# Supplementary Material for How Re-sampling Helps for Long-Tail Learning?

**Anonymous Author(s)**
Affiliation
Address
email

## Contents



## A Training Procedure

The traing procedure of *context-shift augmentation* is summarized in Algorithm 1.

## B Implementation Details for *context-shift augmentation*

For experiments on CIFAR10-LT and CIFAR100-LT, we use ResNet-32 as the backbone network and train it using standard SGD with a momentum of 0.9, a weight decay of $2 \times 10^{-4}$, a batch size of 128. The model is trained for 200 epochs. The initial learning rate is set to 0.2 and is annealed by a factor of 10 at 160 and 180 epochs. We train each model with 1 NVIDIA GeForce RTX 3090.

For experiments on ImageNet-LT, we implement the proposed method on ResNet-10 and ResNet-50. We use standard SGD with a momentum of 0.9, a weight decay of $5 \times 10^{-4}$, and a batch size of 256 to train the whole model for a total of 90 epochs. We use the cosine learning rate decay with an initial learning rate of 0.2. We train each model with 2 NVIDIA Tesla V100 GPUs.

In all experiments, we first warm-up the uniform module for 10 epochs and then train the uniform module and the balanced re-sampling module simultaneously for the rest epochs. For the uniform module, we follow the simple data augmentation used in [1] with only random crop and horizontal flips. For the re-sampling module, we use the proposed context-shift augmentation method. We apply the trick proposed by [2] to disable the augmentation in the balanced re-sampling module at the

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

**Algorithm 1** Training procedure of context-shift augmentation

**Input:** training data $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$; context memory bank $Q$, maximum volume size $V$; model parameters $\phi$, $f^u$, $f^b$; loss functions $\ell^u$, $\ell^b$;

**Procedure:**

Initialize model parameters $\phi$, $f^u$, $f^b$;
Re-sampling a class-balanced dataset $\widetilde{\mathcal{D}} = \{(\tilde{\boldsymbol{x}}_i, \tilde{y}_i)\}_{i=1}^N$;
Empty memory bank $Q$;
**for** epoch $= 1, \ldots, T$ **do**
  **repeat**
    Draw a mini-batch $(\boldsymbol{x}_i, y_i)_{i=1}^B$ from $\mathcal{D}$;
    Draw a mini-batch $(\tilde{\boldsymbol{x}}_i, \tilde{y}_i)_{i=1}^B$ from $\widetilde{\mathcal{D}}$;
    *// uniform module*
    **for** $i = 1, \ldots, B$ **do**
      Calculate $\boldsymbol{z}_i^u = f^u(\phi(\boldsymbol{x}_i))$ and $\mathcal{L}_i^u = \ell^u(\boldsymbol{z}_i^u, y_i)$;
      **if** $p(y = y_i | \boldsymbol{x}_i, \phi, f^u) \geq \delta$ **then**
        Calculate background mask $\boldsymbol{M}_i$ of $\boldsymbol{x}_i$;
        Push $(\boldsymbol{x}_i, \boldsymbol{M}_i)$ into $Q$;
      **end if**
    **end for**
    Calculate $\mathcal{L}^u = \frac{1}{B}\sum_{i=1}^B \mathcal{L}_i^u$;
    *// balanced re-sampling module*
    **if** Size of $Q$ reaches $V$ **then**
      Get backgrounds $(\check{\boldsymbol{x}}_i, \boldsymbol{M}_i)_{i=1}^B$ from $Q$;
      $\lambda \sim \mathrm{Uniform}(0, 1)$;
      **for** $i = 1, \ldots, B$ **do**
        $\tilde{\boldsymbol{x}}_i = \lambda \boldsymbol{M}_i \odot \check{\boldsymbol{x}}_i + (1 - \lambda \boldsymbol{M}_i) \odot \tilde{\boldsymbol{x}}_i$;
        Calculate $\boldsymbol{z}_i^b = f^b(\phi(\tilde{\boldsymbol{x}}_i))$ and $\mathcal{L}_i^b = \ell^b(\boldsymbol{z}_i^b, \tilde{y}_i)$;
      **end for**
      Calculate $\mathcal{L}^b = \frac{1}{B}\sum_{i=1}^B \mathcal{L}_i^b$;
    **else**
      Assign $\mathcal{L}^b = 0$;
    **end if**
    *// total objective function*
    Calculate $\mathcal{L} = \mathcal{L}^u + \mathcal{L}^b$;
    Update model parameters $\phi$, $f^u$, $f^b$ with $\mathcal{L}$;
  **until** all training data are traversed.
**end for**

