Learning Locally, Revising Globally: Global Reviser for Federated Learning with Noisy Labels

Yuxin Tian[1,2], Mouxing Yang[1,2], Yuhao Zhou[1,2], Jian Wang[1,2], Qing Ye[1,2], Tongliang Liu[3], Gang Niu[4], Jiancheng Lv[1,2]*

[1] Sichuan University [2] Engineering Research Center of Machine Learning and Industry Intelligence, Ministry of Education, China

[3] University of Sydney [4] Southeast University, *Corresponding Author

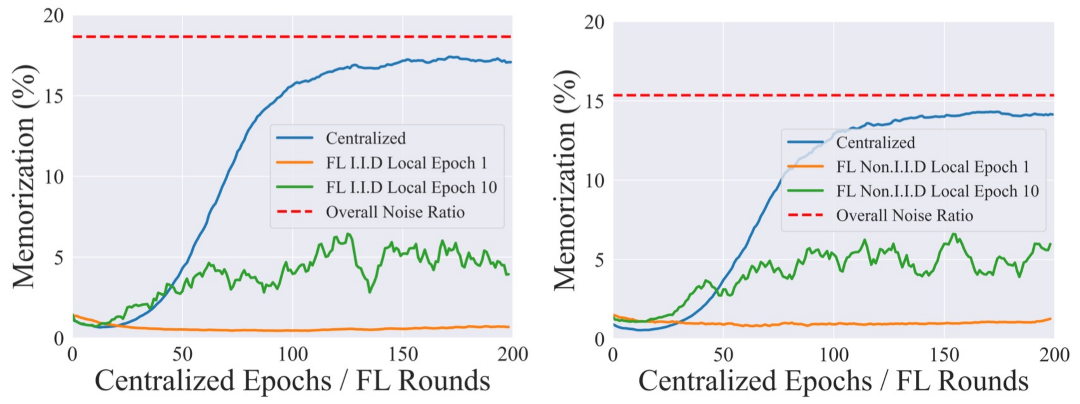
ICML International Conference On Machine Learning

July 6th - 11th, 2026

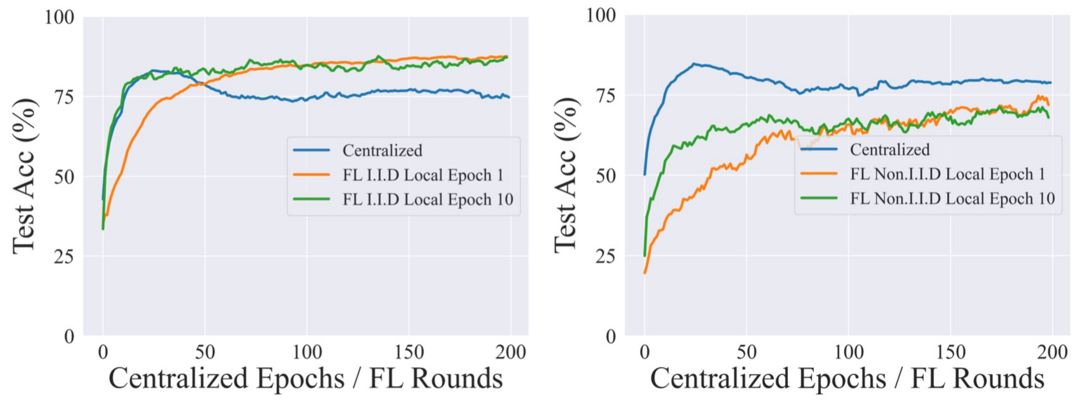
Seoul, South Korea

Problem & Observations

- **Problem:** 1) Federated learning often relies on decentralized labels that are **noisy, heterogeneous, and difficult to audit**; 2) Clients may **differ in data distribution, noise types, and noise ratios**; 3) Traditional C-LNL methods cannot be migrated to this scenario due to **severe class imbalance**
- **Observation**: The global model of FL **delays noisy-label memorization and avoids the test-performance collapse** seen in centralized training.

