	**99.587**	99.548
LF	MSE	99.743	99.909	99.909	99.150	99.733	**99.885**
	Con	99.759	99.913	99.933	**99.382**	**99.817**	99.872
	BD	**99.800**	**99.920**	**99.975**	99.280	99.775	99.877
LN	MSE	99.380	99.720	**99.907**	**98.830**	99.524	99.716
	Con	**99.513**	99.760	99.835	98.667	99.394	**99.758**
	BD	99.493	**99.816**	99.877	98.809	**99.608**	99.748
MF	MSE	98.320	**99.631**	99.756	97.680	**99.295**	99.568
	Con	98.507	99.460	**99.759**	97.578	99.261	**99.626**
	BD	**98.793**	99.402	99.723	**97.764**	99.253	99.586
MN	MSE	**99.473**	99.558	99.828	98.182	**99.504**	99.643
	Con	99.460	99.613	99.875	98.278	99.418	**99.673**
	BD	99.420	**99.724**	**99.905**	**98.456**	99.437	99.636
SF	MSE	99.597	99.847	99.875	**99.461**	99.746	99.839
	Con	99.540	99.860	**99.955**	99.296	99.719	99.843
	BD	**99.600**	**99.875**	99.933	99.301	**99.820**	**99.902**
SN	MSE	99.136	99.720	**99.920**	98.634	**99.589**	**99.696**
	Con	**99.200**	**99.762**	99.848	98.627	99.517	99.617
	BD	99.146	99.724	99.869	**98.716**	99.524	**99.696**

For IF, the optimal accuracy is 99.954% on 5-way, 5-shot recognition using similarity loss and contrastive loss. For IN, the optimal accuracy is 99.909% on 5-way, 5-shot recognition using similarity loss and BD loss. For HF, the optimal accuracy is 99.996% on 15-way, 5-shot recognition using similarity loss and BD loss. For HN, the optimal accuracy is 99.915% on 5-way, 5-shot recognition using similarity loss and contrastive loss. For LF, the optimal accuracy is 99.975% on 5-way, 5-shot recognition using similarity loss and BD loss. For LN, the optimal accuracy is 99.907% on 5-way, 5-shot recognition using similarity loss. For MF, the optimal accuracy

TABLE V
FEW-SHOT RECOGNITION ACCURACIES (%) ON HF DATABASE WITH DIFFERENT HYPERPARAMETERS

loss		5-way			15-way		
		1-shot	3-shot	5-shot	1-shot	3-shot	5-shot
Con	$w=0.1$	98.707	99.433	99.745	97.840	99.196	**99.634**
	$w=0.5$	**98.873**	**99.682**	**99.800**	**98.218**	99.359	99.512
	$w=1$	98.613	99.218	99.739	97.560	**99.424**	99.536
BD	$w=0.1$	98.693	99.424	**99.760**	97.840	99.108	99.464
	$w=0.5$	**98.940**	99.582	99.676	**98.280**	**99.323**	**99.996**
	$w=1$	98.760	**99.598**	99.705	97.113	99.226	99.412

Fig. 7. Some typical samples of XJTU-UP database. (a) and (b) Original images in HF and IN. (c) and (d) ROIs in HF and HN.

is 99.759% on 5-way, 5-shot recognition using similarity loss and contrastive loss. For MN, the optimal accuracy is 99.905% on 5-way, 5-shot recognition using similarity loss and BD loss. For SF, the optimal accuracy is 99.955% on 5-way, 5-shot recognition using similarity loss and contrastive loss. For SN, the optimal accuracy is 99.920% on 5-way, 3-shot recognition using similarity loss. The results of datasets collected in flash light are better than that of databases collected in natural illumination. The flash light can reduce the influence of external light so that the texture and wrinkles are clearer.

D. Ablation Study

1) Effect of Hyperparameter: Here, we conducted several experiments on HF database to show the effect of w, which is used to balance the weights of similarity loss and distance-based loss. We set different values for w and the results are in Table V and Fig. 8. From the results, in most cases, as w increases, the accuracy increases first and then decreases. Though the distance-based losses can improve the performance, their weights cannot be too heavy, which shows the effectiveness of the flexible decision network.

