TRANSDUCTIVE ONE-SHOT LEARNING MEET SUBSPACE DECOMPOSITION

Kyle Stein[1], *Andrew A. Mahyari*[1,2], *Guillermo Francia, III*[3], *Eman El-Sheikh*[3]

[1] Department of Intelligent Systems and Robotics, University of West Florida, Pensacola, FL, USA
[2] Florida Institute For Human and Machine Cognition (IHMC), Pensacola, FL, USA
[3] Center for Cybersecurity, University of West Florida, Pensacola, FL, USA

ABSTRACT

One-shot learning focuses on adapting pretrained models to recognize newly introduced and unseen classes based on a single labeled image. While variations of few-shot and zero-shot learning exist, one-shot learning remains a challenging yet crucial problem due to its ability to generalize knowledge to unseen classes from just one human-annotated image. In this paper, we introduce a transductive one-shot learning approach that employs subspace decomposition to utilize the information from labeled images in the support set and unlabeled images in the query set. These images are decomposed into a linear combination of latent variables representing primitives captured by smaller subspaces. By representing images in the query set as linear combinations of these latent primitives, we can propagate the label from a single image in the support set to query images that share similar combinations of primitives. Through a comprehensive quantitative analysis across various neural network feature extractors and datasets, we demonstrate that our approach can effectively generalize to novel classes from just one labeled image.

***Index Terms*—** Transductive One-Shot Learning, Object Detection, Subspace Decomposition

I. INTRODUCTION

One-shot learning (OSL) enables models to generalize and adapt to new tasks with minimal data [1], [2], [3]. While traditional supervised models perform well with large labeled datasets, collecting and labeling such data is costly, especially in data-scarce fields. OSL allows models to recognize new objects from just one labeled example by leveraging prior knowledge from previously seen classes. This setup typically involves training on a single labeled support sample and evaluating on an unseen query set.

OSL techniques fall into two main categories: inductive and transductive. Inductive methods train a model solely on labeled support data, then apply it independently to predict on query samples [4], [5], [6], [7], [8], [9]. Transductive methods, by contrast, utilize the query set itself, finding feature similarities to labeled support samples to improve prediction accuracy, though they often require significant computational resources [10], [11], [12], [13], [14]. State-of-the-art (SOA) transductive OSL techniques iteratively project query embeddings onto labeled supports for label propagation, yet they rarely exploit latent variables across classes. This can limit generalization on novel classes with similar compositional features.

In this paper, we introduce a data-driven approach based on subspace decomposition that achieves high accuracy while maintaining simplicity. Our method learns subspace bases and extracts latent variables from embeddings to enhance generalization on novel classes. The contributions of our paper include a method that simultaneously learns subspace bases for support and query sets, facilitating the extraction of latent compositional variables and leveraging insights from subspace decomposition and compositional zero-shot learning [15], [16]. Inspired by prior work in subspace decomposition [17], [18], we also develop an unsupervised factorization technique that decomposes embeddings into subspaces representing distinct features, with support embeddings represented as linear combinations of these subspaces. **Unlike state-of-the-art methods that train models to directly classify images of new classes based on their feature vectors, our approach takes a novel turn by decomposing their feature vectors into class labels and subspace bases.**

II. RELATED WORK

Initial efforts in OSL aimed to reduce reliance on extensive annotated data, dividing approaches into metric-based and optimization-based methods. Metric-based methods train models to infer based on similarity measures in embedding spaces. Matching Networks [5] introduced cosine similarity for class embeddings, while Prototypical Networks [4] introduced class prototypes calculated as the mean embedding of support samples, assigning labels based on proximity in Euclidean space.

