Multi-Proxy Learning from an Entropy Optimization Perspective: Appendix

Yunlong Yu[1] , **Dingyi Zhang**[1] , **Yingming Li**[1] and **Zhongfei Zhang**[2]
[1]College of Information Science and Electronic Engineering, Zhejiang University
[2] Computer Science Department, Binghamton University
{yuyunlong, dyzhang, yingming}@zju.edu.cn, zhongfei@cs.binghamton.edu

1 Overview

We provide additional analyses and experiments of our proposed method. In particular:

1. In Section 2, we provide the relations of the proposed approach with the existing approaches.
2. We provide the analysis of the generalization ability improvement from the embedding space metric in Section 3.
3. We evaluate the proposed approach on the Few-Shot Learning (FSL) tasks in Section 4.

2 Relation with existing approaches

In this section, we introduce the difference between the most related works and the proposed MPL in details.

There are several works to exploit the feature distribution with multiple proxies. For example, SoftTriple [Qian *et al.*, 2019] merges the similar centers to capture the local geometry for each class. ProxyGML [Zhu *et al.*, 2020] represents each class with a lot of proxies. Though both SoftTriple and ProxyGML consider the intra-class variation, they still may suffer from the mode collapse issue as no constraints are imposed to enlarge the variance of intra-class proxies. Although using more proxies for each class may alleviate this issue, ProxyGML would suffer from both computation and optimization issues when learning more proxies with limited training data. RepMet [Karlinsky *et al.*, 2019] assumed a multi-modal distribution of each category in the embedding space and correspondingly learned a set of representatives for each training class. They assigned equal weights to each representative in a class, thus encouraging an example to be closed to every representative in its ground truth class indistinguishably, without regard to the class collapse issue. In contrast, our MPL directly regularizes the intra-class variations and mitigates the mode collapse issue.

The other related work is [Dubey *et al.*, 2018] that considered the classic single-center softmax cross-entropy loss and showed that encouraging a large entropy of output class probability distribution would alleviate the overfitting issue for the fine-grained classification. Our work succeeds in the same spirit while extending this idea to a new multi-proxy framework for capturing the intra-class diversity and smoothing the inter-class difference simultaneously, thus bringing in improved performances. Besides, we additionally show the effectiveness of our framework in fine-grained clustering and retrieval tasks.

Data	Model	Intra	Inter	Intra/Inter
CUB	ProxyNCA++	0.616	0.996	0.618
	MPL	0.614	0.862	0.712
Cars196	ProxyNCA++	0.627	0.992	0.633
	MPL	0.668	0.832	0.804

Table 1: Generalization metrics of both ProxyNCA++ and our MPL on both CUB and Cars196 datasets.

3 Embedding space metrics

As reported in [Roth *et al.*, 2020], the embedding space density consistently exhibits significant correlation with generalization performance, we compare the embedding space density of both ProxyNCA++ [Teh *et al.*, 2020] and MPL on both CUB and Cars196 datasets. As shown in Table 1, the proposed MPL obtains larger embedding space densities (i.e., Intra/Inter) and smaller spectral decay (i.e, Inter), thus we conclude that MPL provides a more feature diverse embedding space and the metrics are aligned with the properties of improved generalization in DML.

4 Experiments on FSL

In this section, we evaluate the performances of the proposed approach on the Few-Shot Learning task.

Few-Shot Learning (FSL) [Snell *et al.*, 2017; Sung *et al.*, 2018] aims at recognizing instances from the novel classes with a few support instances, which requires training a transfer model from the base classes. Typically, FSL is addressed via learning a feature embedding space with the base classes and evaluated on the novel classes with a lazy classifier (e.g., kNN) on the feature embeddings. Thus, FSL can be seen as a downstream task of deep metric learning.

We evaluate the proposed MPL on the FSL tasks with three popular datasets, i.e., miniImageNet [Vinyals *et al.*, 2016], CUB [Wah *et al.*, 2011], and Stanford Dogs [Khosla *et al.*, 2011]. The miniImageNet is a subset of ImageNet, which consists of 60,000 instances from 100 classes. Each class

Method	miniImageNet		CUB		Dogs	
	1-Shot	5-Shot	1-Shot	5-Shot	1-Shot	5-Shot
Baseline++	51.87 ± 0.77	75.68 ± 0.63	68.46 ± 0.85	81.02 ± 0.46	58.30 ± 0.35	73.77 ± 0.68
MAML	59.10 ± 1.90	73.10 ± 0.90	71.11 ± 1.00	82.08 ± 0.72	66.56 ± 0.66	79.32 ± 0.35
MatchingNet	63.08 ± 0.80	75.99 ± 0.60	72.62 ± 0.90	84.14 ± 0.50	65.87 ± 0.81	80.70 ± 0.42
ProtoNet	60.37 ± 0.83	78.02 ± 0.57	71.57 ± 0.89	86.37 ± 0.49	65.02 ±0.92	83.69 ± 0.48
MTL	61.20 ± 1.80	75.50 ± 0.80	73.31 ± 0.92	82.29 ± 0.51	54.96 ± 1.03	68.76 ± 0.65
Δ-encoder	59.90 ± n/a†	69.70 ± n/a†	73.91 ± 0.87	85.60 ± 0.62	68.59 ± 0.53	78.60 ± 0.78
MetaOptNet	62.64 ± 0.82	78.63 ± 0.46	75.15 ± 0.46	87.09 ± 0.30	65.48 ± 0.49	79.39 ± 0.25
MPL (Ours)	**66.35 ± 0.41**	**80.22 ± 0.30**	**79.76 ± 0.39**	**91.20 ± 0.21**	**74.81 ± 0.44**	**86.52 ± 0.26**