last 3 epochs to obtain further improvements, which is also applied in other baseline methods [3, 4]. We set the threshold $\delta$ to 0.8.

# C Additional Experimental Results

## C.1 Effects of the context bank

The context bank $Q$ is a novel component of *context-shift augmentation* which receives diverse contexts from the uniform module, and provides them to augment the data in the re-sampling module. To verify the effectiveness of the context bank, we remove it from the framework and train the model on CIFAR100-LT with imbalance ratio of 100. The results are reported in Table 1.

The results show that without context bank $Q$, the performance decrease by a large margin. The performance degradation mostly comes from the medium-shot classes and the few-shot classes, which indicates that the context bank can significantly improve the generalization of tail classes.

Moreover, we study the effect of different variants in the context bank. First, we study the threshold $\delta$ for sample selecting and report the results in Table 2. On the one hand, if $\delta$ is too high, the selected samples will be very few. On the other hand, if $\delta$ is too small, the selected samples might be not well

Table 1: Ablation study on the context bank $Q$.

| | All | Many | Med. | Few |
|---|---|---|---|---|
| Ours | 45.8 | 64.3 | 49.7 | 18.2 |
| Ours w/o $Q$ | 41.2 (-4.6) | 65.1 (+0.8) | 41.9 (-7.8) | 10.7 (-7.5) |

learned. Nevertheless, as the training process progresses, most samples will fit well, so our method is not sensitive to $\delta$. We set $\delta = 0.8$ considering its best performance.

Table 2: Influence of the threshold $\delta$.

| $\delta$ | 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Accuracy | 45.47 | 45.37 | 45.55 | 44.83 | 45.52 | 45.59 | 45.42 | 45.08 | **45.83** | 44.93 |

Second, we study the influence of the volume size $V$ of the context bank $Q$ and report the results in Table 3. Since the bank $Q$ is a first-in-first-out queue, the latest incoming backgrounds are more convincing. When the volume size is too large, the bank might contain more past samples. Besides, a larger size of $V$ would bring more memory overhead. So we set the volume size $V$ equal to the mini-batch size $B$ in our method.

Table 3: Influence of the bank volume size (compared with the mini-batch size $B$).

| Volume | $\times 1$ | $\times 2$ | $\times 4$ | $\times 8$ | $\times 16$ | $\times 32$ | $\times 64$ |
|---|---|---|---|---|---|---|---|
| Accuracy | **45.83** | 45.60 | 45.57 | 45.32 | 45.55 | 45.37 | 45.26 |

### C.2 Influence of augmentation variants

We use a variant $\lambda \sim \mathrm{Uniform}(0, 1)$ for generating novel samples. The value of $\lambda$ can result in different proportions of foreground and background in the novel sample. Also, the size of the sampling space affects the diversity of the novel images. To explore the effect of $\lambda$, we try $\lambda \sim \mathrm{Uniform}(a, b)$ and $\lambda \sim \mathrm{Beta}(a, b)$ to train *context-shift augmentation* on CIFAR100-LT with imbalance ratio 100 and report the results in Figure 1.