2) Roles of Different Losses: Different losses are adopted to optimize the model. In this part, we conducted several experiments to verify the roles that different sub-losses play. HF is also selected, and different losses and their combinations are evaluated. Table VI shows the results, where "Single Con" means only the contrastive loss is adopted, "Single BD" means only BD is used, "MSE + 0.5 × Con" means the similarity loss is combined with contrastive loss, and "MSE + 0.5 × BD" means the similarity loss is combined with BD loss. "MSE + 0.5 × Con + 0.5 × BD" means the similarity loss is combined with both contrastive loss and BD loss as a joint loss. The results are shown in Table VI. It can be observed that the accuracy of joint loss is higher than that of single loss while lower than "MSE + 0.5 × Con" or "MSE + 0.5 × BD." If there is only the BD adopted, the performance is worst, while combined losses can obtain better accuracy. In addition, the accuracy of "MSE" is higher than that of distance-based losses on some datasets. It means the distance-based losses may bring a negative impact, but such scenarios are rare.

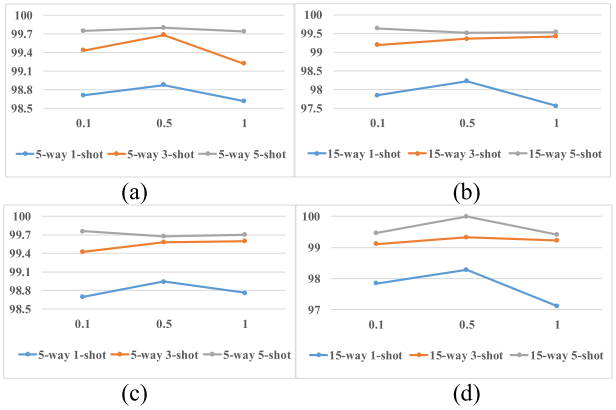

Fig. 8. Results of different hyperparameters. (a) and (b) "Con." (c) and (d) "BD." In each subgraph, the horizontal axis represents w, and the vertical axis represents the recognition accuracy.

TABLE VI
FEW-SHOT RECOGNITION ACCURACIES (%) ON HF DATABASE WITH DIFFERENT LOSSES

loss	5-way			15-way		
	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot
MSE	98.813	99.598	99.725	98.066	99.320	99.991
Single Con	98.860	99.424	99.663	97.098	98.855	99.265
Single BD	97.720	99.202	99.389	96.962	98.850	99.266
MSE+0.5×Con	98.873	**99.682**	**99.800**	98.218	**99.359**	99.512
MSE+0.5×BD	**98.940**	99.582	99.676	**98.280**	99.323	**99.996**
MSE+0.5×Con +0.5×BD	98.870	99.600	99.760	97.757	99.121	99.486

TABLE VII
FEW-SHOT RECOGNITION ACCURACIES (%) ON HF DATABASE USING DIFFERENT DECISION NETWORKS

loss	5-way (w/o convolutional blocks)			5-way (w/ convolutional blocks)		
	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot
MSE	94.270	94.000	96.244	98.813	99.598	99.725
MSE+0.5×Con	94.127	93.410	97.032	98.873	99.682	99.800
MSE+0.5×BD	94.807	93.987	97.865	98.940	99.582	99.676

3) Necessity of Convolutional Blocks in Decision Network: In this article, convolutional blocks are constructed in the decision network to increase the flexibility, instead of pure-stacked FC layers. Here, several experiments are conducted on HF database to show the necessity of convolutional blocks. The results are shown in Table VII. The decision network with or without convolutional blocks is adopted, and other experiment settings are consistent. Form the results, convolutional blocks can extract more potential features for much better performance, though only the 128-dimensional features are extracted.

