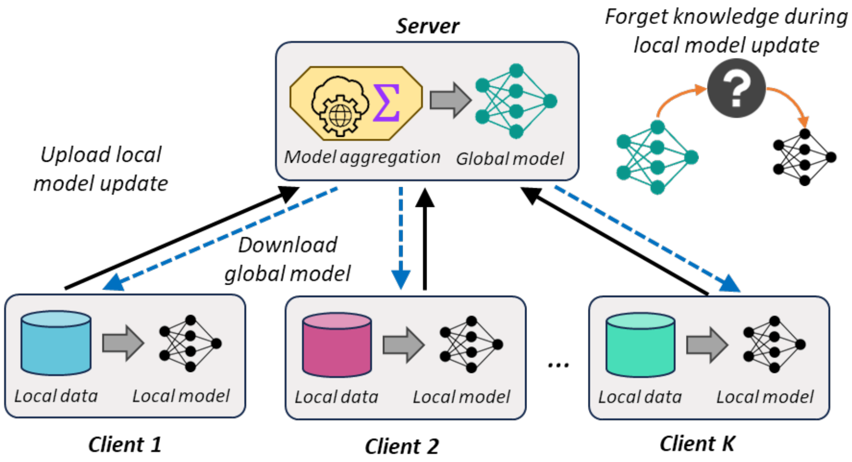
Institute of High Performance Computing
IHPC

CVPR SEATTLE, WA JUNE 17-21, 2024

An Aggregation-Free Federated Learning for Tackling Data Heterogeneity

Dr WANG Yuan
Senior Scientist, Computing & Intelligence, IHPC
June 2024
ARES Classification

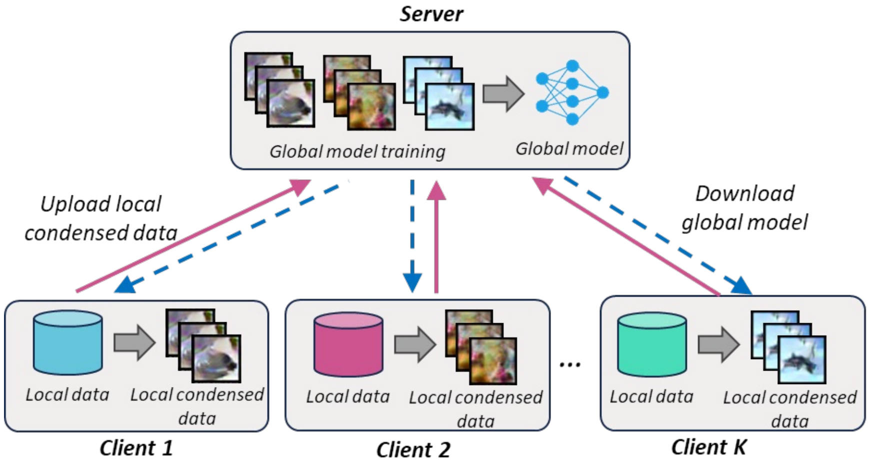
Background and motivation



- Traditional FL requires clients to update local model upon globally-aggregated server model
- Sharing local model/gradients are prone to gradient leakage attack and communication-consuming
- Local update process can lead to forgetting of knowledge learned in previous global model, causing client drift and inferior convergence performance in non-IID scenarios

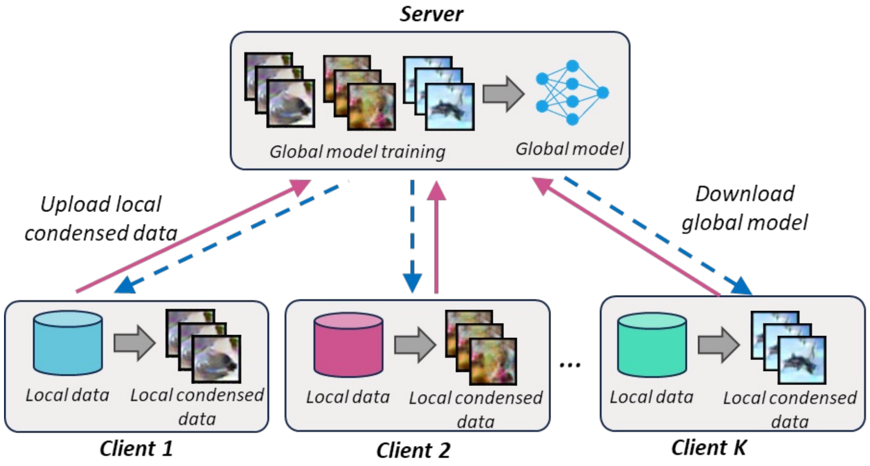
CREATING GROWTH, ENHANCING LIVES

Methodology and contribution



- We employ a novel model-aggregation-free framework to replace traditional model-sharing FL strategies
- Model is trained only at the server, clients instead focus solely on learning and sharing a compact set of synthetic data, i.e., condensed data

