Supplementary material

Mayug Maniparambil[1,2]
mayugmaniparambil@gmail.com
Kevin McGuinness[1,2]
kevin.mcguinness@dcu.ie
Noel O'Connor[1,2]
noel.oconnor@dcu.ie

[1] ML Labs
SFI Centre for Research Training, Dublin, Ireland
[2] Dublin City University, Dublin, Ireland

1 5-shot experiments

For the 5-shot case, we experiment with two different ways of averaging the support instances to form a prototype. Pre-avg averages the support instances before the BaseTransformer. The closest base instances in this case are sampled randomly from the 5 closest base classes using semantic similarity as described in Section 3.4 in the main paper. In contrast, for post-avg we adapt each support instance and its corresponding set of closest base instances independently and the prototype is obtained by averaging the adapted support instances after the BaseTransformer. Table 3 reports the results for both pre-avg and post-avg for 5-shot classification on the mini-ImageNet dataset using a ResNet12 encoder. Here we can see that pre-avg works much better than post-avg for the ResNet12 encoder. We believe that this could be because averaging the support instances results in a more robust input to the BaseTransformer, aiding in its training.

2 Tiered-ImageNet and CUB detailed results

Detailed results with 95% confidence intervals are reported for the tiered-ImageNet and CUB datasets in Table 1 and Table 2 respectively. We have reported tiered-ImageNet results only for the ResNet12 encoder.

3 Comparison with Semantic knowledge baselines

Previous methods that use semantic knowledge [1, 6, 7, 10] use it explicitly to structure the feature space, while we use it only for querying. Despite this, our method outperforms all these methods (see Table 4).

© 2022. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

Table 1: 5-way 1-shot and 5-way 5-shot classification accuracy (%) on tiered-ImageNet dataset for ResNet-12. The numbers in bold are the best performing methods for the corresponding setting.

Setups	1-shot	5-shot
ProtoNets [8]	65.65±0.92	83.40±0.65
SimpleShot [9]	69.75±0.20	85.31±0.15
FEAT [12]	70.80±0.23	84.79±0.16
CAN [3]	69.89±0.51	84.23±0.37
DeepEMD [13]	71.16±0.87	86.030.58
IEPT [14]	72.24±0.50	86.73±0.34
MELR [2]	72.14±0.51	**87.01**±0.35
InfoPatch [4]	71.51±0.52	85.44±0.35
DMF [11]	71.89±0.52	85.96±0.35
META-QDA [15]	69.97±0.52	85.51±0.58
PAL [5]	72.25±0.72	86.95±0.47
BaseTransformer	**72.46**±0.19	84.96±0.18

Table 2: 5-way 1-shot and 5 way 5-shot classification accuracy (%) on CUB dataset. The numbers in bold are the best performing methods for the corresponding setting.

Setups	1-shot		5-shot	
Backbone	Conv4-64	ResNet12	Conv4-64	ResNet12
ProtoNets [8]	51.31±0.89	66.09±0.92	70.77±0.70	82.50±0.58
FEAT [12]	68.87±0.22	-	82.90±0.15	-
DeepEMD [13]	-	75.65±0.83	-	88.69±0.50
IEPT [14]	69.97±0.49	-	84.33±0.33	-
MELR [2]	70.26±0.50	-	**85.01**±0.32	-
BaseTransformer	**72.15**±0.20	**82.27**±0.19	82.12±0.21	**90.64**±0.18

4 Visualization of learnt attention over base datapoints

We visualize the attention maps learnt by the BaseTransformer in Fig. 1. These are obtained by overlaying the resized attention map over the corresponding image of base instance selected by the querying function. We can see that for each support image, BaseTransformer has learnt to attend to visually similar regions of base instances. For example (Fig. 1 quadrant 2), for support instance nematode, the BaseTransformer learns to attend to the tentacle of jelly fish or the legs of harvestman to improve the prototype representation. It is also worth noting that in some cases the BaseTransformer is successful in identifying multiple visually similar features in base instance images when there are multiple instances of the class in one image. For example (Fig. 1 quadrant 4), for golden retriever, the BaseTransformer attends to two instances of gordon setter without being explicitly trained to identify multiple gordon setters.

Table 3: Results for different setups considered for averaging of support instances in 5-shot setting.

Setup	5-shot
Pre-avg	78.38±0.23
Post-avg	82.05±0.19

Table 4: Comparison with semantic knowledge baselines

Method	mini 1-shot	CUB 1-shot	tiered 1-shot
RS_FSL [11]	65.33±0.83	65.66±0.90	-
MS [2]	67.3	76.1	-
AM3 [10]	65.30±0.49	74.1	69.08±0.47
KTN [8]	64.42	-	-
BT (Ours)	70.88±0.17	82.27±0.19	72.46±0.19

Figure 1: Left: support instance; right: the three closest base instances (top) and attention maps overlaid over the closest base instances (bottom). It can be seen that BaseTransformers learns to select visually similar features from the base feature space using the learnt part based correspondences. Warmer color corresponds to higher attention weight.

References

[1] Mohamed Afham, Salman Khan, Muhammad Haris Khan, Muzammal Naseer, and Fahad Shahbaz Khan. Rich semantics improve few-shot learning. *British Machine Vision Conference*, 2021.

[2] Nanyi Fei, Zhiwu Lu, Tao Xiang, and Songfang Huang. Melr: Meta-learning via modeling episode-level relationships for few-shot learning. In *International Conference on Learning Representations*, 2020.

[3] Ruibing Hou, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Cross attention network for few-shot classification. *Advances in Neural Information Processing Systems*, 32, 2019.

[4] Chen Liu, Yanwei Fu, Chengming Xu, Siqian Yang, Jilin Li, Chengjie Wang, and Li Zhang. Learning a few-shot embedding model with contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8635–8643, 2021.

[5] Jiawei Ma, Hanchen Xie, Guangxing Han, Shih-Fu Chang, Aram Galstyan, and Wael Abd-Almageed. Partner-assisted learning for few-shot image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10573–10582, 2021.

[6] Zhimao Peng, Zechao Li, Junge Zhang, Yan Li, Guo-Jun Qi, and Jinhui Tang. Few-shot image recognition with knowledge transfer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 441–449, 2019.

[7] Eli Schwartz, Leonid Karlinsky, Rogerio Feris, Raja Giryes, and Alex Bronstein. Baby steps towards few-shot learning with multiple semantics. *Pattern Recognition Letters*, 160:142–147, 2022.

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

[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] Chen Xing, Negar Rostamzadeh, Boris Oreshkin, and Pedro O O Pinheiro. Adaptive cross-modal few-shot learning. *Advances in Neural Information Processing Systems*, 32, 2019.

[11] Chengming Xu, Yanwei Fu, Chen Liu, Chengjie Wang, Jilin Li, Feiyue Huang, Li Zhang, and Xiangyang Xue. Learning dynamic alignment via meta-filter for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5182–5191, 2021.

[12] Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha. Few-shot learning via embedding adaptation with set-to-set functions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8808–8817, 2020.

[13] 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*, pages 12203–12213, 2020.

[14] Manli Zhang, Jianhong Zhang, Zhiwu Lu, Tao Xiang, Mingyu Ding, and Songfang Huang. Iept: Instance-level and episode-level pretext tasks for few-shot learning. In *International Conference on Learning Representations*, 2020.

[15] Xueting Zhang, Debin Meng, Henry Gouk, and Timothy M Hospedales. Shallow bayesian meta learning for real-world few-shot recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 651–660, 2021.