Supplementary Material for Prototype Rectification for Few-Shot Learning

Jinlu Liu, Liang Song, and Yongqiang Qin*

AInnovation Technology Co., Ltd.
liujinlu, songliang, qinyongqiang@ainnovation.com

1 Implementation Details

WRN-28-10 [12], is used as the main backbone in the experiments. ConvNet-64 [2], ConvNet-128 [3], ConvNet-256 [4] and ResNet-12 [6] are used in ablation study. We remove the last ReLU layer of WRN-28-10 in experiments. The results reported in our experiments are collected by sampling 600 episodes with 95% confidence intervals. We choose SGD as the optimizer with a momentum of 0.9 and a weight decay parameter of 0.0005. The maximum training epoch is set to 60. The initial learning rate is 0.1 and it is reduced after 10, 20, 40 epochs. At the training stage, we use horizontal flip and random crop on the two ImageNet derivatives as in [3, 7, 6].

2 Results on Omniglot and CUB

We conduct extra experiments on another two benchmarks: Omniglot [5] and CUB [11].

2.1 Omniglot

Omniglot has 1623 classes of handwritten characters with 20 samples per class. All images are resized to 28 x 28. The data augmentation techniques proposed by [8, 9] are used in higher-way test, which rotates each image by 90, 180, 270 degrees to form new classes. Therefore, the dataset has total 6492 classes and we use 4112 classes for training, 688 classes for validation and 1692 classes for test as in [9].

2.2 CUB

We use the Caltech-UCSD Birds (CUB) 200-2011 dataset [11] of 200 fine-grained bird species. The dataset is split into 100 training classes, 50 validation classes and 50 test classes as provided in [1].

* Corresponding author.

Table 1. Results on Omniglot.

Omniglot	**1-shot**		**5-shot**	
	CSPN	BD-CSPN	CSPN	BD-CSPN
ConvNet-64	97.40	99.62	99.60	99.76
ConvNet-128	97.33	99.69	99.63	99.75
ConvNet-256	97.85	99.77	99.63	99.76
ResNet-12	98.70	99.80	99.72	99.77
WRN-28-10	99.02	99.08	99.82	99.85

Table 2. Results on CUB.

CUB	**1-shot**		**5-shot**	
	CSPN	BD-CSPN	CSPN	BD-CSPN
ConvNet-64	64.72	75.10	84.21	87.25
ConvNet-128	65.86	76.11	85.97	87.52
ConvNet-256	65.99	75.77	84.74	87.76
ResNet-12	76.24	84.90	88.68	90.22
WRN-28-10	77.80	87.45	90.14	91.74

3 Additional Ablation on miniImageNet and tieredImageNet

We provide supplementary ablation study on miniImageNet and tieredImageNet to show our performance on different backbones.

Table 3. Backbone ablation on miniImageNet.

miniImageNet	**1-shot**	**5-shot**
ConvNet-64	60.48	75.02
ConvNet-256	60.97	75.19

4 Higher-way Results

Results on higher-way tasks are given in Table 5-7 to show the effectiveness of our method in harder tasks.

5 Robust Test

We conduct an experiment as follows to test the robustness of the proposed BD-CSPN. In each 5-way K-shot 15-query episode, we randomly add extra 15×N'

Table 4. Backbone ablation on tieredImageNet.

tieredImageNet	**1-shot**	**5-shot**
ConvNet-64	65.08	78.08
ConvNet-128	66.33	79.57
ConvNet-256	67.09	80.66
ResNet-12	76.17	85.70

Table 5. Higher-way test on miniImageNet.

miniImageNet	**1-shot**	**5-shot**
10-way	51.58	69.35
20-way	36.00	55.23

Table 6. Higher-way test on tieredImageNet.

tieredImageNet	**1-shot**	**5-shot**
10-way	63.39	77.54
20-way	48.48	65.68
50-way	31.67	49.50

Table 7. Higher-way test on Omniglot.