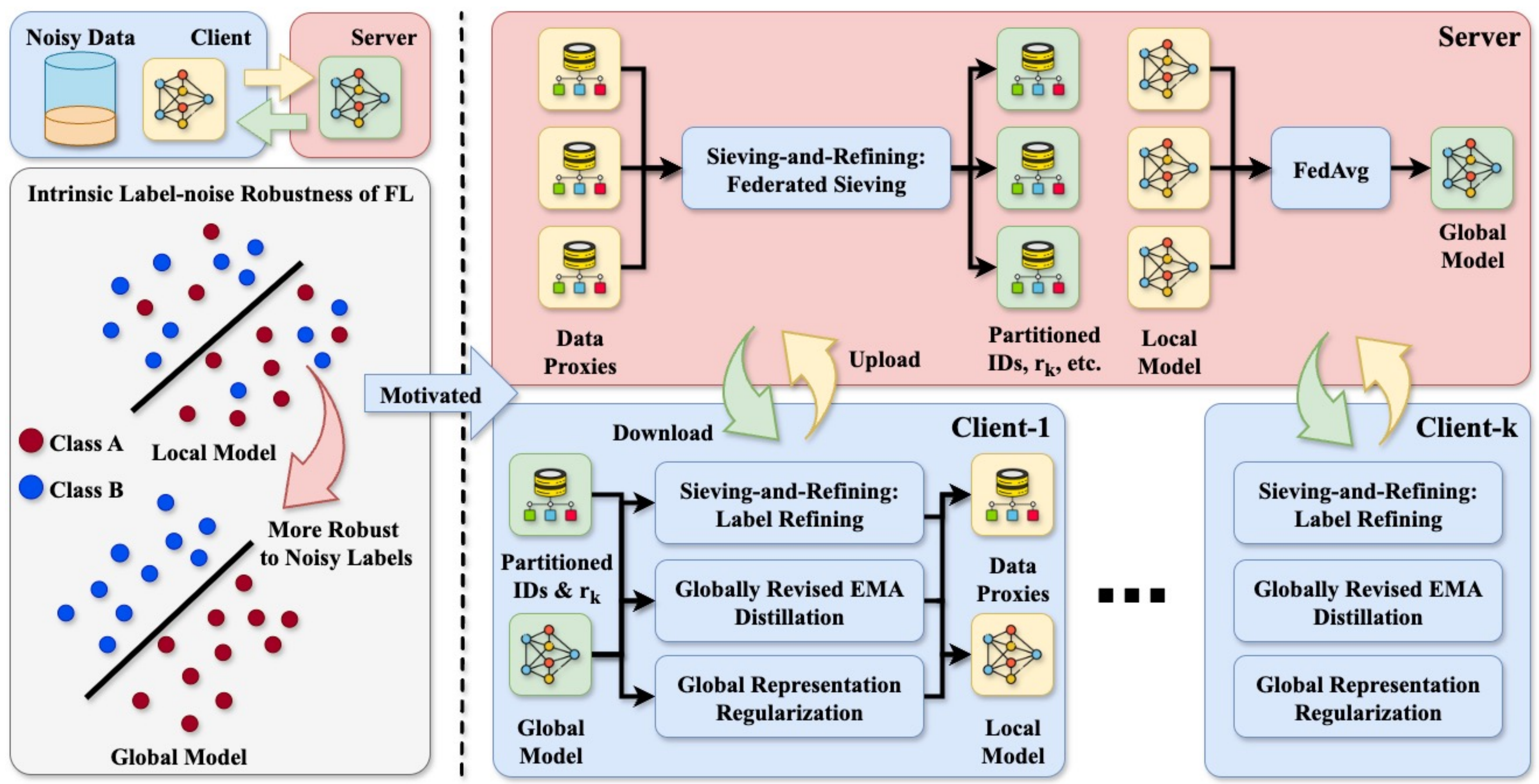


(a) Memorization (I.I.D & Non.I.I.D-Dirichlet-0.3)



(b) Test Acc (I.I.D & Non.I.I.D-Dirichlet-0.3)

Method



$$\min_{\mathbf{w}_g} \mathcal{L}(\mathbf{w}_g) = \sum_{k\in\mathcal{S}} a_k \mathcal{L}_k(\mathbf{w}_k) \quad \text{s.t.} \quad \sum_{k\in\mathcal{S}} a_k = 1, \quad \mathcal{L}_k = \mathcal{L}_k^{SR} + \lambda_{\mathcal{B}}\mathcal{B}_k + \lambda_{\mathcal{R}}\mathcal{R}_k,$$

1. Federated Sieving & Label Refining

Clients upload loss proxies; the server uses a global two-component GMM to detect noisy samples and estimate client noise ratios, then refines labels with global pseudo labels without sharing raw data.