Unlike metric-based approaches, optimization-based approaches in OSL focus on adapting model parameters to new tasks through fine-tuning with minimal updates. These methods aim to develop a model that can quickly adapt to new tasks with only a few gradient updates. Model-Agnostic Meta-Learning (MAML) [6] popularized optimization ap-

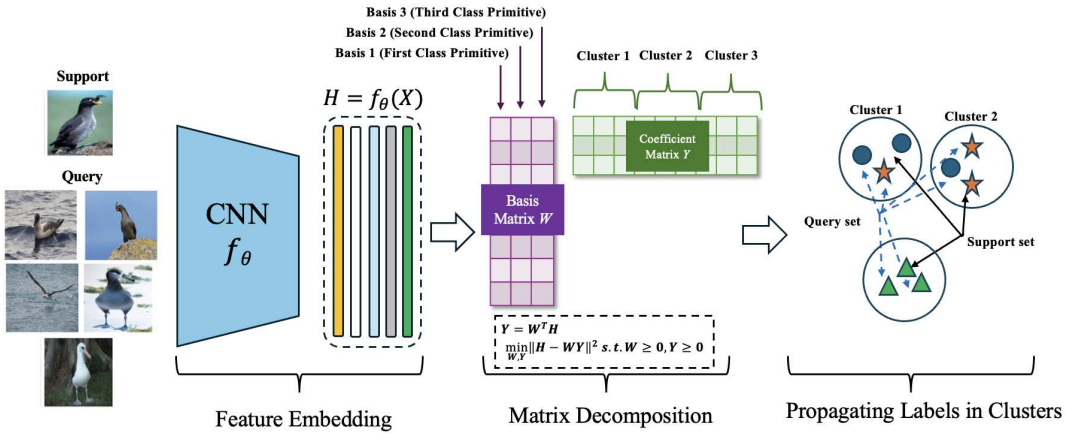

Fig. 1. *Overall architecture of our approach for transductive one-shot learning. A pre-trained CNN extracts the features from the images, forming an embedding matrix. This matrix is then decomposed into a Basis Matrix and a Coefficient Matrix. The Basis Matrix contains fundamental class primitives, while the Coefficient Matrix encodes how these primitives combine to form image embeddings. The optimization process iteratively refines these matrices to minimize the reconstruction error. Finally, the Coefficient Matrix is used to propagate labels from the support set to the query set by classifying images with similar primitives.*

proaches in OSL by training a model's parameters such that a small number of gradient updates will lead to quick adaptation and learning on a new task. Similar methods follow the same route by learning an optimal initialization that allows for efficient fine-tuning on new tasks with limited data. For example, [19] proposed a meta-learning approach that trains meta-learners on related tasks to generalize to new ones using temporal convolutions and soft attention, while [20] introduced an LSTM-based meta-learner designed to learn the specific optimization algorithm for training another neural network classifier. These approaches aim to minimize loss over a diverse set of tasks, training a base model that quickly generalizes and adapts to new scenarios.

OSL can also be broken down into inductive and transductive approaches. Inductive approaches learn functions from support sets, independently predicting on query sets. In contrast, transductive methods access query data at inference, refining predictions. Laplacian Shot [10] employs Laplacian regularization for label consistency, while methods like [11] and Transductive Propagation Network (TPN) [12] use joint feature spaces or graphs for label propagation.

III. PRELIMINARIES AND INSIGHTS

Let $S = \{(x_i, y_i)\}_{i=1}^{L}$ represent L labeled images of the support set, and let $Q = \{(x_i)\}_{i=L+1}^{L+U}$ represent U unlabeled images of the query set. In few-shot learning, we are given K labeled images, K-shot, for N classes, N-way, known as the support set. We are also provided with a backbone feature extractor $f_\theta(\cdot)$ that maps the input raw images to the embedding $\mathbf{h}_i = f_\theta(\mathbf{x}_i)$, where $\mathbf{h}_i \in R^{p\times 1}$. The goal of inductive few-shot learning is to learn a mapping or projection matrix $\mathbf{W} \in R^{p\times N}$ that maps the embedding to the correct labels $\mathbf{y}_i = \mathbf{W}^T\mathbf{h}_i$, where $\mathbf{W}$ is learned from a small support set, and evaluated on the query set. The objective of the inductive few-shot learning is represented:

$$\min_{\mathbf{W}} \sum_{i=1}^{L} \mathcal{L}(y_i, \mathbf{W}^T h_i) \tag{1}$$

In transductive few-shot learning, the relation between the features, *i.e.* embedding, of the support and query sets is leveraged to generalize the projection matrix $\mathbf{W}$ to other unseen samples, and the cost function is given by:

