MDCS: More Diverse Experts with Consistency Self-distillation for Long-tailed Recognition (Supplementary Material)

Qihao Zhao[1,2], Chen Jiang[1], Wei Hu[1], Fan Zhang[1*], Jun Liu[2]
[1] Beijing University of Chemical Technology, China
[2] Singapore University of Technology and Design, Singapore
{zhaoqh,jiangchen,huwei,zhangf}@mail.buct.edu.cn, jun_liu@sutd.edu.sg

1. The efficiency of our Consistency Self-distillation.

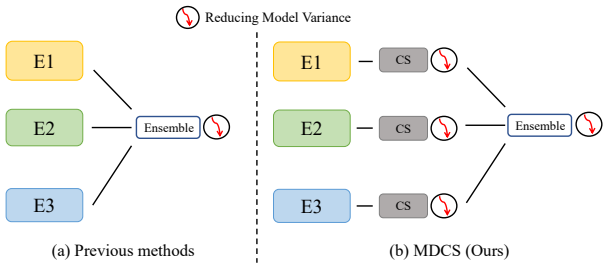

Figure 1.

As illustrated in Fig. 1, previous methods [3, 1, 4] reduced the model variance only by using an ensemble of multiple experts. In contrast, our approach not only reduces the variance by ensemble but also reduces the model variance by CS for each expert. The effect of CS is not only to reduce the model variance. Each expert gets richer constraint information through weakly augmented images, which enhances the expert's own recognition ability. As shown in Table 1, experts with stronger recognition abilities also produce more diverse ensemble models.

2. More details settings for our method.

We implement our method with PyTorch. Following [4, 2], we use ResNeXt-50/ResNet-50 for ImageNet-LT, ResNet-32 for CIFAR100/10-LT, ResNet-152 for Places-LT and ResNet-50 for iNaturalist 2018 as backbones, respectively. Moreover, we adopt the cosine classifier for prediction on all datasets. The details settings for our method are shown in table 2.

References

[1] Jiarui Cai, Yizhou Wang, and Jenq-Neng Hwang. Ace: Ally complementary experts for solving long-tailed recognition in one-shot. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 112–121, 2021.

[2] Jun Li, Zichang Tan, Jun Wan, Zhen Lei, and Guodong Guo. Nested collaborative learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6949–6958, 2022.

[3] Xudong Wang, Long Lian, Zhongqi Miao, Ziwei Liu, and Stella X Yu. Long-tailed recognition by routing diverse distribution-aware experts. *arXiv preprint arXiv:2010.01809*, 2020.

[4] Yifan Zhang, Bryan Hooi, Lanqing Hong, and Jiashi Feng. Self-supervised aggregation of diverse experts for test-agnostic long-tailed recognition. *Advances in Neural Information Processing Systems*, 3, 2022.

Method	E1 Acc	E2 Acc	E3 Acc	All Acc	σ
w/o CS	38.8	45.2	31.4	50.7	53.4
w/ CS	40.6(+1.8)	50.8(+5.6)	36.0(+4.6)	56.1(+5.4)	60.4(+7.0)

Table 1. The efficiency of Consistency Self-distillation. With CS, not only is the model variance reduced, but also the expert recognition ability and the final model diversity are improved.

Items	CIFAR100/10-LT	ImageNet-LT	Places-LT	iNaturalist 2018
Network Architectures				
network backbone	ResNet-32	ResNeXt-50/ResNet-50	ResNet-152	ResNet-50
Training Phase				
epochs	200/400	180/400	30	100/400
batch size	64	256	64	512
learning rate (lr)	0.1	0.1	0.01	0.2
lr schedule	linear decay	cosine decay	linear decay	linear decay
λ	-0.5, 1, 2.5	-0.5, 1, 2.5	-0.5, 1, 2.5	-0.5, 1, 2.5
weight decay factor	$5 * 10^{-4}$	$5 * 10^{-4}$	$5 * 10^{-4}$	$5 * 10^{-4}$
momentum factor	0.9			
optimizer	SGD optimizer with nesterov			

Table 2. Statistics of the used network architectures and hyper-parameters in our experiments.