Transductive Few-shot Learning with Prototype-based Label Propagation by Iterative Graph Refinement (Supplementary Materials)

Hao Zhu[†,§], Piotr Koniusz[*,§,†]
[†]The Australian National University [§]Data61♥CSIRO
allenhaozhu@gmail.com, firstname.lastname@anu.edu.au

Below are additional details regarding our model and experiments.

6. Datasets

mini-ImageNet. This dataset [8, 10] is widely used in few-shot classification. It contains 100 randomly chosen classes from ImageNet [9]. There are 64 training (base) classes, 16 validation (novel) classes, and 20 test (novel) classes among the 100 classes. There are 600 images in each class. We adopt the split provided in [8].

tiered-ImageNet. ImageNet with a hierarchical structure was used to create the tiered-ImageNet. Categories of classes are divided into 34 categories, each of which contains 351, 97, and 160 classes for training, validation, and testing, respectively. Please note the training and test classes are semantically disjoint. We follow the common split in [1] and 84 by 84px resolution.

CIFAR-FS. This dataset has 100 classes, each with 600 examples in CIFAR-100 [4], on which this dataset is based on. We use the 64 training, 16 validation, and 20 test classes provided by [2].

CUB. There are 200 classes, each representing a bird species, in this fine-grained dataset. Following the setting in [2] , we divide our classes into three groups: 100 training, 50 validation, and 50 testing classes.

7. Feature Extraction and Pre-processing

ResNet-12A is the pre-trained backbone network used in [12]. For all of our transductive and semi-supervised experiments using this network, we adopt exactly the same pre-processing as [12], which includes normalizing feature vectors by their ℓ_2 norms.

WRN-28-10 is the pre-trained network used in [7] and [3]. To provide fair comparisons with PT+MAP [3], we adopt exactly the same pre-processing as [3]. In the all experiments, we apply the power transform and normalize feature vectors by their ℓ_2 norms.

DenseNet is the pre-trained network used in [6] and [11]. To provide fair comparisons with TASFFL [6], we adopt exactly the same pre-processing. In the all experiments, we decentre feature vectors by using the center in the training set and normalize decentred feature vectors by their ℓ_2 norms.

MobilieNet is the pre-trained network used in [11]. To provide fair comparison with LaplacianShot [14], we adopt exactly the same pre-processing. In the all experiments, we follow the setting as same as other backbones.

8. Hyper-parameters

In Eq. (7), parameter λ is in charge of regularizing the graph. For the balanced class setting, we simply set $\lambda = 1$. For the unbalanced class setting, we set $\lambda = 0.5$.

Hyper-parameter α is used to control updating prototypes $\boldsymbol{C}$ in Eq. (16). We empirically found that α basically does not impact the final result but the convergence speed. In this paper, we set it as 0.2 for all experiments.

9. Note on Fair Comparisons

During our experimental studies, we noticed the importance of fair comparisons by ensuring the common testbed. Below we talk about common issues.

- **Some comparisons use different networks**. In some papers, ResNet-18 is compared directly with other methods using ResNet-12. In Table 1 (our main paper), we show that ResNet-18 has some advantages over ResNet-12 in some cases.

- **Methods are mainly compared under the class-balanced prior**. For ease of understanding, transductive FSL evolves directly from the setting of inductive

[*]The corresponding author. Code: `https://github.com/allenhaozhu/protoLP`

FSL. Class-balanced queries are not an issue for inductive FSL because these methods treat queries one-by-one and ignore the class distribution of queries. However, the same setting in transductive setting introduces class-balanced prior. Many approaches over the last few years have focused on how to exploit this prior. We argue that this prior is unreasonable for transductive FSL, since it is rare for queries to follow a uniform class distribution. We can see the power of optimal transport under this prior i numerous works [3,5,13]. We can also see that this technique has the negative effect for queries that do not conform to the uniform class prior (*e.g.*, PT-MAP [3] drop 17% in mini-ImageNet under the unbalanced setting). Without the optimal transport, these methods also loose performance in the balanced setting.

We hope that bringing attention to these evaluation issues will help researchers avoid following the unrealistic settings and move toward fairer evaluation protocols and models. We encourage the community to compare different methods with under the same testbed.

