FEATURE AUGMENTATION LEARNING FOR FEW-SHOT PALMPRINT IMAGE RECOGNITION WITH UNCONSTRAINED ACQUISITION

Kunlei Jing[†,‡,*], *Xinman Zhang*[†], *Zhiyuan Yang*[‡], *Bihan Wen*[‡,*]

[†]School of Automation Science and Engineering, Xi'an Jiaotong University, China
[‡]School of Electrical and Electronic Engineering , Nanyang Technological University, Singapore

ABSTRACT

Few-shot learning is challenging in unconstrained palmprint recognition, where the palmprint images are collected by unconstrained acquisitions, *i.e.*, different imaging sensors, backgrounds, palm postures, and illumination conditions. Furthermore, due to the lack of unconstrained palmprint databases and sufficient intra-class samples, it is difficult to apply the classic few-shot techniques, such as pre-training, fine-tuning, and sample augmentation, to generalize the model. In this work, we propose a novel feature augmentation network (FAN) for few-shot unconstrained palmprint recognition. Without any external databases, FAN aims to simultaneously remove the image variations caused by the unconstrained acquisitions and augment their feature representation from only a few support samples. To this end, the proposed *deep self-expression module* first decouples the support images into their principle and variation features. Assuming that the variations are *translational* across palmprint samples, the *variation-sharing module* achieves feature augmentations by swapping and combining all pairs of principle and variation features. The augmented palmprint features generated by FAN enable more general representations of categorical prototypes for few-shot unconstrained palmprint recognition. Experimental results on the standard palmprint databases show that FAN can effectively represent the prototypes of palmprint images from only a few available samples, thus outperforming the state-of-the-art methods in unconstrained palmprint recognition.

***Index Terms*—** Few-shot learning, unconstrained palmprint recognition, feature augmentation network, self-expression module, variation-sharing module

1. INTRODUCTION

Recent advances in deep learning have achieved great success in computer vision and natural language processing applications [1–3] by learning from large-scale annotated databases. However, it remains challenging to apply deep learning in many supervision-starving problems (*e.g.*, medical and biological applications), for which it is expensive or even intractable to acquire large-scale data with annotations [4]. Unconstrained palmprint recognition is one of such type, which has only a few publicly available databases with maximally 20 samples per class [5, 6]. However, it is more practical in real-world applications than constrained palmprint recognition. In this problem, the palmprint images are obtained with unconstrained acquisitions, *i.e.*, images are captured with different sensors, backgrounds, palm postures, or illumination conditions. Figure. 1(a) shows some palmprint images with illumination and posture variations. Learning discriminative feature embeddings and categorical prototypes for such tasks from only a few annotated samples is widely known as *few-shot learning* (FSL). In this work, we focus on the few-shot unconstrainted palmprint recognition (FS-UPR), with minimal support samples and no external spare database for exploiting transferable knowledge.

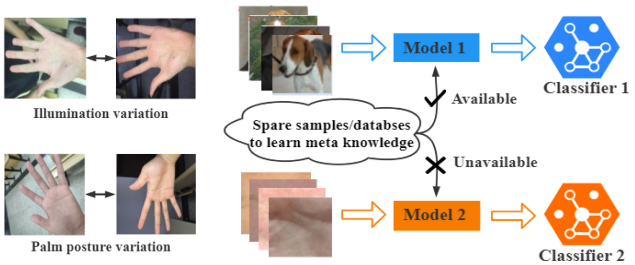


Fig. 1. Few-shot unconstrained palmprint recognition: (a) Palmprint images with variations in illumination and posture by unconstrained acquisitions; (b) Difference settings between the classic FSL (top) and FS-UPR (bottom).