First, when $\lambda$ is close to 0, the background merely takes effect, and the performance decreases a lot. Second, when $\lambda = 1$, the background image might cover the important content in the foreground image. Also, the diversity of new samples is limited. Although the performance is better than that of $\lambda = 0$, it is still unsatisfactory. Overall, choosing $\lambda \sim \mathrm{Uniform}(0, 1)$ or $\lambda \sim \mathrm{Beta}(1, 1)$ lead to the best performance.

### C.3 Comparison between different modules

In our framework, the uniform sampling module is only enabled in the training phase. While in the inference phase, we use the balanced re-sampling module to predict unseen instances. To verify the superiority of the re-sampling module, we compare the performance of these two modules as well as their ensemble. We report the results in Table 4. The results show that the re-sampling module is superior to the uniform module, and even achieves higher accuracy than the ensembled results. This also indicates the superiority of the proposed context-shift re-sampling method.

Moreover, we study the influence of different balance ratios on our re-sampling module and compare it with the vanilla re-sampling method. We report the results in Figure 2. For vanilla re-sampling, adopting a more balanced re-sampling would yield more severe performance degradation. In contrast, our method achieves higher performance through class-balanced re-sampling.

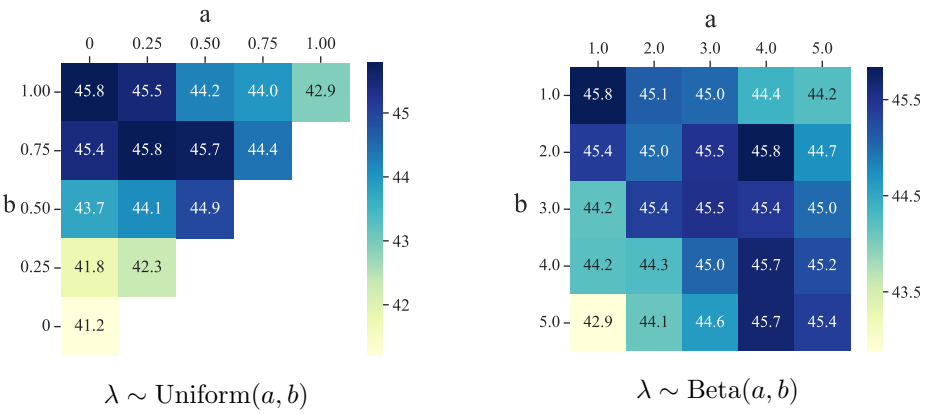


Figure 1: Influence of variants $\lambda$.

Table 4: Comparison between the uniform module and the re-sampling module in *context-shift augmentation*.

| | All | Many | Med. | Few |
|---|---|---|---|---|
| Uniform module | 39.4 | **68.3** | 37.3 | 6.1 |
| Re-sampling module | **45.8** | 64.3 | **49.7** | **18.2** |
| Ensemble results | 43.0 | 67.5 | 44.1 | 11.4 |

### C.4 The influence of Grad-CAM

The Grad-CAM [5] is utilized to extract background in previous works such as open-set learning and adversarial learning [6, 7]. Also, we use Grad-CAM in the *context-shift augmentation* to generate diverse backgrounds for tail-class data. We compare Grad-CAM with CAM [8]. The results shown in Table 5 demonstrate that Grad-CAM is superior to CAM when applied to our method.

Table 5: Comparison between CAM and Grad-CAM in *context-shift augmentation*

| | All | Many | Med. | Few |
|---|---|---|---|---|
| Ours w/ CAM | 45.1 | 63.8 | 48.1 | 18.0 |
| Ours w/ Grad-CAM | **45.8** | 64.3 | **49.7** | **18.2** |

### C.5 Combination with self-supervised learning

It is interesting to combine self-supervised methods with *context-shift augmentation*. Inspired by this, we follow the self-supervised + fine-tune method, i.e., SimSiam+rwSAM in [9] and conduct extensive experiments on CIFAR10-LT dataset. The results are shown in Table 6.

