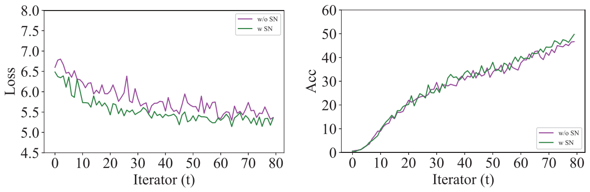
Appendix

A. Stable Training

The main difficulty with generator model training is that it requires a large number of iterations and is prone to model collapse. Therefore, in order to stabilize the model training, we adopted spectral normalization (SN) to stabilize the training of the model, essentially imposing Lipschitz constraints on the weights of the generator, which restricts the weights to a controllable interval. From Fig. 8 below, during the first 80 iterations, when adopting the SN, the loss decreases more stable, and also accuracy improves obviously. Therefore, a stable generator benefits the whole training of DFKD.



(a) Loss of generator. (b) Accuracy of student.

Figure 8. Effect of loss and accuracy, when exploiting the SN to optimize the generator.

B. Extended Experiments

To further verify the generalization of our approach to coarse-grained classification, we conduct extended experiments on the CIFAR-10 and CIFAR-100 datasets. As can be seen from Tab. 7, the performance of our method on both datasets can also outperform other reported methods. Although our method is designed for fine-grained classification, it can also perform well on coarse-grained classification, demonstrating our combined approaches, i.e., spatial-wise attention, mixed high-order attention mechanism, and semantic feature contrastive learning, are beneficial for data-free distillation.

Table 7. Results on CIFAR-10/CIFAR-100 compared with different data-free distillation methods.

	CIFAR10		CIFAR100	
Method	FLOPs	Accuracy	FLOPs	Accuracy
Teacher	1.16G	95.70	1.16G	78.05
Student	557M	95.20	558M	77.10
ZSKD	360M	69.56	558M	56.63
ZSKT	329M	89.71	558M	67.52
ADI	557M	93.26	558M	60.83
DAFL	557M	92.22	558M	74.47
DFAD	557M	93.30	558M	67.70
DFQ	557M	94.61	558M	77.01
CMI	557M	94.84	558M	77.04
MAD	557M	94.90	558M	77.31
Our	557M	**94.96**	558M	**77.33**