DnA: Improve Few-Shot Transfer Learning with Low-Rank Decompose and Align

Ziyu Jiang[1,2†], Tianlong Chen[3], Xuxi Chen[3], Yu Cheng[1], Luowei Zhou[1], Lu Yuan[1], Ahmed Awadallah[1], and Zhangyang Wang[3‡]

[1] Microsoft Corporation
[2] Texas A&M University
[3] University of Texas at Austin

This supplementary material contains the following details that we could not include in the main paper due to space restrictions.

- **(Sec. 1)** Details of the computing infrastructure.
- **(Sec. 2)** Comparison with LoRA [1].

1 Details of computing infrastructure

Our codes are based on Pytorch [2], and all models are trained with NVIDIA A100 Tensor Core GPU.

2 Comparison with LoRA [1]

LoRA, a closely related work, shows that formalizing the weight changing as a low-rank matrix can also improve the fine-tuning performance. Therefore, we compare with Align+LoRA to verify the effectiveness of the proposed Decomposition method. As illustrated at Table 1, by applying the LoRA with Align, the performance could improve by 0.4%. However, it is still weaker than the proposed DnA with an obvious margin of 0.5%.

Table 1: Compare with LoRA [1] in terms of the 5-shot performance on iNaturalist-1k.

Method	Accuracy
Align	47.4
Align+LoRA	47.8
DnA (Ours)	48.3

† Work done during an intership at Microsoft Corporation
‡ Correspondence to: Zhangyang Wang (atlaswang@utexas.edu)

References

1. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)
2. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)