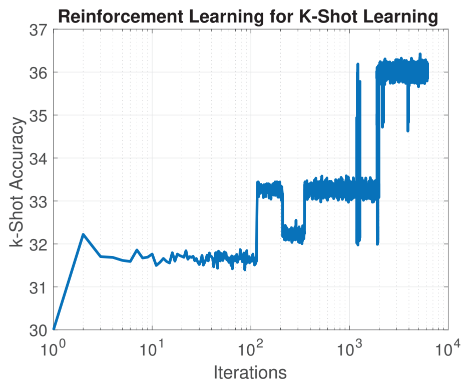
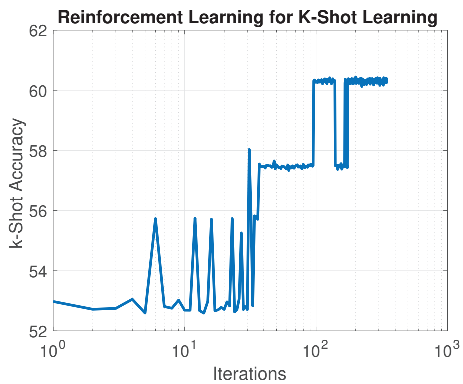


(a) one-shot



(b) 10-shot

Figure 6: Progression of hyper-parameter search through reinforcement learning through the iterations for K-shot transfer learning from CIFAR-100 to CIFAR-10

in performance with our clustering approach in comparison to the standard fine-tuning procedure.

Effect of Clustering Across Layers

It is commonly believed that most deep convolutional neural networks have highly redundant filters at the initial layers only. If this is indeed the case, applying our clustering method to layers other than the initial few layers should not be helpful. To test this hypothesis, we perform clustering to increasing number of layers, starting at the initial layers of the network. For this experiment we considered a pre-trained ResNet-18 network trained on a few categories in the CIFAR-10 dataset and used the other categories as the k-shot learning task. The results of GNA in Table 4 surprisingly does not confirm our hypothesis. We found that all layers of the deep network did consist of redundant filters for the k-shot learning task. In fact, applying our method to all the layers of the network resulted in the best performance. This experiment suggests that large convolutional neural networks could potentially consist of redundant parameters even in the higher layers, necessitating search over the entire hyper-parameter space of parameter groupings. This motivates the need for efficient techniques to search the hyper-parameter space, like the one we proposed in this paper.

Table 3: k-shot classification performance as a function of number of samples per category

The number	Accuracy (%)	
of clustering	w/o clustering	w/ clustering
25 shot	80.41	84.48
20 shot	76.90	82.09
15 shot	81.63	84.01
10 shot	70.64	72.88
5 shot	68.84	68.93
1 shot	52.25	53.77

(a) Accuracy

The number	Standard deviation	
of clustering	w/o clustering	w/ clustering
25 shot	5.68	0.76
20 shot	2.66	1.44
15 shot	5.16	0.79
10 shot	5.31	3.79
5 shot	6.95	2.72
1 shot	10.02	10.66

(b) Standard deviation

Conclusion

In this paper we proposed a new regularization method for fine-tuning a pre-trained network for k-shot learning. The key idea of our approach was to effectively reduce the dimensionality of the network parameter space, by clustering the weights in each layer while ensuring intra-group similarity and inter-group orthogonality. To provide additional supervision to the k-shot learning problem we introduce a triplet loss to maximize the separation between the k-shot samples. Lastly, we introduced a reinforcement learning based approach to efficiently search over the hyper-parameters of our clustering approach. The experimental results demonstrate that our proposed regularization technique can significantly improve the performance of fine-tuning based k-shot learning approaches.

Acknowledgments

This work was sponsored in part by JST CREST grant (JPMJCR14E1) and NSF NRI grant (1637927).

References

Bergstra, J., and Bengio, Y. 2012. Random search for hyper-parameter optimization. *Journal Machine Learning Research* 13:281–305.

Daume III, H. 2009. Frustratingly easy domain adaptation. *arXiv*.

Fei-Fei, L.; Fergus, R.; and Perona, P. 2006. One-shot learning of object categories.

Hansen, S. 2016. Using Deep Q-learning to Control Optimization Hyperparameters. *Optimization and Control*.