DGS: Dual Gradient and Semantic-Shift Guided Low-Rank Adaptation for Class Incremental Learning

Supplementary Material

In this supplementary material, we first prove the generalizability of the proposed method to other types of PTMs (appendix A). Then we discuss the sensitivity to the choice of hyperparameter α (appendix B).

1. Appendix A

In this section, we provide additional data on the generalizability of **DGS** to other types of PTMs. The core principle of our method is to leverage the knowledge already embedded in pre-trained models to construct a task-relevant knowledge space from the training data. Table.1 reports the results on ImageNet-R using three different backbones, including ResNet50, ResNet101, and ViT-DINO.

In the table, consistent and notable performance improvements are observed across backbones initialized with different pre-trained parameters, indicating that the effectiveness of the proposed approach under diverse pre-training settings.

Method	A_{avg}	A_{last}	F_{last}
ResNet50 LoRA	46.77	63.92	34.70
ResNet50 with DGS	**52.97**	**66.16**	**25.76**
ResNet101 LoRA	57.48	69.96	24.01
ResNet101 with DGS	**60.78**	**71.09**	**18.59**
ViT-DINO LoRA	71.22	78.81	9.39
ViT-DINO with DGS	**75.20**	**80.99**	**6.82**

Table 1. Ablation study on different types of PTMs on ImageNet-R.

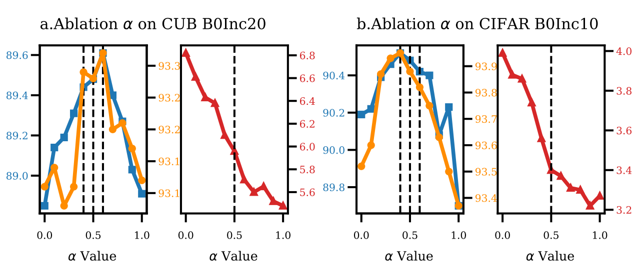

Figure 1. Performance variation on CUB and CIFAR100 with respect to hyperparameter $\boldsymbol{\alpha}$.

2. Appendix B

In this section, we present the sensitivity to the choice of the hyperparameter α in more CIL datasets, including CUB and CIFAR-100.

The larger α emphasizes the projected gradient, reducing forgetting but slightly sacrificing accuracy, indicating stronger stability but weaker plasticity. The smaller α prioritizes the raw task-specific gradient, improving accuracy while increasing forgetting. As shown in Fig.4 of the main paper and Fig.1, $\alpha = 0.5$ lies in a stable performance region across multiple datasets. Thus, it serves as a practical default that effectively balances pre-trained and task-specific knowledge without task-specific tuning. Similar sensitivity behavior has been observed in prior continual learning methods.