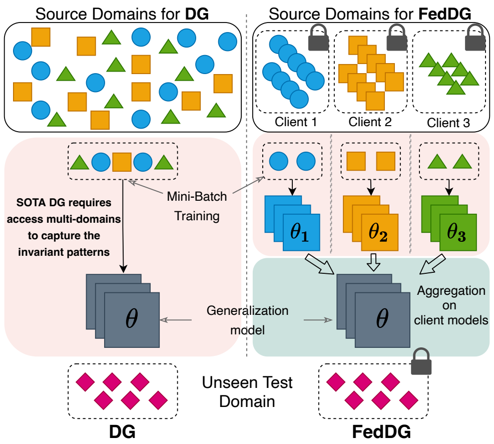
Poster Tag: TUE-AM-377

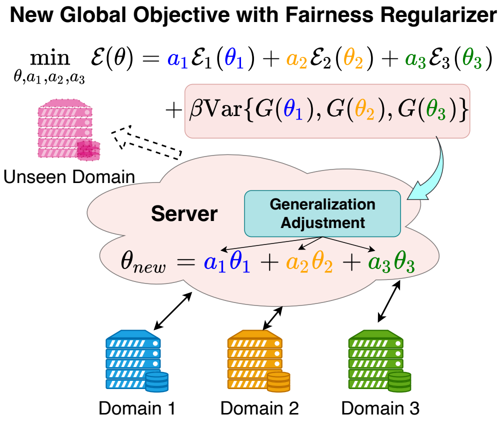
Federated Domain Generalization with Generalization Adjustment

Ruipeng Zhang[1,2], Qinwei Xu[1,2], Jiangchao Yao[1,2], Ya Zhang[1,2], Qi Tian[3], Yanfeng Wang[1,2]

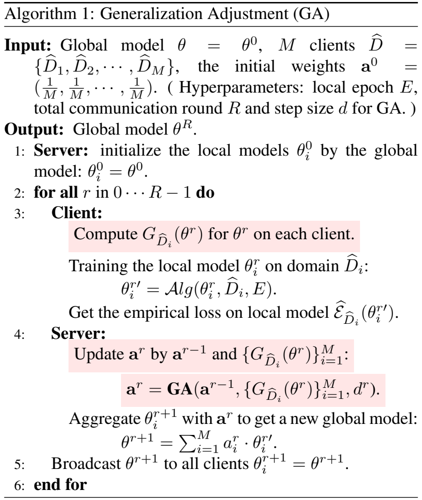
[1] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University
[2] Shanghai AI Laboratory [3] Huawei Cloud & AI

Summary of highlights





(b) Generalization Adjustment



```
Algorithm 1: Generalization Adjustment (GA)
Input: Global model $\theta = \theta^0$, $M$ clients $\widehat{D} = \{\widehat{D}_1, \widehat{D}_2, \cdots, \widehat{D}_M\}$, the initial weights $\mathbf{a}^0 = (\frac{1}{M}, \frac{1}{M}, \cdots, \frac{1}{M})$. ( Hyperparameters: local epoch $E$, total communication round $R$ and step size $d$ for GA. )
Output: Global model $\theta^R$.
Server: initialize the local models $\theta_i^0$ by the global model: $\theta_i^0 = \theta^0$.
for all $r$ in $0 \cdots R-1$ do
    Client:
        Compute $G_{\widehat{D}_i}(\theta^r)$ for $\theta^r$ on each client.
        Training the local model $\theta_i^r$ on domain $\widehat{D}_i$:
            $\theta_i^{r\prime} = \mathcal{A}lg(\theta_i^r, \widehat{D}_i, E)$.
        Get the empirical loss on local model $\widehat{\mathcal{E}}_{\widehat{D}_i}(\theta_i^{r\prime})$.
    Server:
        Update $\mathbf{a}^r$ by $\mathbf{a}^{r-1}$ and $\{G_{\widehat{D}_i}(\theta^r)\}_{i=1}^M$:
            $\mathbf{a}^r = \mathbf{GA}(\mathbf{a}^{r-1}, \{G_{\widehat{D}_i}(\theta^r)\}_{i=1}^M, d^r)$.
        Aggregate $\theta_i^{r+1}$ with $\mathbf{a}^r$ to get a new global model:
            $\theta^{r+1} = \sum_{i=1}^M a_i^r \cdot \theta_i^{r\prime}$.
    Broadcast $\theta^{r+1}$ to all clients $\theta_i^{r+1} = \theta^{r+1}$.
end for
```