V. Evaluation and Analyses

A. Comparisons Between Different Settings

1) Comparisons Between Different "Ways" and "Shots": In the experiments, 5-way/1-shot, 5-way/3-shot, 5-way/5-shot, 15-way/1-shot, 15-way/3-shot, and 15-way/5-shot recognition tasks are performed. From the results, the accuracies of 5-way are better than that of 15-way. In this article, meta-learning strategies are adopted. In each episode-based training iteration, it seems that the network tries to classify the images in the sample/query sets. In 15-way few-shot recognition, the knowledge acquired by sampling tasks when training may not be specific enough for handling classification problem among larger amounts of classes. The categories in 5-way setting are less, so the classification task is easier and the accuracy is higher. Furthermore, if there are more labeled images in sample set (more "shots"), most of the results will be also better, because more knowledge of a certain class is obtained by MSN. It can also be observed that the accuracy of 3-shot is higher than that of 5-shot in some datasets, such as Red in Table II. It may be because there is more variation between images, and more "shots" increase the difficulty of learning.

2) Comparisons Between Different Databases: In this article, three benchmarks are adopted. PolyU multispectral database and Tongji contactless database are constrained databases, which are collected in closed space with additional illuminations. So their qualities are better and easier to identify. XJTU-UP database consists of images collected by mobile phones in an unconstrained manner, so they contain more noise. From the results, the performances of constrained images outperform the unconstrained images, but the latter are also good. However, the unconstrained acquisition is more suitable for mobile terminal application scenarios.

B. Comparisons With Other Models

For comparison, we present the results of some baseline methods in the 5-way, 1-shot recognition, namely SN [17], model-agnostic meta-learning (MAML) [39], Prototypical Nets (P-Net) [37], DHN [57], Matching Net (M-Net) [49], GNN [29], DRCC [25], ALDC [58], local discriminant direction binary pattern (LDDBP) [59], DDBPD [26], PCANet [30], [60], PalmNet [11], TPN [44], LGM-Net [47], ABLM [46], lifted structure (LS) loss [61], and multisimilarity (MS) loss [62]. The results are shown in Tables VIII and IX, and the top 1 accuracy is highlighted with bold.

1) SN [17] uses two same networks to extract features and a decision network to get the similarity scores of matched images.
2) MAML [39] is based on meta-learning and aims to explicitly train a network on a number of learning tasks so that it can adapt to new learning tasks.
3) P-Net [37] adopts distance-based loss to learn a metric space and achieves classification by obtaining the distances to prototype representations of every category.
4) M-Net [49] is based on deep metric learning and augments neural networks with external memories to adapt new tasks.
5) DHN [57] converts palmprint images into binary codes, which can improve the efficiency of authentication and obtain the state of the arts on the traditional palmprint recognition scenario.
6) GNN [29] uses nodes to represent image features and edge to represent their positive or negative relation.
7) DRCC [25] adopts a more accurate dominant orientation representation of palmprint by weighting the orientation information of a neighbor area.
8) ALDC [58] extracts the apparent and latent direction features and pools them as the histogram feature descriptor.
9) LDDBP [59] adopts a novel exponential and Gaussian fusion model (EGM) to present the discriminative power of different directions of palmprint.
10) DDBPD [26] concatenates several binary feature DDBC codes as a global feature vector to perform recognition.
11) PCANet [30], [60] applies cascaded PCA, binary hashing, and block-wise histograms to extract features.
12) PalmNet [11] combines Gabor responses and CNN, and is trained by an unsupervised procedure.
13) TPN [44] learns a graph construction module to propagate labels from labeled instances to unlabeled test instances.
14) LGM-Net [47] learns transferable knowledge across different tasks and produces network parameters for similar unseen tasks through TargetNet and MetaNet.
15) ABLM [46] amortizes hierarchical variational inference across tasks and learns a prior distribution over neural network weights.
16) LS loss [61] samples an equal number of negative pairs as the positive pairs to take full advantage of training batches.
17) MS loss [62] adopts two iterative steps with sampling and weighting to improve the performance.