2. Globally Revised EMA Distillation

Global pseudo labels refine noisy labels, while robust global parameters revise local EMA teachers before distillation to reduce noisy update accumulation and provide reliable soft targets.

3. Representation Regularization

Weak global representations guide strongly augmented local representations, keeping clients aligned with the robust global model and reducing noisy-label overfitting under heterogeneous data distributions.

Motivation & Contributions & Takeaway

- **Motivation**: Since the global FL model memorizes noisy labels more slowly, making it a useful server-side reviser for stabilizing local training.
- **New perspective**: Reveal the intrinsic label-noise robustness of the FL global model.
- **New method**: Propose FedGR, a self-contained global reviser for federated learning with noisy labels.
- **Strong performance**: Achieve the state-of-the-art label-noise robustness on CIFAR-10, CIFAR-100, and Clothing1M with I.I.D., Non-I.I.D., and dual-heterogeneity noisy-label settings
- **Takeaway**: FedGR turns the global model into an active reviser that identifies noisy samples, corrects labels, and stabilizes local training without sharing raw data, substantially improving robustness to severe label noise in FL.

Main Results & Ablations

- **FedGR achieves the best performance in almost all noisy-label settings, with especially large gains under severe noise setups.**

Table 2. Results on CIFAR-100. The 1st/2nd-best results are in a gray box w/. and w/o. boldface.

Data Partition	I.I.D							
Noise Type	Clean	Sym		Asym		Mixed		Avg
ϕ	0.0	0.6	1.0	0.6	1.0	0.6	1.0	
$\mathcal{U}(\rho_{min}, \rho_{max})$	0.0-0.0	0.5-1.0	0.5-1.0	0.2-0.4	0.2-0.4	0.2-0.4	0.2-0.4	
FedAvg	64.85 ±0.33	33.97 ±1.82	13.00 ±1.85	54.51 ±0.83	44.18 ±1.72	52.37 ±0.40	43.02 ±0.44	40.18 ±1.18
FedProx	56.85 ±0.55	30.22 ±1.68	11.94 ±1.91	45.97 ±0.92	37.83 ±0.76	45.08 ±0.50	36.82 ±0.47	34.64 ±1.04
FL-Coteaching	-	41.19 ±1.04	25.98 ±1.87	58.84 ±0.63	50.56 ±1.26	58.13 ±0.36	53.42 ±0.68	48.02 ±1.02
FL-DivideMix	-	49.48 ±0.52	**35.35 ±1.15**	59.26 ±0.25	55.12 ±0.31	59.40 ±0.25	57.53 ±0.20	52.69 ±0.45
FedCorr [CVPR22]	70.77 ±2.11	58.29 ±1.30	27.54 ±2.04	67.04 ±0.49	59.58 ±0.70	66.56 ±0.67	61.41 ±1.03	56.74 ±1.04
FedNoRo [IJCAI23]	-	29.61 ±0.40	16.00 ±0.39	35.48 ±0.27	30.57 ±0.52	34.56 ±0.36	30.81 ±0.98	29.50 ±0.49
FedDiv [AAAI24]	-	37.14 ±4.20	4.13 ±3.25	**69.37 ±0.42**	60.26 ±1.56	66.65 ±0.66	62.64 ±0.42	59.21 ±1.25
FedFixer [AAAI24]	-	34.46 ±1.03	13.93 ±0.60	53.17 ±0.56	44.11 ±0.33	53.86 ±0.98	46.08 ±0.07	40.94 ±0.60
FedGR [Ours]	**71.64 ±0.22**	**63.19 ±0.91**	35.28 ±1.47	69.10 ±0.20	**62.97 ±0.44**	**68.73 ±0.46**	**64.56 ±0.32**	**60.64 ±0.63**