$$\min_{\mathbf{W}} \sum_{i=1}^{L} \mathcal{L}(y_i, \mathbf{W}^T h_i) + \sum_{i=1}^{L}\sum_{j=1}^{U} d(x_i, x_j), \tag{2}$$

where $d(\cdot,\cdot)$ is a similarity metric capturing the relationship among the samples of the support and query set. Our method builds on this by embedding both support and query samples into a shared feature space. We initialize the basis matrix $\mathbf{W}$ using the embedding of a labeled support sample and then construct a subspace that captures the relationship between embeddings of both support and query sets. Instead of relying on a predefined similarity metric, the relationships are infered by decomposing embeddings into latent components. The label from the support set is then propagated to the query set by comparing the coefficient vectors in the learned subspace.

IV. METHODOLOGY

In this section, we examine the OSL problem from a subspace analysis perspective. Our approach aims to derive equations for learning subspaces that effectively represent the primitives of images in both support and query sets. By leveraging the subspace structure, we facilitate the classifying of images based on similar primitive combinations, enabling efficient label propagation in a transductive learning

setting. While our method naturally groups similar features together in the subspace, we refer to this process as classifying, since the primary objective is to assign labels by aligning query images with the most relevant prototypes formed from a support sample.

To address the OSL problem, we derive equations for learning subspaces that best represent primitives of images across both support and query sets, and use these primitives for classification. Since the transductive approach leverages information from unlabeled samples in the query set, we combine the support and query sets into $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_L, x_{L+1}, \ldots, \mathbf{x}_{L+U}]^T$. Our unsupervised method assumes that the labels for this set $\mathbf{Y}$ are unknown, even though the label of one sample per class is known. We obtain the embeddings of the images in the set by passing them through a backbone feature extractor, resulting in $\mathbf{H} = f_\theta(\mathbf{X}) = [\mathbf{h}_1, \ldots, \mathbf{h}_{L+U}]^T$.

Similar to prior works [21], we assume that the labels of the support and query sets can be predicted by a linear projection of the embeddings onto the output manifold, which we capture with $\mathbf{Y} = \mathbf{W}^T\mathbf{H}$. Since $\mathbf{Y}$ and $\mathbf{W}$ are unknown in this equation, we rearrange the subspace projection equation as $\mathbf{H} = \mathbf{W}\mathbf{Y}$, where $\mathbf{W}$ is orthonormal. The label matrix $\mathbf{Y}$ is sparse; thus, this equation is interpreted as a sparse representation of the embeddings $\mathbf{H}$, where the columns of $\mathbf{W}$ are the basis of a subspace, and $\mathbf{Y}$ represents the coefficient vectors for this basis. The matrix $\mathbf{H}$ is thus projected onto the subspace defined by the columns of $\mathbf{W}$. Ideally, each embedding vector $\mathbf{h}_i$ is represented by one basis (i.e., one column) of the basis matrix (i.e., the projection matrix) $\mathbf{W}$. In the special case of OSL, each column of $\mathbf{W}$ could be equivalent to the embedding vector of the support sample, $\mathbf{h}_i = \mathbf{w}_i$. This simplifies to the average of the embeddings of the samples per class in the supporting set, resulting in a Protoypical network [4].

The embedding matrix derived from input images through the backbone feature extractor is the result of ReLU operations, thus ensuring that the embeddings are always non-negative, $\mathbf{H} \geq 0$. This is consistent with common practices in deep learning architectures, where ReLU activation functions are incorporated to introduce non-linearity while avoiding the vanishing gradient problem [22], [23]. Similarly, the coefficient matrix, which represents the labels or the distribution over the classes, is also non-negative, and each row ($\mathbf{y}_i$) represents the output distribution, therefore we have two additional conditions: $\mathbf{Y} \geq 0$ and $\sum_j softmax(\mathbf{y}_i)_j = 1$. The second constraint, *i.e.,* $\sum_j softmax(\mathbf{y}_i)_j = 1$, ensures the the estimated coefficients sum up to one after passing through a *softmax* operator, representing the categorical distribution over classes. Incorporating these constraints into the linear relationship between the embeddings and the output labels shapes our primary objective function:

$$\begin{aligned} &\min_{\mathbf{Y},\mathbf{W}} \|\mathbf{H} - \mathbf{W}\mathbf{Y}\|_F^2 \\ s.t.\ \ &\mathbf{W} \geq 0, \mathbf{Y} \geq 0, \textstyle\sum_j softmax(\mathbf{y}_i)_j = 1 \end{aligned} \tag{3}$$