Many FSL methods have been recently proposed for image classification, including metric learning, meta-learning, and sample augmentation [7, 8]. Metric-based methods [9–13] strive to train encoders to be discriminative via similarity metric loss functions, through which the deep embeddings of intra-class samples are clustered while inter-class embeddings become far apart [8]. Meta-learning methods [14–16] make efforts to transfer the knowledge distilled from plenty of tasks to novel conceptions with several gradient descent-based updates. However, most of the metric- and meta-learning methods could not generate well to novel classes with scarce knowledge extracted from a few support samples. Some of them train to transfer knowledge from external spare databases of similar distribution, which are not available for FS-UPR problem. Figure. 1(b) compares the different settings between the classic FSL and FS-UPR. Confronted with the challenges of metric- and meta-based methods, the augmentation-based methods generate more training samples via image deformation and noise-appending operation [17, 18] or GAN-like generators [19]. Whereas, the synthesized samples being highly similar in deformation lack of variants and miss principle properties of real samples [20]. Most GAN-based methods require training from related external large-scale databases or samples split from the base set to learn the underlying data distribution, which renders them inapplicable to FS-UPR.

In this work, we propose a novel feature augmentation network (FAN) to tackle the challenges in FS-UPR. Without any external palmprint images for either transfer learning or GAN training, FAN aims to achieve feature augmentation, assuming that the palm pos-

*Corresponding author.

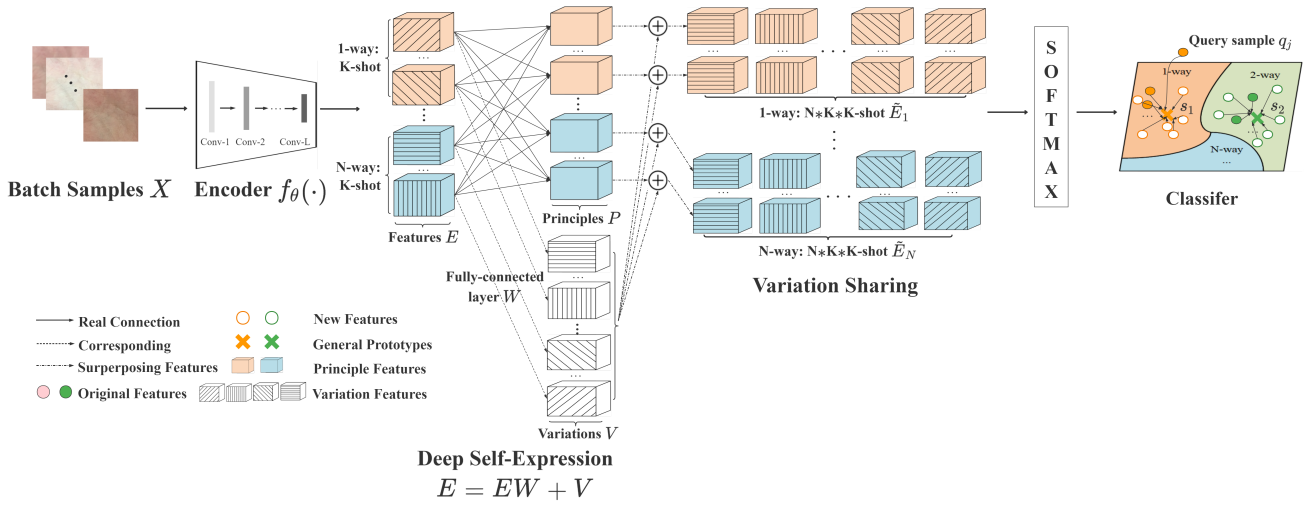


Fig. 2. The FAN architecture comprises a convolutional layer-based encoder, self-expression, variation-sharing, and classifier modules.

ture and illumination variations are **translational** across palmprint support samples without altering their identities [21, 22]. To this end, we propose a *deep self-expression module* to decouple the embedded support palmprint images into their principle and variation features using a sparsely-connected linear layer [23]. After that, our *variation-sharing module* generates feature augmentations by translating the extracted variation features to combine with different principle features, *e.g.*, an N-way K-shot support set can be augmented to a set of N-way N×K×K-shot large samples. Different from the classic self-expression methods [21, 22] based on variation extraction over proxy losses, FAN is trained in an end-to-end and task-driven manner. To the best of our knowledge, FAN is the first deep learning method for self-expression data augmentation in FS-UPR. We have conducted extensive experiments over publicly available FS-UPR datasets, namely XJTU-UP and MPD, showing the promise of FAN in both 5-way 5-shot and 5-way 1-shot classification tasks.

