

# Supplementary Materials: One-shot-but-not-degraded federated learning

Anonymous Authors

## A MORE EXPERIMENTAL RESULTS

### A.1 Implementation details

We adopt the same local training setting as [2]. We use the SGD optimizer with a learning rate of 0.01 and a fixed momentum of 0.9. Each local model is trained on the client’s local data for  $E = 400$  rounds, and the client number is 5. Following the setting of [1, 2], we train the generator  $g$  with a DNN. We adopt the Adam optimizer with a learning rate of 0.001 and training for  $T_g = 20$  iterations. For the training of the MoE network, we use the SGD optimizer with a learning rate  $\eta_G = 0.01$  and momentum of 0.9 and train the gating network for  $T = 40$  iterations.

### A.2 Metrics

We utilize the test accuracy as the prime metric over all baselines and the proposed IntactOFL. For results with error bars, we run five repeated experiments with different random seeds.

### A.3 Horizontal scalability analysis

We provide more horizontal scalability analysis on SVHN (see Figure 1), which is the test accuracy across different numbers of clients. The conclusion is the same with the test on CIFAR-10. With the increasing number of clients, the local data become more sparser and more fragmented. Consequently, the local models trained from such data are highly prone to overfitting, resulting in inferior performance. In summary, we also still conclude that the proposed IntactOFL is scalable across diverse distributed networks of varying sizes.

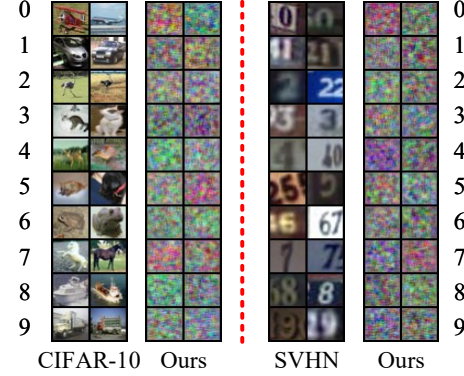
**Table 1: Test accuracy of the server model on SVHN across different numbers of clients  $m = \{5, 10, 25, 50, 100\}$ .**

$m$	5	10	25	50	100
MA-Echo	80.23	67.12	59.89	56.23	47.54
FedAvg	57.61	45.22	42.18	33.98	30.15
FedDF	72.11	61.09	58.89	52.32	44.18
F-ADI	77.62	64.39	60.98	59.63	48.39
F-DAFL	74.55	62.39	60.88	53.69	47.95
Ensemble	81.22	67.51	63.12	56.68	48.25
DENSE	80.03	68.98	62.39	59.99	53.76
Co-Boosting	81.34	69.71	63.85	60.01	55.15
Ours	<b>84.81</b>	<b>72.95</b>	<b>68.44</b>	<b>65.64</b>	<b>59.19</b>

### A.4 Visualization of Generated Auxiliary Data

We provide the visualization of the generated data on CIFAR-10 and SVHN in Figure 1. It is worth noting that the goal of the data generator is designed to generate the data which is similar in utilization not in visualization. The generated data looks different from

the original data, which can mitigate the risk of leaking sensitive information. Meanwhile, it plays an important role in training the MoE network, which extracts the information from local models and helps achieve higher performance than other baselines.



**Figure 1: Visualization of generated data on CIFAR-10 and SVHN.**

### A.5 Impact of Gating Network Architecture

We investigate the impact of different gating network architectures on performance. The gating network outputs the weights of each expert according to the input, which can be viewed as a function. We vary the different architectures of the gating network, including MLP, CNN, and ResNet, with the primary difference lying in their capabilities for information processing. The results are shown in Table 2. We conclude that different gating network architecture has a limited impact on the MoE network performance.

**Table 2: Test accuracy of the MoE network on CIFAR-10 across different gating network architectures (MLP, CNN, and ResNet).**

gating network	MLP	CNN	ResNet
CIFAR-10	79.93	79.23	<b>80.04</b>
CIFAR-100	46.78	46.55	<b>46.88</b>
SVHN	84.81	84.77	<b>85.00</b>
Tiny-ImageNet	35.09	34.69	<b>35.11</b>

## REFERENCES

- [1] Hanting Chen, Yunhe Wang, Chang Xu, Zhaohui Yang, Chuanjian Liu, Boxin Shi, Chunjing Xu, Chao Xu, and Qi Tian. 2019. Data-free learning of student networks. In *Proc. of ICCV*. 3514–3522.
- [2] Jie Zhang, Chen Chen, Bo Li, Lingjuan Lyu, Shuang Wu, Shouhong Ding, Chunhua Shen, and Chao Wu. 2022. Dense: Data-free one-shot federated learning. In *Proc. of the Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 35. 21414–21428.