Note that all of modules are implemented using similar hyperparameters with a slight difference in each model to reach the best performance, respectively. LS loss and MS loss are adopted to train deep metric model to extract discriminative features, and Resnet 18 is used as the backbone [63]. The experiment settings are kept as consistent, such as the split of the training data and test data and evaluation method. From the tables, our model can achieve competitive results compared with several popular low-shot recognition methods, namely M-Net, P-Net, MAML, GNN, TPN, LGM-Net, and ABLM. Compared with the state-of-the-art palmprint recognition models using traditional training strategies without special design for low-shot recognition, our model performs better in all datasets. ALDC, LDDBP, and DDBPD are handcrafted palmprint recognition methods. Though they can obtain relatively high accuracy on constrained database, but they are not as good as deep learning-based methods, such as PalmNet. LS loss and MS loss can obtain satisfactory performance when there are enough training data. However, compared with MSN, due to the lack of labeled training data, their performances are also limited. SN and DHN are supervised algorithms, but there are not enough labeled samples here, so their performances have dropped significantly. In the SSS palmprint recognition

TABLE VIII
COMPARATIVE RESULTS (%) OF FEW-SHOT RECOGNITION ON DIFFERENT MODELS

Database	SN	MAML	P-Net	M-Net	GNN	DHN	LS loss	MS loss	MSE (ours)	Con (ours)	BD (ours)
Blue	74.000	76.659	99.935	99.850	99.848	84.836	88.000	96.473	99.956	**99.973**	**99.973**
Green	80.000	76.120	99.775	99.750	99.732	83.654	89.964	95.564	99.917	99.907	**99.933**
NIR	90.000	76.364	99.938	**100**	99.846	67.345	86.727	94.727	99.987	99.993	99.967
Red	81.000	76.989	99.931	99.900	99.858	74.018	91.818	94.691	99.967	**99.986**	99.980
Tongji	81.000	80.132	99.927	99.750	99.886	82.365	84.537	91.495	99.890	**99.933**	99.893
IF	64.750	79.072	97.693	98.950	97.576	65.249	91.220	92.332	99.323	**99.500**	99.380
IN	46.500	76.633	95.744	97.250	96.478	66.440	77.440	86.789	99.243	**99.353**	99.273
HF	41.750	87.491	97.876	98.819	96.740	59.147	75.893	87.444	98.813	98.873	**98.940**
HN	47.500	87.131	97.860	99.300	96.058	52.000	76.789	87.224	99.293	**99.520**	99.293
LF	60.000	85.279	98.416	99.350	97.632	66.552	89.111	91.893	99.743	99.759	**99.800**
LN	49.750	84.009	97.607	97.150	97.632	51.111	83.783	88.334	99.380	**99.513**	99.493
MF	85.000	80.825	98.338	98.450	97.786	60.195	85.443	89.783	98.320	98.507	**98.793**
MN	40.250	80.984	97.322	99.050	97.756	49.074	76.892	87.332	**99.473**	99.460	99.420
SF	53.350	86.492	97.958	98.750	97.876	63.484	85.112	88.673	99.597	99.540	**99.600**
SN	51.750	77.613	98.069	97.800	96.992	47.320	78.224	88.000	99.136	**99.200**	99.146
NUIG_Palm1	83.833	81.320	85.349	91.200	90.892	55.501	61.654	60.902	94.673	**94.880**	94.587
CASIA	94.000	89.092	98.930	97.150	96.544	83.534	92.885	93.080	**99.097**	99.077	99.050
IITD	93.430	91.450	98.120	97.700	95.345	62.935	85.766	93.080	99.063	99.073	**99.147**

