Supplementary Material for
Towards Realistic Long-Tailed Semi-Supervised Learning: Consistency Is All You Need

1. More Experimental Details

Our experiments are run on NVIDIA GeForce RTX 3090. In the implementation of the proposed method, we use PyTorch 1.12.0 and Torchvision 0.13.0.

Datasets. In our experiments, we compare the proposed algorithm ACR with existing state-of-the-art methods on several datasets under various unlabeled data class distributions, e.g., *consistent*, *uniform*, and *reversed*.

- **CIFAR-10-LT and CIFAR-100-LT**: There are 5000 and 500 samples in CIFAR-100 and CIFAR-100 originally, with the image size of 32×32. The entire CIFAR dataset can be downloaded from https://www.cs.toronto.edu/kriz/cifar.html. We need to manually construct the imbalanced dataset according to N_1, M_1 (M_C for *reversed*), and the imbalance ratio as we describe in our paper. We calculate the top-1 accuracy for measuring the performance of each method.

- **STL10-LT**: STL-10 includes 5000 labeled samples and 100k samples without labels initially, so ground-truth labels of unlabeled data in STL-10 are unknown. We train the model with a long-tailed labeled dataset constructed artificially and the whole $100k$ unlabeled data. The image size of each sample is 96×96, and the dataset can be downloaded at https://cs.stanford.edu/acoates/stl10/.

- **ImageNet-127**: ImageNet127 is a naturally long-tailed dataset. The dataset can be downloaded from https://image-net.org/download-images. Following CoSSL [1], we use the image size of 32 × 32 and 64 × 64 for training.

2. Pseudo-code for Our ACR Algorithm

Algorithm 1 summarizes the whole framework of the proposed ACR. *We also include the source code for ACR in the supplementary*.

Algorithm 1 Adaptive Consistency Regularizer (ACR)

Input: labeled dataset $\mathcal{D}^l$ and unlabeled dataset $\mathcal{D}^u$, standard classifier f and balanced classifier $\tilde{f}$, class prior of labeled dataset π, estimated unlabeled dataset class distribution π_{est}, number of iterations in each epoch K, scaling parameter τ, three anchor distributions.
Require: Strong augmentations $\mathcal{A}$.
Calculate the distance between π_{est} and three anchor distributions and obtain the adaptive $\tau(t)$ according to Equation (7)
for $k = 1$ **to** K **do**
  $\{(x_i^{(l)}, y_i^{(l)})\}_{i=0}^{B-1} \leftarrow$ Sample a batch of labeled data
  $\{x_j^{(u)}\}_{j=0}^{B-1} \leftarrow$ Sample a batch of unlabeled data
  $q(x_j^{(u)}) \leftarrow \arg\max f\left(x_j^{(u)}\right) - \tau(t) \cdot \log \pi$
  $\tilde{q}(x_j^{(u)}) \leftarrow \arg\max f\left(x_j^{(u)}\right) - \tau \cdot \log \pi$
  Update estimated class distribution π_{est} according to pseudo-labels $\tilde{f}(x_j^{(u)})$
  Generate sample masks M and $\tilde{M}$ for standard and balanced branch respectively based on Equation (8)
  $\mathcal{L}_{\text{labeled}} \leftarrow \sum_{i=0}^{B-1} \ell(f(x_i^{(l)}), y_i^{(l)})$
  $\mathcal{L}_{\text{b-labeled}} \leftarrow \sum_{i=0}^{B-1} \ell(\tilde{f}(x_i^{(l)}), y_i^{(l)})$
  $\mathcal{L}_{\text{con}} \leftarrow \sum_{j=0}^{B-1} M(x_j^{(u)})\ell\left(f(\mathcal{A}(x_j^{(u)})), q(x_j^{(u)})\right)$
  $\mathcal{L}_{\text{b-con}} \leftarrow \sum_{j=0}^{B-1} \tilde{M}(x_j^{(u)})\ell\left(\tilde{f}(\mathcal{A}(x_j^{(u)})), \tilde{q}(x_j^{(u)})\right)$
  $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{labeled}} + \mathcal{L}_{\text{con}} + \mathcal{L}_{\text{b-labeled}} + \mathcal{L}_{\text{b-con}}$
  Update f and $\tilde{f}$ based on $\nabla \mathcal{L}_{\text{total}}$ using SGD
end for

3. Additional Experimental Results

In this section, we report more results to show the effectiveness of ACR and detailed analysis to help better understand the approach.

