SUBSPACE REPRESENTATION LEARNING FOR FEW-SHOT IMAGE CLASSIFICATION

Ting-Yao Hu, Zhi-Qi Cheng, and Alexander G. Hauptmann
Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

In this paper, we propose a subspace representation learning (SRL) framework to tackle few-shot image classification tasks. It exploits a subspace in local CNN feature space to represent an image, and measures the similarity between two images according to a weighted subspace distance (WSD). When $K$ images are available for each class, we develop two types of template subspaces to aggregate $K$-shot information: the prototypical subspace (PS) and the discriminative subspace (DS). Based on the SRL framework, we extend metric learning based techniques from vector to subspace representation. While most previous works adopted global vector representation, using subspace representation can effectively preserve the spatial structure, and diversity within an image. We demonstrate the effectiveness of the SRL framework on three public benchmark datasets: MiniImageNet, TieredImageNet and Caltech-UCSD Birds-200-2011 (CUB), and the experimental results illustrate competitive/superior performance of our method compared to the previous state-of-the-art.

1. INTRODUCTION

Deep neural networks (DNN) has enabled huge advances in many computer vision tasks, such as image classification [1] and object detection [2]. However, the astounding success of DNN is conditioned on the availability of large scale datasets with thorough manual annotation, which is usually too expensive for real-world applications. In contrast, the human visual system is capable of learning a new visual concept with only a few annotated examples. This phenomenon inspired the development of few-shot learning [3, 4, 5, 6, 7, 8, 9, 10, 11], which aims at obtaining a reliable prediction model that can easily be generalized to unseen concepts.

One of the mainstream approaches to few-shot image classification is based on metric learning [3, 4, 5, 6, 7, 8, 9, 10, 11]. This type of method focuses on learning a good metric function from known concepts with sufficient labeled data, and transferring the learned metric to unseen concepts. Specifically, they exploit a Convolutional Neural Network (CNN) to extract a feature vector for each image, and measure the similarity between two images in hidden feature space based on distance functions, such as Euclidean and cosine distance. While receiving state-of-the-art performance, metric learning based methods are still not able to handle some unseen visual concepts with large intra-class variation and cluttered background [6]. One major reason is that an image-level feature vector ignores the spatial structure and diversity of an image. In the context of few-shot image classification, this problem becomes more severe since there is not enough supervision signal to guide the network focusing on the correct local regions.

In this paper, we propose a novel subspace representation learning (SRL) framework for few-shot image classification tasks. The SRL framework represents an image as a subspace extracted from its local CNN features. Then, the similarity between two images is measured by a subspace-to-subspace distance. While many strategies [12] have been proposed to compute the distance between two subspaces, we choose the weighted subspace distance (WSD) [13], which considers the importance of each dimension and reflects the original distribution of local CNN features. The SRL framework supports end-to-end training using a loss function related to distance based classifier [14]. When $K$-shot examples are available for a class, SRL utilizes a template subspace to summarize the information from $K$ images. To do so, we adapt two popular strategies developed for vector space to SRL framework. The first one is to obtain a class-specific subspace prototype by calculating the average of subspaces, while the second one is to learn a set of task-specific discriminative subspaces. Both strategies can be formulated as an optimization problem on a Stiefel manifold. In comparison with an image-level vector, a subspace is a compact representation capable of capturing the spatial structure and diversity of an image.

