One-Shot Image Recognition Using Prototypical Encoders with Reduced Hubness Supplementary Materials

Chenxi Xiao
Purdue University
West Lafayette, 47906, USA
xiao237@purdue.edu

Naveen Madapana
Purdue University
West Lafayette, 47906, USA
nmadapan@purdue.edu

Juan Wachs
Purdue University
West Lafayette, 47906, USA
jpwachs@purdue.edu

1. Details on Data Augmentation

For a fair comparison, the results reported in this work on two experimental scenarios (Belga –>Flickr32, and Belga –>Toplogos) given in Table 1 and 2 are acquired using the same data augmentation techniques as [3], which only includes random rotation and horizontal flipping.

Since the self-supervised contrastive learning benefits from data augmentation [1], the performance can be further boosted by incorporating additional data variations. The following ablation study is conducted on Belga–>Flickr32 and Belga –>Toplogos scenarios. We applied color jittering on top of the existing data augmentation techniques, including 1) Uniformly jittering image brightness, contrast, saturation from -80% to 80%, and 2) Uniformly jittering image hue from -20% to 20%. We compare the performance before and after applying the additional data augmentation in Table 5. Results show that the additional data augmentation can lead to superior results.

Table 5. Comparison of one-shot learning experiments on brand logo datasets with and w/o additional data augmentations.

	Belga–>Flickr32		Belga –>Toplogos	
Split	All	Unseen	All	Unseen
No. classes	32	28	11	6
No. support set	32-way		11-way	
VPE	56.6	53.53	58.65	57.75
Ours	65.54	62.56	65.57	70.27
Ours+aug	**69.55**	**67.33**	**65.98**	**71.83**

2. General One-shot Learning w/o Prototypes

While this approach targets one-shot learning tasks when class prototypes are available, the question about how to extend the proposed methodology to few-shot learning tasks when the class prototypes are absent remains unanswered. We demonstrate the performance of 1-shot 5-way experiments on a CIFAR-100 dataset [4]. We utilized 60/100 categories for training, 20/100 classes for validation and 20/100 classes for testing. The data partitions are subject to the ones proposed by [5]. For each class, we use a random image sampled within the class as a prototypical image. In the 1-shot 5-way classification task, our approach achieves a classification accuracy of 37.42%. Result shows that the performance is better than VPE but marginally lower (1-3%) than those algorithms that are designed for general few-shot learning tasks (refer to Table 6).

Table 6. Performance on CIFAR-100 for 1-shot 5-way experiments.

Approach	Test Acc %
MAML [2]	38.1
TADAM [5]	40.1
VPE [3]	32.4
Ours	37.4

In the above FSL experiment conducted on CIFAR-100 dataset, we noticed that using only one prototype image per class cannot fully represent categories with large variations. For example, a task that is more complex than recognizing 2D logos and gestures is to classify multi-view images of some 3D objects, which may require more than one image prototype to accurately characterize and recognize them. As a part of the future work, we will extend the VPE++ architecture to the scenarios where more than one prototypes per category need to be utilized.

References

[1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020.

[2] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Con-*

ference on Machine Learning-Volume 70, pages 1126–1135. JMLR. org, 2017.

[3] Junsik Kim, Tae-Hyun Oh, Seokju Lee, Fei Pan, and In So Kweon. Variational prototyping-encoder: One-shot learning with prototypical images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

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

[5] Boris Oreshkin, Pau Rodríguez López, and Alexandre Lacoste. Tadam: Task dependent adaptive metric for improved few-shot learning. In *Advances in Neural Information Processing Systems*, pages 721–731, 2018.