TABLE IX
COMPARATIVE RESULTS (%) OF FEW-SHOT RECOGNITION ON DIFFERENT METHODS

Database	DRCC	ALDC	LDDBP	DDBPD	PCANet	PalmNet	LGM-Net	ABLM	TPN	MSE (ours)	Con (ours)	BD (ours)
Blue	95.950	98.550	99.600	99.380	98.040	98.950	99.688	99.691	99.970	99.956	**99.973**	**99.973**
Green	95.780	97.780	99.530	99.350	98.173	99.350	99.219	99.931	99.910	99.917	99.907	**99.933**
NIR	96.220	99.240	99.820	99.380	98.387	99.270	99.688	99.988	99.800	99.987	**99.993**	99.967
Red	96.190	99.450	99.930	99.670	98.253	99.750	99.844	99.661	99.980	99.967	**99.986**	99.980
Tongji	90.423	93.530	97.510	96.580	96.324	98.740	94.219	98.652	99.730	99.890	**99.933**	99.893
IF	79.724	92.000	96.670	96.440	91.041	95.890	94.219	98.652	99.179	99.323	**99.500**	99.380
IN	73.897	88.560	95.440	93.782	86.611	96.890	89.750	94.119	98.773	99.243	**99.353**	99.273
HF	76.613	93.440	97.110	95.780	87.934	97.220	94.219	92.530	98.230	98.813	98.873	**98.940**
HN	76.833	91.670	96.780	95.440	84.675	97.220	94.844	91.726	99.353	99.293	**99.520**	99.293
LF	75.892	96.670	98.560	97.110	90.210	96.220	93.438	95.678	99.720	99.743	99.759	**99.800**
LN	73.220	90.670	97.440	96.560	87.394	98.220	89.531	95.489	**99.780**	99.380	99.513	99.493
MF	78.614	96.333	97.780	96.670	90.088	96.111	94.375	97.581	98.111	98.320	98.507	**98.793**
MN	77.615	89.440	95.440	96.222	86.792	98.000	95.625	97.033	98.669	**99.473**	99.460	99.420
SF	80.226	92.780	95.560	93.000	91.440	80.110	94.219	98.848	99.518	99.597	99.540	**99.600**
SN	76.392	90.330	96.000	95.560	86.773	97.440	93.438	97.478	99.190	99.136	**99.200**	99.146
NUIG_Palm1	63.940	56.180	49.340	60.920	78.046	51.970	71.766	90.288	88.909	94.673	**94.880**	94.587
CASIA	64.520	89.400	88.410	90.110	95.433	94.350	95.469	93.322	92.184	**99.097**	99.077	99.050
IITD	55.330	83.700	72.930	87.720	92.340	85.220	88.953	92.434	97.270	99.063	99.073	**99.147**

scenario (only a few labeled samples can be used for training and registration), which is more common in practical applications, these previous methods do not work well, and this shows the effectiveness of our proposed methods. From the results, few-shot learning-based methods can generally achieve better performance, especially for constrained databases. However, our MSN combined similarity loss with distance loss can obtain higher accuracy on few-shot palmprint recognition.

VI. CONCLUSION

In this article, a novel few-shot model, MSN, is proposed for palmprint recognition only using a few labeled images. On the basis of classical SN, meta-episode training is introduced for better generalization performance. Two weight-shared networks are adopted to extract the features, which are measured by a decision network to obtain their similarities. In the similarity metric learning stage, the initial model is modified to compare two palmprint images flexibly by introducing CNN

blocks. Inspired by meta-learning, the training set is split into sample/query sets and the test set is split into support/testing sets. Furthermore, two distance-based losses are adopted to assist the optimization, which makes the positive matchings closer and negative matchings farther in the feature space. Finally, the model learns the ability to measure the similarity between two palmprint images on the learning tasks during training, which can adapt the new recognition tasks in the test set. Experiments on several benchmarks, including constrained and unconstrained palmprint databases, show that our algorithms can outperform other methods to be the state of the arts, and the accuracies can be up to 100%. And, our model is very suitable for hand-based practical personal authentication scenarios when the size of the training or registration set is small or when only a small part of palmprint images are labeled in the acquisition stage. In the future, we will extend our model to zero-shot recognition scenario by introducing the semantic features of palmprint.

REFERENCES

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[2] M. Kopaczka, R. Kolk, J. Schock, F. Burkhard, and D. Merhof, "A thermal infrared face database with facial landmarks and emotion labels," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 5, pp. 1389–1401, May 2019.

[3] K. Cao and A. K. Jain, "Automated latent fingerprint recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 788–800, Apr. 2019.

[4] A.-S. Ungureanu, S. Salahuddin, and P. Corcoran, "Toward unconstrained palmprint recognition on consumer devices: A literature review," *IEEE Access*, vol. 8, pp. 86130–86148, 2020.

[5] W. M. Matkowski, T. Chai, and A. W. K. Kong, "Palmprint recognition in uncontrolled and uncooperative environment," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1601–1615, 2020.

[6] L. Zhang, L. Li, A. Yang, Y. Shen, and M. Yang, "Towards contactless palmprint recognition: A novel device, a new benchmark, and a collaborative representation based identification approach," *Pattern Recognit.*, vol. 69, pp. 199–212, Sep. 2017.