To evaluate the proposed SRL framework, we conduct experiments on three popular benchmarks for few shot image classification: MiniImageNet, TieredImageNet and Caltech-UCSD Birds-200-2011 (CUB). Experimental results show that our method achieves competitive/superior performance compared to state-of-the-art few-shot learning approaches. In summary, our main contributions are three fold:

- we tackle the few-shot image classification task by proposing the idea of subspace representation.
- we propose and compare two types of template subspace to aggregate $K$-shot information.
- our SRL framework achieves state-of-the-art performance on three public benchmarks.
2. RELATED WORK

Previous approaches to few-shot image classification can be roughly divided into three categories: (1) Distance metric based methods [3,9,10] construct a proper hidden feature space, whose distance metric is used to determine the image-class or image-image similarity. The distance metric can be a non-parametric function [3,9] or a parametric network module [10]. (2) Optimization based methods [4,15,16] aim at learning a good initialization for the model so that it can be quickly fine-tuned to a target task with limited amount of data. (3) Hallucination based methods [17,18,19,20] solve data scarcity by generating more training samples. The generation process is done in either hidden feature space or raw image space. While achieving promising performance, these methods adopt a global feature vector to represent an image.

As an extension of distance metric based methods, some recent works [6,7] rely on a local feature set representation to preserve the spatial structure of an image, and define their own metric to measure the similarity between two feature sets. Li et al. [7] calculates cosine distance between local feature pairs, and aggregates them by a k-NN classifier. Zhang et al. [6] adopt earth mover’s distance (EMD) to discover an optimal matching between local feature sets. In comparison, our method extracts a subspace representation for each local CNN feature set, and computes the weighted subspace distance between two sets.

The concept of subspace learning has been utilized to solve few-shot image classification [21,22]. However, these works consider the subspace in a global, image-level feature space. Also, they make use of a subspace projection operation to compute subspace-point distance and measure class-to-image similarity. In comparison, our method represents an image as a subspace in local feature space, and calculates subspace-to-subspace distance.

3. PROPOSED METHOD

The goal of few-shot image classification task is to build a prediction model that can be quickly adapted to unseen classes with limited amount of annotated examples. Most previous works validate their approaches to this task using N-way K-shot classification as testing scenario. Specifically, we are given a support set $S = \{(x_{i,j}, y_i)\}_{j=1}^{K} \}_{i=1}^{N_s}$ and query set $Q = \{(x_q, y_q)\}_{q=1}^{N_q}$, $y_q \in \{1, \ldots, N\}$, where $(x, y)$ is the pair of raw image and class label. The prediction model is trained/adapted based on $S$, and evaluated on the classification results of $Q$.

In this work, we propose subspace representation learning (SRL) framework to tackle this task. The overall architecture is illustrated in Fig. 1. Concretely, our method represents an image as a subspace, which is estimated from the reconstruction of local CNN features of this image. The dissimilarity between two images can be determined by a weighted subspace distance (WSD) between two subspaces. In the rest of this section, we first introduce the concept of subspace representation. Then, we elaborate the WSD adopted in SRL framework, and the end-to-end training process in the context of few-shot image classification. Finally, we describe two types of template subspace, which can effectively summarize information about a specific class from K-shot examples.

3.1. Subspace Representation

Our method exploits a subspace to represent each training/testing image. Given an image $x$, we extract the hidden feature map $(h \times w \times d)$ tensor using a backbone CNN with parameter $\Phi$, and collect the $d$-dimensional local feature vectors at all the spatial locations to form a matrix $H \in \mathbb{R}^{d \times (h \times w)}$. Then, our method finds the best-fit $s$-dimensional subspace $U \in \mathbb{R}^{d \times s}$, which minimizes the reconstruction error with respect to $H$:

$$
\min_U \|H - UU^TH\|_F \\
\text{s.t. } U^TU = I
$$

where $\| \cdot \|_F$ is the Frobenius norm of a matrix. This optimization problem can be solved by singular value decomposition (SVD) of $H$, and the optimal $U$ is obtained by the top-$s$ right-singular vectors with the largest singular values.

Similar to other works [7,6] leveraging local CNN features, our SRL framework is able to capture the spatial structure of an image. However, there are two additional reasons why we choose to construct a subspace representation. First, a subspace encourages the preservation of diversity because of the orthonormal constraint in eq. (1). Second, using a subspace results in a compact representation for image, since we can set a small $s$ without sacrificing the performance. More analysis of these two properties will be elaborated in the experiment section.