Data Partition	Non.I.I.D-Dirichlet (0.3)							
Noise Type	Clean	Sym		Asym		Mixed		Avg
ϕ	0.0	0.6	1.0	0.6	1.0	0.6	1.0	
$\mathcal{U}(\rho_{min}, \rho_{max})$	0.0-0.0	0.5-1.0	0.5-1.0	0.2-0.4	0.2-0.4	0.2-0.4	0.2-0.4	
FedAvg	63.86 ±0.61	29.04 ±1.64	10.23 ±1.27	51.74 ±0.60	41.36 ±1.16	49.29 ±0.68	39.16 ±1.60	36.80 ±1.16
FedProx	58.83 ±0.63	26.19 ±1.52	9.31 ±1.27	46.12 ±0.87	36.86 ±1.16	44.26 ±0.80	35.40 ±1.23	37.77 ±0.98
FL-Coteaching	-	36.64 ±1.60	24.81 ±1.03	57.70 ±0.54	50.71 ±0.95	56.80 ±0.40	52.35 ±1.21	46.50 ±0.96
FL-DivideMix	-	44.25 ±0.81	27.29 ±2.50	57.93 ±0.35	52.53 ±1.33	57.70 ±0.23	55.47 ±0.95	49.20 ±1.03
FedCorr [CVPR22]	64.46 ±3.54	55.83 ±0.40	19.02 ±1.52	61.29 ±0.90	52.18 ±1.61	61.45 ±1.88	53.94 ±0.69	50.62 ±1.17
FedNoRo [IJCAI23]	-	26.66 ±0.57	12.43 ±0.61	34.00 ±0.37	27.00 ±0.61	32.64 ±0.22	27.02 ±0.33	26.63 ±0.45
FedDiv [AAAI24]	-	10.18 ±1.05	1.11 ±0.18	39.49 ±7.41	39.59 ±5.80	32.72 ±9.67	33.20 ±5.33	27.19 ±0.42
FedFixer [AAAI24]	-	29.26 ±0.22	11.61 ±0.34	51.43 ±0.71	43.01 ±0.58	51.67 ±1.30	43.63 ±0.38	38.44 ±0.59
FedGR [Ours]	69.38 ±0.52	**57.76 ±0.54**	**30.30 ±0.96**	**65.57 ±0.65**	**56.49 ±0.49**	**65.57 ±0.64**	**59.68 ±0.40**	**55.90 ±0.61**

Table 3. Results on Clothing1M. The 1st/2nd-best results are in a gray box w/. and w/o. boldface.

Methods	FedAvg	FedProx	FL-Coteaching	FL-DivideMix	FedCorr [CVPR22]	FedDiv [AAAI24]	FedFixer [AAAI24]	FedGR [Ours]
I.I.D	69.60±0.27	69.69±0.25	69.48±0.20	69.13±0.22	69.84±0.20	67.71±0.30	70.61±0.28	**71.19±0.42**
Non.I.I.D-Dirichlet (0.3)	67.90±1.31	68.18±1.13	68.16±0.82	63.88±1.21	60.42±7.23	65.79±2.77	65.16±0.95	**68.52±1.11**

- **Ablation & Further Analysis.**

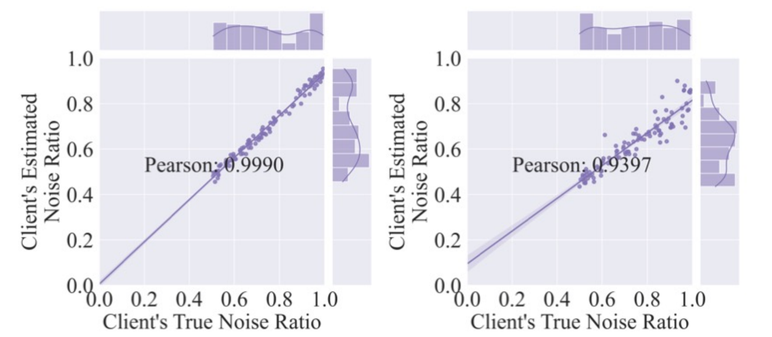
Table 4. Ablation studies.