More results compared with DASO. To further illustrate the superior performance of our method in the case of *uniform* and *reversed* class distributions of unlabeled dataset, we combine our ACR with DASO w/ LA and DASO w/ ABC on three datasets. The results are reported in Ta-

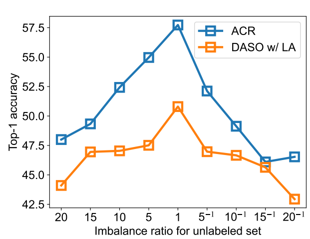
Algorithm	CIFAR10-LT($\gamma_l \neq \gamma_u$)				STL10-LT($\gamma_u = N/A$)			
	$\gamma_u = 1$(uniform)		$\gamma_u = 1/100$(reversed)		$\gamma_l = 10$		$\gamma_l = 20$	
	$N_1 = 500$ $M_1 = 4000$	$N_1 = 1500$ $M_1 = 3000$	$N_1 = 500$ $M_1 = 4000$	$N_1 = 1500$ $M_1 = 3000$	$N_1 = 150$ $M_1 = 100k$	$N_1 = 450$ $M_1 = 100k$	$N_1 = 150$ $M_1 = 100k$	$N_1 = 450$ $M_1 = 100k$
FixMatch [5]	73.0 ±3.81	81.5 ±1.15	62.5 ±0.94	71.8 ±1.70	56.1 ±2.32	72.4 ±0.71	47.6 ±4.87	64.0 ±2.27
w/ DASO [4]	86.6 ±0.84	88.8 ±0.59	71.0 ±0.95	80.3 ±0.65	70.0 ±1.19	78.4 ±0.80	65.7 ±1.78	75.3 ±0.44
w/ LA [3]+DASO	84.6 ±2.04	86.8 ±0.76	72.6 ±0.38	78.5 ±1.31	72.7 ±1.45	79.7 ±0.44	66.8 ±0.62	75.8 ±0.50
w/ ABC [2]+DASO	85.2 ±1.56	88.4 ±0.82	70.1 ±1.25	79.8 ±0.21	71.8 ±1.17	78.4 ±0.58	67.3 ±2.06	75.9 ±0.43
w/ ACR (ours)	**92.1** ±0.18	**93.5** ±0.11	**85.0** ±0.09	**89.5** ±0.17	**77.1** ±0.24	**83.0** ±0.32	**75.1** ±0.70	**81.5** ±0.25

Table 1. Test accuracy of DASO combined with LA or ABC and our proposed ACR under *uniform* and *reversed* class distributions, i.e., $\gamma_l \neq \gamma_u$, on CIFAR10-LT and STL10-LT datasets. The γ_l is fixed to 100 for CIFAR10-LT, while it is set to 10 and 20 for STL10-LT dataset. The best results are in **bold**.

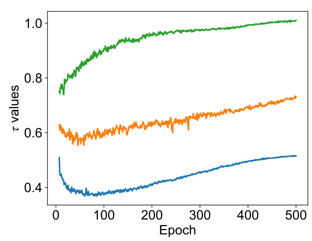
Algorithm	CIFAR100-LT($\gamma_l \neq \gamma_u$)			
	$\gamma_u = 1$(uniform)		$\gamma_u = 1/10$(reversed)	
	$N_1 = 50$ $M_1 = 400$	$N_1 = 150$ $M_1 = 300$	$N_1 = 50$ $M_1 = 400$	$N_1 = 150$ $M_1 = 300$
FixMatch [5]	45.5 ±0.71	58.1 ±0.72	44.2 ±0.43	57.3 ±0.19
w/ DASO [4]	53.9 ±0.66	61.8 ±0.98	51.0 ±0.19	60.0 ±0.31
w/ LA [3]+DASO	54.7 ±0.40	62.4 ±1.06	51.1 ±0.12	60.5 ±0.23
w/ ABC [2]+DASO	53.4 ±0.53	62.4 ±0.61	51.2 ±0.19	60.8 ±0.39
w/ ACR (ours)	**66.0** ±0.25	**73.4** ±0.22	**57.0** ±0.46	**67.6** ±0.12

Table 2. Test accuracy on CIFAR100-LT dataset under *uniform* and *reversed* class distributions. DASO combines LA and ABC to enhance its re-balancing ability. The best results are in **bold**.

ble 1 and Table 2. As we can see, ACR significantly outperforms competing methods with an average 7.9% performance gain, indicating the strong performance of ACR.



