# Rethinking Class-Balanced Methods for Long-Tailed Visual Recognition from a Domain Adaptation Perspective
## *- Supplementary Materials -*

## Outline

- Multiple runs of experiments on CIFAR-LT-10 under different imbalance factors (IFs). (Section A).
- Detailed comparison of various methods on large-scale long-tailed datasets (Section B).

## A. Multiple runs on CIFAR-LT-10

In this experiment, we further validate our approach by running each setting 5 times with different random seeds. Table 1 shows the mean top-1 errors (%) and the standard deviations under the imbalance factor of 200, 100, and 50. We can see that the mean error rates are consistent with the results provided in Table 2 of the main paper.

## B. Detailed comparison of various methods on large-scale long-tailed datasets

In this section, we present the top-1 errors (%) of various methods for ImageNet-LT, Places-LT, and iNaturalist 2018[1]. As the experiment setups of the existing works vary by network initialization, the sampling strategy of mini-batches, losses, trainable layers of a network, etc., it is hard to have a fair comparison by the end results. Hence, besides their top-1 errors, we also report the experiment setups for each method. Tables 2, 3, and 4 show the results of the different methods on ImageNet-LT, Places-LT, and iNaturalist 2018, respectively. Our approach outperforms the class-balanced weighting scheme for both the cross-entropy loss and the focal loss, as we observed in the main paper. Moreover, our results are on par with best reported ones except on ImageNet-LT. Finally, we stress that almost all existing methods employ a class-balanced weighting or sampling strategy no matter what their main techniques are to tackle the long-tailed problem. Hence, given our consistent improvements over the class-balanced weighting, we expect the methods which have benefited from the class-balancing can gain further from our two-component weighting.

[1] While we include the comparison on iNaturalist 2018 due to that most existing related works report results on this dataset, we reiterate that we advocate the use of iNaturalist 2017, instead of 2018, in this and future work due to the extremely small validation set of iNaturalist 2018.

## References


[1] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In *Advances in Neural Information Processing Systems*, pages 1565–1576, 2019. 3

[2] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge J. Belongie. Class-balanced loss based on effective number of samples. *arXiv:1901.05555*, 2019. 2, 3

[3] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. *arXiv:1910.09217*, 2019. 2, 3

[4] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *TPAMI*, 2018. 2

[5] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X. Yu. Large-scale long-tailed recognition in an open world. In *CVPR*, 2019. 2

[6] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. *arXiv:1803.09050*, 2018. 2

[7] Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. Meta-weight-net: Learning an explicit mapping for sample weighting. In *NeurIPS*, 2019. 2

[8] Xiao Zhang, Zhiyuan Fang, Yandong Wen, Zhifeng Li, and Yu Qiao. Range loss for deep face recognition with long-tailed training data. In *ICCV*, pages 5419–5428, 2017. 2

Table 1. Multiple runs of our approach by using the cross-entropy loss on CIFAR-LT-10. The results are top-1 errors% on the test sets.

| Imbalance factor | 200 | 100 | 50 |
|---|---|---|---|
| Cross-entropy training | 34.32 | 29.64 | 25.19 |
| Class-balanced cross-entropy loss [2] | 31.11 | 27.63 | 21.95 |
| Class-balanced fine-tuning | 33.76 | 28.66 | 22.56 |
| Class-balanced fine-tuning* | 33.92 | 28.67 | 22.58 |
| L2RW [6] | 33.75 | 27.77 | 23.55 |
| L2RW [6]* | 33.49 | 25.84 | 21.07 |
| Meta-weight net [7] | 32.8 | 26.43 | 20.9 |
| Ours with cross-entropy loss | **29.32 $\pm$ 0.23** | **23.71 $\pm$ 0.22** | **19.45 $\pm$ 0.28** |

Table 2. Test top-1 errors (%) of different methods on ImageNet-LT. * indicates the re-run results.

| Methods | NN | Initialization | Sampling | Loss | Stage-1 Trainable Variables | Stage-2 Trainable Variables | Results |
|---|---|---|---|---|---|---|---|
| Vanilla Model | ResNet-10 | No-pretrain | Class-Balanced | CE | All | - | 80.0 |
| Vanilla Model [4] | ResNet-10 | No-pretrain | Class-Balanced | Focal | All | - | 69.8 |
| Vanilla Model | ResNet-10 | No-pretrain | Class-Balanced | Lifted | All | - | 69.2 |
| Vanilla Model [8] | ResNet-10 | No-pretrain | Class-Balanced | Range | All | - | 69.3 |
| Joint [3] | ResNet-10 | No-pretrain | Class-Balanced | CE | All | All | 65.2 |
| NCM [3] | ResNet-10 | No-pretrain | Class-Balanced | CE | All | Classifier layer | 64.5 |
| cRT [3] | ResNet-10 | No-pretrain | Class-Balanced | CE | All | Classifier layer | **58.2** |
| $\tau$-normalized [3] | ResNet-10 | No-pretrain | Class-Balanced | CE | All | Classifier layer | <u>59.4</u> |
| OLTR* [5] | ResNet-10 | No-pretrain | Class-Balanced | CE | All | All | 65.6 |
| OLTR [5] | ResNet-10 | No-pretrain | Class-Balanced | CE | All | All | 64.4 |
| Ours | ResNet-10 | No-pretrain | None | CE | All | Classifier layer | 63.5 |
| Ours | ResNet-10 | No-pretrain | None | Focal | All | Classifier layer | 63.3 |
| Vanilla Model | ResNet-50 | No-pretrain | None | CE | All | - | 59.0 |
| CB [2] | ResNet-50 | No-pretrain | None | CE | All | - | 58.2 |
| Joint [3] | ResNet-50 | No-pretrain | Class-Balanced | CE | All | All | 58.4 |
| NCM [3] | ResNet-50 | No-pretrain | Class-Balanced | CE | All | Classifier layer | 55.7 |
| cRT [3] | ResNet-50 | No-pretrain | Class-Balanced | CE | All | Classifier layer | <u>52.7</u> |
| $\tau$-normalized [3] | ResNet-50 | No-pretrain | Class-Balanced | CE | All | Classifier layer | 53.3 |
| Ours | ResNet-50 | No-pretrain | None | CE | All | Classifier layer | **52.0** |