2. PRELIMINARY

2.1. Problem Setting

The classic few-shot learning (FSL) aims to generalize knowledge from training data to recognize new categories with a few support samples. Given a training dataset $\mathcal{T}_{train}$ with base classes $\mathcal{C}_{base}$, FSL aims to predict novel classes $\mathcal{C}_{novel}$ from a testing dataset $\mathcal{T}_{test}$, *i.e.*, $\mathcal{C}_{base} \cap \mathcal{C}_{novel} = \emptyset$. The samples in $\mathcal{T}_{train}$ and $\mathcal{T}_{test}$ are assembled into support sets, $\mathcal{S}_{base}$ and $\mathcal{S}_{novel}$, and query sets, $\mathcal{Q}_{base}$ and $\mathcal{Q}_{novel}$. The support sets contain N classes with K labeled samples, where N and K are small; The query sets contain samples from N classes for prediction. Such FSL paradigm is known as N-way K-shot learning. Based on the classic FSL setting, the FS-UPR problem requires that the intra-class samples in $\mathcal{T}_{train}$ and $\mathcal{T}_{test}$ are just enough for necessary supporting and querying, and no external databases are available for pre-training, fine-tuning, and sample augmentation. The proposed FAN learns to generate feature augmentation merely from the N-way K-shot support samples from $\mathcal{S}_{base}$.

2.2. Sparse Optimization for Self-Expression

Classic methods [23] exploited the self-expression property using subspace-sparse representation, *i.e.*, a data in a union of subspaces can be represented by a combination of other points from the dataset. We consider a data matrix $X = [x_1, \cdots, x_M] \in \mathbb{R}^{I \times M}$ being formed by M I-dimensional vectorized samples from $\mathcal{S}_{base}$, where $M = N \times K$ and each sample x_m is resulted from corrupting a noise-free sample $x_m^0 \in \mathbb{R}^I$ by a noise vector $z_m \in \mathbb{R}^I$ as

$$x_m = x_m^0 + z_m. \tag{1}$$

The subspace-sparse representation [23] suggested that under mild conditions, the corrupted x_m can be represented by its self-class samples by solving the following minimization problem:

$$\begin{aligned} \left\{\hat{W}, \hat{Z}\right\} &= \arg\min_{W,Z} \left\{\|W\|_1 + \alpha \|Z\|_F^2\right\} \\ s.t.\, X &= XW + Z, \mathrm{diag}(W) = 0, \end{aligned} \tag{2}$$

where the non-zero entries in the column vector $\hat{w}_m$ of the self-expression matrix $\hat{W}$ correspond the other samples from the same subspace as x_m. The matrix $\hat{Z} \in \mathbb{R}^{I \times M}$ is composed of M corrected noise vectors z_m, α is a balance parameter, and $X^0 = XW$ contains the recovered samples. Different from the sparse optimization (2) [23], we propose a deep self-expression module in FAN, which optimizes W in an end-to-end manner via task-driven learning. More details are presented in Section 3.1.