Note that in [9], the models are pre-trained with the long-tailed dataset, while fine-tuned with the balanced in-domain dataset. However, it is hard to achieve a balanced in-domain version of a long-tailed dataset in real-world scenarios. So we use the long-tailed dataset with balancing method including class-balanced re-sampling (CB-RS), and re-sampling with *context-shift augmentation*. The results show that our method is superior to CB-RS.

### C.6 Combination with the logit adjustment

Since our work aims to study the effectiveness of re-sampling in long-tail learning, we use the Class-Balanced Re-Sampling (CB-RS) in our method. We consider combining our method with other re-balancing methods, such as Logit Adjustment (LA) [10]. Specifically, we change the class-balanced re-sampling to the uniform sampling while adopting *context-shift augmentation*. Moreover, we consider combining class-balanced re-sampling and LA simultaneously. The comparison results

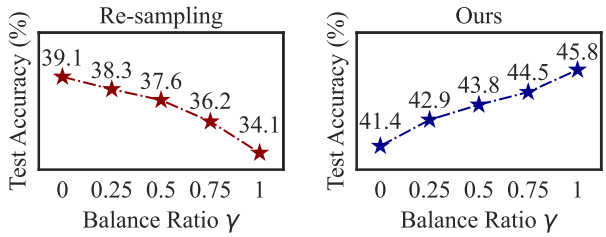


Figure 2: Comparison of re-sampling and our method under different balance ratios $\gamma$.

Table 6: Combing self-supervised method SimSiam+rwSAM with different re-balancing methods

| Pre-train model | Fine-tune method | Accuracy (%) |
|---|---|---|
| SimSiam+rwSAM | CE | 69.5 |
| SimSiam+rwSAM | CB-RS | 72.8 |
| SimSiam+rwSAM | Ours | **75.5** |

are shown in Table 7. The results show that our method can be combined with logit adjustment to yield a more higher performance. However, by applying the balanced loss and the balanced sampling, the model puts much focus on tail classes and results in a deterioration of overall accuracy.

Table 7: Combination with Logit Adjustment (LA)

| | All | Many | Med. | Few |
|---|---|---|---|---|
| w/ CB-RS | 45.8 | **64.3** | 49.7 | 18.2 |
| w/ LA | **47.2** | 62.4 | 48.8 | 26.1 |
| w/ both | 45.0 | 48.7 | **51.0** | **31.4** |

# References


[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016.

[2] Zhuoxun He, Lingxi Xie, Xin Chen, Ya Zhang, Yanfeng Wang, and Qi Tian. Data augmentation revisited: Rethinking the distribution gap between clean and augmented data. *CoRR*, abs/1909.09148, 2019.

[3] Yongshun Zhang, Xiu-Shen Wei, Boyan Zhou, and Jianxin Wu. Bag of tricks for long-tailed visual recognition with deep convolutional neural networks. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Event, February 2-9, 2021*, pages 3447–3455, 2021.

[4] Seulki Park, Youngkyu Hong, Byeongho Heo, Sangdoo Yun, and Jin Young Choi. The majority can help the minority: Context-rich minority oversampling for long-tailed classification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pages 6877–6886, 2022.

[5] Bolei Zhou, Aditya Khosla, Àgata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *CVPR*, pages 2921–2929. IEEE Computer Society, 2016.

[6] Jie-Jing Shao, Xiao-Wen Yang, and Lan-Zhe Guo. Open-set learning under covariate shift. *Machine Learning*, pages 1–17, 2022.

[7] Qibing Ren, Yiting Chen, Yichuan Mo, Qitian Wu, and Junchi Yan. DICE: domain-attack invariant causal learning for improved data privacy protection and adversarial robustness. In *KDD*, pages 1483–1492. ACM, 2022.

[8] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 618–626, 2017.

[9] Hong Liu, Jeff Z. HaoChen, Adrien Gaidon, and Tengyu Ma. Self-supervised learning is more robust to dataset imbalance. In *ICLR*. OpenReview.net, 2022.

[10] Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. Long-tail learning via logit adjustment. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.