国防科技大学
National University of Defense Technology

武漢大學
WUHAN UNIVERSITY

ICML
International Conference On Machine Learning

Does One-shot Give the Best Shot? Mitigating Model Inconsistency in One-shot Federated Learning

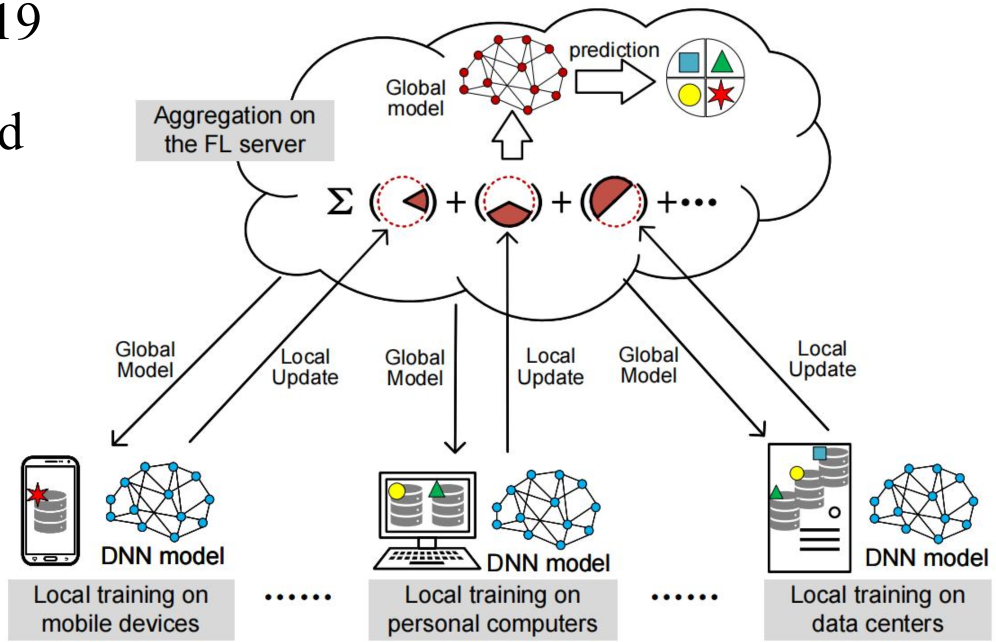
[1]Hui Zeng, [2]Wenke Huang, [1]Tongqing Zhou, [1]Xinyi Wu, [2]Guancheng Wan, [1]Yingwen Chen, [1]Zhiping Cai

[1]National University of Defense Technology, [2]Wuhan University

Background & Motivation

Heavy communication overhead in vanilla Federated Learning

- Heavy communication burden: More than 250GB for a VGG19
- High communication time: More than 194 hours for one round transmission of GPT-3



Existing One-shot Federated Learning:

- Optimization-based methods
- Distillation-based methods
- Generative methods
- Selective ensemble methods

Bottleneck: significant performance gap ⇨ Garbage in, garbage out

ICML
International Conference
On Machine Learning

Background & Motivation

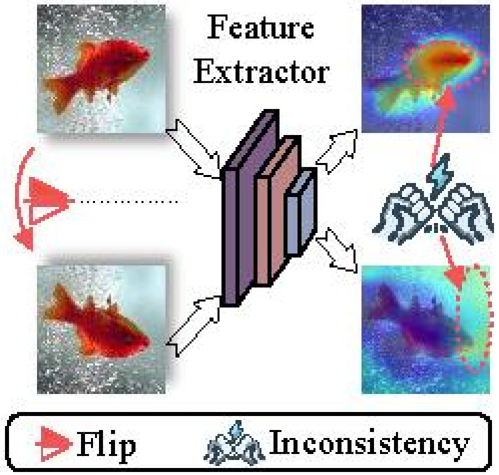
Model inconsistencies:

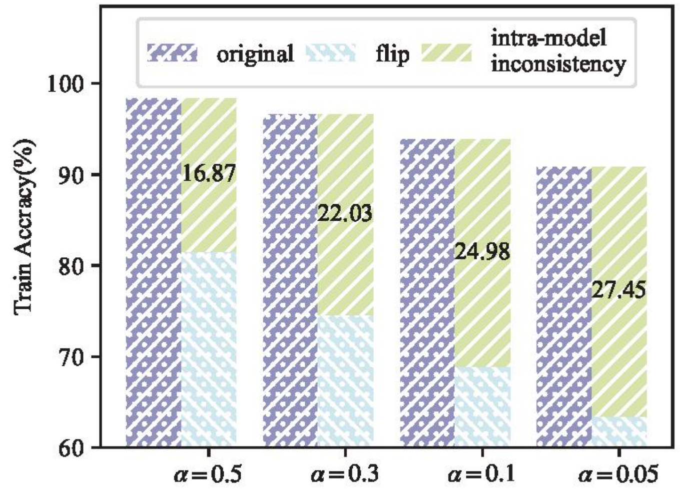
■ Intra-model inconsistency:

Intra-model inconsistency. The intra-model inconsistency of the one-shot local model on the original samples (x, y) and augmented samples $(A(x), y)$ can be represented as:

$$\|\Delta_{intra}\|^2 \geq \| \left(p \cdot \nabla g_a \cdot \nabla A\right)^T (x - A(x))\|^2 > 0,$$

where $p = \sum_{c=1}^{C}(z_c - y_c)$, z is the prediction of w_i activated by softmax function with the augmented samples $A(x)$, ∇g_a is the gradient of the local model w_i, and ∇A is the gradient of the data augmentation function.



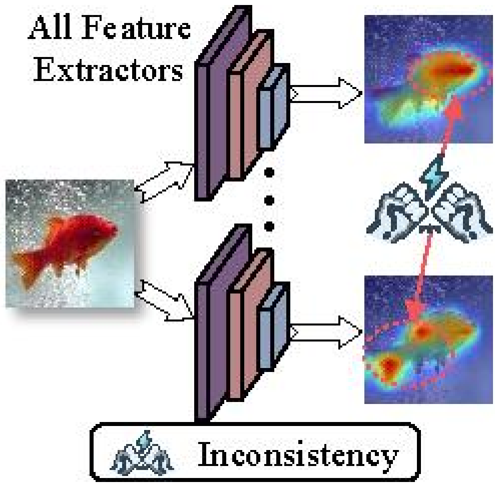


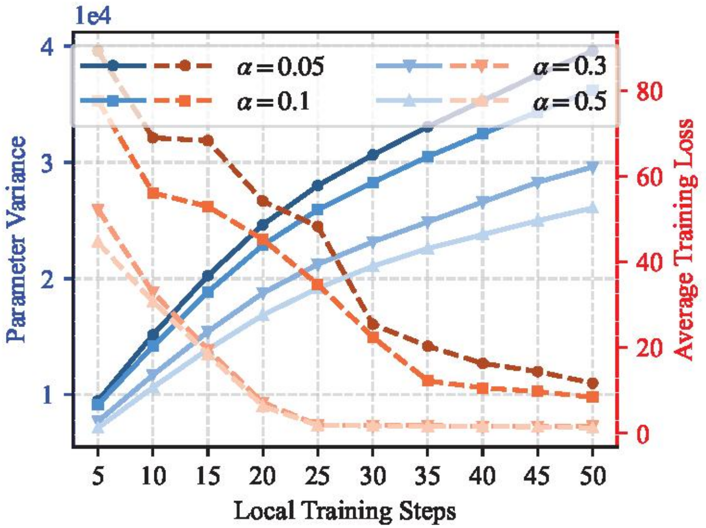
■ Inter-model inconsistency:

Inter-model inconsistency. For any two client u and v with the same quantity of samples $n_u = n_v$, the one-step model deviation between the two clients $\Delta_{inter} = \nabla w_u - \nabla w_v$ can be represented as:

$$\begin{aligned}
&\|\Delta_{inter}\|^2 \\
&= \|\frac{\eta}{n_u}[(n_{u,c}(1 - \overline{z}_{u,c})\overline{x}_{u,c} - n_{v,c}(1 - \overline{z}_{v,c})\overline{x}_{v,c}) \\
&- (\sum_{c' \in [C_u] \backslash c} n_{u,c'}\overline{z}_{u,c'}\overline{x}_{u,c'} - \sum_{c' \in [C_v] \backslash c} n_{v,c'}\overline{z}_{v,c'}\overline{x}_{v,c'})]\|^2 \\
&> 0,
\end{aligned}$$

where η is the learning rate, $n_{u,c}$ and $n_{v,c}$ is the sample quantity of c-th class, c' is the negative classes except c.





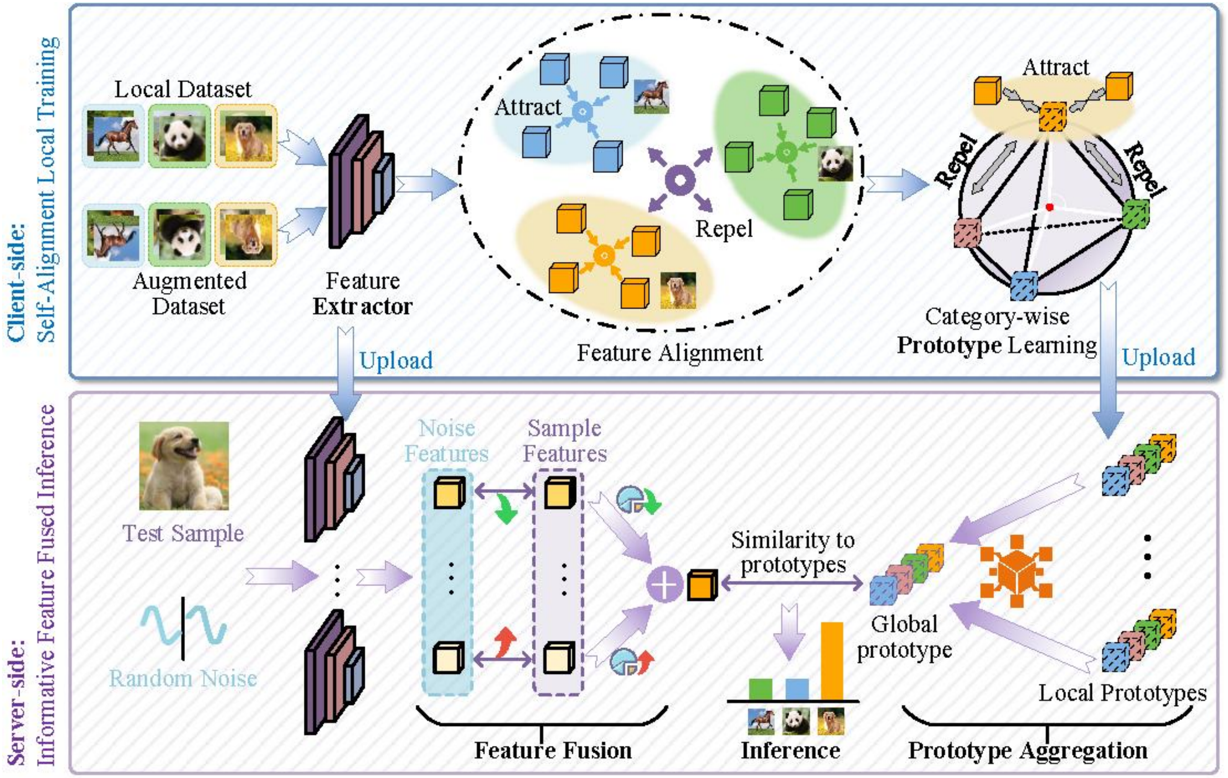
Challenges & Method

Challenges #1: How can we construct a model capable of capturing invariant features and achieving stable predictions under such heterogeneous conditions?