[7] L. Fei, B. Zhang, S. Teng, Z. Guo, S. Li, and W. Jia, "Joint multiview feature learning for hand-print recognition," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9743–9755, Dec. 2020.

[8] N. B. Mahfoudh, Y. B. Jemaa, and F. Bouchhima, "A robust palmprint recognition system based on both principal lines and Gabor wavelets," *Int. J. Image, Graph. Signal Process.*, vol. 5, no. 7, pp. 1–8, 2013.

[9] D. Zhong, X. Du, and K. Zhong, "Decade progress of palmprint recognition: A brief survey," *Neurocomputing*, vol. 328, pp. 16–28, Feb. 2019.

[10] A. Kong, D. Zhang, and M. Kamel, "A survey of palmprint recognition," *Pattern Recognit.*, vol. 42, no. 7, pp. 1408–1418, Jul. 2009.

[11] A. Genovese, V. Piuri, K. N. Plataniotis, and F. Scotti, "PalmNet: Gabor-PCA convolutional networks for touchless palmprint recognition," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 12, pp. 3160–3174, Dec. 2019.

[12] H. Shao and D. Zhong, "Towards cross-dataset palmprint recognition via joint pixel and feature alignment," *IEEE Trans. Image Process.*, vol. 30, pp. 3764–3777, 2021, doi: 10.1109/TIP.2021.3065220.

[13] C. Yan, B. Shao, H. Zhao, R. Ning, Y. Zhang, and F. Xu, "3D room layout estimation from a single RGB image," *IEEE Trans. Multimedia*, vol. 22, no. 11, pp. 3014–3024, Nov. 2020.

[14] H. Shao and D. Zhong, "One-shot cross-dataset palmprint recognition via adversarial domain adaptation," *Neurocomputing*, vol. 432, pp. 288–299, Apr. 2021.

[15] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recognit. Lett.*, vol. 24, no. 16, pp. 3079–3087, Dec. 2003.

[16] C. Lemke, M. Budka, and B. Gabrys, "Metalearning: A survey of trends and technologies," *Artif. Intell. Rev.*, vol. 44, no. 1, pp. 117–130, Jun. 2015.

[17] D. Zhong, Y. Yang, and X. Du, "Palmprint recognition using siamese network," in *Proc. Chin. Conf. Biometric Recognit.*, Urumqi, China, 2018, pp. 48–55.

[18] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Jun. 2005, pp. 539–546.

[19] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Stockholm, Sweden, Aug. 2014, pp. 1–11.

[20] X. Du, D. Zhong, and P. Li, "Low-shot palmprint recognition based on meta-siamese network," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Shanghai, China, Jul. 2019, pp. 79–84.

[21] F. Yue, W. Zuo, D. Zhang, and K. Wang, "Orientation selection using modified FCM for competitive code-based palmprint recognition," *Pattern Recognit.*, vol. 42, no. 11, pp. 2841–2849, Nov. 2009.

[22] Z. Guo, D. Zhang, L. Zhang, and W. Zuo, "Palmprint verification using binary orientation co-occurrence vector," *Pattern Recognit. Lett.*, vol. 30, no. 13, pp. 1219–1227, Oct. 2009.

[23] L. Zhang, H. Li, and J. Niu, "Fragile bits in palmprint recognition," *IEEE Signal Process. Lett.*, vol. 19, no. 10, pp. 663–666, Oct. 2012.

[24] L. Fei, Y. Xu, W. Tang, and D. Zhang, "Double-orientation code and nonlinear matching scheme for palmprint recognition," *Pattern Recognit.*, vol. 49, pp. 89–101, Jan. 2016.

[25] Y. Xu, L. Fei, J. Wen, and D. Zhang, "Discriminative and robust competitive code for palmprint recognition," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 2, pp. 232–241, Feb. 2018.

[26] L. Fei, B. Zhang, Y. Xu, Z. Guo, J. Wen, and W. Jia, "Learning discriminant direction binary palmprint descriptor," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 3808–3820, Aug. 2019.