- **Introduce a novel global optimization objective for FedDG with a new variance reduction regularizer that can constrain the fairness of the global model.**

- **Design an FL-friendly method named Generalization Adjustment (GA) to optimize the above objective by reweighting the aggregation weights among training clients.**

- **Conduct extensive experiments on four benchmark datasets, demonstrating consistent improvement when combining GA with different federated learning algorithms.**

Poster Tag: TUE-AM-377

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Federated Domain Generalization with Generalization Adjustment

Ruipeng Zhang[1,2] , Qinwei Xu[1,2], Jiangchao Yao[1,2], Ya Zhang[1,2], Qi Tian[3], Yanfeng Wang[1,2]

[1] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University
[2] Shanghai AI Laboratory [3] Huawei Cloud & AI

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

HUAWEI

Domain Generalization (DG)

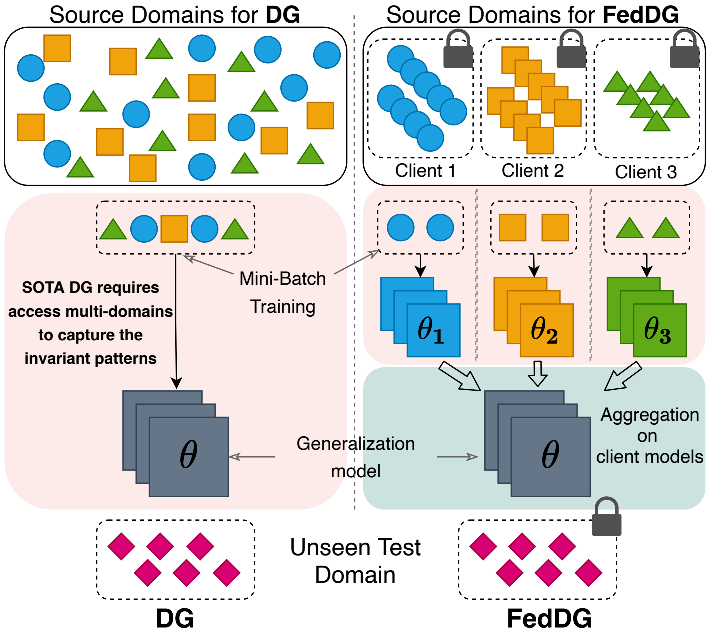
- **Learn a model under domain shifts to unseen domains.**

Federated Domain Generalization (FedDG)

- **Learn a global model in a federated learning system with domain shifts across clients to generalize on clients with unseen domains.**

Problems of existing studies

- **Data heterogeneity in the form of domain shift is very common in federated learning and needs to be addressed.**
- **Most current DG methods cannot be applied on FedDG scenarios.**



FedDG disallows direct data sharing among clients!

Objective of FedAvg:

$$\min_{\theta} \mathcal{E}_{\mathcal{D}}(\theta) \approx \sum_{i=1}^{M} p_i \widehat{\mathcal{E}}_{\widehat{D}_i}(\theta) = \sum_{i=1}^{M} p_i \sum_{j=1}^{N_i} \mathcal{L}\left(f(x_j^i;\theta), y_j^i\right)$$

$$s.t.\ p_i = \frac{N_i}{\sum_{i'=1}^{M} N_{i'}}.$$

$p_\mathrm{i} \neq$ **data distribution!**

Observation:

- **Above objective cannot describe the generalizable goal under domain shifts.**
- **the flatness (fair performance) of each domain can be reflected by the generalization gaps.**

$$G_{\widehat{D}_i}(\theta) = G_{\widehat{D}_i}(\textstyle\sum_j a_j \theta_j^*) = \widehat{\mathcal{E}}_{\widehat{D}_i}(\textstyle\sum_j a_j \theta_j^*) - \widehat{\mathcal{E}}_{\widehat{D}_i}(\theta_i^*)$$

Assume that a global model with fair performance among all clients may lead to better generalization performance.

Objective with fairness:

$$\min_{\theta_1,\dots,\theta_M,\mathbf{a}} \widehat{\mathcal{E}}_{\widehat{D}}(\theta) = \sum_{i=1}^{M} a_i \widehat{\mathcal{E}}_{\widehat{D}_i}(\theta) + \beta \mathrm{Var}(\{G_{\widehat{D}_i}(\theta)\}_{i=1}^{M})$$

$$s.t. \sum_{i=1}^{M} a_i = 1,\ \theta = \sum_{i=1}^{M} a_i \cdot \theta_i,\ \text{and}\ \forall i,\ a_i \geq 0.$$

Considering the variance of generalization gaps among local clients to guarantee the flatness of the optimal global model on all domains.