Table 3. Test top-1 errors (%) of different methods on Places-LT. * indicates the re-run results.

| Methods | NN | Initialization | Sampling | Loss | Stage-1 Trainable Variables | Stage-2 Trainable Variables | Results |
|---|---|---|---|---|---|---|---|
| Vanilla Model | ResNet-152 | ImageNet | Class-Balanced | CE | FC layers<br>Last Block + FC | -<br>- | 72.1<br>69.7 |
| Vanilla Model [4] | ResNet-152 | ImageNet | Class-Balanced | Focal | FC layers<br>Last Block + FC | -<br>- | 67.0<br>66.5 |
| Vanilla Model | ResNet-152 | ImageNet | Class-Balanced | Lifted | FC layers | - | 64.8 |
| Vanilla Model [8] | ResNet-152 | ImageNet | Class-Balanced | Range | FC layers | - | 64.9 |
| Joint [3] | ResNet-152 | ImageNet | Class-Balanced | CE | Last block + FC | Last block + FC | 69.8 |
| NCM [3] | ResNet-152 | ImageNet | Class-Balanced | CE | Last block + FC | Classifier layer | 63.7 |
| cRT [3] | ResNet-152 | ImageNet | Class-Balanced | CE | Last block + FC | Classifier layer | 63.3 |
| $\tau$-normalized [3] | ResNet-152 | ImageNet | Class-Balanced | CE | Last block + FC | Classifier layer | **62.1** |
| OLTR* [5] | ResNet-152 | ImageNet | Class-Balanced | CE | Last block + FC | FC + memory | 64.8 |
| OLTR [5] | ResNet-152 | ImageNet | Class-Balanced | CE | Last block + FC | FC + memory | 64.1 |
| Ours | ResNet-152 | ImageNet | None | CE | Last block + FC | Classifier layer | 62.9 |
| Ours | ResNet-152 | ImageNet | None | Focal | Last block + FC | Classifier layer | <u>62.2</u> |

Table 4. Test top-1 errors (%) of different methods on iNaturalist 2018.

| Methods | NN | Initialization | Sampling | Loss | Stage-1 Trainable Variables | Stage-2 Trainable Variables | Results |
|---|---|---|---|---|---|---|---|
| Vanilla Model | ResNet-50 | No-pretrain | None | CE | All | - | 42.9 |
| Vanilla Model | ResNet-50 | ImageNet+iNat'17 | None | CE | All | - | 36.2 |
| LDAM [1] | ResNet-50 | No-pretrain | None | LDAM | All | - | 35.4 |
| LDAM-DRW [1] | ResNet-50 | No-pretrain | None | LDAM | All | - | **32.0** |
| CB [2] | ResNet-50 | ImageNet+iNat'17 | None | CE | All | - | 34.7 |
| CB [2] | ResNet-50 | No-pretrain | None | Focal | All | - | 38.9 |
| Joint [3] | ResNet-50 | No-pretrain | Class-Balanced | CE | All | All | 38.3 |
| NCM [3] | ResNet-50 | No-pretrain | Class-Balanced | CE | All | Classifier layer | 41.8 |
| cRT [3] | ResNet-50 | No-pretrain | Class-Balanced | CE | All | Classifier layer | 34.8 |
| $\tau$-normalized [3] | ResNet-50 | No-pretrain | Class-Balanced | CE | All | Classifier layer | 34.4 |
| Ours | ResNet-50 | ImageNet+iNat'17 | None | CE | All | All | <u>32.4</u> |
| Ours | ResNet-50 | ImageNet+iNat'17 | None | Focal | All | All | <u>32.3</u> |
| Vanilla Model | ResNet-101 | ImageNet+iNat'17 | None | CE | All | - | 34.3 |
| CB [2] | ResNet-101 | ImageNet+iNat'17 | None | CE | All | - | 32.7 |
| Ours | ResNet-101 | ImageNet+iNat'17 | None | CE | All | All | **31.5** |