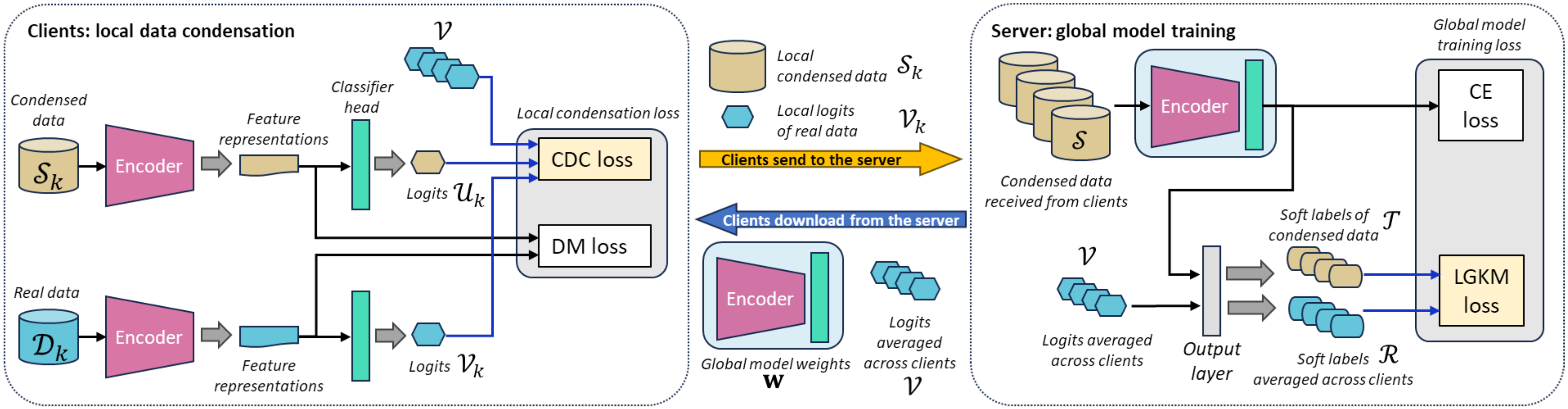
Methodology and contribution



- ❑ Learning local condensed data is enhanced by utilizing broader peer knowledge through Collaborative Data Condensation
- ❑ Global model training is enhanced with Local-Global Knowledge Matching, utilizing more global insights other than condensed data only, improving the learning performance
- ❑ Improves convergence performance over traditional FL methods in the context of non-IID cross-client data distribution

CREATING GROWTH, ENHANCING LIVES

Method in detail – FedAF framework



$$\mathcal{L}_{\text{loc}}(\mathcal{S}_k, \{\mathcal{D}_k\}_{k=0}^{K-1}) = \mathcal{L}_{\text{DM}}(\mathcal{S}_k, \mathcal{D}_k) + \lambda_{\text{loc}} \sum_{c=0}^{C-1} \mathcal{F}\Big(\mathbf{u}_{k,c}(\mathcal{S}_k), \mathbf{v}_c(\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K)\Big),$$

$$\mathcal{L}_{\text{glob}}(\mathbf{w}, \mathcal{S}) = \mathcal{L}_{\text{CE}}(\mathbf{w}, \mathcal{S}) + \lambda_{\text{glob}} \mathcal{L}_{\text{LGKM}}(\mathbf{w}, \mathcal{S})$$

$$\mathcal{L}_{\text{LGKM}}(\mathbf{w}, \mathcal{S}) = \frac{1}{2}\Big(D_{\text{KL}}(\mathcal{R} \| \mathcal{T}) + D_{\text{KL}}(\mathcal{T} \| \mathcal{R})\Big)$$

Clients:

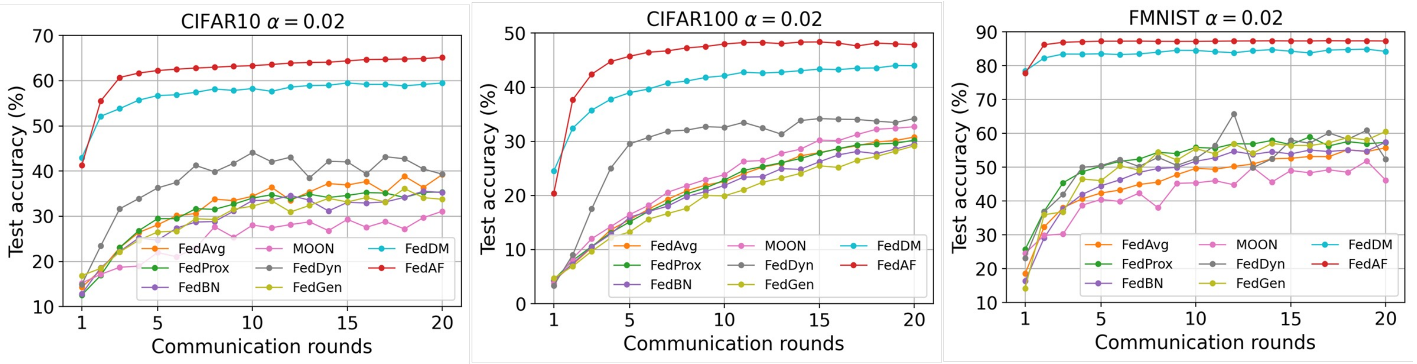
1) download global model and class-wise mean logits
2) update local condensed data using Distribution Matching loss regularized by Collaborative Data Condensation loss

Sever:

1) receive logits from clients and compute soft labels
2) update the global model using cross-entropy loss regularized by Local-Global Knowledge Matching loss

Experiment results

Methods	$\alpha = 0.02$ FMNIST	CIFAR10	CIFAR100	$\alpha = 0.05$ FMNIST	CIFAR10	CIFAR100	$\alpha = 0.1$ FMNIST	CIFAR10	CIFAR100
FedAvg	56.50±5.55	39.71±1.15	30.80±2.20	69.14±5.84	46.51±3.07	33.37±0.75	82.19±5.67	56.15±4.62	39.97±1.53
FedProx	60.38±5.00	36.46±5.39	30.82±0.80	69.33±4.12	45.83±2.23	36.61±1.44	81.56±4.52	58.54±1.87	40.45±1.53
FedBN	58.26±4.28	36.53±2.52	29.73±1.73	72.91±4.69	45.13±2.18	33.73±2.15	77.33±3.07	57.67±3.21	39.84±0.20
MOON	51.33±7.00	33.32±1.13	33.41±0.70	71.41±4.08	47.41±4.59	37.90±0.80	81.61±2.68	57.62±4.99	40.24±0.68
FedDyn	69.79±5.04	45.73±3.98	35.01±2.07	75.19±5.49	57.68±1.84	39.10±0.34	84.73±2.74	59.97±2.20	41.81±1.46
FedGen	61.44±2.07	36.61±1.06	29.20±2.09	75.48±1.83	42.72±2.11	33.56±3.91	82.29±2.53	58.17±2.84	40.23±1.06
FedDM	85.36±0.96	60.28±0.82	44.15±0.30	86.08±0.68	62.97±0.96	46.27±0.98	86.65±0.31	64.88±0.35	47.05±0.13
FedAF	**87.53±0.32**	**65.15±0.86**	**48.71±0.33**	**87.29±0.23**	**67.50±0.76**	**49.49±0.33**	**87.91±0.41**	**69.11±0.86**	**50.61±0.26**

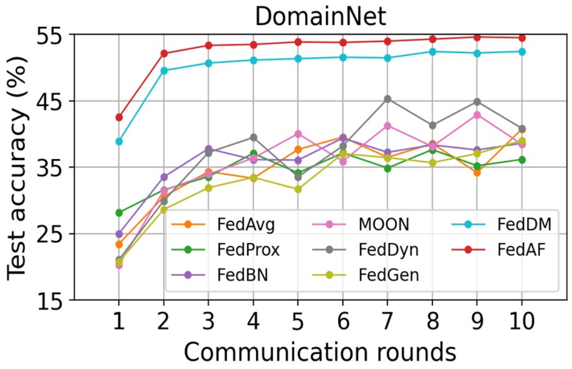


Performance for label-skew non-IID scenarios (FMNIST, CIFAR10, CIFAR100)