Table 2: Few-shot classification accuracy (%) on the miniImageNet, CUB, and Stanford Dogs dataset. All experiments are from 5-way classification with the same backbone network (ResNet12) except the results of Δ-encoder on the miniImageNet. † denotes the results with ResNet18. The best performance is marked in bold.

consists of 600 instances. Following [Ravi and Larochelle, 2016], we split 64/16/20 classes for training, validation, and test, respectively. For the CUB dataset, we following [Chen *et al.*, 2019] and split the 200 classes into 100, 50, 50 for training, validation, and test, respectively. Stanford Dogs dataset [Khosla *et al.*, 2011] consists of 120 classes of dogs with a total number of 20,580 samples. We follow the data split protocol applied in the previous work [Li *et al.*, 2019] and take 70, 20 and 30 classes for training (auxiliary), validation and test, respectively. The instances from all the datasets are resized as 84×84. Both CUB and Stanford Dogs are fine-grained datasets and miniImageNet is a coarse-grained dataset.

For the feature extraction network, we leverage the most popular backbone, i.e., ResNet12, which consists of four residual blocks. Each residual block is composed of three convolutional layers with 3×3 kernels. A 2×2 maxpooling layer is applied at the end of each residual block. The whole model is trained from scratch in an end-to-end manner without pre-trained on the ImageNet. The model is trained with the SGD optimizer for 50 epochs. The initial learning rate is set to 0.05 and decreases by 0.5 at epoch 40 and 45. During the test stage, each few-shot classification task is provided with a support set and a query set. For an N-Way K-Shot classification task, the support set consists of N novel classes, each of which contains K samples. The support set is used to classify the query set. For the model inference, we apply the strategy proposed in [Snell *et al.*, 2017] to construct the prototypes of the support set with the learned feature embeddings and use the Cosine distance as the metric. We report both the mean accuracy and the 95% confidence interval of 600 randomly sampled few-shot tasks. We select seven approaches for comparison. They are Baseline++ [Chen *et al.*, 2019], MAML [Finn *et al.*, 2017], MatchingNet [Vinyals *et al.*, 2016], ProtoNet [Snell *et al.*, 2017], MTL [Sun *et al.*, 2020], Δ-encoder [Schwartz *et al.*, 2018], and MetaOptNet [Lee *et al.*, 2019].

Table 2 shows the few-shot classification results on the three datasets. All the results of the competitors are from the existing literature. From the results, we observe that the proposed MPL performs very competitively on both fine-grained and coarse-grained datasets. Specifically, MPL achieves a large margin improvement than the second-best approaches, which indicates the effectiveness of the proposed approaches on the downstream FSL tasks.

References

[Chen *et al.*, 2019] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. In *ICLR*, 2019.

[Dubey *et al.*, 2018] Abhimanyu Dubey, Otkrist Gupta, Ramesh Raskar, and Nikhil Naik. Maximum-entropy fine grained classification. In *NeurIPS*, 2018.

[Finn *et al.*, 2017] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, pages 1126–1135, 2017.

[Karlinsky *et al.*, 2019] Leonid Karlinsky, Joseph Shtok, Sivan Harary, Eli Schwartz, Amit Aides, Rogerio Feris, Raja Giryes, and Alex M Bronstein. Repmet: Representative-based metric learning for classification and few-shot object detection. In *CVPR*, pages 5197–5206, 2019.

[Khosla *et al.*, 2011] Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao, and Fei-Fei Li. Novel dataset for fine-grained image categorization: Stanford dogs. In *CVPR Workshop*, 2011.

[Lee *et al.*, 2019] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-learning with differentiable convex optimization. In *CVPR*, pages 10657–10665, 2019.

[Li *et al.*, 2019] 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 *CVPR*, pages 7260–7268, 2019.

[Qian *et al.*, 2019] Qi Qian, Lei Shang, Baigui Sun, Juhua Hu, Hao Li, and Rong Jin. Softtriple loss: Deep metric learning without triplet sampling. In *CVPR*, pages 6450–6458, 2019.

[Ravi and Larochelle, 2016] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *ICLR*, 2016.

[Roth *et al.*, 2020] Karsten Roth, Timo Milbich, Samarth Sinha, Prateek Gupta, Bjorn Ommer, and Joseph Paul Cohen. Revisiting training strategies and generalization performance in deep metric learning. In *ICML*, pages 8242–8252, 2020.

[Schwartz *et al.*, 2018] Eli Schwartz, Leonid Karlinsky, Joseph Shtok, Sivan Harary, Mattias Marder, Abhishek Kumar, Rogerio Feris, Raja Giryes, and Alex Bronstein. Delta-encoder: an effective sample synthesis method for few-shot object recognition. In *NeurIPS*, pages 2845–2855, 2018.

[Snell *et al.*, 2017] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *NeurIPS*, pages 4077–4087, 2017.

[Sun *et al.*, 2020] Qianru Sun, Yaoyao Liu, Zhaozheng Chen, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning through hard tasks. *TPAMI*, 2020.

[Sung *et al.*, 2018] 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 *CVPR*, pages 1199–1208, 2018.

[Teh *et al.*, 2020] Eu Wern Teh, Terrance DeVries, and Graham W Taylor. Proxynca++: Revisiting and revitalizing proxy neighborhood component analysis. In *ECCV*, 2020.

[Vinyals *et al.*, 2016] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one-shot learning. In *NeurIPS*, pages 3630–3638, 2016.

[Wah *et al.*, 2011] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

[Zhu *et al.*, 2020] Yuehua Zhu, Muli Yang, Cheng Deng, and Wei Liu. Fewer is more: A deep graph metric learning perspective using fewer proxies. In *NeurIPS*, 2020.