Supplementary of SAFA: Sample-Adaptive Feature Augmentation for Long-Tailed Image Classification

Yan Hong[1], Jianfu Zhang*[1], Zhongyi Sun[2], and Ke Yan*[2]

[1] MoE Key Lab of Artificial Intelligence, Shanghai Jiao Tong University, China
yanhong.sjtu@gmail.com, c.sis@sjtu.edu.cn
[2] Tencent Youtu Lab, China
{zhongyisun, kerwinyan}@tencent.com

In this document, we provide additional material to support our main submission. In Section 1, we detail the split setting of datasets used in our paper. In Section 2, we describe the structure of our sample-adaptive generator including delta feature extractor E, sample-specific delta generator D, sample-adaptive feature generator G, and contrastive module Q. In Section 3, we report the comparison results between our SAFA and previous sample generation methods. In section 4, report additional experimental results on CIFAR-LT-10/CIFAR-LT-100 by integrating SAFA into different layers. In Section 5, we show t-SNE visualization of embeddings extracted from our ablated methods. In Section 6, we introduce the baselines used in our paper.

1 Datasets

In this section, we conclude the statistics of datasets in Table 1 and Table 2. The imbalance factor, maximum(*resp.,* minimum) number of images in classes, and categories of CIFAR-LT-10 (*resp.,* CIFAR-LT-100) [5] are listed in Table 1. In Table 2, we also report the imbalance factor, maximum (*resp.,* minimum) number of images in classes, categories, and the number of training images of ImageNet-LT [8] (*resp.,* Places-LT [16], and iNaturalist 2018 [11]).

Dataset	CIFAR-LT-10					CIFAR-LT-100				
Imbalance Factor ρ	**10**	**20**	**50**	**100**	**200**	**10**	**20**	**50**	**100**	**200**
Max. Number	5000	5000	5000	5000	5000	500	500	500	500	500
Min. Number	500	250	100	50	25	50	25	10	5	2
Category	10	10	10	10	10	100	100	100	100	100

Table 1. Statistics of CIFAR-LT-10 and CIFAR-LT-100 datasets. We present the maximum and minimum numbers of training images in the classes under different imbalance factor ρ.

* Corresponding authors.

Dataset	ImageNet-LT	Places-LT	iNaturalist 2018
Imbalance Factor $\boldsymbol{\rho}$	1280/5	4980/5	1000/2
Max. Number	1280	4980	1000
Min. Number	5	5	2
Category	1000	365	8142
Images Numbers	115,846	62,500	435,713

Table 2. Statistics of ImageNet-LT, Places-LT and iNaturalist 2018 datasets. We present the maximum and minimum numbers of training images in the classes, categories, and imbalance factor ρ.

2 Architecture

SAFA consists of a delta feature extractor E, a sample-specific delta generator D, a sample-adaptive feature generator G, and a contrastive module Q. The delta feature extractor E (*resp.,* sample-specific delta generator D, the sample-adaptive feature generator G) is composed of a Conv-BN-ReLU block, in which each block contains 1 convolutional layer with batch normalization and ReLU. Given feature $\boldsymbol{F} \in \mathbb{R}^{C\times W\times H}$ extracted from $N-1$-th layer in deep network, delta $\boldsymbol{\Delta}$ (*resp.,* sample-specific delta $\boldsymbol{\Delta}_t$) $\in \mathbb{R}^{C_\Delta\times W\times H}$, where $C_\Delta = C$. The contrastive module Q is also built upon Conv-BN-ReLU block followed an additional FC layer.

Dataset	CIFAR-LT-10		CIFAR-LT-100	
Imbalance factor	200	50	200	50
Delta-encoder [9]	29.93	23.76	63.51	54.91
Imaginary [13]	31.59	23.99	64.95	55.08
FTL [14]	31.87	23.56	65.12	55.24
CE-RSG [12]	29.56	20.25	62.94	54.44
CE-SAFA (Ours)	**25.11**	**18.86**	**61.34**	**52.31**

Table 3. Comparison results among our SAFA and other sample generation methods on CIFAR-LT-10 and CIFAR-LT-100 with imbalance factor $\rho = \{200, 50\}$. All of them are based on ResNet-32 combined with cross-entropy loss (CE) for a fair comparison.

3 Comparison with Previous Sample Generation Methods

SAFA is also compared to existing sample generation techniques [9,13,14,12]. As shown in Table 3, we report top-1 error of these methods on CIFAR-LT datasets with different imbalance factors. Our proposed SAFA outperforms earlier approaches by obvious margins, demonstrating that the proposed SAFA can overcome the shortcomings of previous generation methods and increase long-tail classification performance.