Data Partition	I.I.D			Non.I.I.D-Dirichlet (0.3)		
Noise Type	Sym	Asym	Mixed	Sym	Asym	Mixed
ϕ	1.0	1.0	1.0	1.0	1.0	1.0
$\mathcal{U}(\rho_{min}, \rho_{max})$	0.5-1.0	0.2-0.4	0.2-0.4	0.5-1.0	0.2-0.4	0.2-0.4
FedGR	83.91 ±1.32	91.64 ±0.38	92.27 ±0.18	63.64 ±5.39	83.67 ±5.02	84.65 ±2.38
w/o. FS	54.59 ±3.04	87.49 ±0.12	91.71 ±0.13	45.48 ±7.81	83.45 ±4.49	84.01 ±0.31
w/o. LR	75.23 ±2.86	90.42 ±0.13	90.46 ±0.10	59.48 ±5.85	82.92 ±3.43	83.21 ±1.89
w/o. $\mathcal{R}_k$	81.49 ±1.07	91.22 ±0.34	91.84 ±0.30	58.23 ±4.08	81.80 ±2.32	82.70 ±0.63
w/o. $\mathcal{B}_k$	78.14 ±1.48	91.54 ±0.24	91.24 ±0.13	51.07 ±5.16	80.07 ±3.36	79.44 ±2.79

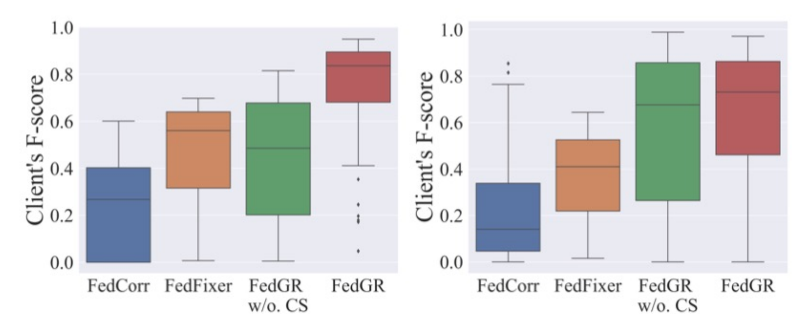
Table 5. Further analysis.

Data Partition	I.I.D			Non.I.I.D-Dirichlet (0.3)		
Noise Type	Sym	Asym	Mixed	Sym	Asym	Mixed
ϕ	1.0	1.0	1.0	1.0	1.0	1.0
$\mathcal{U}(\rho_{min}, \rho_{max})$	0.5-1.0	0.2-0.4	0.2-0.4	0.5-1.0	0.2-0.4	0.2-0.4
FedGR	83.91 ±1.32	91.64 ±0.38	92.27 ±0.18	63.64 ±5.39	83.67 ±5.02	84.65 ±2.38
w/o. weak aug	55.73 ±12.97	89.93 ±0.29	90.37 ±0.20	25.32 ±4.36	72.48 ±6.03	78.54 ±4.22
w/o. strong aug	34.51 ±3.26	80.53 ±0.41	78.72 ±0.68	18.43 ±1.47	59.74 ±2.01	58.60 ±1.09
Online EMA distill	82.80 ±1.44	91.34 ±0.27	92.04 ±0.29	62.18 ±4.13	82.50 ±5.79	83.56 ±3.63
Randomly sample clients w/. replacement in warmup	82.56 ±1.15	91.21 ±0.46	92.08 ±0.22	62.03 ±5.85	82.91 ±2.23	83.19 ±2.52

- **FS accurately captures noise structure across the clients.**



(a) Pearson Coefficient Analysis (I.I.D & Non.I.I.D-Dirichlet-0.3)



(b) Clients' F-score Distributions (I.I.D & Non.I.I.D-Dirichlet-0.3)