3.2. Weighted Subspace Distance

To conduct metric learning in the space of subspace representation, we need to calculate subspace-to-subspace distance. While several types of distance have been proposed, our SRL framework utilizes the weighted subspace distance (WSD) introduced in [13]: Given two subspaces $U_1$ and $U_2$ representing two images $x_1$ and $x_2$, WSD is expressed as:

$$
D(U_1, U_2) = \sqrt{1 - \sum_{i=1}^{s} \sum_{j=1}^{s} \lambda_{i,j}^1 \lambda_{2,j}^2 (u_{1,i}^T u_{2,j})^2 }
$$

$$
\lambda_{i,j}^1 = \frac{\lambda_{1,i}^1}{\sum_{i=1}^{s} \lambda_{1,i}^1}, \quad \lambda_{2,j}^2 = \frac{\lambda_{2,j}^2}{\sum_{i=1}^{s} \lambda_{2,j}^2}
$$

(2)
Fig. 1: The overall architecture of proposed subspace representation learning (SRL) framework. The backbone CNN extracts the local feature map \((h \times w \times d)\) from each query and support image. After reshaping the feature map into a matrix \(H \in \mathbb{R}^{d \times (h \cdot w)}\), whose columns are the CNN features at every spatial location, we extract the subspace representation by conducting SVD. The similarity between two images is determined by a weighted subspace distance (WSD). The end-to-end training is guided by the loss function of a distance based classifier.

where \(u_{1,i(2,j)}\) and \(\lambda_{1,i(2,j)}\) are the \(i(j)\)-th column of \(U_{1(2)}\), and the corresponding singular value obtained from SVD operation, respectively. Comparing to other types of subspace distance, WSD considers the relative importance of each basis component in a subspace, so it can better capture the distribution of original local feature set.

3.3. End-to-End Training

The training process of our SRL framework follows the episodic learning mechanism [9], which mimics the situation of a testing phase. In each training iteration, we sample a pair of support and query sets \((S, Q)\) from training dataset. Then, we extract the subspace representation for every image in \(S\) and \(Q\), and plug the WSD (eq. (2)) between the two subspaces into a distance based classifier [14]. Thus, the training objective of SRL can be formulated as:

\[
\mathcal{L}_{e2e}(\Phi) = \sum_{q} \log \left( \frac{\exp(-D(U_{q}, U_{q}))}{\sum_{i=1}^{N} \exp(-D(U_{i}, U_{q}))} \right) \tag{3}
\]

where \(U_{q}\) is the subspace representation of query, \(U_{i}\) is the subspace representation of the supported image of class \(y_{i}\) if \(K = 1\). In the case of \(K > 1\), \(U_{i}\) stands for the subspace that summarizes the \(K\)-shot information from class \(y_{i}\). The estimation of template subspace will be discussed in Sec. 3.4. By minimizing eq. (3), the parameter set \(\Phi\) of backbone CNN can be learned in an end-to-end manner.

3.4. Template Subspace for K-shot Learning

To cope with the \(K\)-shot learning scenario, the SRL framework computes a template subspace \(U_{\text{temp}}\) to aggregate the information from \(K\) subspaces for each class. In this work, we propose two types of template subspace to represent a class: A prototypical subspace and a discriminative subspace.