Contribution #1: we design Self-Alignment Local Training(SALT), which employs contrastive learning and prototypes to mitigate the intra-model inconsistency.

Challenges #2: How can we effectively leverage models with parameter discrepancies in a one-shot manner?

Contribution #2: we design Informative Feature Fused Inference(IFFI), which performs feature-level fusion to mitigate the inter-model inconsistency.



Experiment Results

■ Effectiveness

Methods	CIFAR-10				CIFAR-100				Tiny-ImageNet			
	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$
MA-Echo	36.77$_{\pm 0.91}$	51.23$_{\pm 0.28}$	60.14$_{\pm 0.21}$	64.21$_{\pm 0.23}$	19.54$_{\pm 0.45}$	29.11$_{\pm 0.26}$	37.77$_{\pm 0.24}$	41.94$_{\pm 0.21}$	15.46$_{\pm 0.66}$	22.23$_{\pm 0.56}$	23.46$_{\pm 0.19}$	28.21$_{\pm 0.42}$
O-FedAvg	12.13$_{\pm 2.11}$	17.43$_{\pm 0.51}$	28.07$_{\pm 0.89}$	35.42$_{\pm 0.67}$	4.77$_{\pm 0.21}$	6.45$_{\pm 0.71}$	10.67$_{\pm 0.31}$	12.13$_{\pm 0.05}$	5.67$_{\pm 0.45}$	8.31$_{\pm 0.21}$	13.61$_{\pm 0.10}$	13.71$_{\pm 0.16}$
FedFisher	40.03$_{\pm 1.11}$	47.01$_{\pm 1.81}$	49.33$_{\pm 1.52}$	50.34$_{\pm 1.32}$	16.56$_{\pm 2.67}$	18.98$_{\pm 2.09}$	27.24$_{\pm 1.92}$	31.44$_{\pm 1.87}$	15.65$_{\pm 1.54}$	17.89$_{\pm 1.46}$	19.54$_{\pm 1.31}$	20.77$_{\pm 1.15}$
FedDF	35.53$_{\pm 0.67}$	41.58$_{\pm 0.80}$	44.78$_{\pm 0.60}$	54.58$_{\pm 0.73}$	15.07$_{\pm 0.74}$	27.17$_{\pm 0.55}$	31.23$_{\pm 0.79}$	35.39$_{\pm 0.47}$	11.45$_{\pm 0.40}$	16.32$_{\pm 0.33}$	17.79$_{\pm 0.57}$	27.55$_{\pm 0.66}$
F-ADI	35.93$_{\pm 1.56}$	48.35$_{\pm 1.23}$	52.66$_{\pm 1.44}$	58.78$_{\pm 1.67}$	14.65$_{\pm 0.98}$	28.13$_{\pm 1.24}$	33.18$_{\pm 0.67}$	39.44$_{\pm 1.11}$	13.92$_{\pm 1.99}$	19.00$_{\pm 1.78}$	26.01$_{\pm 1.44}$	29.98$_{\pm 1.34}$
F-DAFL	38.32$_{\pm 1.40}$	46.34$_{\pm 1.12}$	54.03$_{\pm 1.71}$	59.09$_{\pm 2.23}$	16.31$_{\pm 0.33}$	26.80$_{\pm 1.33}$	34.89$_{\pm 1.45}$	37.88$_{\pm 1.34}$	15.12$_{\pm 1.34}$	19.01$_{\pm 1.11}$	23.78$_{\pm 1.23}$	27.98$_{\pm 1.10}$