10. The motivation of Parameterized Label Prediction

From a theoretical point of view, if we do not use parameterized label prediction, we need to optimize:

$$\min_{\widetilde{\boldsymbol{Y}}} \frac{1}{2}\|\widetilde{\boldsymbol{Y}}_L - \boldsymbol{Y}_L\|_F^2 + \frac{\lambda}{2}\operatorname{Tr}(\widetilde{\boldsymbol{Y}}^\top \left(\boldsymbol{I} - \lambda \boldsymbol{Z}\boldsymbol{\Lambda}^{-1}\boldsymbol{Z}^\top\right)\widetilde{\boldsymbol{Y}}). \quad (19)$$

The closed form solution is $\left(\boldsymbol{I} - \lambda \boldsymbol{Z}\boldsymbol{\Lambda}^{-1}\boldsymbol{Z}^\top\right)^{-1}\boldsymbol{Y}$. The computational complexity of the inverse matrix is n^3. and the $rank(\boldsymbol{Z}\boldsymbol{\Lambda}^{-1}\boldsymbol{Z}^\top) = 5$ (for 5-way problem). Thus we can use linear function $\boldsymbol{A}$ to do propagation on $\boldsymbol{Y}$ and reduce computational complexity from n^3 to c^3 where $n \gg c$. Projection to low-dimensional space limits the number of unnecessary parameters. Experiments in Table 10 confirm the parameterized LP works better.

Table 10. Test accuracy on standard LP and parameterized LP setting (ResNet-12 backbone).

	mini-ImageNet		tiered-ImageNet		CUB	
Methods (ResNet-18)	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Ours (non-paramerized)	69.41	79.52	83.60	88.56	88.50	91.70
Ours (parameterized)	70.04	79.80	84.04	88.72	88.85	91.93

11. Inference Time

Table 11 provides the average inference time (1× AMD 3600 CPU) for the 1- and 5-shot tasks on mini-ImageNet (ResNet-12 and WRN-28-10).

Table 11. Average inference time (in seconds) for the 1-shot and 5-shot tasks in mini-ImageNet dataset with different backbones.

Backbone	ResNet-12		WRN-28-10	
Shot	1	5	1	5
iLPC [5]	4.5e-2	5.6e-2	5.5e-2	7.0e-2
ICI [12]	3.4e-2	4.2e-2	4.1e-2	5.2e-2
protoLP	4.7e-3	5.8e-3	6.2e-3	6.8e-3

References

[1] Chaofan Chen, Xiaoshan Yang, Changsheng Xu, Xuhui Huang, and Zhe Ma. Eckpn: Explicit class knowledge propagation network for transductive few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6596–6605, 2021. 1

[2] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. In *International Conference on Learning Representations*, 2019. 1

[3] Yuqing Hu, Vincent Gripon, and Stéphane Pateux. Leveraging the feature distribution in transfer-based few-shot learning. In *Artificial Neural Networks and Machine Learning–ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part II 30*, pages 487–499. Springer, 2021. 1, 2

[4] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 1

[5] 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*, pages 8751–8760, 2021. 2

[6] Moshe Lichtenstein, Prasanna Sattigeri, Rogerio Feris, Raja Giryes, and Leonid Karlinsky. Tafssl: Task-adaptive feature sub-space learning for few-shot classification. In *European Conference on Computer Vision*, pages 522–539. Springer, 2020. 1

[7] Puneet Mangla, Nupur Kumari, Abhishek Sinha, Mayank Singh, Balaji Krishnamurthy, and Vineeth N Balasubramanian. Charting the right manifold: Manifold mixup for few-shot learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2218–2227, 2020. 1

[8] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International Conference on Learning Representations*, 2017. 1

[9] 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*, 115(3):211–252, 2015. 1

[10] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016. 1

[11] 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. 1

[12] Yikai Wang, Chengming Xu, Chen Liu, Li Zhang, and Yanwei Fu. Instance credibility inference for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12836–12845, 2020. 1, 2

[13] 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*, pages 9078–9088, 2022. 2

[14] Imtiaz Ziko, Jose Dolz, Eric Granger, and Ismail Ben Ayed. Laplacian regularized few-shot learning. In *International Conference on Machine Learning*, pages 11660–11670. PMLR, 2020. 1