Please see our papers for the complete set of learning curves

Experiment results

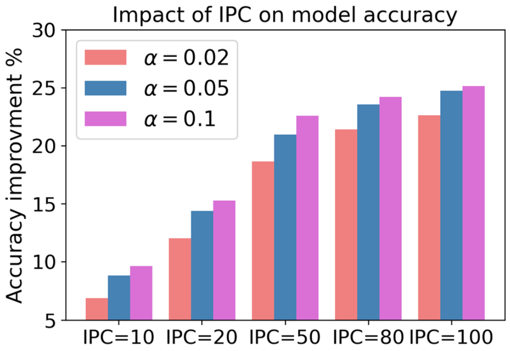
Methods	DomainNet C	I	P	Q	R	S	Avg
FedAvg	43.03	40.76	59.16	39.60	41.03	28.46	42.01
FedProx	44.81	43.76	60.22	38.13	41.55	29.18	42.94
FedBN	46.07	34.27	52.01	43.10	47.33	29.72	42.08
MOON	48.80	37.97	56.26	48.07	42.02	29.72	43.81
FedDyn	48.04	**60.03**	**67.46**	37.73	41.77	32.67	47.95
FedGen	42.77	37.88	54.37	37.33	42.86	25.69	40.15
FedDM	**52.28**	41.38	60.58	62.37	52.45	46.69	52.62
FedAF	51.2	47.05	62.53	**64.6**	**52.64**	**50.06**	**54.68**



Performance for feature-skew non-IID scenarios (DomainNet)

Experiment results

Configuration	α=0.02	α=0.05	α=0.1
IPC=10	53.39±2.09	55.33±0.81	56.15±0.42
IPC=20	58.56±0.55	60.89±0.11	61.79±0.59
IPC=50	65.15±0.86	67.50±0.76	69.11±0.86
IPC=80	67.94±1.18	70.07±0.45	70.72±0.37
IPC=100	69.14±0.56	71.27±0.58	71.66±0.37



Impact of different Image-Per-Class (IPC) values, see our paper for more ablation studies

Configuration	α=0.02	α=0.05	α=0.1
FedAF	65.15±0.86	67.50±0.76	69.11±0.86
w/o CDC	64.16±0.83	65.88±0.93	67.90±0.53
w/o LGKM	64.12±0.85	66.27±1.31	68.14±0.81
FedDM	60.28±0.82	62.97±0.96	64.88±0.35

Impact of individual modules

$$\mathbf{w} \leftarrow \gamma \mathbf{w} + (1-\gamma)\tilde{\mathbf{w}}$$

γ	0.2	0.5	0.8	0.9	1.0
Accuracy	61.30	63.52	66.15	66.97	64.92

Impact of model resampling

Experiment results

Methods	α=0.02	α=0.05	α=0.1
FedAvg	26.48±0.58	32.72±2.47	35.85±3.73
FedProx	26.86±2.69	32.73±2.45	36.25±2.96
FedBN	27.00±2.49	30.29±3.38	35.48±3.45
MOON	29.59±3.57	33.11±3.74	37.26±2.66
FedDyn	22.67±1.54	29.89±4.48	35.38±1.56
FedGen	26.63±2.07	32.48±3.04	38.85±2.00
FedDM	39.18±0.29	39.47±0.66	40.83±0.67
FedAF	**41.10±0.50**	**41.40±0.66**	**42.93±0.29**

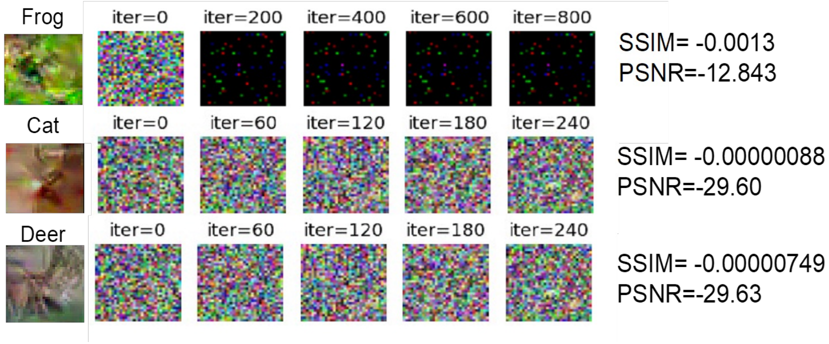
Results with ResNet18

Dataset	α	CNN		ResNet18	
		FedAvg	FedAF	FedAvg	FedAF
FMNIST	0.02	1.46 MB	0.06 MB	42.65 MB	0.06 MB
	0.05		0.09 MB		0.09 MB
	0.1		0.14 MB		0.14 MB
CIFAR10	0.02	1.46 MB	0.22 MB	42.65 MB	0.22 MB
	0.05		0.31 MB		0.31 MB
	0.1		0.44 MB		0.44 MB
CIFAR100	0.02	1.46 MB	1.93 MB	42.65 MB	1.93 MB
	0.05		2.46 MB		2.46 MB
	0.1		3.22 MB		3.22 MB

Comparison for communication cost

Visual privacy and attack robustness

Illustrative examples of learned condensed data



Results from reconstruction attack

CREATING GROWTH, ENHANCING LIVES

Institute of High Performance Computing

IHPC

Thank you so much!

Contact: wang_yuan@ihpc.a-star.edu.sg