3. FEATURE AUGMENTATION NETWORK

The proposed Feature Augmentation Network (FAN) aims to tackle the FS-UPR challenges by learning a feature level self-expression. Figure 2 shows the architecture of FAN, which consists of an encoder, deep self-expression, variation-sharing, and classifier modules. The encoder is based on a multi-layer Convolutional Neural Network(CNN), which embeds a raw palmprint image $x_m \in \mathbb{X}$ into a D-dimensional deep feature space, *i.e.*, $e_m = f_\theta(x_m) : \mathbb{X} \to \mathbb{R}^D$, where θ denotes the learnable parameters of the encoder f. Next, the proposed deep self-expression module decouples the embedded e_m into its principle and variation features, namely p_m and v_m. In FS-UPR applications, the variation features are assumed to be translational, as similar variations such as posture or illumination changes in $\mathcal{S}_{base}$ are usually observed in query samples. This assumption is reasonable since palm posture and illumination vary randomly among real-world acquisitions. After that, under the same variation-translational assumption, the variation-sharing module generates new features by re-combining the extracted p_ms and

v_ms. The augmented features are finally used to train a classifier using prototype learning [11]. We explain more details on the deep self-expression, variation-sharing and classifier modules in the following.

3.1. Deep Self-Expression

Inspired by the subspace-sparse representation, we propose FAN to exploit the self-expression property [23] to learn the principle and variation features of palmprint images via the designed deep self-expression module. The classic sparse optimization Eq. (2) assumes the variations in X to be additive residual in a spatial domain, and the self-expression to be modeled as a linear combination. However, such a linear model is too restrictive in FS-UPR applications, *e.g.*, the real palm images involve posture and illumination changes which are not additive corruption. Besides, the sensor noise being multiplicative also cannot be modeled as the additive residual in Eq. (2).

To this end, the proposed deep self-expression generalizes the sparse self-expression approach [23] towards the embedded feature space. The self-expression property is exploited in a learned latent semantic space via the encoder f_θ, where the principle features corresponding to the clear samples and the variation features can be effectively disentangled by using a sparse self-expression approach. Subsequently, new features can be generated by the superposition operation similar to Eq. (1), in this latent space. In detail, the embedding features in the latent space $E = [e_1, \cdots, e_M] \in \mathbb{R}^{D\times M}$ span their union subspaces. During each episode, to capture the batch-wise variations v_m in E, the feature-domain self-expression can be obtained by solving the followig problem

$$\begin{aligned} \{W^*, V^*\} &= \arg\min_{W,V} \left\{ \|W\|_1 + \alpha \|V\|_F^2 \right\} \\ s.t.\, E &= EW + V, \mathrm{diag}(W) = 0. \end{aligned} \tag{3}$$

Note that Eq. (3) is a constrained optimization problem, which is hard to be optimized by gradient descent methods such as ADAM [24]. We further eliminate the constraint in Eq. (3), and convert to the following unconstrained problem with a self-expression penalty:

$$W^* = \arg\min_{W} \left\{ \|W\|_1 + \alpha \|E - EW\|_F^2 \right\}, \mathrm{diag}(W) = 0. \tag{4}$$

Here, we denote $P^* = EW^*$ as the *principle features*. Correspondingly, the *variation features* can be obtained as $V^* = E - EW^*$.

The feature-domain self-expression is implemented as a sparsely-connected linear transformation, without bias and activation functions in FAN. Figure. 2 shows the deep self-expression module behind the encoder representing the self-expression matrix W. The output of the deep self-expression module consists of both the principle and variation features. Note that the actual parameter W in FAN will not be solved by optimizing Eq. (4), but by further combining the task-specific loss (*e.g.*, the classification loss in this work). We present and conclude the total training loss of FAN in Section 3.3.

3.2. Variation Sharing

As discussed in Section 2.1, we assume that there are no external databases for FS-UPR. Moreover, the self-expression approach requires the batch-wise features to be arranged by class. However, we should not force the samples in query set to be sequential. Thus, all variation features are generated only from the support set. According to Eq. (4), matrix V contains M variation columns respectively corresponding to the principle columns in P. We can generate new features by randomly superposing column p_m in P and v_m in V. In the latent space, the new sample $\tilde{e}_{m'}$ is generated by

$$\tilde{e}_{m'} = p_m + v_m. \tag{5}$$

From Eq. (5), the new set $\tilde{E}$ accommodates up to $M \times M$ features. The new features entirely derived from real data are full of variants and $N \times K$ times larger than the original set E, which we believe characterize the N categorical distributions better.