Dataset	CIFAR-LT-10		CIFAR-LT-100	
Imbalance factor	200	50	200	50
1st down-sampling	23.59	17.81	58.87	51.87
2nd down-sampling	**22.47**	**16.43**	**57.53**	**49.98**
3rd down-sampling (GAP)	23.16	17.63	58.12	51.31

Table 4. Comparison results of our SAFA integrated into different layers on CIFAR-LT-10 and CIFAR-LT-100 with imbalance factor $\rho = \{200, 50\}$. All of them are based on ResNet-32 combined with LDAM-DRW. Note that GAP represent global average pooling.

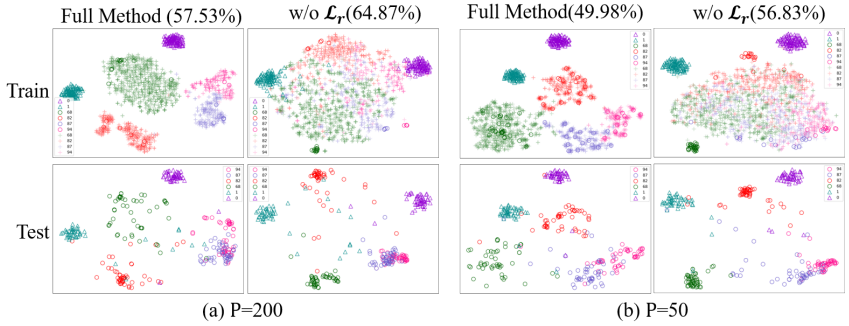

Fig. 1. Visualization comparison between our full method and ablated method without cycle reconstruction loss $\mathcal{L}_r$ on CIFAR-LT-100 dataset. The number in brackets denotes top-1 error. (a): comparison results with $\rho = 200$; (b): comparison results with $\rho = 50$. From top to down: visualization of real head-class features ($\triangle$), real tail-class features ($\circ$), and augmented tail-class features ($+$) on imbalanced train set, visualization of real head-class features ($\triangle$) and real tail-class features ($\circ$) on balanced test set

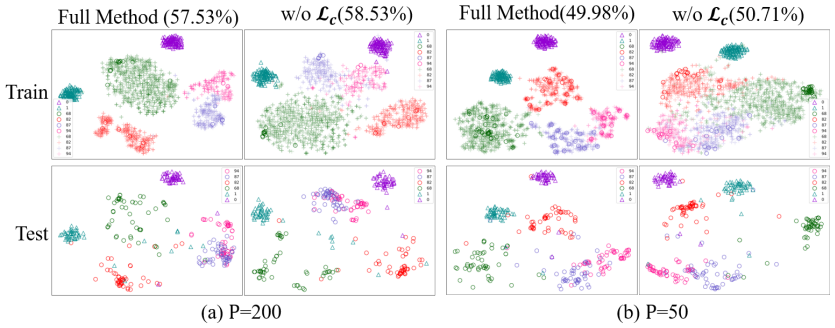

Fig. 2. Visualization comparison between our full method and ablated method without contrastive loss $\mathcal{L}_c$ on CIFAR-LT-100 dataset. The number in brackets denotes top-1 error. (a): comparison results with $\rho = 200$, (b): comparison results with $\rho = 50$. From top to down: visualization of real head-class features ($\triangle$), real tail-class features ($\circ$), and augmented tail-class features ($+$) on imbalanced train set, visualization of real head-class features ($\triangle$) and real tail-class features ($\circ$) on balanced test set

4 SAFA Works in Different Layers

In this section, we used SAFA in front of several ResNet-32 layers to explore which level of feature is best for creating new samples from our SAFA. The top-1 error of comparison results in Table 4 demonstrates that when SAFA is employed before the second-to-last down-sampling layer, the best results are obtained.

5 Analysis of Ablated Methods

In this section, we further analyze the impacts of each loss term in Eqn. (1) in main paper, and provide t-SNE comparison visualization among different ablated methods by removing specific loss item on CIFAR-LT-100 dataset with imbalance factor $\rho = \{200, 50\}$. We refer to our SAFA optimized with total loss as "Full Method", while ablated method by removing $\mathcal{L}_r$ (*resp.,* $\mathcal{L}_c$, $\mathcal{L}^h_{ms}$, and $\mathcal{L}^t_{ms}$) as "w/o $\mathcal{L}_r$" (*resp.,* "w/o $\mathcal{L}_c$", " w/o $\mathcal{L}^h_{ms}$", and "$\mathcal{L}^t_{ms}$").