(a) More settings

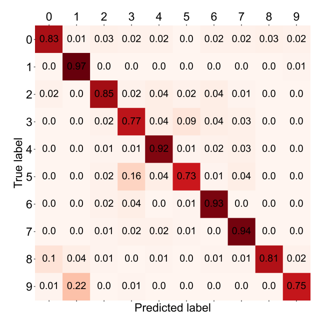


(b) τ values

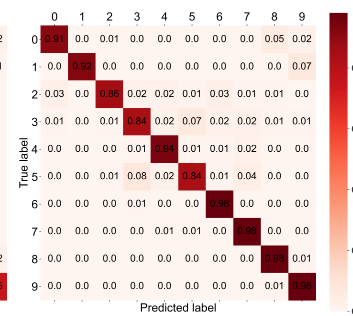
Figure 1. (1a): More realistic LTSSL settings on CIFAR100-LT. (1b): The adaptive τ values when training CIFAR100-LT.

More settings on CIFAR100-LT. We evaluate our proposed algorithm ACR under more types of class distributions of unlabeled data on CIFAR100-LT dataset. The result is illustrated in Figure 1a. In this experiment, we fix $\gamma_l = 20$ and vary γ_u from *consistent* to *reversed* by a step size of 5. We set $N_1 = 50$ and $M_1 = 400$ ($M_C = 400$ for *reversed*), and compare the performance with DASO w/ LA. It can be seen that ACR outperforms DASO w/ LA in all settings, showing the superior performance of ACR when dealing with various unlabeled class distributions.

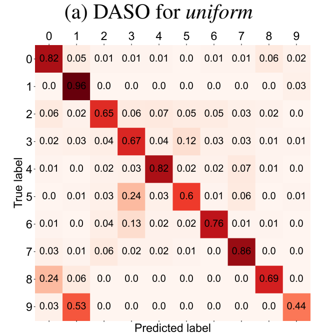
Adaptive τ values on CIFAR100-LT. We investigate the adaptive change of τ values on CIFAR100-LT dataset in Figure 1b. We can see that the scaling parameter τ increases its value from *consistent* to *reversed* settings, which coincides with observation on CIFAR10-LT illustrated in the main paper.



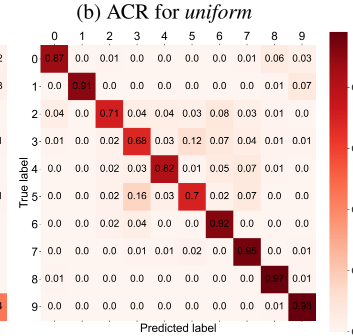
(a) DASO for *uniform*



(b) ACR for *uniform*



(c) DASO for *reversed*



(d) ACR for *reversed*

Figure 2. The confusion matrices of the test set for DASO and ACR on CIFAR-10-LT dataset in *uniform* and *reversed* settings.

Confusion matrices on test set. We compare ACR and DASO [4] under *uniform* and *reversed* unlabeled data class distributions by calculating their confusion matrices in Figure 2. We have two key observations: 1) ACR achieves higher overall accuracy for classes and 2) ACR significantly improves the minority class accuracy. On the contrary, DASO produces biased predictions toward the majority class.

Precision and Recall of pseudo-labels. In this experiment,

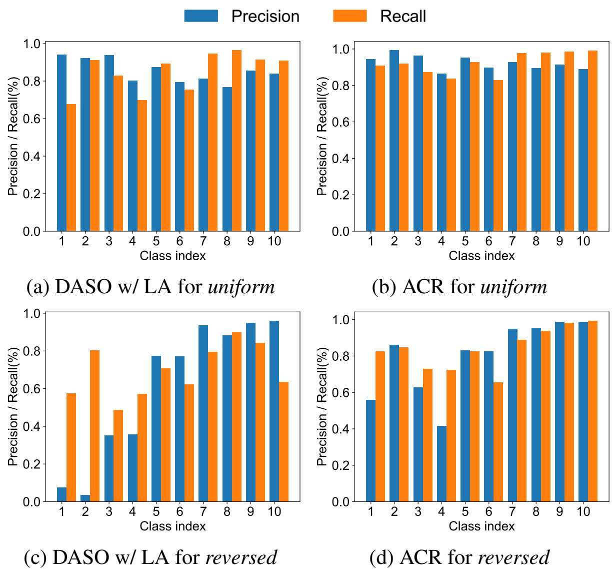


Figure 3. The precision and recall of pseudo-labels for DASO and ACR on CIFAR-10-LT dataset in *uniform* and *reversed* settings.

we investigate the quality of pseudo-labels generated by our method. The results are illustrated in Figure 3. Interestingly, ACR produces pseudo-labels of high quality in both *uniform* and *reversed* settings. As we can see, ACR significantly improves the precision and recall of pseudo-labels compared with DASO w/ LA, showing the effectiveness of the proposed adaptive logit adjustment.