3.3. Classifier

In addition to ensuring the reality of the new features, another subtlety of our augmentation approach is that $\tilde{E}$ also contains the original features, as can be verified by Eq. (5). We prefer to represent class n with its prototypes, where all the intra-class features participate in the characterization and the intra-class embeddings tend to be closer. This enforces the new features to keep the category-specific properties. In our work, the prototype of n-th class is computed by

$$s_n = \frac{1}{NK^2} \sum_{k=1}^{NK^2} \tilde{E}_{n,k}, \tag{6}$$

where $\tilde{E}_{n,k}$ denotes k-th sample from n-th class. From Eq. (6), s_n will be more general to forecast the variations of upcoming query samples than ProtoNet [11], thus leading to better generality of FAN. We adopt the Euclidean distance as the similarity metric between a query sample and each prototypes. Using the softmax operation, the probability distribution over N classes for j-th query sample q_j is

$$r_{\theta,j}(y = n|q_j) = \frac{\exp(d(f_\theta(q_j), s_n))}{\sum_n^N \exp(d(f_\theta(q_j), s_n))}, \tag{7}$$

which yields the following similarity loss for classification:

$$L_s = \frac{1}{|\mathcal{Q}_{novel}|} \sum_{j\in\mathcal{Q}_{novel}} -\log(r_{\theta,j}(y = g(q_j)|q_j)), \tag{8}$$

where $g(q_j)$ denotes the ground-truth label of q_j.

The proposed FAN is trained in an end-to-end manner. Therefore, both the loss and constraint from the self-expression Eq. (4) and the classification loss from Eq. (8) will be combined to form our total training loss. We formulate the FAN training loss as

$$L = \|W\|_1 + \alpha \|E - EW\|_F^2 + \beta L_s, \mathrm{diag}(W) = 0. \tag{9}$$

The training of FAN minimizes Eq. (9) over the trainable parameters W and θ. Thus, the deep embedding and variations are concurrently optimized via a gradient descent method. FAN is easy to optimize and can adapt to new tasks without additional learning processes.

4. EXPERIMENTS

4.1. Databases

We perform experiments on two latest public unconstrained palmprint databases: XJTU-UP [5] and MPD databases [6]. The XJTU-UP database includes ten subsets collected by five brands of mobile phones under natural light and flashlight. One hundred volunteers acquired both palms ten times with palm postures and backgrounds vary. Each subset contains 2,000 palmprint images belonging to 200 categories. Here, we choose the iPhone-natural-light subset with the images cropped with a size of 224×224 pixels. The first five samples per subject are used to train, and the left five are to test.

Table 1. Few-shot palmprint recognition accuracies on XJTU-UP and MPD databases

Method	XJTU-UP test Accuracy		MPD test Accuracy	
	5-way 5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot
MAML[15]	$96.44 \pm 0.92\%$	$87.08 \pm 1.79\%$	$93.21 \pm 0.96\%$	$74.65 \pm 0.21\%$
Matching Net[10]	$98.44 \pm 0.22\%$	$92.08 \pm 0.51\%$	$96.60 \pm 0.26\%$	$81.38 \pm 0.72\%$
RelationNet[12]	$98.71 \pm 0.30\%$	$95.83 \pm 0.74\%$	$96.29 \pm 0.32\%$	$85.69 \pm 0.69\%$
ProtoNet[11]	$98.32 \pm 0.34\%$	$94.00 \pm 0.82\%$	$96.35 \pm 0.42\%$	$82.86 \pm 0.80\%$
DSN[13]	$98.99 \pm 0.28\%$	$92.87 \pm 0.96\%$	$97.00 \pm 0.32\%$	$83.29 \pm 0.73\%$
PMN[18]	$98.11 \pm 0.35\%$	$96.14 \pm 0.46\%$	$94.62 \pm 0.41\%$	$82.53 \pm 0.73\%$
ADHN[25]	$99.00 \pm 0.92\%$	$90.17 \pm 2.77\%$	$96.71 \pm 0.96\%$	$77.79 \pm 2.27\%$
FAN(Ours)	$\mathbf{99.45 \pm 0.25\%}$	$\mathbf{97.80 \pm 0.59\%}$	$\mathbf{97.84 \pm 0.25\%}$	$\mathbf{88.25 \pm 0.61\%}$