[27] Y.-T. Luo *et al.*, "Local line directional pattern for palmprint recognition," *Pattern Recognit.*, vol. 50, pp. 26–44, Feb. 2016.

[28] W. Jia *et al.*, "Palmprint recognition based on complete direction representation," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4483–4498, Sep. 2017.

[29] H. Shao and D. Zhong, "Few-shot palmprint recognition via graph neural networks," *Electron. Lett.*, vol. 55, no. 16, pp. 890–891, Aug. 2019.

[30] A. Meraoumia, F. Kadri, H. Bendjenna, S. Chitroub, and A. Bouridane, "Improving biometric identification performance using PCANet deep learning and multispectral palmprint," in *Biometric Security and Privacy*. Cham, Switzerland: Springer, 2017, pp. 51–69.

[31] S. Zhao and B. Zhang, "Deep discriminative representation for generic palmprint recognition," *Pattern Recognit.*, vol. 98, pp. 1–11, Feb. 2020.

[32] H. Shao, D. Zhong, and Y. Li, "PalmGAN for cross-domain palmprint recognition," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Shanghai, China, Jul. 2019, pp. 1390–1395.

[33] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 1199–1208.

[34] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "Meta-learning with memory-augmented neural networks," in *Proc. Int. Conf. Mach. Learn.*, New York, NY, USA, 2016, pp. 1842–1850.

[35] T. Munkhdalai and H. Yu, "Meta networks," in *Proc. 34th Int. Conf. Mach. Learn.*, Sydney, NSW, Australia, 2017, pp. 2554–2563.

[36] L. Bertinetto, J. F. Henriques, J. Valmadre, P. H. S. Torr, and A. Vedaldi, "Learning feed-forward one-shot learners," in *Proc. 30th Conf. Neural Inf. Process. Syst. (NIPS)*, Barcelona, Spain, 2016, pp. 523–531.

[37] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 4080–4090.

[38] M. Ren *et al.*, "Meta-learning for semi-supervised few-shot classification," in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, 2018, pp. 1–15.

[39] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Mach. Learn.*, Sydney, NSW, Australia, 2017, pp. 1126–1135.

[40] B. Kang, Z. Liu, X. Wang, F. Yu, J. Feng, and T. Darrell, "Few-shot object detection via feature reweighting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 8419–8428.

[41] H. Yao *et al.*, "Graph few-shot learning via knowledge transfer," in *Proc. 34th AAAI Conf. Artif. Intell.*, New York, NY, USA, 2020, pp. 6656–6663.

[42] V. G. Satorras and J. B. Estrach, "Few-shot learning with graph neural networks," in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, Vancouver, BC, Canada, 2018, pp. 1–13.