DENSE	38.37$_{\pm 1.08}$	50.26$_{\pm 0.24}$	59.76$_{\pm 0.45}$	62.19$_{\pm 0.12}$	18.37$_{\pm 2.43}$	32.03$_{\pm 0.44}$	37.33$_{\pm 0.48}$	38.84$_{\pm 0.39}$	18.77$_{\pm 0.67}$	22.25$_{\pm 0.33}$	28.14$_{\pm 0.34}$	32.34$_{\pm 0.32}$
Ensemble	41.36$_{\pm 0.67}$	45.43$_{\pm 0.32}$	62.18$_{\pm 0.34}$	61.61$_{\pm 0.23}$	20.46$_{\pm 0.62}$	26.23$_{\pm 0.55}$	38.01$_{\pm 0.67}$	41.61$_{\pm 0.77}$	13.28$_{\pm 0.67}$	15.38$_{\pm 0.23}$	17.53$_{\pm 0.31}$	28.50$_{\pm 0.46}$
Co-Boosting	39.20$_{\pm 0.81}$	58.49$_{\pm 1.24}$	67.21$_{\pm 1.76}$	70.24$_{\pm 2.34}$	20.19$_{\pm 1.44}$	27.59$_{\pm 1.35}$	39.30$_{\pm 1.30}$	42.67$_{\pm 1.40}$	19.00$_{\pm 1.45}$	21.90$_{\pm 1.20}$	29.24$_{\pm 1.32}$	30.78$_{\pm 2.01}$
FuseFL	54.42$_{\pm 0.41}$	73.79$_{\pm 0.34}$	84.58$_{\pm 0.91}$	84.34$_{\pm 0.88}$	29.12$_{\pm 0.23}$	36.86$_{\pm 0.38}$	45.12$_{\pm 0.51}$	49.30$_{\pm 0.32}$	22.15$_{\pm 2.11}$	29.28$_{\pm 2.04}$	33.04$_{\pm 1.79}$	34.34$_{\pm 1.81}$
IntactOFL	48.22$_{\pm 0.43}$	61.13$_{\pm 0.63}$	70.21$_{\pm 0.60}$	79.93$_{\pm 0.23}$	27.99$_{\pm 0.67}$	39.15$_{\pm 0.46}$	41.86$_{\pm 0.60}$	46.78$_{\pm 0.78}$	20.45$_{\pm 0.34}$	28.43$_{\pm 0.17}$	30.15$_{\pm 0.12}$	35.09$_{\pm 0.14}$
Ours	**71.84**$_{\pm 1.53}$	**77.83**$_{\pm 1.32}$	**84.76**$_{\pm 0.46}$	**88.74**$_{\pm 0.11}$	**31.02**$_{\pm 1.17}$	**45.48**$_{\pm 1.01}$	**56.65**$_{\pm 0.91}$	**61.07**$_{\pm 0.55}$	**36.96**$_{\pm 0.92}$	**43.62**$_{\pm 0.77}$	**53.32**$_{\pm 0.50}$	**56.48**$_{\pm 0.32}$
Δ	↑ 17.42	↑ 6.04	↑ 0.18	↑ 4.40	↑ 1.90	↑ 6.33	↑ 11.53	↑ 11.77	↑ 14.81	↑ 14.34	↑ 20.28	↑ 21.39

■ Scalability

Methods	Client scales m				
	5	10	25	50	100
MA-Echo	64.21	52.64	48.36	45.35	38.54
O-FedAvg	35.42	32.09	28.03	28.24	27.14
FedFisher	50.34	45.67	34.66	29.09	28.89
FedDF	54.58	48.88	35.44	29.91	25.66
F-ADI	59.34	46.33	31.83	27.66	24.89
F-DAFL	58.59	45.45	32.88	29.98	28.91
DENSE	62.19	54.67	49.32	48.67	43.34
Ensemble	61.61	60.44	58.44	52.51	45.72
Co-Boosting	55.34	51.11	49.32	44.56	42.45
FuseFL	84.34	78.28	62.12	42.18	37.11
IntactOFL	79.93	69.11	64.32	59.45	53.21
Ours	**88.74**	**86.96**	**85.25**	**81.32**	**75.37**

■ Efficiency