How to optimize it under federated learning?

Objective with fairness:

$$\min_{\theta_1,\ldots,\theta_M,\mathbf{a}} \widehat{\mathcal{E}}_{\widehat{D}}(\theta) = \sum_{i=1}^{M} a_i \widehat{\mathcal{E}}_{\widehat{D}_i}(\theta) + \beta \mathrm{Var}(\{G_{\widehat{D}_i}(\theta)\}_{i=1}^{M})$$

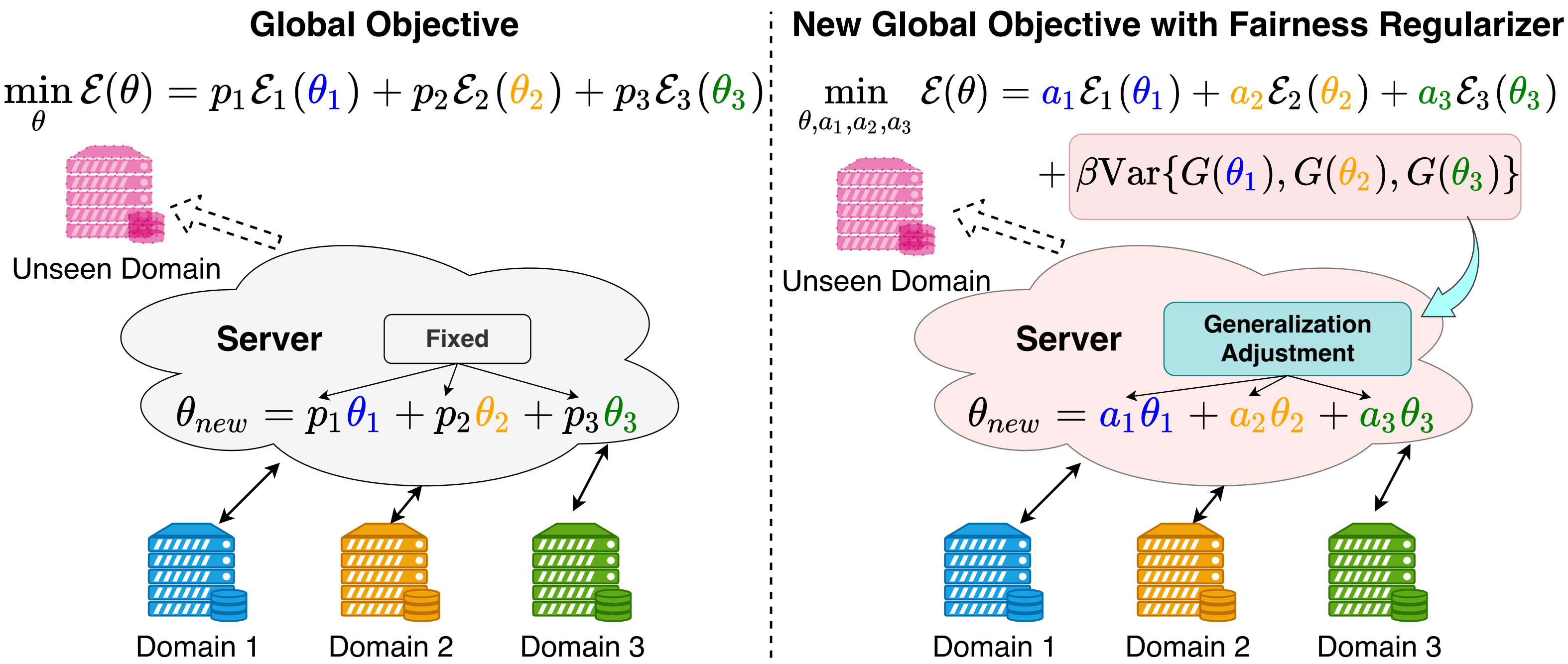
$$s.t. \sum_{i=1}^{M} a_i = 1,\ \theta = \sum_{i=1}^{M} a_i \cdot \theta_i,\ \text{and}\ \forall i,\ a_i \geq 0.$$

