Appendix to "No One Left Behind: Improving the Worst Categories in Long-Tailed Learning"

Yingxiao Du, Jianxin Wu*
State Key Laboratory for Novel Software Technology, Nanjing University
Nanjing, China, 210023
duyx@lamda.nju.edu.cn, wujx2001@gmail.com

1. More Details about the Datasets Used

The imbalance ratio of a dataset is defined as the number of training images of the most frequent class divide by the number of images of the least frequent class.

CIFAR100-LT. CIFAR100 [3] is a balanced dataset for image recognition, which has 50,000 training images and 10,000 test images from 100 categories. The CIFAR100-LT dataset used in our experiments are obtained by downsampling the original training set while keeping the test set unchanged. Following Zhou *et al*. [7], we use the exponential function $N_i = N_0 \times \mu^i$ to determine the number of training images for each category, where $N_0 = 500$. By varying μ^i, we are able to construct datasets with different imbalance ratios. In our experiments, we only use the one with imbalance ratio 100.

ImageNet-LT and Places-LT. ImageNet [2] and Places [8] are also two balanced dataset. Unlike CIFAR, these two datasets have larger scale and are more difficult. ImageNet-LT and Places-LT are their long-tailed version constructed by Liu *et al*. [4]. The number of training images for each class is determined using the Pareto distribution with a power value $\alpha = 6$. Their original test sets are left unchanged.

2. Implementation Details of the Fine-tuning Stage

CIFAR100-LT. For data augmentation, we randomly crop a 32×32 patch from the original image or its horizontal flip with 4 pixels padded on each side. We use the stochastic gradient descent (SGD) to optimize the network with momentum of 0.9 and weight decay of 5×10^{-4}. We train the model for 40 epochs. The initial learning rate is set to 5×10^{-2} and decrease it at the 10^{th} epoch by 0.2. We use a batch size of 128.

Methods	G-Mean	H-Mean	Lowest Recall
CE + GML	**36.59**	**31.26**	**6.00**
CE + CE (re-weighting)	35.30	27.55	4.00
CE + CE (re-sampling)	30.84	18.52	2.00

Table 1. Comparing with better baselines.

ImageNet-LT and Places-LT. For data augmentation, we resize the image by setting the shorter side to 256 and then take a random crop of 224×224 from it or its horizontal flip. Finally, color jittering is applied. We train our model for 40 epochs with a batch size of 512. We use stochastic gradient descent (SGD) with momentum of 0.9 and weight decay of 5×10^{-4}. The initial learning rate is set to 5×10^{-2} and is decreased at the 20^{th} epoch by 0.2. For Places-LT, when applied to MiSLAS [6], since MiSLAS is a two-stage method, we find it fairer to also apply their proposed label aware smoothing loss in the fine-tuning stage. So when computing the loss function, we combine two loss functions together as $\mathcal{L} = \lambda * \mathcal{L}_{\text{GML}} + \mathcal{L}_{\text{LAS}}$. During the experiment, we simply use $\lambda = 1$ without tuning it.

3. Additional Ablation Studies

We present some additional ablation studies here.

3.1. Better Baselines

Since our method requires re-training the classifier, the model is essentially trained longer. To better understand the performance improvement, here we conduct some experiments on CIFAR100-LT that serve as better baselines. Specifically, in the fine-tuning stage of our method, instead of using the proposed GML, we use either balanced cross-entropy or pure cross-entropy but combined with a balanced sampler. All the other settings remain unchanged. The results are shown in Tab. 1. As we can see, our proposed GML is better than them in terms of the harmonic mean of recall and the lowest recall value.

*J. Wu is the corresponding author. This research was partly supported by the National Natural Science Foundation of China under Grant 62276123 and Grant 61921006.

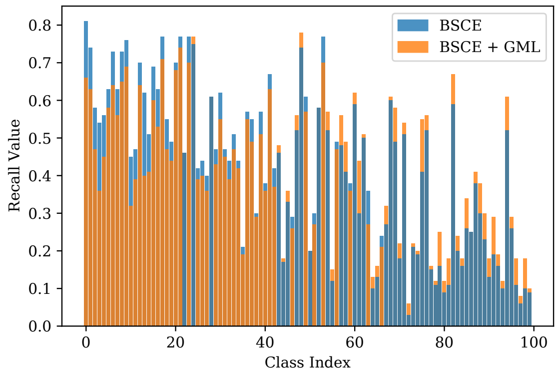

Figure 1. Bar plot of per-class recall on the imbalanced CIFAR100 (with imbalance ratio 100) before and after the fine-tuning when GML is applied to BSCE [5].

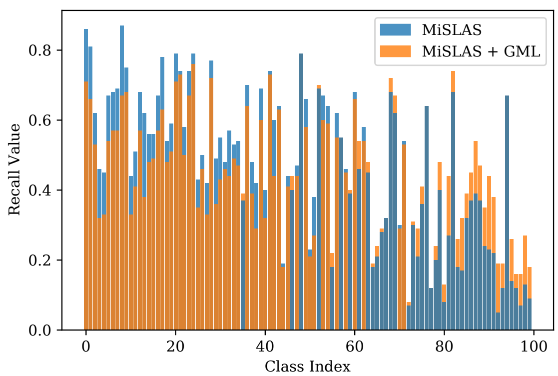

Figure 2. Bar plot of per-class recall on the imbalanced CIFAR100 (with imbalance ratio 100) before and after the fine-tuning when GML is applied to MiSLAS [6].

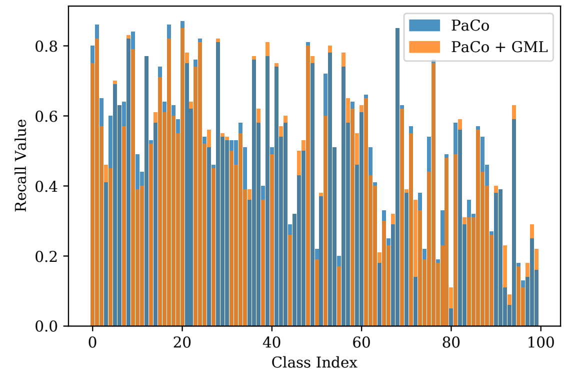

Figure 3. Bar plot of per-class recall on the imbalanced CIFAR100 (with imbalance ratio 100) before and after the fine-tuning when GML is applied to PaCo [1].

3.2. More Visualizations of the Per-Class Recall

Here we present more visualization results of the per-class recall when GML is applied to different methods. All experiments are conducted on CIFAR100-LT (with imbalance ratio 100). The results are shown in Fig. 1, Fig. 2 and Fig. 3.

References

[1] Jiequan Cui, Zhisheng Zhong, Shu Liu, Bei Yu, and Jiaya Jia. Parametric contrastive learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 715–724, October 2021. 2

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009. 1

[3] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. `https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf`. 1

[4] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X. Yu. Large-scale long-tailed recognition in an open world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2537–2546, June 2019. 1

[5] Jiawei Ren, Cunjun Yu, Shunan Sheng, Xiao Ma, Haiyu Zhao, Shuai Yi, and Hongsheng Li. Balanced meta-softmax for long-tailed visual recognition. In *Advances in Neural Information Processing Systems 33*, pages 4175–4186, December 2020. 2

[6] Zhisheng Zhong, Jiequan Cui, Shu Liu, and Jiaya Jia. Improving calibration for long-tailed recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16489–16498, June 2021. 1, 2

[7] Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9719–9728, June 2020. 1

[8] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6):1452–1464, 2018. 1