The prototypical subspace (PS) is the "average" of all \(K\) subspaces, following the spirit of ProtoNet [3] in vector space. Specifically, given \(K\) subspaces \(U_{1}, U_{2}, ..., U_{K}\) representing \(K\) images of the same class, the prototypical subspace is obtained by minimizing the summation of the distances between \(U_{\text{temp}}\) and \(U_{i}\):

\[
\mathcal{L}_{\text{ps}}(U_{\text{temp}}) = \sum_{j=1}^{K} D(U_{\text{temp}}, U_{j}) \tag{4}
\]

On the other hand, the discriminative subspace (DS) can be calculated by training a distance-based classifier with respect to a support set \(S\) of a \(N\)-way, \(K\)-shot classification task. Given the set of \(NK\) subspaces, \(\{\{U_{i,j}\}_{j=1}^{K}\}_{i=1}^{N}\), extracted from \(S\), the set of \(N\) template subspaces \(U_{\text{temp}} = \{U_{\text{temp}(i)}\}_{i=1}^{N}\) for \(N\) classes would be the minimizer of the following loss function:

\[
\mathcal{L}_{\text{ds}}(U_{\text{temp}}) = \sum_{i=1}^{N} \sum_{j=1}^{K} \log \left( \frac{\exp(-D(U_{i,j}, U_{\text{temp}(i)}))}{\sum_{l=1}^{N} \exp(-D(U_{i,j}, U_{\text{temp}(l)}))} \right) \tag{5}
\]

Please note that \(\{U_{\text{temp}(i)}\}_{i=1}^{N}\) are task-specific, since they are derived jointly from a \(N\)-way, \(K\)-shot task. In contrast, PS
is class-specific because the optimal $U_{temp}$ in eq. (4) is independent to other $(N - 1)K$ images in $S$. Fig. 2 compares PS and DS, along with their correspondence in vector space.

While minimizing $L_{ps}$ and $L_{ds}$, $U_{temp}$ is subject to the orthonormal constraint ($U^T U = I$), which prevents a closed-form solution of both problems. To solve this type of optimization problem with a SGD-like algorithm, we exploit the Cayley transform [23], projecting the gradient to the tangent space of a Stiefel manifold. We describe the update rule of estimating PS as an example. Let $Z = \partial L_{ps}/\partial U_t$, where $U_t$ is the current estimated $U_{temp}$ in eq. (4). The calculation of the next $U_{temp}$ estimation $U_{t+1}$ can be expressed as:

$$W = \hat{W} - \hat{W}^T, \hat{W} = ZU_t - \frac{1}{2}U_t(U_t^T ZU_t^T)$$

$$U_{t+1} = (I - \frac{\alpha}{2} W)^{-1}(I + \frac{\alpha}{2} W)U_t$$

(6)

where $\alpha$ is a hyper-parameter analogous to the learning rate in SGD-like algorithms.

4. EXPERIMENT

4.1. Implementation

For a fair comparison, we select the commonly used ResNet-12 as the backbone network of our SRL framework. The softmax layer and the spatial average pooling are removed from the backbone. All the images are resized to $84 \times 84$ pixels, and become $5 \times 5 \times 640$ tensors after analyzed by the backbone network. To optimize the parameters of this backbone ResNet-12, we conduct a two-step training process. In the first step, we pre-train the parameters by minimizing a cross entropy loss function of a standard classification task using training classes. In the second step, we perform the episodic training mechanism described in Sec. 3. We adopt SGD optimizer for 10k iterations with an initial learning rate 0.002, which decreases by a factor 0.1 for every 2k iteration. No data augmentation methods are applied during the episodic training.

The PS and DS are initialized by the subspaces extracted from the union of $K$ local CNN feature sets. Then, we update these template subspaces using SGD with Cayley transform for 50 iterations. The learning rate $\alpha$ for PS and DS are 0.1 and 0.01, respectively.

In our experiments, we follow the standard 5-way, 1-shot and 5-shot classification protocols, and sample 5,000 tasks with 15 query images for each target class. The category of each query image is predicted independently (inductive scenario). We report the average accuracy with the 95% confidence interval of all the sampled tasks.