MPD [6] includes two subsets collected by Huawei and Xiaomi mobile phones. Two hundred volunteers captured both palms ten times in two sessions under varying conditions. Each subset contains 8,000 images from 400 categories. Here, we adopt the Huawei-subset, where the images were cropped with 224 × 224 pixels. For each subject, the first five samples are carried out for training, and the left 15 samples are for testing.

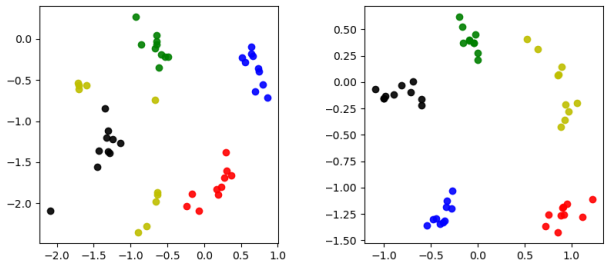
4.2. Training Details

Following the standard FSL paradigm [10], we evaluate FAN on 5-way 5-shot and 5-way 1-shot settings and train it from scratch over total 5000 episodes. The backbone is ConvNet-4 with an input dimension of 3×84×84 and output dimension of 64×19×19. For the 1-shot case, to ensure other homologous features for self-expression, we augment each base set by adding mild Gaussian noise. We split the databases by 5:2:3 to obtain training, validation, and test sets. We adopt the ADAM [24] optimizer with an initial learning rate of 0.005 and decrease it in half per 1000 episodes. We report the average accuracy and 95% confidence interval over 600 episodes when testing. The hyperparameters α and β are set as 0.08 and 50, according to the best validation performance. For the hyperparameters of competing methods, we fine-tune them to report the best accuracies. Since the self-expression layer is trained in a meta-learning way, the learned model can be directly used for new tasks in the test stage. It means no extra learning processes are required for a testing task.

4.3. Experimental Results

This paper focuses on FS-UPR without sufficient intra-class samples and external databases. The augmentation methods [17, 19] generating new samples from large external data are improper to be carried for comparisons. Here, we choose the other augmentation-based methods [18, 25]. Moreover, we choose some meta- and metric-based methods. Table 1 lists the accuracy comparisons, where the results of our method are bolded. We can see that FAN performs better on both 5-way 5-shot and 5-way 1-shot cases. Especially, considering the sample-starved 1-shot case on the XJTU-UP database, FAN exceeds the second-best RelationNet 1.97%, more prominent than its addition 0.41% over the second-best DSN on the 5-shot case. This suggests that the proposed feature augmentation approach renders FAN more general, especially when the base samples are limited.

An ablation study can be found by comparing FAN and ProtoNet [11], as it bases on the prototypical similarity. Considering the accuracies on the MPD database, FAN surpasses ProtoNet with 1.49% and 5.39%, affirming our proposal that new features derived from the variations of a few real samples promote FAN to embed more general prototypes to characterize category-specific distributions.



(a) (b)

Fig. 3. The t-SNE visualizations of (a) the embeddings from ProtoNet and (b) Our FAN introducing feature augmentation.

4.4. Feature Visualization

In addition to learning more general prototypes, the advantage of FAN can also be justified from another view. Generating new data around the original data can enforce the encoder to sharpen class representation, so that the categorical clusters are more discriminative, which leads to broader boundaries to accommodate homologous query samples with unforeseen variations. We still take our baseline, ProtoNet, for visualization comparison, as depicted in Figure3. By ProtoNet, some categorical clusters are somewhat mixed and without clear boundaries. After introducing our feature augmentation approach, the categorical clusters are more discriminative, hence leading to better generalization ability of FAN.

