Nested Collaborative Learning for Long-Tailed Visual Recognition — Supplementary Material —

Jun Li[1,2*], Zichang Tan[3,4*], Jun Wan[1,2†], Zhen Lei[1,2,5], Guodong Guo[3,4]

[1]CBSR&NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China

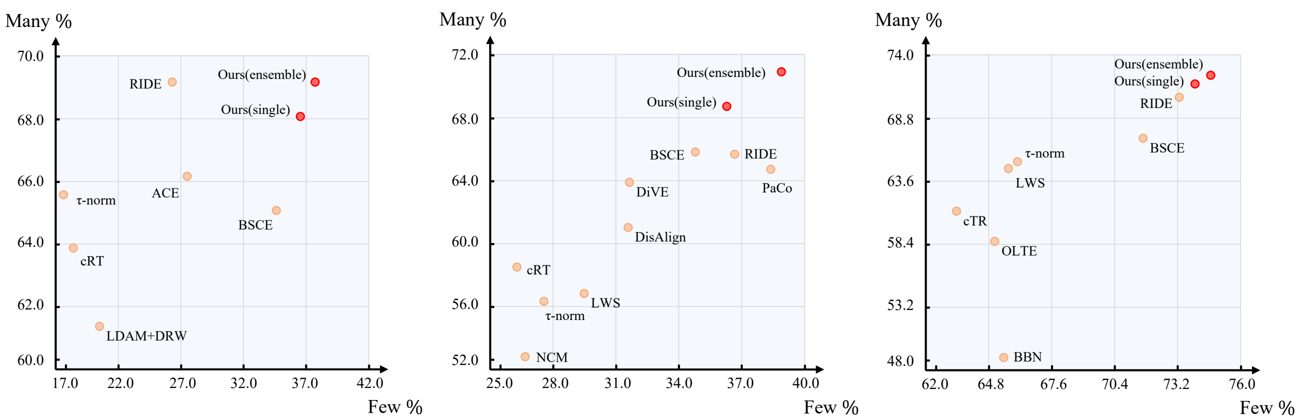
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China

[3]Institute of Deep Learning, Baidu Research, Beijing, China

[4]National Engineering Laboratory for Deep Learning Technology and Application, Beijing, China

[5]Centre for Artificial Intelligence and Robotics, Hong Kong Institute of Science&Innovation, Chinese Academy of Sciences, Hong Kong, China

{lijun2021,jun.wan}@ia.ac.cn, zlei@nlpr.ia.ac.cn, {tanzichang, guoguodong01}@baidu.com



(a) CIFAR100-LT (b) ImageNet-LT (c) iNaturalist 2018

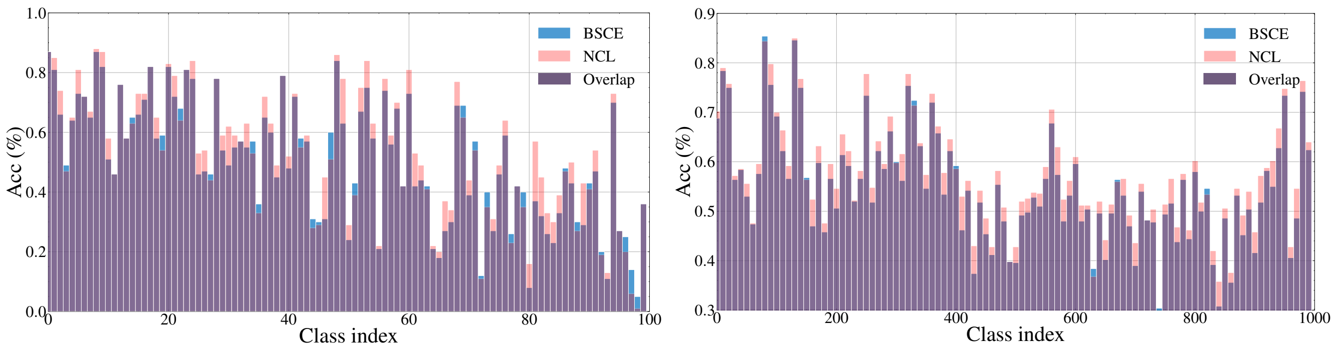
Figure 1. Comparisons on (a) CIFAR100-LT, (b) ImageNet-LT with ResNet-50 and (c) iNaturalist 2018 of our proposed method and some representative methods over many and few splits.