Eq. 3 depicts a problem of simultaneous sparse representation and dictionary learning, where $\mathbf{W}$ functions as an unknown dictionary and $\mathbf{Y}$ as the sparse representation of the embeddings relative to this dictionary $\mathbf{W}$. Although various dictionary learning methods could be employed to determine $\mathbf{W}$ and $\mathbf{Y}$, we opt for matrix decomposition to address this optimization challenge.

In Eq. 3, the matrix decomposition approach breaks down the embedding matrix $\mathbf{H}$ into two components: the unknown projection matrix $\mathbf{W}$ and the coefficient matrix $\mathbf{Y}$. Each embedding vector $\mathbf{h_i}$ in $\mathbf{H}$ is approximated as a combination of the basis vectors in $\mathbf{W}$, weighted by the coefficients in the corresponding column in $\mathbf{Y}$. Each column of $\mathbf{W}$ serves as a latent feature vector, encapsulating a primitive within the embedding matrix [26]. Given that the number of columns in $\mathbf{W}$ is significantly fewer than the dimension of the embedding vectors, this decomposition method characterizes each class primarily by one dominant primitive. Consequently, these primitives contain the main distinguishing feature of each class, allowing images from both the support and query sets to be classified based on their dominant primitive [26].

The coefficient matrix $\mathbf{Y}$ represents the relationship between the basis vectors in $\mathbf{W}$ to the representation of the embedding vector $\mathbf{h_i}$. Each column of $\mathbf{Y}$ represents a coefficient vector which captures the combination of primitives that are shared among images. These coefficient vectors indicate how different embeddings are represented within the learned subspace $\mathbf{W}$. The similarity between these coefficient vectors allows us to classify the query sample features to those of a single labeled support sample, leveraging the shared subspace for label propagation. Eq. 3 is convex with respect to either $\mathbf{W}$ or $\mathbf{Y}$. To solve it, we employ gradient descent, iteratively estimating $\mathbf{W}$ and $\mathbf{Y}$. It is worth noting that this optimization involves only a few samples, allowing us to perform the calculations in one step without requiring stochastic gradient descent. Furthermore, overfitting is not a concern since we are not learning parameters but instead decomposing the matrix $\mathbf{H}$ into the product of two matrices, akin to matrix decomposition techniques. However, because the closed-form solution for Equation 3 cannot be derived, we rely on gradient descent to approximate the solution.

To initialize $\mathbf{Y}$, known labels from the support set are one-hot encoded. During optimization, the alternating update of $\mathbf{W}$ and $\mathbf{Y}$ iteratively adjusts the representation of both the projection matrix and coefficient representation, minimizing the reconstruction error $\|\mathbf{H} - \mathbf{W}\mathbf{Y}\|_F^2$. By minimizing the reconstruction error, the model identifies the most discriminative features in the data, encouraging similar samples in the latent space to classify based on shared primitives. As optimization converges, $\mathbf{Y}$ captures the label distribution for both support and query samples. The predicted label for each sample is determined by the position of the maximum value in its corresponding column of $\mathbf{Y}$.

