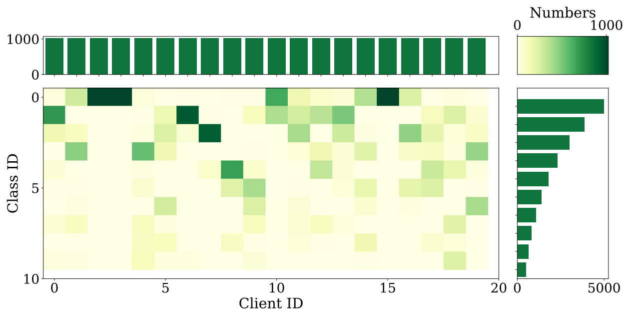
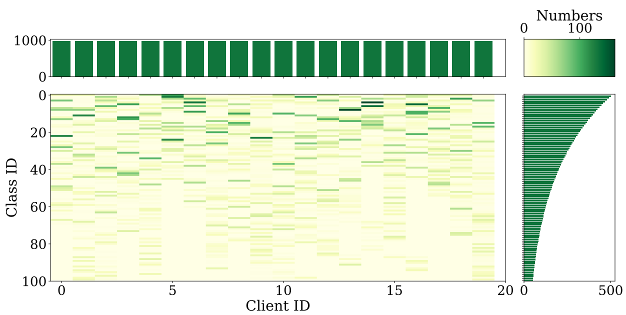
You Are Your Own Best Teacher: Achieving Centralized-level Performance in Federated Learning under Heterogeneous and Long-tailed Data

Supplementary Material



(a) CIFAR10-LT, $\alpha = 0.5$



(b) CIFAR100-LT, $\alpha = 0.5$

Figure 8. The data distributions for CIFAR10-LT and CIFAR100-LT with IF = 100. The lower-left plot shows each client's data distribution across 20 clients. The upper-left plot displays sample counts per client, and the lower-right shows sample counts per class. The color bar in the upper right represents the distribution intensity across clients.

A. More Experimental Setup

Fig. 8 illustrates the data distributions of CIFAR10-LT and CIFAR100-LT with an imbalance factor (IF) of 100. Lighter colors indicate fewer samples, highlighting the sparsity of data distribution under the federated long-tailed setting.

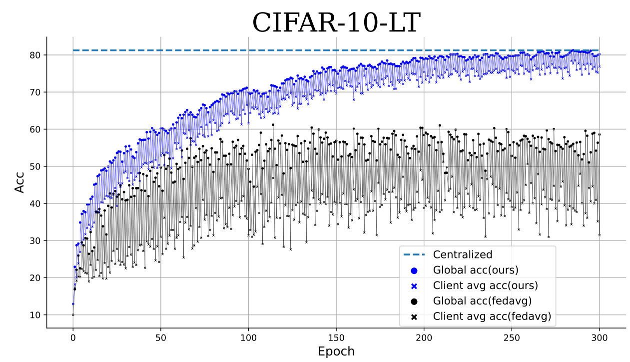
B. More Analysis of FedYoYo

B.1. Global-to-local Model Gap

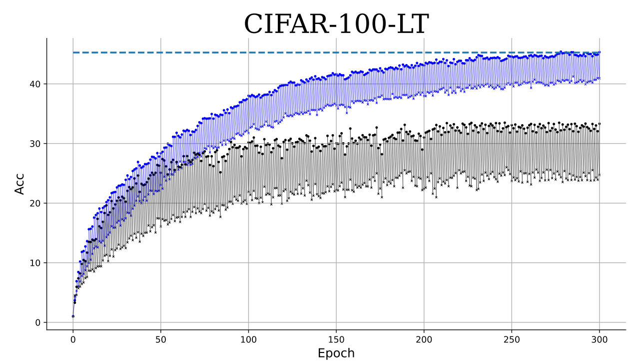
As demonstrated in Fig. 9, our method effectively reduces the Global-to-local model gap in long-tailed data distribution scenarios, significantly enhancing the global model's accuracy. This improvement is especially pronounced on the CIFAR100-LT dataset, where our approach consistently outperforms the baseline methods and achieves higher convergence rates, thereby demonstrating superior generalization across clients.

B.2. Feature Consistency

In Fig. 10, we illustrate the average cosine similarity between local models and the global model across different class categories: Many, Medium, Few, and All. Our method consistently achieves the highest similarity on both CIFAR-10-LT and CIFAR-100-LT datasets. This higher similarity reflects a reduced discrepancy between local and global models, indicating that our approach effectively mitigates client drift. These findings confirm that our method can better align local models with the global model, thereby enhancing the overall consistency and performance in federated long-tailed learning scenarios.

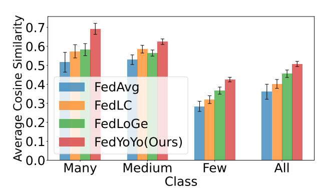


(a) CIFAR-10-LT.

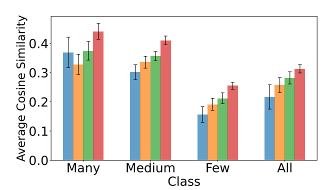


(b) CIFAR-100-LT.

Figure 9. Comparison of global-client gap.

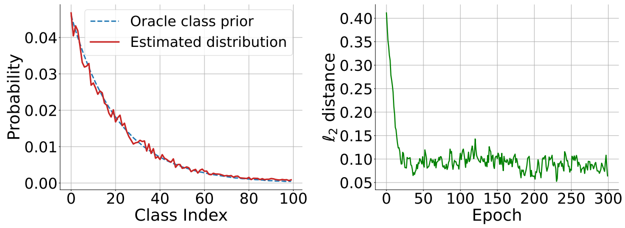


(a) CIFAR-10-LT



(b) CIFAR-100-LT

Figure 10. **Comparison of feature similarity between global model and local models.**



(a) The estimated distribution. (b) ℓ_2 distance during training.

Figure 11. **(a) The global long-tailed distribution obtained through estimated. (b) ℓ_2 distance between the statiscal distribution and the oracle class prior during training.** Experiments is conducted on CIFAR-100-LT with IF $= 100$ and $\alpha = 0.5$.

Table 6. **Top-1 test accuracy (%) of different data augment policys.**

Policys	CIFAR-10-LT			CIFAR-100-LT		
	RA	AA	TA	RA	AA	TA
Many	81.48	81.95	82.92	60.26	59.6	58.66
Medium	83.30	80.90	78.63	47.00	47.51	47.91
Few	79.57	81.20	80.60	28.63	28.4	29.37
All	81.45	81.41	80.94	46.13	46.01	46.11

B.3. Effectiveness of Estimated Global Distribution

As shown in Fig. 11b, the distance steadily decreases, indicating that our parameterized class distribution converges toward the oracle prior over time. These results suggest that our model captures robust feature representations, as improved feature extraction directly enhances the accuracy of the distribution estimation, forming a mutually reinforcing cycle.

C. More Ablation Experiments

To evaluate the impact of different data augmentation strategies, we compare RandAugment (RA) [5], AutoAugment (AA) [4], and TrivialAugment (TA) [23]. As shown in Tab. 6, RA achieves the best performance on both CIFAR-10-LT and CIFAR-100-LT, particularly in the Many and Medium class categories. However, the differences between these augmentation strategies are not significant, suggesting that the choice of augmentation method has limited impact on the overall performance. This indicates that the effectiveness of our method does not rely heavily on the specific augmentation strategy, but rather on the robustness of the approach itself.