Frequency-Aware Self-Supervised Long-Tailed Learning
Supplementary Material

Table A. Evaluation on Places-LT with the standard setting (i.e., use of all labeled data for fine-tuning).

Method	All	*Rare*	*Medium*	*Frequent*	*Std*
SDCLR [15]	21.50	7.18	18.58	38.74	13.05
FASSL (Ours)	**22.89**	**7.98**	**20.35**	**40.34**	13.34

Table B. Performance of our proposed FASSL with/without alternate training on CIFAR100-LT.

Method	All	*Rare*	*Medium*	*Frequent*
FASSL (w/ alternate training)	55.06	52.58	**54.94**	57.68
FASSL (w/o alternate training)	**55.27**	**53.55**	54.52	**57.74**

A. Additional Experiments

A.1. Experiments on Places-LT

Places-LT is a long-tailed dataset sampled from Places [A]. It contains 365 categories with a total of $62,500$ images. The amount of data in each class ranges from $4,980$ to 5. In Table A, we show that SDCLR [15] achieved the accuracy of 21.50% while our FASSL reported **22.89%** on Places-LT and is therefore preferable.

A.2. Ablation Studies

Alternate Training. To address the long-tailed data learning problem without label supervision, we propose a *Frequency-Aware Self-Supervised Learning* (FASSL) scheme, which is composed of two learning stages: *Frequency-Aware Prototype Learning* and *Prototypical Re-balanced Self-Supervised Learning*. In Table B, we demonstrate the results of our FASSL with or without alternating between the above two learning stages. We see that alternate training would result in degraded performance. This is because alternate optimization tends to hinder the prototypes from describing long-tailed data distributions (and also increases the training time). Therefore, we choose not to alternate between the two stages.

Different Model Architectures. In Table C, we provide experimental results and show that deeper CNN models (e.g., ResNet-34) are not preferable on CIAFR100-LT due to possible overfitting problems. Thus, we choose to use ResNet-18 on CIAFR100-LT as existing works [15] did.

Table C. Performance of our proposed FASSL when using different model architectures on CIFAR100-LT.

Method	Model	All	*Rare*	*Medium*	*Frequent*
FASSL	ResNet-34	54.19	51.45	**54.88**	56.24
FASSL	ResNet-18	**55.27**	**53.55**	54.52	**57.74**

Table D. Performance of our proposed FASSL with different initialization on CIFAR10-LT.

Method	All	*Rare*
FASSL (w/o initialization)	76.11	71.87
SimCLR [5]	75.37	69.33
FASSL (init. from SimCLR)	76.42	72.70
SDCLR [15]	80.49	75.10
FASSL (init. from SDCLR)	**80.69**	**78.80**

Table E. Comparison with semi-supervsied learning works when using 30% labeled data on CIFAR100-LT.

Method	All
FixMatch w/ CReST+ [B]	42.0
FASSL (Ours)	**52.1**

Model Initialization. In Table D, we observe that if we train our model from scratch without any initialization, FASSL only achieves 71.87% on rare categories. This is because our FASSL performs *data-distribution-level* contrastive learning instead of *image-level* one to identify the imbalanced data distribution, and therefore image-level patterns/features may not be well captured. To address this issue, we choose to initialize our CNN model from image-level SSL methods, and we see that the rare-class accuracy would improve to 72.70% and 78.80% when initialized from SimCLR [5] and SDCLR [15], respectively. This demonstrates that when using any image-level SSL methods for initialization, our FASSL consistently improves the performance on long-tailed data.

A.3. Comparison with Semi-Supervised Learning Works

Since labeled data is also required in linear evaluation phase (i.e., finetuning a linear classifier), we also compare our method with semi-supervised works [B], as shown in Table E. By using the same amount (30%) of labeled data, our FASSL achieved the averaged accuracy of **52.1%** while [B] only reported 42.0% on CIFAR100-LT. Thus, the use of our scheme to properly weigh and regularize long-tailed data for SSL would be desirable.

References

[A] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6), 1452-1464.

[B] Wei, C., Sohn, K., Mellina, C., Yuille, A., & Yang, F. (2021). Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10857-10866).