5. CONCLUSION

We proposed a deep feature augmentation network (FAN) to tackle the FS-UPR problem without sufficient intra-class samples and external databases. FAN can concurrently learn deep embedding and feature variations merely using a few samples from the support set, but can generate more effective prototypes and discriminative embedding than the existing methods. Learning generality from a few available samples renders our network applicable to FSL tasks lacking of sufficient samples and databases to acquire knowledge.

6. ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China (No. 61673316) and the China Scholarship Council Scholarship (No. 202006280327). This work was carried out at Rapid-Rich Object Search (ROSE) Lab, while Kunlei Jing was visiting Nanyang Technological University, Singapore.

References

[1] Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernockỳ, and Sanjeev Khudanpur, "Recurrent neural network based language model.," in *INTERSPEECH*, 2010, pp. 1045–1048.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[3] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, 2012.

[4] Hong-Gyu Jung and Seong-Whan Lee, "Few-shot learning with geometric constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4660–4672, 2020.

[5] Huikai Shao, Dexing Zhong, and Xuefeng Du, "Efficient deep palmprint recognition via distilled hashing coding," in *CVPR Workshops*, 2019.

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

[7] Zeyang Zhu and Xin Lin, "Kan: Knowledge-augmented networks for few-shot learning," in *ICASSP*. IEEE, 2021, pp. 1735–1739.

[8] Debasmit Das and CS George Lee, "A two-stage approach to few-shot learning for image recognition," *IEEE Trans. Image Process*, pp. 3336–3350, 2019.

[9] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al., "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, 2015.

[10] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al., "Matching networks for one shot learning," in *NIPS*, 2016, pp. 3630–3638.

[11] Jake Snell, Kevin Swersky, and Richard S Zemel, "Prototypical networks for few-shot learning," in *NIPS*, 2017, pp. 4077–4087.

[12] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales, "Learning to compare: Relation network for few-shot learning," in *ICCV*, 2018, pp. 1199–1208.

[13] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi, "Adaptive subspaces for few-shot learning," in *CVPR*, 2020, pp. 4136–4145.

[14] Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell, "Meta-learning with latent embedding optimization," *arXiv preprint arXiv:1807.05960*, 2018.

[15] Chelsea Finn, Pieter Abbeel, and Sergey Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *ICML*, 2017, pp. 1126–1135.

[16] Sachin Ravi and Hugo Larochelle, "Optimization as a model for few-shot learning," in *ICLR*, 2017.

[17] Zitian Chen, Yanwei Fu, Yu-Xiong Wang, Lin Ma, Wei Liu, and Martial Hebert, "Image deformation meta-networks for one-shot learning," in *CVPR*, 2019, pp. 8680–8689.

[18] Yu-Xiong Wang, Ross Girshick, Martial Hebert, and Bharath Hariharan, "Low-shot learning from imaginary data," in *CVPR*, 2018, pp. 7278–7286.

[19] Bharath Hariharan and Ross Girshick, "Low-shot visual recognition by shrinking and hallucinating features," in *ICCV*, 2017, pp. 3018–3027.

[20] Nan Lai, Meina Kan, Chunrui Han, Xingguang Song, and Shiguang Shan, "Learning to learn adaptive classifier-predictor for few-shot learning," *IEEE Trans. Neural Netw. Learn. Syst.*, 2020.

[21] Li Fei-Fei, Rob Fergus, and Pietro Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–611, 2006.

[22] Antonio Torralba, Kevin P Murphy, and William T Freeman, "Sharing visual features for multiclass and multiview object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 854–869, 2007.

[23] Ehsan Elhamifar and René Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, 2013.

[24] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015.

[25] Kai Li, Yulun Zhang, Kunpeng Li, and Yun Fu, "Adversarial feature hallucination networks for few-shot learning," in *CVPR*, 2020, pp. 13470–13479.