Up to this point, we have approached the OSL task as an

Method	Setting	Backbone	mini-ImageNet (1-shot)	tiered-ImageNet (1-shot)
MAML [6]	Inductive	ResNet-18	49.61 ± 0.92	-
RelationNet [7]	Inductive	ResNet-18	52.48 ± 0.86	-
MatchingNet [5]	Inductive	ResNet-18	52.91 ± 0.88	-
ProtoNet [4]	Inductive	ResNet-18	54.16 ± 0.82	-
DeepEMD [8]	Inductive	ResNet-18	65.91 ± 0.82	-
TPN [12]	Transductive	ResNet-12	55.51 ± 0.86	59.91 ± 0.94
Transductive Tuning [3]	Transductive	ResNet-12	62.35 ± 0.66	-
DSN-MR [24]	Transductive	ResNet-12	64.60 ± 0.72	67.39 ± 0.82
CAN-T [25]	Transductive	ResNet-12	67.19 ± 0.55	73.21 ± 0.58
EASE [21]	Transductive	ResNet-12	57.00 ± 0.26	69.74 ± 0.31
Proposed	Transductive	ResNet-12	**67.55 ± 0.24**	**81.06 ± 0.49**
ProtoNet [4]	Inductive	WRN-28-10	62.60 ± 0.20	-
MatchingNet [5]	Inductive	WRN-28-10	64.03 ± 0.20	-
SimpleShot [9]	Inductive	WRN-28-10	65.87 ± 0.20	70.90 ± 0.22
Transductive Tuning [3]	Transductive	WRN-28-10	65.73 ± 0.68	73.34 ± 0.71
TIM [2]	Transductive	WRN-28-10	77.80	82.10
EPNet [14]	Transductive	WRN-28-10	70.74 ± 0.85	78.50 ± 0.91
LaplacianShot [10]	Transductive	WRN-28-10	74.86 ± 0.19	80.18 ± 0.21
Oblique Manifold [11]	Transductive	WRN-28-10	**80.64 ± 0.34**	**85.22 ± 0.34**
EASE [21]	Transductive	WRN-28-10	67.42 ± 0.27	75.87 ± 0.29
Proposed	Transductive	WRN-28-10	76.96 ± 0.60	84.55 ± 0.33

Table I. Test accuracy vs. the state-of-the-art (1-shot classification) on mini-ImageNet and tiered-ImageNet.

unsupervised task, aiming to classify query images based on similar primitives with a support image. After establishing the classes, we propagate the known label from a single support image to all query images within the same class.

V. EXPERIMENTS AND RESULTS

V-A. Datasets and Benchmarks

Multiple datasets were chosen to validate our method, notably: mini-ImageNet [5] and tiered-ImageNet [28]. These datasets are commonly inferred upon in the OSL community due to their complexity and diversity, which make them ideal for evaluating the generalization capabilities of these models. **MiniImageNet** consists of 60,000 colour images with 100 classes, each having 600 examples. **Tiered-ImageNet** represents a larger subset of classes from ILSVRC-12 than mini-Imagenet, with 608 classes. Not only do more classes exist, but this dataset also provides a more structured hierarchy, which ensures that all of the training classes are sufficiently distinct from the testing classes.

Our method is compared to other inductive and transductive SOA results present in the literature: MAML [6], RelationNet [7], MatchingNet [5], ProtoNet [4], DeepEMD [8], TPN [12], Transductive Tuning [3], DSN-MR [24], CAN-T [25], EASE [21], SimpleShot [9], TIM [2], Boosting [29], EPNet [14], LaplacianShot [10], and Oblique Manifold [11].

V-B. Experimental Setup

Episodic training is a widely utilized technique in few-shot learning, particularly in OSL scenarios. This method mimics the test conditions where the model is exposed to a limited number of labeled samples S and is expected to generalize to unlabeled examples from Q. Each training episode involves a N-way, K-shot task. This task is set up by selecting a subset of N classes from the training set. From each class in this subset, K samples are randomly chosen to create the labeled support set S. Additional random samples from these N classes are selected to form the query set Q. During each episode, the feature extractor $f_\theta(\cdot)$ processes both S and Q to generate embeddings. The embeddings from S are utilized to train $\mathbf{W}$, which is then applied to the embeddings of Q for label prediction. The accuracy of these predictions is assessed by comparing them with the true labels of the query set. The discrepancy, measured as loss, is used to refine the parameters θ of $f_\theta(\cdot)$.

Our initial experiment, in Table I, involves 10,000 randomly generated episodes, each following a 5-way, 1-shot format with 15 query samples per episode. To conduct further analysis, we extend the experimental setup in Table II to a more challenging 10-way, 1-shot scenario, while keeping the number of episodes the same and reducing the number of query samples to 10. We conduct each experiment 5 times, calculating the mean between the experiments and 95% confidence intervals for consistency. For our analysis, we employ pre-trained feature extractors: ResNet-12 [23] and WRN-28-10 [22] as $f_\theta(\cdot)$ to extract embeddings from input images.