Impact of Cycle Reconstruction Loss By removing cycle reconstruction loss $\mathcal{L}_r$ from total optimization functions, the visualization comparison between the ablated method "w/o $\mathcal{L}_r$" and "Full Method" are shown in Fig. 1. It can be seen that the augmented tail-class features from ablated method "w/o $\mathcal{L}_r$" are scattered in large feature space, even far from the real tail-class feature. Obviously, for ablated method "w/o $\mathcal{L}_r$", the distribution gap between real tail-class features and the augmented tail-class features exists, failing to improve generalization on test set. It indicates that the cycle reconstruction loss is the basis of our sample-specific augmentation.

Impact of Contrastive Loss To investigate the impact of our contrastive loss $\mathcal{L}_c$, we show comparison results between ablated method "w/o $\mathcal{L}_c$" and "Full

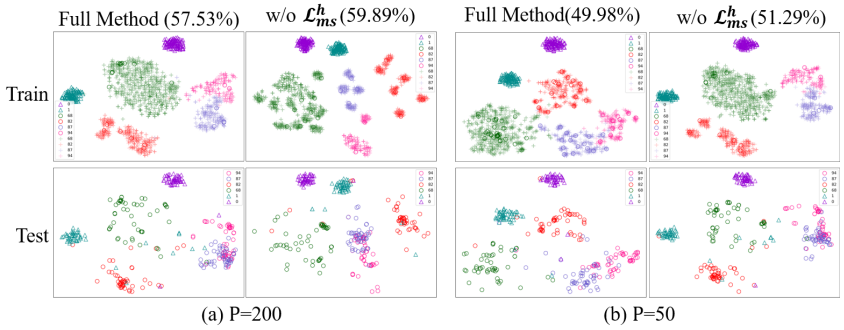

Fig. 3. Visualization comparison between our full method and ablated method without head mode seeking loss $\mathcal{L}_{ms}^{h}$ on CIFAR-LT-100 dataset. The number in brackets denotes top-1 error.(a): comparison results with $\rho = 200$, (b): comparison results with $\rho = 50$. From top to down: visualization of real head-class features ($\triangle$), real tail-class features ($\circ$), and augmented tail-class features ($+$) on imbalanced train set, visualization of real head-class features ($\triangle$) and real tail-class features ($\circ$) on balanced test set

Method" in Figure 2. By comparison, the overlap among augmented features from different tail class in ablated method "w/o $\mathcal{L}_c$" indicates that removing contrastive loss may lead class information from head class to augmented tail class, results in class confusion among augmented features.

Impact of Head Mode Seeking Loss As is analyzed in Section 4.3 in main paper, applying the head mode seeking loss in extremely imbalanced setting ($\rho = 200$) can improve the diversity of generated tail-class features. In Figure 3 (a) where imbalance factor $\rho = 200$, the tail-class features generated from "w/o $\mathcal{L}_{ms}^{h}$" gather in a density feature space near to real tail-class feature, and the diversity of generated features is limited, while the diversity of generated features in Figure 3 (b) is less compromised in relative balanced setting ($\rho = 50$).

Impact of Tail Mode Seeking Loss To investigate the impact of tail mode seeking loss $\mathcal{L}_{ms}^{t}$, we conduct experiment on CIFAR-LT-100 with imbalance factor ρ ranging from $\{200, 50\}$, and show visualization results in Figure 4. By comparison, we can see that the diversity of tail-class features generated from ablated method "w/o $\mathcal{L}_{ms}^{t}$" in $\rho = 50$ setting is obviously compromised.

6 Baselines

In this section, we briefly introduce our selected baselines in this paper.

Cross-entropy Training is the baseline method in long-tailed visual recognition, which trains CNNs with standard softmax with cross-entropy loss, which is denoted as "CE loss" in this paper.

Class-Level Re-Weighting Methods This type of method assigns weights to training examples in class level, which includes Class-Balanced Cross-Entropy loss [2] referred as "CB-CE loss", and LDAM-DRW [1]. Class-balanced loss proposes effective number to measure the sample size of each class and the class-level

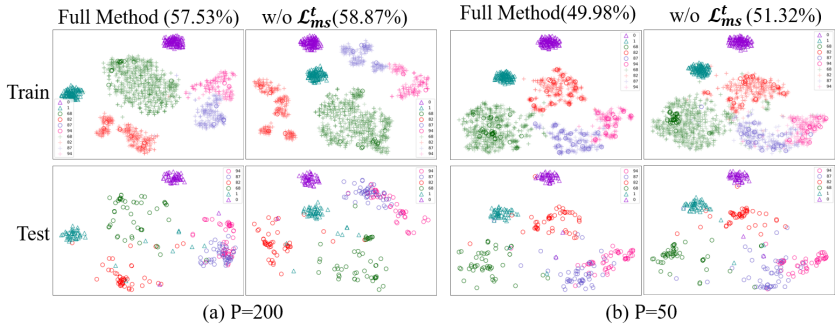

Fig. 4. Visualization comparison between our full method and ablated method without tail mode seeking loss $\mathcal{L}_{ms}^{t}$ on CIFAR-LT-100 dataset. The number in brackets denotes top-1 error. (a): comparison results with $\rho = 200$, (b): comparison results with $\rho = 50$. From top to down: visualization of real head-class features ($\triangle$), real tail-class features ($\circ$), and augmented tail-class features ($+$) on imbalanced train set, visualization of real head-class features ($\triangle$) and real tail-class features ($\circ$) on balanced test set