Omniglot		**1shot**		**5shot**	
		CSPN	BD-CSPN	CSPN	BD-CSPN
10-way	ConvNet-128	92.83	98.46	98.67	99.02
	ConvNet-256	93.82	98.65	98.90	99.14
	ResNet-12	96.38	98.97	99.11	99.22
	WRN-28-10	96.62	99.12	99.35	99.40
200-way	ConvNet-64	75.44	89.08	93.21	94.72
1000-way	ConvNet-64	56.85	71.18	82.72	85.87

samples of N' classes that do not belong to the 5 classes. The extra samples are treated as unlabeled data. Our model shows good robustness (aka little performance drop) in 5-shot cases. The accuracy decreases to some extents when the unlabeled data increases.

Table 8. Robust test on miniImageNet. Acc: the accuracy of the labeled 5×15 query data. mAP: it is computed from top-15 confidently predicted data of each class.

miniImageNet	**N'=1**	**N'=5**
1-shot Acc	66.88 (3.43↓)	64.58 (5.73↓)
5-shot Acc	80.31 (1.58↓)	79.25 (2.64↓)
1-shot mAP	76.08	64.35
5-shot mAP	89.03	81.06

6 Results on Meta-Dataset

Meta-Dataset [10] is a new benchmark for few-shot learning. It is large-scale and consists of diverse datasets for training and evaluating models. We show our results in Table 9 and the ranks of our 5-shot model. For detailed comparison, please refer to *Table 1 (top) in [10]*.

Table 9. Results on Meta-Dataset. Avg. rank of our 5-shot model is **1.9**.

Test Source	1-shot	5-shot
ILSVRC	45.57	59.80 (1)
Omniglot	66.77	78.29 (1)
Aircraft	32.85	43.42 (7)
Birds	49.41	67.22 (3)
Textures	40.64	54.82 (1)
Quick Draw	45.52	58.80 (1)
Fungi	44.65	61.56 (1)
VGG Flower	69.97	83.88 (4)
Traffic Signs	53.93	68.68 (1)
MSCOCO	40.06	52.69 (1)

References

1. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B.: A closer look at few-shot classification. In: ICLR (2019)

2. Dhillon, G.S., Chaudhari, P., Ravichandran, A., Soatto, S.: A baseline for few-shot image classification. In: ICLR (2020)
3. Gidaris, S., Komodakis, N.: Dynamic few-shot visual learning without forgetting. In: CVPR. pp. 4367–4375 (2018)
4. Kim, J., Kim, T., Kim, S., Yoo, C.D.: Edge-labeling graph neural network for few-shot learning. In: CVPR. pp. 11–20 (2019)
5. Lake, B.M., Salakhutdinov, R., Gross, J., Tenenbaum, J.B.: One shot learning of simple visual concepts. Cognitive Science **33**(33) (2011)
6. Lee, K., Maji, S., Ravichandran, A., Soatto, S.: Meta-learning with differentiable convex optimization. In: CVPR. pp. 10657–10665 (2019)
7. Qiao, S., Liu, C., Shen, W., Yuille, A.L.: Few-shot image recognition by predicting parameters from activations. In: CVPR. pp. 7229–7238 (2018)
8. Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., Lillicrap, T.: Meta-learning with memory-augmented neural networks. In: ICML. pp. 1842–1850 (2016)
9. Snell, J., Swersky, K., Zemel, R.S.: Prototypical networks for few-shot learning. In: NIPS. pp. 4077–4087 (2017)
10. Triantafillou, E., Zhu, T., Dumoulin, V., Lamblin, P., Evci, U., Xu, K., Goroshin, R., Gelada, C., Swersky, K., Manzagol, P.A., Larochelle, H.: Meta-dataset: A dataset of datasets for learning to learn from few examples. In: ICLR, https://github.com/google-research/meta-dataset
11. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset. California Institute of Technology (2011)
12. Zagoruyko, S., Komodakis, N.: Wide residual networks. In: BMVC (2016)