V-C. Results

The experimental results on mini-ImageNet and tiered-ImageNet are shown in Table I. We show SOA performance for OSL across both datasets. We can observe that the proposed method outperforms the SOA methods for image classification on the tiered-ImageNet when using extracted

Model	Setting	mini-ImageNet Accuracy (%)	tiered-ImageNet Accuracy (%)
MAML [6]	Inductive	31.27 ± 1.15	34.44 ± 1.19
MAML+Transduction [6]	Transductive	31.83 ± 0.45	34.78 ± 1.18
ProtoNet [4]	Inductive	32.88 ± 0.47	37.35 ± 0.56
RelationNet [7]	Inductive	34.86 ± 0.48	38.62 ± 0.57
TPN [12]	Transductive	36.62 ± 0.50	40.93 ± 0.61
Simple CNAPS [27]	Transductive	37.10 ± 0.50	48.10 ± 0.70
Transductive CNAPS [1]	Transductive	42.80 ± 0.70	54.60 ± 0.80
Proposed Method	Transductive	**47.03 ± 0.18**	**63.26 ± 0.19**

Table II. 1-shot 10-way accuracy results with 10 query samples for various models on mini-ImageNet and tiered-ImageNet.

features from ResNet. We improve accuracy by nearly 8% over the nearest method using one labeled support sample. On mini-ImageNet, with features extracted using ResNet, we also obtain the highest classification accuracy. When employing the features extracted from WRN-28-10, we can see overall improved performance of our method when compared to ResNet. However, WRN-28-10's larger feature space ($p = 640$) expands the search space for our non-convex subspace decomposition, making the alternating updates more prone to settle in suboptimal stationary points rather than find the global optimum. This complicates the task of identifying appropriate primitives (columns of $\mathbf{W}$) within a more expansive and complex search space. The performance variability demonstrates the challenges of different architectures and suggests that all methods have specific strengths and limitations depending on the experimental setup.

To test the robustness of our model across different scenarios, we increased the number of classes during inference from 5 to 10, while reducing the number of query samples to 10, following the approach presented in [1]. Table II displays our method's results using ResNet-12 in comparison with other SOA OSL methods. We observe that our method outperforms previous methods in the 10-way classification scenario. Specifically, our model improves accuracy by over 4% on mini-ImageNet and achieves an impressive increase of more than 9% on tiered-ImageNet. To the best of our knowledge, these are SOA results for 10-way accuracy on both mini-ImageNet and tiered-ImageNet. This performance increase shows the robustness of our method, even when tasked with handling a more complex classification task.

The efficient performance of our model can be attributed to subspace decomposition, which provides a more refined representation of data in the latent space. This method enables us to effectively utilize the information from a single labeled support sample to extend the labels to the query samples within the same subspace. This process ensures that the embedding vectors projected onto the subspace establish a clearer connection between the support and query samples. This achievement is facilitated by the concurrent learning of the basis and coefficient matrices. Additionally, the cost function plays a crucial role in enhancing the stability and overall performance of the model. These results confirm that our subspace decomposition method, regardless of the feature extractor, enables efficient label propagation and classification in challenging one-shot scenarios.

VI. CONCLUSION

In this paper, we introduced a novel transductive OSL approach that identifies primitives of images by decomposing the embeddings of images from both support and query sets into representative subspaces. While our method demonstrates high accuracy, further extensive research is necessary to explore this data-driven approach, particularly to understand the impact of hidden factors and their connections to both seen and unseen classes. The empirical study revealed that the variability in performance demonstrates the inherent challenges posed by different architectures, suggesting that each method has specific strengths and limitations influenced by both the experimental setup and the nature of the datasets. Future efforts in this area will aim to expand this data-driven subspace decomposition methodology to zero-shot learning, linking attribute vectors to the primitives extracted through subspace factorization techniques.

VII. REFERENCES

[1] Peyman Bateni, Jacob Barber, Jan-Willem van de Meent, and Frank Wood, "Enhancing few-shot image classification with unlabelled examples," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2022, pp. 1597–1606.

[2] Malik Boudiaf, Imtiaz Ziko, Jérôme Rony, José Dolz, Pablo Piantanida, and Ismail Ben Ayed, "Information maximization for few-shot learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 2445–2457, 2020.