Mechanism of relationship between a_k and $\mathbf{G}_{\widehat{\mathbf{D}_k}}(\boldsymbol{\theta}_\mathbf{k})$

$$\theta = \theta_k + \Delta\theta,\ \text{where}\ \ \Delta\theta = (1 - a_k)\theta_k + \textstyle\sum_{i \neq k} a_i \theta_i.$$

$$a_k \uparrow \Longrightarrow \Delta\theta \downarrow \Rightarrow \mathbf{G}_{\widehat{\mathbf{D}_k}}(\boldsymbol{\theta}_\mathbf{k}) \downarrow$$

$$a_k \downarrow \Longrightarrow \Delta\theta \uparrow \Rightarrow \mathbf{G}_{\widehat{\mathbf{D}_k}}(\boldsymbol{\theta}_\mathbf{k}) \uparrow$$



Global Objective

$$\min_{\theta} \mathcal{E}(\theta) = p_1 \mathcal{E}_1(\theta_1) + p_2 \mathcal{E}_2(\theta_2) + p_3 \mathcal{E}_3(\theta_3)$$

$$\theta_{new} = p_1 \theta_1 + p_2 \theta_2 + p_3 \theta_3$$

New Global Objective with Fairness Regularizer

$$\min_{\theta, a_1, a_2, a_3} \mathcal{E}(\theta) = a_1 \mathcal{E}_1(\theta_1) + a_2 \mathcal{E}_2(\theta_2) + a_3 \mathcal{E}_3(\theta_3)$$

$$+ \beta \mathrm{Var}\{G(\theta_1), G(\theta_2), G(\theta_3)\}$$

$$\theta_{new} = a_1 \theta_1 + a_2 \theta_2 + a_3 \theta_3$$

(a) FedAvg (b) Generalization Adjustment

Algorithm 1: Generalization Adjustment (GA)

```
Input: Global model $\theta = \theta^0$, $M$ clients $\widehat{D} = \{\widehat{D}_1, \widehat{D}_2, \cdots, \widehat{D}_M\}$, the initial weights $\mathbf{a}^0 = (\frac{1}{M}, \frac{1}{M}, \cdots, \frac{1}{M})$. ( Hyperparameters: local epoch $E$, total communication round $R$ and step size $d$ for GA. )
Output: Global model $\theta^R$.
Server: initialize the local models $\theta_i^0$ by the global model: $\theta_i^0 = \theta^0$.
for all $r$ in $0 \cdots R-1$ do
    Client:
        Compute $G_{\widehat{D}_i}(\theta^r)$ for $\theta^r$ on each client.
        Training the local model $\theta_i^r$ on domain $\widehat{D}_i$:
            $\theta_i^{r\prime} = \mathcal{A}lg(\theta_i^r, \widehat{D}_i, E)$.
        Get the empirical loss on local model $\widehat{\mathcal{E}}_{\widehat{D}_i}(\theta_i^{r\prime})$.
    Server:
        Update $\mathbf{a}^r$ by $\mathbf{a}^{r-1}$ and $\{G_{\widehat{D}_i}(\theta^r)\}_{i=1}^M$:
            $\mathbf{a}^r = \mathbf{GA}(\mathbf{a}^{r-1}, \{G_{\widehat{D}_i}(\theta^r)\}_{i=1}^M, d^r)$.
        Aggregate $\theta_i^{r+1}$ with $\mathbf{a}^r$ to get a new global model:
            $\theta^{r+1} = \sum_{i=1}^M a_i^r \cdot \theta_i^{r\prime}$.
    Broadcast $\theta^{r+1}$ to all clients $\theta_i^{r+1} = \theta^{r+1}$.
end for
```

Client-side:

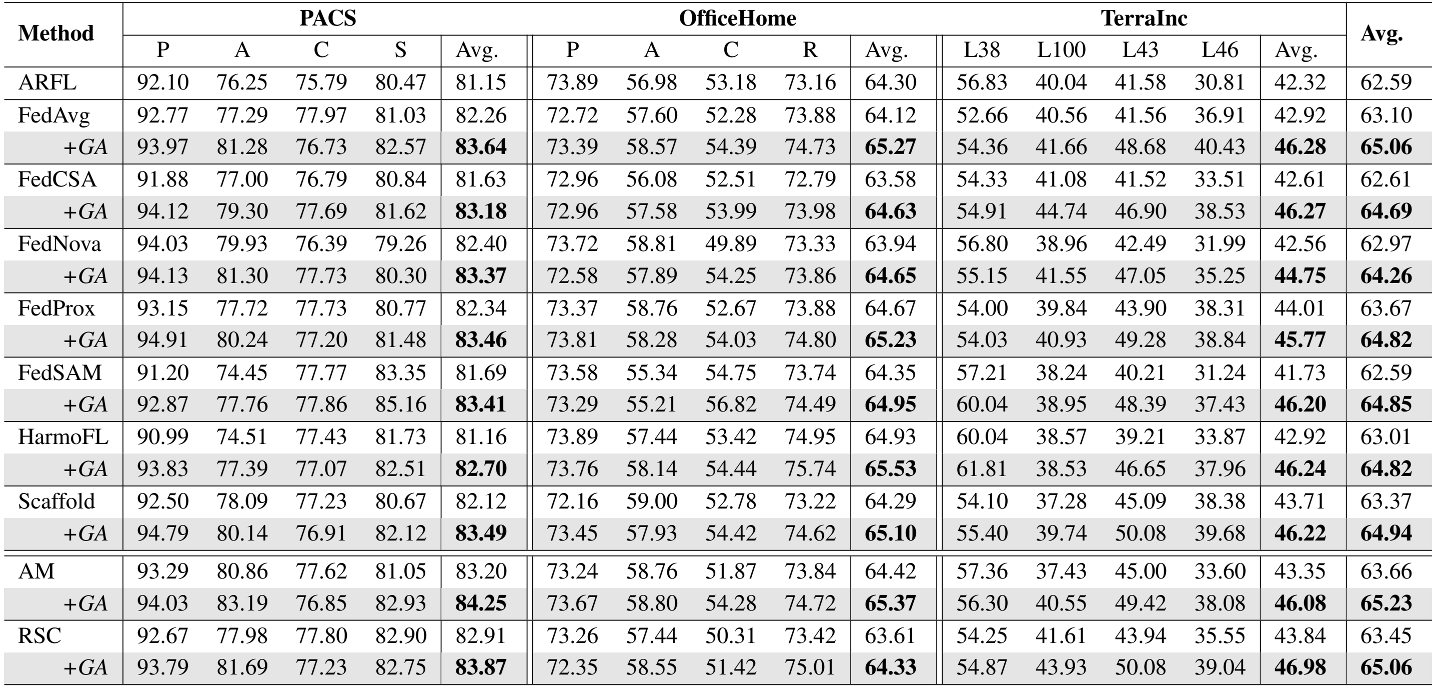
calculate the generalization gap

$$G_{\widehat{D}_i}(\theta^r) = \widehat{\mathcal{E}}_{\widehat{D}_i}(\theta^r) - \widehat{\mathcal{E}}_{\widehat{D}_i}(\theta_i^{r-1\prime}),\ i = 1, 2, \ldots, M.$$

Server-side:

adjust the aggregation weights $\mathbf{a_i^r}$ by $\mathbf{G_{\widehat{D}_i}(\theta^r)}$

$$a_i^{r\prime} = \frac{(G_{\widehat{D}_i}(\theta^r) - \mu) * d^r}{\max_j(G_{\widehat{D}_j}(\theta^r) - \mu)} + a_i^{r-1},\ a_i^r = \frac{a_i^{r\prime}}{\sum_{i=1}^M a_i^{r\prime}},$$