A. More Comparisons

Following previous works [1, 7], the dataset can be divided into many (with more than 100 images), medium (with 20 $\sim$ 100 images) and few (with less than 20 images) splits. In this subsection, we further report the accuracies on many, medium and few splits for detailed comparisons. The comparisons on CIFAR100-LT and ImageNet-LT are shown in Table 2 and Table 1, respectively. For iNaturalist 2018, its results of many, medium and few splits have been reported in the article. As we can see, the proposed method also outperforms all previous methods in all splits. Moreover, compared with the baseline method BSCE, the proposed method also can achieve comprehensive improvements on all splits. It indicates that all categories can benefit from the proposed NCL, where the knowledge transferring is allowed over all categories. To better illustrate the superiority of our proposed method, we visualize the performance of our method and, as comparison some representative previous methods in terms of many and few splits for comparisons as shown in Fig. 1. It can be witnessed that the proposed method achieves remarkable improvements on both many and few splits, compared with previous methods.

B. Improvements on All Categories

In this section, we analyze the improvements on all categories over CAFAR100-LT and ImageNet-LT. We visualize the accuracy on each category of the baseline model BSCE and our proposed NCL as shown in Fig. 2. Some previous methods [7, 11, 12] can improve the overall performance for long-tailed visual recognition but mostly at the sacrifice of



(a) CIFAR100-LT (b) ImageNet-LT

Figure 2. Detailed analysis of improvements on all categories on (a) CIFAR100-LT (IF=100) and (b) ImageNet-LT datasets. Due to ImageNet-LT containing 1,000 categories, we report the average accuracy of every 10 adjacent categories for comparisons to avoid confusion caused by too many histograms. The number of each category decreases as the category index increases. For example, category-0 and category-99 have the largest and smallest number of training images in CIFAR100-LT dataset, respectively.

Method	Ref.	ResNet-50				ResNeXt-50			
		Many	Medium	Few	All	Many	Medium	Few	All
BBN [19]	CVPR'20	–	–	–	48.3	–	–	–	49.3
NCM [7]	ICLR'20	53.1	42.3	26.5	44.3	56.6	45.3	28.1	47.3
cRT [7]	ICLR'20	58.8	44.0	26.1	47.3	61.8	46.2	27.4	49.6
τ-norm [7]	ICLR'20	56.6	44.2	27.4	46.7	59.1	46.9	30.7	49.4
LWS [7]	ICLR'20	57.1	45.2	29.3	47.7	60.2	47.2	30.3	49.9
RIDE [14]	ICLR'21	66.2	52.3	36.5	55.4	68.2	53.8	36.0	56.8
DisAlign [16]	CVPR'21	61.3	52.2	31.4	52.9	62.7	52.1	31.4	53.4
DiVE [5]	ICCV'21	64.1	50.4	31.5	53.1	–	–	–	–
SSD [9]	ICCV'21	–	–	–	–	66.8	53.1	35.4	56.0
ACE [1]	ICCV'21	–	–	–	54.7	–	–	–	56.6
PaCo [3]	ICCV'21	65.0	55.7	38.2	57.0	67.5	56.9	36.7	58.2
BSCE (baseline)	–	66.1	51.9	34.6	53.9	66.8	51.1	34.4	53.6
Ours (single)	–	**69.0**	54.0	36.1	**57.4**	**70.4**	54.9	**37.0**	**58.4**
Ours (ensemble)	–	**71.2**	**56.3**	**38.7**	**59.5**	**72.5**	**57.2**	**38.8**	**60.5**

Table 1. Comparisons on ImageNet-LT dataset.

