SUPPLEMENTARY MATERIAL OF DYN-ADAPTER

Anonymous authors
Paper under double-blind review

A DATASET

We have conducted experiments based on a large and diverse set of datasets. All the datasets leveraged are listed in Table 1.

Table 1: Datasets we leveraged for multiple experiments.

	Dataset	Classes	Train	Val	Test
	VTAB-1k (Zhai et al., 2019)				
Natural	CIFAR 100 (Krizhevsky et al., 2009)	100	800/1000	200	10000
	Caltech101 (Fei-Fei et al., 2004)	102			6084
	DTD (Cimpoi et al., 2014)	47			1880
	Oxford-Flowers 101 (Nilsback & Zisserman, 2006)	102			6149
	Oxford-Pets (Parkhi et al., 2012)	37			3669
	SVHN (Netzer et al., 2011)	10			26032
	Sun397 (Xiao et al., 2010)	397			21750
Specialized	Patch Camelyon	2	800/1000	200	32768
	EuroSAT (Helber et al., 2019)	10			5400
	Resisc45 (Cheng et al., 2017)	45			6300
	Retinopathy (Graham, 2015)	5			42670
Structured	Clevr/count (Johnson et al., 2017)	8	800/1000	200	15000
	Clevr/distance Johnson et al. (2017)	6			15000
	DMLab (Beattie et al., 2016)	6			22735
	KITTI-Dist (Geiger et al., 2013)	4			711
	dSprites/location (Matthey et al., 2017)	16			73728
	dSprites/orientation (Matthey et al., 2017)	16			73728
	SmallNORB/azimuth (LeCun et al., 2004)	18			12150
	SmallNORB/elevation (LeCun et al., 2004)	18			12150
	Few-shot Learning				
	Food-101 (Bossard et al., 2014)	101	16 per class	20200	30300
	Stanford Cars (Krause et al., 2013)	196		1635	8041
	Oxford-Flowers (Nilsback & Zisserman, 2006)	102		1633	2463
	FGVC-Aircraft (Maji et al., 2013)	100		3333	3333
	Oxford-Pets (Parkhi et al., 2012)	37		736	3669
	Domain Generalization				
	ImageNet (Deng et al., 2009)	1000	16 per class	50000	50000
	ImageNetv2 (Recht et al., 2019)	1000	-	-	10000
	ImageNet-Sketch (Wang et al., 2019)	1000	-	-	50000
	ImageNet-A (Hendrycks et al., 2021b)	200	-	-	7500
	ImageNet-R (Hendrycks et al., 2021a)	200	-	-	30000

B DISCUSSION

B.1 LIMITATION

Although we think deeply about the optimization process of *Dyn-Adapter* from a theoretical perspective, conducting extensive experiments and achieving inspiring results, there still lacks profound mathematical modeling for this joint optimization problem to elucidate from a more fundamental standpoint which direction is more optimal. Moreover, while the addition of an early head significantly enhances inference efficiency and ensures accuracy, the increment of the head inevitably introduces a minor portion of training parameters. Contemplating more efficient learning strategies, such as employing parameter sharing strategies to allow them to share some knowledge, is a direction that worth thinking in the future. We expect that future improvements based on our concise and efficient method will yield more desirable benefits.

B.2 Broader Impact

This work can benefit the wide application scenarios of PETL methods, and further reduce the inference efficiency with a large gap. Under resource-limited circumstances, our method provide more possibilities for the widespread utilization of PETL and save hardware resources in practice.

B.3 Future Work

The aforementioned limitations demonstrate the directions of our future work. Moreover, investigating exit decision-making process in *Dyn-Adapter* is meaningful, as the right decision precedes the efficient early exit. Currently, the threshold of early stopping relies on thresholds induced from the training set, without explicit modeling of the network's inherent attributes, features of the pre-trained model, and transferred information within the adapter. There may exists several secrets to explore.

References

Charles Beattie, Joel Z Leibo, Denis Teplyashin, Tom Ward, Marcus Wainwright, Heinrich Küttler, Andrew Lefrancq, Simon Green, Víctor Valdés, Amir Sadik, et al. Deepmind lab. *arXiv preprint arXiv:1612.03801*, 2016. 1

Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101–mining discriminative components with random forests. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13*, pp. 446–461. Springer, 2014. 1

Gong Cheng, Junwei Han, and Xiaoqiang Lu. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 105(10):1865–1883, 2017. 1

Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3606–3613, 2014. 1

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009. 1

Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *2004 conference on computer vision and pattern recognition workshop*, pp. 178–178. IEEE, 2004. 1

Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. 1

Ben Graham. Kaggle diabetic retinopathy detection competition report. *University of Warwick*, 22, 2015. 1

Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019. 1

Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, et al. The many faces of robustness: A critical analysis of out-of-distribution generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8340–8349, 2021a. 1

Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. Natural adversarial examples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15262–15271, 2021b. 1

Justin Johnson, Bharath Hariharan, Laurens Van Der Maaten, Li Fei-Fei, C Lawrence Zitnick, and Ross Girshick. Clevr: A diagnostic dataset for compositional language and elementary visual reasoning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2901–2910, 2017. 1

Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision workshops*, pp. 554–561, 2013. 1

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

Yann LeCun, Fu Jie Huang, and Leon Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pp. II–104. IEEE, 2004. 1

Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013. 1

Loic Matthey, Irina Higgins, Demis Hassabis, and Alexander Lerchner. dsprites: Disentanglement testing sprites dataset. https://github.com/deepmind/dsprites-dataset/, 2017. 1

Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011. 1

M-E Nilsback and Andrew Zisserman. A visual vocabulary for flower classification. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, volume 2, pp. 1447–1454. IEEE, 2006. 1

Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, and CV Jawahar. Cats and dogs. In *2012 IEEE conference on computer vision and pattern recognition*, pp. 3498–3505. IEEE, 2012. 1

Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. Do imagenet classifiers generalize to imagenet? In *International conference on machine learning*, pp. 5389–5400. PMLR, 2019. 1

Haohan Wang, Songwei Ge, Zachary Lipton, and Eric P Xing. Learning robust global representations by penalizing local predictive power. *Advances in Neural Information Processing Systems*, 32, 2019. 1

Jianxiong Xiao, James Hays, Krista A Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pp. 3485–3492. IEEE, 2010. 1

Xiaohua Zhai, Joan Puigcerver, Alexander Kolesnikov, Pierre Ruyssen, Carlos Riquelme, Mario Lucic, Josip Djolonga, Andre Susano Pinto, Maxim Neumann, Alexey Dosovitskiy, et al. A large-scale study of representation learning with the visual task adaptation benchmark. *arXiv preprint arXiv:1910.04867*, 2019. 1