	Method	PACS					OfficeHome					TerraInc					Avg.
		P	A	C	S	Avg.	P	A	C	R	Avg.	L38	L100	L43	L46	Avg.	
	ARFL	92.10	76.25	75.79	80.47	81.15	73.89	56.98	53.18	73.16	64.30	56.83	40.04	41.58	30.81	42.32	62.59
baseline	FedAvg	92.77	77.29	77.97	81.03	82.26	72.72	57.60	52.28	73.88	64.12	52.66	40.56	41.56	36.91	42.92	63.10
	+GA	93.97	81.28	76.73	82.57	**83.64**	73.39	58.57	54.39	74.73	**65.27**	54.36	41.66	48.68	40.43	**46.28**	**65.06**
FL SOTA	FedCSA	91.88	77.00	76.79	80.84	81.63	72.96	56.08	52.51	72.79	63.58	54.33	41.08	41.52	33.51	42.61	62.61
	+GA	94.12	79.30	77.69	81.62	**83.18**	72.96	57.58	53.99	73.98	**64.63**	54.91	44.74	46.90	38.53	**46.27**	**64.69**
	FedNova	94.03	79.93	76.39	79.26	82.40	73.72	58.81	49.89	73.33	63.94	56.80	38.96	42.49	31.99	42.56	62.97
	+GA	94.13	81.30	77.73	80.30	**83.37**	72.58	57.89	54.25	73.86	**64.65**	55.15	41.55	47.05	35.25	**44.75**	**64.26**
	FedProx	93.15	77.72	77.73	80.77	82.34	73.37	58.76	52.67	73.88	64.67	54.00	39.84	43.90	38.31	44.01	63.67
	+GA	94.91	80.24	77.20	81.48	**83.46**	73.81	58.28	54.03	74.80	**65.23**	54.03	40.93	49.28	38.84	**45.77**	**64.82**
	FedSAM	91.20	74.45	77.77	83.35	81.69	73.58	55.34	54.75	73.74	64.35	57.21	38.24	40.21	31.24	41.73	62.59
	+GA	92.87	77.76	77.86	85.16	**83.41**	73.29	55.21	56.82	74.49	**64.95**	60.04	38.95	48.39	37.43	**46.20**	**64.85**
	HarmoFL	90.99	74.51	77.43	81.73	81.16	73.89	57.44	53.42	74.95	64.93	60.04	38.57	39.21	33.87	42.92	63.01
	+GA	93.83	77.39	77.07	82.51	**82.70**	73.76	58.14	54.44	75.74	**65.53**	61.81	38.53	46.65	37.96	**46.24**	**64.82**
	Scaffold	92.50	78.09	77.23	80.67	82.12	72.16	59.00	52.78	73.22	64.29	54.10	37.28	45.09	38.38	43.71	63.37
	+GA	94.79	80.14	76.91	82.12	**83.49**	73.45	57.93	54.42	74.62	**65.10**	55.40	39.74	50.08	39.68	**46.22**	**64.94**
DG SOTA	AM	93.29	80.86	77.62	81.05	83.20	73.24	58.76	51.87	73.84	64.42	57.36	37.43	45.00	33.60	43.35	63.66
	+GA	94.03	83.19	76.85	82.93	**84.25**	73.67	58.80	54.28	74.72	**65.37**	56.30	40.55	49.42	38.08	**46.08**	**65.23**
	RSC	92.67	77.98	77.80	82.90	82.91	73.26	57.44	50.31	73.42	63.61	54.25	41.61	43.94	35.55	43.84	63.45
	+GA	93.79	81.69	77.23	82.75	**83.87**	72.35	58.55	51.42	75.01	**64.33**	54.87	43.93	50.08	39.04	**46.98**	**65.06**

GA achieves consistent improvements on top of SOTA methods on four benchmarks!

DomainNet

Method	**C**	**I**	**P**	**Q**	**R**	**S**	**Avg.**
ARFL	69.93	31.32	60.68	57.45	67.76	60.62	57.96
FedAvg	67.92	32.77	60.27	52.90	68.72	61.15	57.29
+*GA*	71.86	34.40	63.25	57.50	67.26	67.15	**60.24**
FedCSA	68.94	33.67	61.66	58.25	67.25	61.15	58.49
+*GA*	70.34	33.71	64.22	56.85	66.06	67.33	**59.72**
FedNova	68.45	32.95	61.70	59.05	67.21	61.57	58.49
+*GA*	73.29	34.09	64.38	57.05	68.08	65.25	**60.36**
FedProx	68.55	32.12	60.79	55.63	68.17	61.20	57.75
+*GA*	69.39	33.26	63.25	57.15	67.50	65.79	**59.39**
FedSAM	66.47	34.97	56.90	51.11	66.28	55.82	55.26
+*GA*	72.62	36.30	64.62	57.75	69.35	65.43	**61.01**
HarmoFL	71.39	34.70	61.23	57.50	67.01	56.77	58.10
+*GA*	72.81	35.77	63.73	59.30	68.08	64.35	**60.67**
Scaffold	68.04	33.20	60.60	54.39	67.32	60.88	57.41
+*GA*	71.67	34.93	62.20	57.70	67.46	66.97	**60.16**
AM	71.91	32.54	63.70	56.87	67.80	69.42	60.37
+*GA*	74.33	35.31	64.54	58.55	68.61	72.02	**62.23**
RSC	70.96	34.25	60.31	55.20	66.91	63.84	58.58
+*GA*	71.96	35.62	62.52	56.95	67.13	64.98	**59.86**

GA achieves consistent improvements on top of SOTA methods on four benchmarks!