Method	Ref.	CIFAR100-LT			
		Many	Medium	Few	All
CB Focal loss [4]	CVPR'19	65.2	37.1	9.1	38.3
LDAM+DRW [2]	NeurIPS'19	61.5	41.7	20.2	42.0
τ-norm [7]	ICLR'20	65.7	43.6	17.3	43.2
cRT [7]	ICLR'20	64.0	44.8	18.1	43.3
BBN [19]	CVPR'20	47.2	49.4	19.8	39.4
LFME [15]	ECCV'20	–	–	–	42.3
CAM [17]	AAAI'21	–	–	–	47.8
Logit Adj. [10]	ICLR'21	–	–	–	43.9
RIDE [14]	ICLR'21	69.3	49.3	26.0	49.1
LDAM+M2m [8]	CVPR'21	–	–	–	43.5
MiSLAS [18]	CVPR'21	–	–	–	47.0
LADE [6]	CVPR'21	–	–	–	45.4
Hybrid-SC [13]	CVPR'21	–	–	–	46.7
DiVE [5]	ICCV'21	–	–	–	45.4
SSD [9]	ICCV'21	–	–	–	46.0
ACE [1]	ICCV'21	66.3	52.8	27.2	49.6
PaCo [3]	ICCV'21	–	–	–	52.0
BSCE (baseline)	–	65.2	50.4	34.3	50.6
Ours (single)	–	**67.8**	**53.7**	**36.5**	**53.3**
Ours (ensemble)	–	**68.3**	**55.0**	**38.2**	**54.2**

Table 2. Comparisons on CIFAR100-LT with an IF of 100.

the accuracy on head classes. However, by comparison with the baseline method on both datasets, our NCL can achieve the accuracy improvements almost on all categories, which manifests the advantages of the collaborative learning.

References

[1] Jiarui Cai, Yizhou Wang, and Jenq-Neng Hwang. Ace: Ally complementary experts for solving long-tailed recognition in one-shot. In *ICCV*, 2021. 1, 2

[2] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In *NeurIPS*, 2019. 2

[3] Jiequan Cui, Zhisheng Zhong, Shu Liu, Bei Yu, and Jiaya Jia. Parametric contrastive learning. In *ICCV*, 2021. 2

[4] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *CVPR*, 2019. 2

[5] Yin-Yin He, Jianxin Wu, and Xiu-Shen Wei. Distilling virtual examples for long-tailed recognition. In *ICCV*, 2021. 2

[6] Youngkyu Hong, Seungju Han, Kwanghee Choi, Seokjun Seo, Beomsu Kim, and Buru Chang. Disentangling label distribution for long-tailed visual recognition. In *CVPR*, 2021. 2

[7] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. *arXiv preprint arXiv:1910.09217*, 2019. 1, 2

[8] Jaehyung Kim, Jongheon Jeong, and Jinwoo Shin. M2m: Imbalanced classification via major-to-minor translation. In *CVPR*, 2020. 2

[9] Tianhao Li, Limin Wang, and Gangshan Wu. Self supervision to distillation for long-tailed visual recognition. In *ICCV*, 2021. 2

[10] Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. Long-tail learning via logit adjustment. *arXiv preprint arXiv:2007.07314*, 2020. 2

[11] Jiawei Ren, Cunjun Yu, Shunan Sheng, Xiao Ma, Haiyu Zhao, Shuai Yi, and Hongsheng Li. Balanced meta-softmax for long-tailed visual recognition. *arXiv preprint arXiv:2007.10740*, 2020. 1

[12] Jiaqi Wang, Wenwei Zhang, Yuhang Zang, Yuhang Cao, Jiangmiao Pang, Tao Gong, Kai Chen, Ziwei Liu, Chen Change Loy, and Dahua Lin. Seesaw loss for long-tailed instance segmentation. In *CVPR*, 2021. 1

[13] Peng Wang, Kai Han, Xiu-Shen Wei, Lei Zhang, and Lei Wang. Contrastive learning based hybrid networks for long-tailed image classification. In *CVPR*, 2021. 2

[14] Xudong Wang, Long Lian, Zhongqi Miao, Ziwei Liu, and Stella X Yu. Long-tailed recognition by routing diverse distribution-aware experts. In *ICLR*, 2021. 2

[15] Liuyu Xiang, Guiguang Ding, and Jungong Han. Learning from multiple experts: Self-paced knowledge distillation for long-tailed classification. In *ECCV*, 2020. 2

[16] Songyang Zhang, Zeming Li, Shipeng Yan, Xuming He, and Jian Sun. Distribution alignment: A unified framework for long-tail visual recognition. In *CVPR*, 2021. 2

[17] Yongshun Zhang, Xiu-Shen Wei, Boyan Zhou, and Jianxin Wu. Bag of tricks for long-tailed visual recognition with deep convolutional neural networks. In *AAAI*, 2021. 2

[18] Zhisheng Zhong, Jiequan Cui, Shu Liu, and Jiaya Jia. Improving calibration for long-tailed recognition. In *CVPR*, 2021. 2

[19] Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In *CVPR*, 2020. 2