4.2. Dataset

To evaluate the efficacy of SRL, we conduct experiments on three benchmark datasets: MiniImageNet [8], TieredImage-
Table 1: Results on MiniImageNet and TieredImageNet. All the methods use ResNet-12 as the backbone network. For SRL, we set subspace basis size, $s = 5$. *: Our re-implementation. †: Using local CNN feature.

| Methods      | MiniImageNet         | TieredImageNet       |
|--------------|----------------------|----------------------|
|              | 5-way, 1-shot        | 5-way, 5-shot         |
|              | 5-way, 1-shot        | 5-way, 5-shot         |
| Baseline++ [14] | 53.97 ± 0.79%    | 61.49 ± 0.51%    |
|              | 75.90 ± 0.61%    | 82.37 ± 0.67%    |
| ProtoNet* [3]    | 63.56 ± 0.34%    | 69.53 ± 0.36%    |
|              | 81.08 ± 0.18%    | 84.02 ± 0.23%    |
| MetaOpt-SVM [5] | 62.64 ± 0.82%    | 65.99 ± 0.72%    |
|              | 78.63 ± 0.46%    | 81.56 ± 0.53%    |
| MatchNet [9]    | 63.08 ± 0.80%    | 68.50 ± 0.92%    |
|              | 75.99 ± 0.60%    | 82.06 ± 0.71%    |
| DSN-MR [22]     | 64.60 ± 0.62%    | 73.79 ± 0.82%    |
|              | 79.51 ± 0.50%    | 82.85 ± 0.56%    |
| FEAT [11]      | 66.78 ± 0.20%    | 70.80 ± 0.23%    |
|              | 82.05 ± 0.15%    | 84.79 ± 0.16%    |
| Seq-distill [23]| 64.80 ± 0.60%    | 71.52 ± 0.69%    |
|              | 82.14 ± 0.43%    | 86.03 ± 0.49%    |
| DN4† [7]       | 63.72 ± 0.32%    | 70.23 ± 0.33%    |
|              | 81.54 ± 0.20%    | 84.01 ± 0.24%    |
| DeepEMD† [6]   | 65.91 ± 0.82%    | 71.16 ± 0.87%    |
|              | 82.43 ± 0.56%    | 86.03 ± 0.58%    |
| SRL, DS (ours)†| 67.00 ± 0.27%    | 71.88 ± 0.32%    |
|              | 82.68 ± 0.18%    | 86.24 ± 0.22%    |

Table 2: Results on CUB. All the methods use ResNet-12 as the backbone network. For SRL, we set subspace basis size, $s = 5$. *: Our re-implementation. †: Using local CNN feature.

| Methods      | 5-way, 1-shot       | 5-way, 5-shot       |
|--------------|---------------------|---------------------|
|              | 72.45 ±0.34%       | 85.94 ±0.23%       |
| ProtoNet*    | 71.87 ±0.85%       | 85.08 ±0.57%       |
| MatchNet     | 69.55 ±0.89%       | 85.17 ±0.50%       |
| Baseline++   | 72.30 ±0.32%       | 85.23 ±0.23%       |
| DN4†         | 75.65 ±0.83%       | 88.69 ±0.50%       |
| DeepEMD†     | 75.32 ±0.27%       | 88.81 ±0.21%       |

Fig. 3: Sensitivity analysis with respect to the size of subspace basis. The results show that the accuracy saturates around $s = 6$.

4.3. Analysis for SRL Design

To better understand the property of SRL and validate our design choices, we conduct two quantitative studies: (1) sensitivity analysis for basis size of subspace representation (2) validating the choice of WSD.

In the first study, we adjust subspace basis size $s$, which is also the number of columns of $U$ in eq. (1), and report the performance of 5-way, 1-shot classification task in MiniImageNet and TieredImageNet. From the results shown in Fig 3, we find that the performance gets better when $s$ becomes larger, but saturates around $s = 6$. This observation is expected since the later included subspace basis components are with smaller singular values, and less important according to the definition of WSD. The results also suggest that subspace representation with basis size $s = 6$ is enough to preserve...
both PS and DS can outperform two naive baselines, and DS of MiniImageNet and TieredImageNet. The methods are evaluated on the 5-way, 5-shot classification task in terms of WSD. All the samples are re-weights the importance of each basis component, reflecting the distribution of local CNN features.