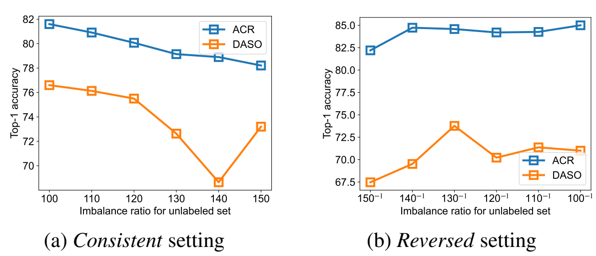


Figure 4. Accuracy on CIFAR10-LT of various imbalance ratios.

What if the target distribution is more imbalanced? We fix the labeled data imbalance ratio as $\gamma_l = 100$ and design the unlabeled data to be more class-imbalanced. Results in Figure 4 indicate that our method (ACR) consistently outperforms DASO by large margins, showing its capability to generalize to various target distributions.

Fine-grained results. Here, we report the fine-grained accuracy on the test set in Table 3. ACR achieves impressive improvements on few-shot classes compared with previous state-of-the-art method (DASO) at the cost of negligible drop under the *consistent* setting. For other settings, ACR outperforms DASO in all cases, indicating its strong generalization ability for both head and tail classes.

Setup	Consistent ($\gamma_u = 100$)				Uniform ($\gamma_u = 1$)				Reversed ($\gamma_u = 1/100$)			
	Many	Medium	Few	All	Many	Medium	Few	All	Many	Medium	Few	All
DASO	**95.1**	78.6	60.4	78.1	89.6	84.4	85.7	86.3	84.0	71.6	68.2	74.3
Ours	93.9	**81.6**	**75.3**	**83.4**	**92.8**	**90.6**	**97.9**	**93.5**	**90.7**	**83.8**	**96.4**	**89.7**

Table 3. Fine-grained accuracy on CIFAR10-LT ($\gamma_l = 100$).

Effect of the class imbalance ratio and the amount of labeled data. We illustrate the performance by varying the imbalance ratio and amount of labeled data in Figure 4 and Table 4, respectively. ACR outperforms DASO in all cases even under very high class imbalance ratios. Also, more results can be found in Figure 3 (main paper) and Figure 1a. Concerning the amount of labeled data, ACR improves DASO by an average 4.7% test accuracy.

Dataset	CIFAR10-LT			CIFAR100-LT		
$\lvert D^l \rvert$	0.5×	0.75×	1×	0.5×	0.75×	1×
DASO	66.6	77.3	79.1	48.3	56.7	59.2
Ours	**73.9**	**80.3**	**84.1**	**51.3**	**60.3**	**65.6**

Table 4. Accuracy on two datasets with various amount of labeled data (e.g., 0.5× means a half of the initial labeled data is used).

Self-supervised pre-training baseline. Following [6], we train SimCLR using all data and fine-tune the model on the labeled set. We find that its performance is even worse than FixMatch (8.8% accuracy drop). This is because the limited labeled data is not enough for model fine-tuning. Moreover, by generating high-quality pseudo-labels, methods based on FixMatch leverage unlabeled data more effectively than self-supervised learning.

Dataset	CIFAR10-LT			CIFAR100-LT		
γ_u	100	1	1/100	10	1	1/10
SSP	64.9	70.6	68.7	48.4	49.6	47.9
FixMatch	77.5	81.5	71.8	56.5	58.1	57.3

Table 5. Performance comparison between SSP and FixMatch. The labeled imbalance ratio γ_l is fixed 10 and 100 for CIFAR10-LT and CIFAR100-LT, respectively

References

[1] Yue Fan, Dengxin Dai, Anna Kukleva, and Bernt Schiele. Cossl: Co-learning of representation and classifier for imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14574–14584, 2022. 1

[2] Hyuck Lee, Seungjae Shin, and Heeyoung Kim. Abc: Auxiliary balanced classifier for class-imbalanced semi-supervised learning. *Advances in Neural Information Processing Systems*, 34:7082–7094, 2021. 2

[3] Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar.

Long-tail learning via logit adjustment. In *International Conference on Learning Representations*, 2020. 2

[4] Youngtaek Oh, Dong-Jin Kim, and In So Kweon. Daso: Distribution-aware semantics-oriented pseudo-label for imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9786–9796, 2022. 2

[5] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems*, 33:596–608, 2020. 2

[6] Yuzhe Yang and Zhi Xu. Rethinking the value of labels for improving class-imbalanced learning. *Advances in neural information processing systems*, 33:19290–19301, 2020. 3