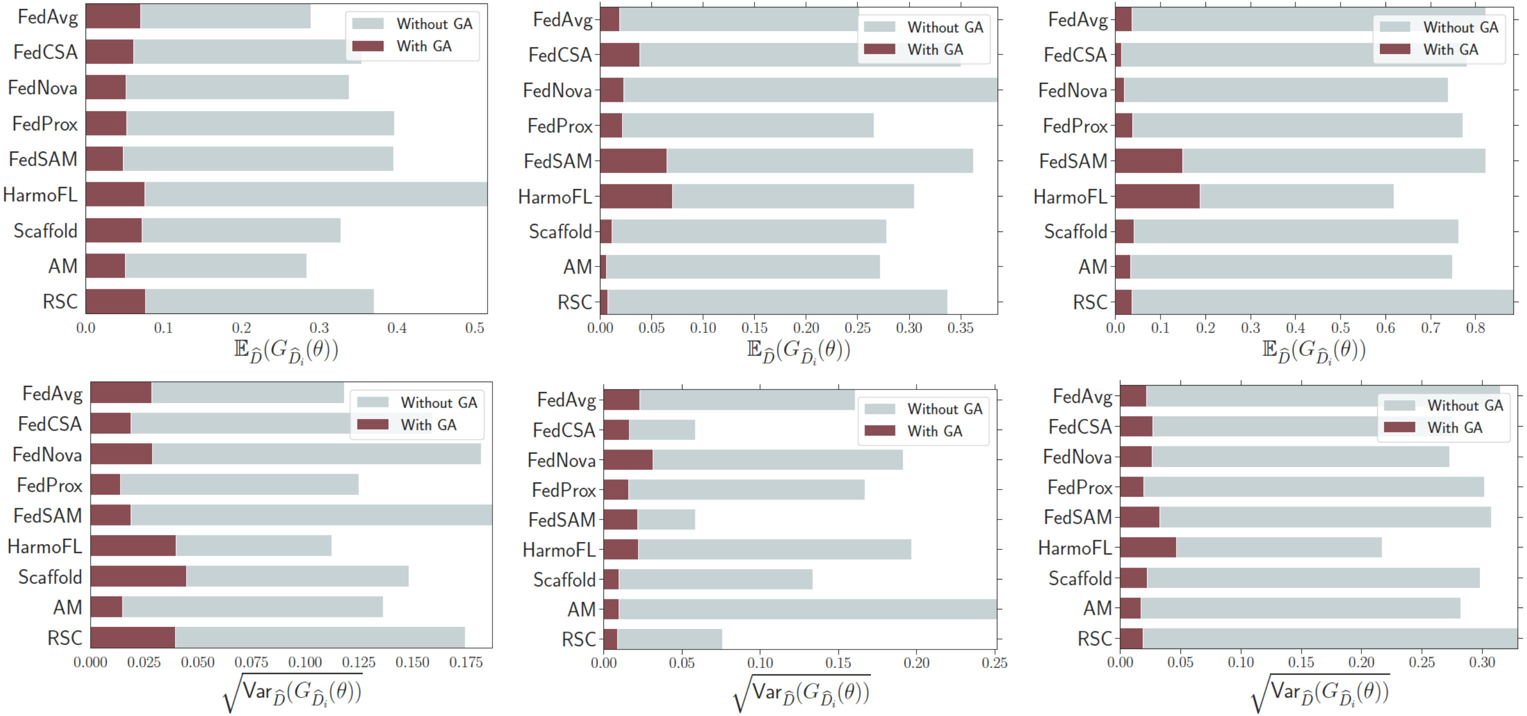
Method	PACS					OfficeHome				
	P	A	C	S	Avg,	P	A	C	R	Avg.
DSBN	96.26	82.23	80.99	77.50	84.25	73.34	56.49	53.64	73.03	64.13
+*GA*	96.56	83.18	81.21	80.11	**85.26**	72.91	57.50	54.99	73.85	**64.81**
Tent	96.92	85.94	83.06	91.39	86.83	74.63	57.95	56.48	74.67	65.92
+*GA*	97.16	86.77	83.98	83.28	**87.80**	74.53	59.79	56.61	75.36	**66.57**

GA also shows improvements on top of test-time adaptation methods.

Table 1. Results with more clients & with more advanced SOTAs.