4.4. Analysis for Template Subspace

In this experiment, we compare two types of template subspace, PS and DS, along with two other naive methods, baseline (union) and baseline (NN), in K-shot learning scenario. Baseline (union) extracts the subspace from the union of all the local CNN features from K-shot images. It is also adopted as the initialization step of PS and DS. Baseline (NN) computes the subspace-subspace distance from query image to the nearest neighbor support image in terms of WSD. All the methods are evaluated on the 5-way, 5-shot classification task of MiniImageNet and TieredImageNet.

From the results illustrated in Table 3, we can see that both PS and DS can outperform two naive baselines, and DS receives the best performance. A possible explanation is that DS is optimized based on task-specific information, while PS is only conditioned on the intra-class information. Thus, when two confusing unseen concepts appear in the same sampled task, DS has a better chance to distinguish them.

4.5. Comparison with State-of-the-art

We compare the performance of our SRL framework with two approaches using local feature set, DeepEMD and DN4, as well as other previous state-of-the-art methods. The experiment results are summarized in Table 1 and 2. From this set of results, we have the following observations. First, ProtoNet from our implementation receives competitive performance on three datasets, serving as a strong baseline. Second, for 1-shot, 5-way task, our proposed SRL performs the best on MiniImageNet and TieredImageNet, and is only slightly worse than DeepEMD on CUB dataset. Third, SRL outperforms all the other methods on the 5-way, 5-shot task of all three datasets.

4.6. Visualization

To understand the underline information captured by subspace representation, we visualize the subspace basis components extracted from images of MiniImageNet and CUB datasets. Specifically, we calculate the cosine similarity between each basis component and the local CNN feature of each 5 x 5 regions. Fig. 5 shows the visualization results of the first two basis components. The brightness of each spatial region is proportional to the cosine similarity between the components and the local CNN feature. From the results, we can see that the first component contains shared information among all the features, while the second focuses on some specific local regions. These local regions reflect the characteristics of the corresponding class, which would be useful for classification task. However, observing the third and fourth example images from MiniImageNet, we find that their second basis components are highly correlated to the background regions. It indicates that the proposed SRL sometimes suffers from some dataset bias, and fails to represent the object.

5. CONCLUSION

In this paper, we propose a subspace representation learning (SRL) framework for few shot image classification task. It represents an image as a subspace in local CNN feature space, and compares two images by calculating a weighted subspace distance (WSD). When K-shot information is available, we propose two types of class template representation for SRL: a prototypical subspace (PS) and a discriminative subspace (DS). The estimation of PS and DS can be formulated as an optimization problem in a Stiefel manifold. The experiment
results on three public benchmark datasets show that our SRL framework achieves state-of-the-art performance.

The proposed SRL framework extends the concept of deep metric learning from vector to subspace, and serves as a general tool for modeling the local structure of an image. It can easily be applied to other end-to-end learning network architectures. Our future work will focus on exploring the applicability of this approach to other tasks and domains.

6. REFERENCES

[1] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al., “Imagenet large scale visual recognition challenge,” International journal of computer vision, 2015.

[2] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” IEEE TPAMI, vol. 39, no. 6, pp. 1137–1149, 2017.

[3] Jake Snell, Kevin Swersky, and Richard Zemel, “Prototypical networks for few-shot learning,” in Proc. NeurIPS, 2017, pp. 4077–4087.

[4] Chelsea Finn, Pieter Abbeel, and Sergey Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” arXiv preprint arXiv:1703.03400, 2017.

[5] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto, “Meta-learning with differentiable convex optimization,” in Proc. CVPR, 2019, pp. 10657–10665.