weights. Class-balanced focal loss denoted as "CB Focal loss" [2] and "CB-CE" [2] refer to applying class-balanced loss on focal loss and cross-entropy loss, respectively. LDAM-DRW allocates label-aware margins to the examples based on the label distribution, and adopts deferred re-weighting strategy for better performance on tail classes.

Sample-Level Re-Weighting Methods Re-weighting method assign weights to samples according to the instance characteristic [7,10,6]. For example, focal loss [6] determine the weights for samples based on the sample difficulty. L2RW [7] is designed to assign weights to examples sample-wisely based on the gradient. Meta-weight [10] assign weights to examples sample-wisely.

Two-stage Methods This type of methods [4,1,17,3] adopt two-stage learning to learn representation firstly, and then finetune the classifier learning. CB finetuning[3] finetune classifier on the basis of fixed backbone. Differently, BBN [17] unifies the two-stage learning with a curriculum learning strategy. LDAM-DRW-SSP [4] applies reweighting technique into the second-stage for classifier learning. LDAM-DRS [1] leverages resampling method to reshape the decision boundary of classifier in the second stage.

Augmentation-based Methods These methods [15,9,12] adopt various augmentation techniques to augment tail class to balance dataset. In fact, knowledge from head class is transferred to tail classes in Delta-encoder [9] and RSG[12]. As mentioned in Sec 1 in main paper, Delta-encoder [9] takes two-stage method to extract intra-class variance from head class, and then apply those variance to tail classes without end-to-end training. RSG [12] relies on estimation of class centers to provide variance which are combined with tail-class samples to produce augmented samples.

References

1. Cao, K., Wei, C., Gaidon, A., Arechiga, N., Ma, T.: Learning imbalanced datasets with label-distribution-aware margin loss. NeurIPS (2019)
2. Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S.: Class-balanced loss based on effective number of samples. In: CVPR (2019)
3. Cui, Y., Song, Y., Sun, C., Howard, A., Belongie, S.: Large scale fine-grained categorization and domain-specific transfer learning. In: CVPR (2018)
4. Jamal, M.A., Brown, M., Yang, M.H., Wang, L., Gong, B.: Rethinking class-balanced methods for long-tailed visual recognition from a domain adaptation perspective. In: CVPR (2020)
5. Krizhevsky, A., et al.: Learning multiple layers of features from tiny images (2009)
6. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: ICCV (2017)
7. Ren, M., Zeng, W., Yang, B., Urtasun, R.: Learning to reweight examples for robust deep learning. In: ICML (2018)
8. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet: large scale visual recognition challenge. IJCV **115**(3), 211–252 (2015)
9. Schwartz, E., Karlinsky, L., Shtok, J., Harary, S., Marder, M., Kumar, A., Feris, R., Giryes, R., Bronstein, A.: Delta-encoder: an effective sample synthesis method for few-shot object recognition. NeurIPS (2018)
10. Shu, J., Xie, Q., Yi, L., Zhao, Q., Zhou, S., Xu, Z., Meng, D.: Meta-weight-net: Learning an explicit mapping for sample weighting. NeurIPS (2019)
11. Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., Belongie, S.: The inaturalist species classification and detection dataset. In: CVPR (2018)
12. Wang, J., Lukasiewicz, T., Hu, X., Cai, J., Xu, Z.: Rsg: A simple but effective module for learning imbalanced datasets. In: CVPR (2021)
13. Wang, Y.X., Girshick, R., Hebert, M., Hariharan, B.: Low-shot learning from imaginary data. In: CVPR (2018)
14. Yin, X., Yu, X., Sohn, K., Liu, X., Chandraker, M.: Feature transfer learning for deep face recognition with under-represented data. arXiv preprint arXiv:1803.09014 (2018)
15. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: ICLR (2018)
16. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 million image database for scene recognition. PAMI **40**(6), 1452–1464 (2017)
17. Zhou, B., Cui, Q., Wei, X.S., Chen, Z.M.: Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In: CVPR (2020)