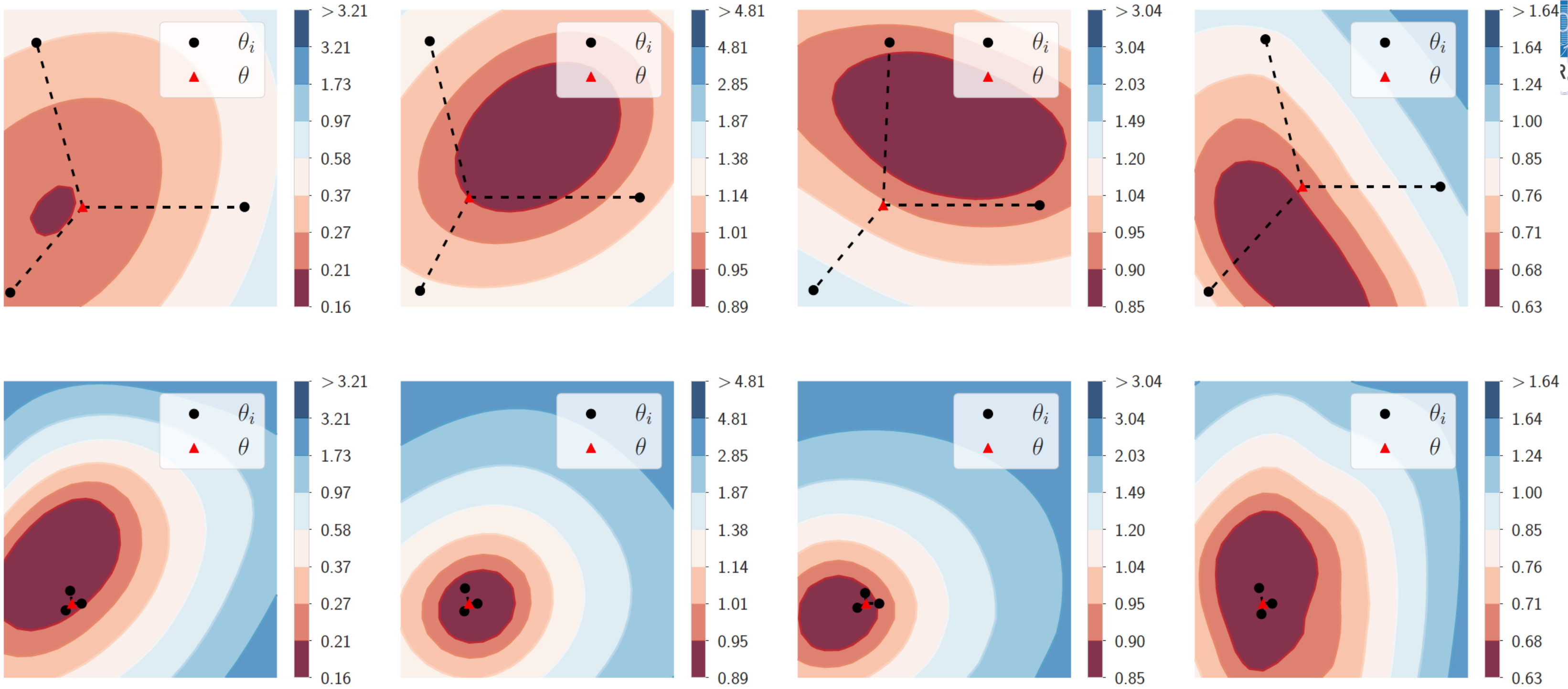
Dataset	*more clients*		*suggested SOTAs*			
	FedAvg	Best[1]	ELCFS	FedSR	FedASAM*	CCST
PACS	80.33	81.62	84.07	83.70	82.04	83.48
with GA	**81.99**	**82.72**	**84.88**	**84.66**	**83.57**	**84.35**
OfficeHome	63.38	64.08	62.88	64.29	64.32	64.25
with GA	**64.40**	**65.04**	**64.60**	**64.65**	**64.80**	**65.42**

GA also shows improvements with SOTA FedDG methods.



(a) PACS (b) OfficeHome (c) TerraInc

GA can effectively reduce the variance of the generalization gap!



(a) Photo (b) Art (c) Cartoon (d) Sketch

GA will lead the global model to converge on more flatness minima of unseen domain's loss surface.

Thanks for your watching!

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

未来媒体网络协同创新中心
Cooperative Medianet Innovation Center

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

HUAWEI

Github repo