[6] Chi Zhang, Yujun Cai, Guosheng Lin, and Chunhua Shen, “Deepemd: Few-shot image classification with differentiable earth mover’s distance and structured classifiers,” in Proc. CVPR, 2020, pp. 12203–12213.

[7] Wenbin Li, Lei Wang, Jinglin Xu, Jing Huo, Yang Gao, and Jiebo Luo, “Revisiting local descriptor based image-to-class measure for few-shot learning,” in Proc. CVPR, 2019, pp. 7260–7268.

[8] Sachin Ravi and Hugo Larochelle, “Optimization as a model for few-shot learning,” in Proc. ICLR, 2017.

[9] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra, “Matching networks for one shot learning,” in Proc. NeurIPS, 2016, pp. 3637–3645.

[10] 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 Proc. CVPR, 2018, pp. 1199–1208.

[11] Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha, “Few-shot learning via embedding adaptation with set-to-set functions,” in Proc. CVPR, 2020, pp. 8808–8817.

[12] Guido Zuccon, Leif A Azzopardi, and CJ Van Rijsbergen, “Semantic spaces: Measuring the distance between different subspaces,” in International Symposium on Quantum Interaction. Springer, 2009, pp. 225–236.

[13] Fei Li, Qionghai Dai, Wenli Xu, and Guihua Er, “Weighted subspace distance and its applications to object recognition and retrieval with image sets,” IEEE Signal Processing Letters, vol. 16, no. 3, pp. 227–230, 2009.

[14] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang, “A closer look at few-shot classification,” in Proc. ICLR, 2019.
[15] Muhammad Abdullah Jamal and Guo-Jun Qi, “Task agnostic meta-learning for few-shot learning,” in CVPR, 2019, pp. 11719–11727.

[16] Tsendsuren Munkhdalai, Xingdi Yuan, Soroush Mehri, and Adam Trischler, “Rapid adaptation with conditionally shifted neurons,” in Proc. ICML. PMLR, 2018, pp. 3664–3673.

[17] Bharath Hariharan and Ross Girshick, “Low-shot visual recognition by shrinking and hallucinating features,” in Proc. CVPR, 2017, pp. 3018–3027.

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

[19] Chia-Ching Lin, Yu-Chiang Frank Wang, Chin-Laung Lei, and Kuan-Ta Chen, “Semantics-guided data hallucination for few-shot visual classification,” in Proc. ICIP, IEEE, 2019, pp. 3302–3306.

[20] Jaekyeom Kim, Hyoungseok Kim, and Gunhee Kim, “Model-agnostic boundary-adversarial sampling for test-time generalization in few-shot learning,” in Proc. ECCV, 2020.

[21] Sung Whan Yoon, Jun Seo, and Jaekyun Moon, “TapNet: Neural network augmented with task-adaptive projection for few-shot learning,” in ICML, 2019, pp. 7115–7123.

[22] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrta Harandi, “Adaptive subspaces for few-shot learning,” in Proc. CVPR, 2020, pp. 4136–4145.

[23] Yasunori Nishimori and Shotaro Akaho, “Learning algorithms utilizing quasi-geodesic flows on the stiefel manifold,” Neurocomputing, vol. 67, pp. 106–135, 2005.

[24] Yonglong Tian, Yue Wang, Dilip Krishnan, Joshua B Tenenbaum, and Phillip Isola, “Rethinking few-shot image classification: a good embedding is all you need?,” arXiv preprint arXiv:2003.11539, 2020.

[25] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B. Tenenbaum, Hugo Larochelle, and Richard S. Zemel, “Meta-learning for semi-supervised few-shot classification,” in Proc. ICLR, 2018.

[26] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie, “The caltech-ucsd birds-200-2011 dataset,” 2011.

[27] Alan Edelman, Tomás A Arias, and Steven T Smith, “The geometry of algorithms with orthogonality constraints,” SIAM journal on Matrix Analysis and Applications, vol. 20, no. 2, pp. 303–353, 1998.