[3] Guneet S Dhillon, Pratik Chaudhari, Avinash Ravichandran, and Stefano Soatto, "A baseline for few-shot image classification," *arXiv preprint arXiv:1909.02729*, 2019.

[4] Jake Snell, Kevin Swersky, and Richard Zemel, "Prototypical networks for few-shot learning," *Advances in neural information processing systems*, vol. 30, 2017.

[5] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al., "Matching networks for one shot learning," *Advances in neural information processing systems*, vol. 29, 2016.

[6] Chelsea Finn, Pieter Abbeel, and Sergey Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 1126–1135.

[7] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1199–1208.

[8] Chi Zhang, Yujun Cai, Guosheng Lin, and Chunhua Shen, "Deepemd: Few-shot image classification with differentiable earth mover's distance and structured classifiers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12203–12213.

[9] Yan Wang, Wei-Lun Chao, Kilian Q Weinberger, and Laurens van der Maaten, "Simpleshot: Revisiting nearest neighbor classification for few-shot learning," *arXiv preprint arXiv:1911.04623*, 2019.

[10] Imtiaz Ziko, Jose Dolz, Eric Granger, and Ismail Ben Ayed, "Laplacian regularized few-shot learning," in *International Conference on Machine Learning*. 2020, pp. 11660–11670, PMLR.

[11] Guodong Qi, Huimin Yu, Zhaohui Lu, and Shuzhao Li, "Transductive few-shot classification on the oblique manifold," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 8412–8422.

[12] Yanbin Liu, Juho Lee, Minseop Park, Saehoon Kim, Eunho Yang, Sung Ju Hwang, and Yi Yang, "Learning to propagate labels: Transductive propagation network for few-shot learning," *arXiv preprint arXiv:1805.10002*, 2018.

[13] Michalis Lazarou, Tania Stathaki, and Yannis Avrithis, "Iterative label cleaning for transductive and semi-supervised few-shot learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 8751–8760.

[14] Pau Rodríguez, Issam Laradji, Alexandre Drouin, and Alexandre Lacoste, "Embedding propagation: Smoother manifold for few-shot classification," in *European Conference on Computer Vision*. 2020, pp. 121–138, Springer.

[15] Dat Huynh and Ehsan Elhamifar, "Compositional zero-shot learning via fine-grained dense feature composition," *Advances in Neural Information Processing Systems*, vol. 33, pp. 19849–19860, 2020.

[16] Muhammad Gul Zain Ali Khan, Muhammad Ferjad Naeem, Luc Van Gool, Alain Pagani, Didier Stricker, and Muhammad Zeshan Afzal, "Learning attention propagation for compositional zero-shot learning," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 3828–3837.

[17] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[18] Megasthenis Asteris, Dimitris Papailiopoulos, and Alexandros G Dimakis, "Orthogonal nmf through subspace exploration," *Advances in neural information processing systems*, vol. 28, 2015.

[19] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel, "Meta-learning with temporal convolutions," *arXiv preprint arXiv:1707.03141*, vol. 2, no. 7, pp. 23, 2017.

[20] Sachin Ravi and Hugo Larochelle, "Optimization as a model for few-shot learning," *International Conference on Learning Representations (ICLR)*, 2017.

[21] Hao Zhu and Piotr Koniusz, "Ease: Unsupervised discriminant subspace learning for transductive few-shot learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9078–9088.

[22] Sergey Zagoruyko, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[24] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi, "Adaptive subspaces for few-shot learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4136–4145.

[25] Ruibing Hou, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen, "Cross attention network for few-shot classification," *arXiv preprint arXiv:1910.07677*, 2019.

[26] Yu-Xiong Wang and Yu-Jin Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on knowledge and data engineering*, vol. 25, no. 6, pp. 1336–1353, 2012.

[27] Peyman Bateni, Raghav Goyal, Vaden Masrani, Frank Wood, and Leonid Sigal, "Improved few-shot visual classification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 14493–14502.

[28] 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," *arXiv preprint arXiv:1803.00676*, 2018.

[29] Spyros Gidaris, Andrei Bursuc, Nikos Komodakis, Patrick Pérez, and Matthieu Cord, "Boosting few-shot visual learning with self-supervision," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 8059–8068.