[43] C. Jiang, H. Xu, X. Liang, and L. Lin, "Hybrid knowledge routed modules for large-scale object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada 2018, pp. 1552–1563.
[44] Y. Liu *et al.*, "Learning to propagate labels: Transductive propagation network for few-shot learning," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, New Orleans, LA, USA, 2019, pp. 1–14.
[45] Y. Wang, C. Xu, C. Liu, L. Zhang, and Y. Fu, "Instance credibility inference for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 12833–12842.
[46] S. Ravi and A. Beatson, "Amortized Bayesian meta-learning," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, New Orleans, LA, USA, 2019, pp. 1–14.
[47] H. Li, W. Dong, X. Mei, C. Ma, F. Huang, and B. Hu, "LGM-Net: Learning to generate matching networks for few-shot learning," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, Long Beach, CA, USA, 2019, pp. 3825–3834.
[48] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proc. ICML Deep Learn. Workshop*, Lille, France, 2015, pp. 1–30.
[49] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, Barcelona, Spain, 2016, pp. 3630–3638.
[50] E. Ustinova and V. S. Lempitsky, "Learning deep embeddings with histogram loss," in *Proc. Adv. Neural Inf. Process. Syst.*, Barcelona, Spain, 2016, pp. 4170–4178.
[51] D. Zhang, Z. Guo, G. Lu, L. Zhang, and W. Zuo, "An online system of multispectral palmprint verification," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 2, pp. 480–490, Feb. 2010.
[52] H. Shao, D. Zhong, and X. Du, "Efficient deep palmprint recognition via distilled hashing coding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Long Beach, CA, USA, Jun. 2019, pp. 714–723.
[53] H. Shao, D. Zhong, and X. Du, "Deep distillation hashing for unconstrained palmprint recognition," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021, doi: 10.1109/TIM.2021.3053991.
[54] A.-S. Ungureanu, S. Thavalengal, T. E. Cognard, C. Costache, and P. Corcoran, "Unconstrained palmprint as a smartphone biometric," *IEEE Trans. Consum. Electron.*, vol. 63, no. 3, pp. 334–342, Aug. 2017.
[55] Z. Sun, T. Tan, Y. Wang, and S. Li, "Ordinal palmprint represention for personal identification [represention read representation]," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Orlando, FL, USA, Jun. 2005, pp. 279–284.
[56] A. Kumar and S. Shekhar, "Personal identification using multibiometrics rank-level fusion," *IEEE Trans. Syst., Man, Cybern., C (Appl. Rev.)*, vol. 41, no. 5, pp. 743–752, Sep. 2011.
[57] D. Zhong, H. Shao, and X. Du, "A hand-based multi-biometrics via deep hashing network and biometric graph matching," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 12, pp. 3140–3150, Dec. 2019.
[58] F. Ma, X. Zhu, C. Wang, H. Liu, and X.-Y. Jing, "Multi-orientation and multi-scale features discriminant learning for palmprint recognition," *Neurocomputing*, vol. 348, pp. 169–178, Jul. 2019.
[59] L. Fei, B. Zhang, Y. Xu, D. Huang, W. Jia, and J. Wen, "Local discriminant direction binary pattern for palmprint representation and recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 2, pp. 468–481, Feb. 2020.
[60] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.
[61] H. O. Song, Y. Xiang, S. Jegelka, and S. Savarese, "Deep metric learning via lifted structured feature embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2016, pp. 4004–4012.
[62] X. Wang, X. Han, W. Huang, D. Dong, and M. R. Scott, "Multi-similarity loss with general pair weighting for deep metric learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 5017–5025.
[63] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2016, pp. 770–778.

Huikai Shao (Graduate Student Member, IEEE) received the B.Sc. degree from Chongqing University, Chongqing, China, in 2017. He is currently pursuing the Ph.D. degree with the School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, China.

His main research interests are biometrics and computer vision.

Dexing Zhong (Member, IEEE) received the B.Sc. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 2005 and 2010, respectively.

He was a Visiting Scholar with the University of Illinois at Urbana–Champaign, Champaign, IL, USA. He is currently an Associate Professor with the School of Automation Science and Engineering, Xi'an Jiaotong University. His main research interests are biometrics and computer vision.

Xuefeng Du received the B.Eng. degree from the School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, China, in 2020. He is currently pursuing the Ph.D. degree (major in computer science) with the University of Wisconsin–Madison, Madison, WI, USA.

His main research interests are computer vision and deep learning.

Shaoyi Du (Member, IEEE) received the dual bachelor's degrees in computational mathematics and computer science, the M.S. degree in applied mathematics, and the Ph.D. degree in pattern recognition and intelligence systems from Xi'an Jiaotong University, Xi'an, China, in 2002, 2005, and 2009, respectively.

He is currently a Professor with Xi'an Jiaotong University. His current research interests include computer vision, machine learning, and pattern recognition.

Raymond N. J. Veldhuis (Senior Member, IEEE) received the degree from the University of Twente, Twente, The Netherlands, in 1981, and the Ph.D. degree from Nijmegen University, Nijmegen, The Netherlands, on a thesis entitled Adaptive Restoration of Lost Samples in Discrete-Time Signals and Digital Images, in 1988.

From 1982 to 1992, he was a Researcher with Philips Research Laboratories, Eindhoven, The Netherlands, in various areas of digital signal processing. From 1992 to 2001, he was involved in the field of speech processing. He is currently a Full Professor in biometric pattern recognition with the University of Twente, where he is leading a research team in this field. The main research topics are face recognition (2-D and 3-D), fingerprint recognition, vascular pattern recognition, multibiometric fusion, and biometric template protection. The research is both applied and fundamental.