ICLR 2025 Poster

Federated Domain Generalization with Data-free On-server Matching Gradient

Trong-Binh Nguyen[1], Minh-Duong Nguyen[1], Jinsun Park[1], Quoc-Viet Pham[2], Won Joo Hwang[1]

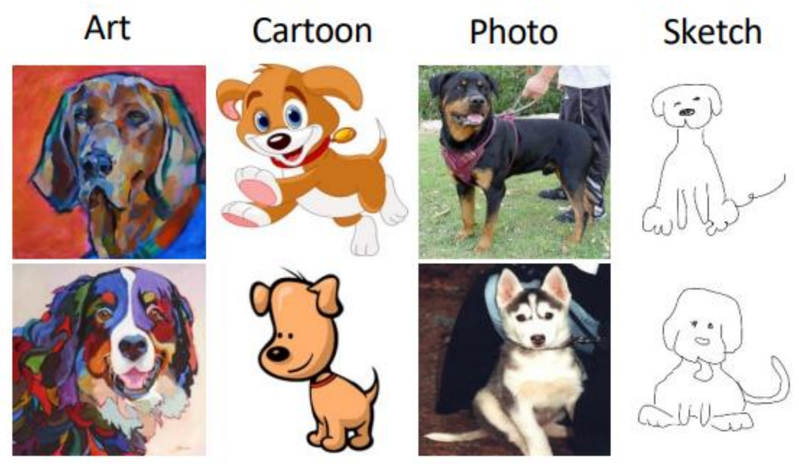
[1]Pusan National University, Republic of Korea; [2]Trinity College Dublin, Ireland

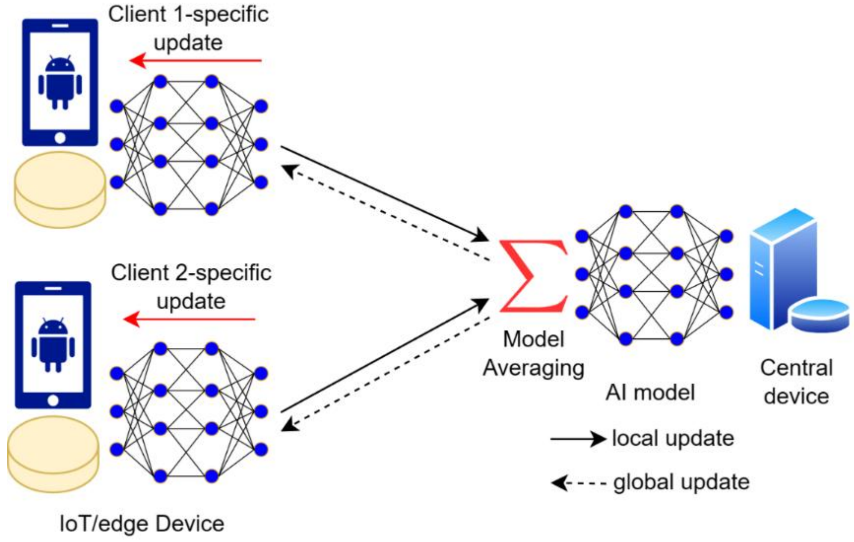
부산대학교 PUSAN NATIONAL UNIVERSITY

QuAIL Quantum AI Lab

Overview

Federated Domain Generalization

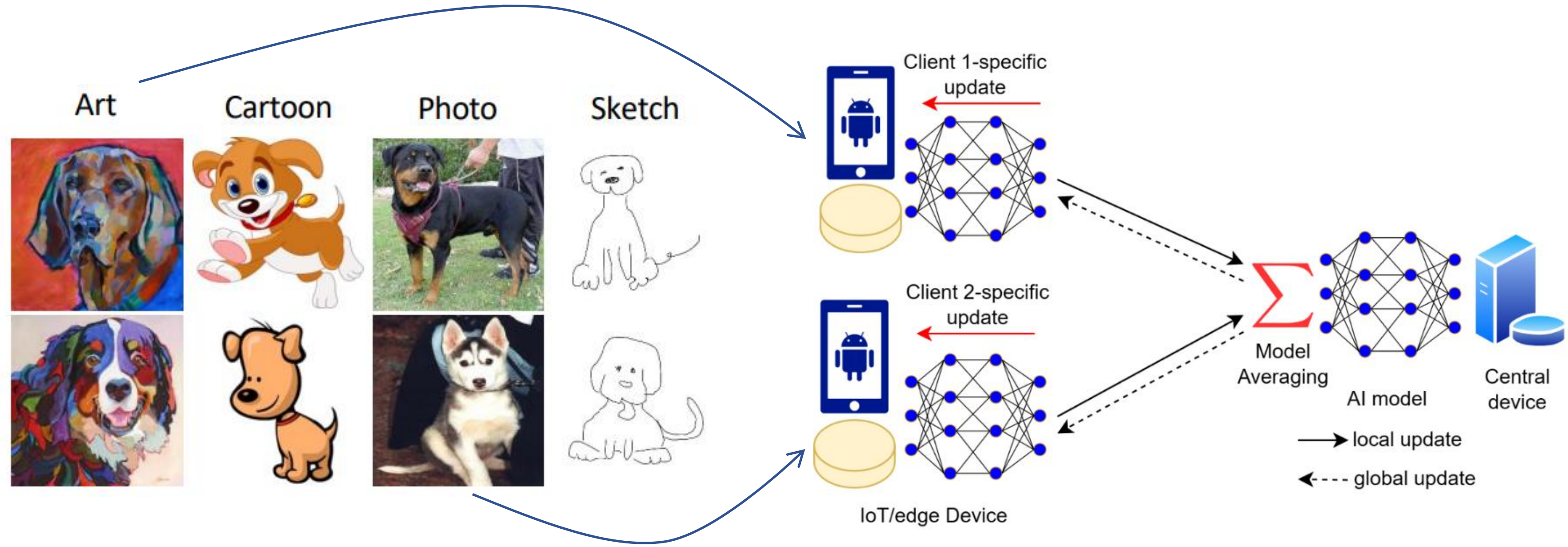




[*] PACS dataset, https://paperswithcode.com/dataset/pacs

Overview

Federated Domain Generalization

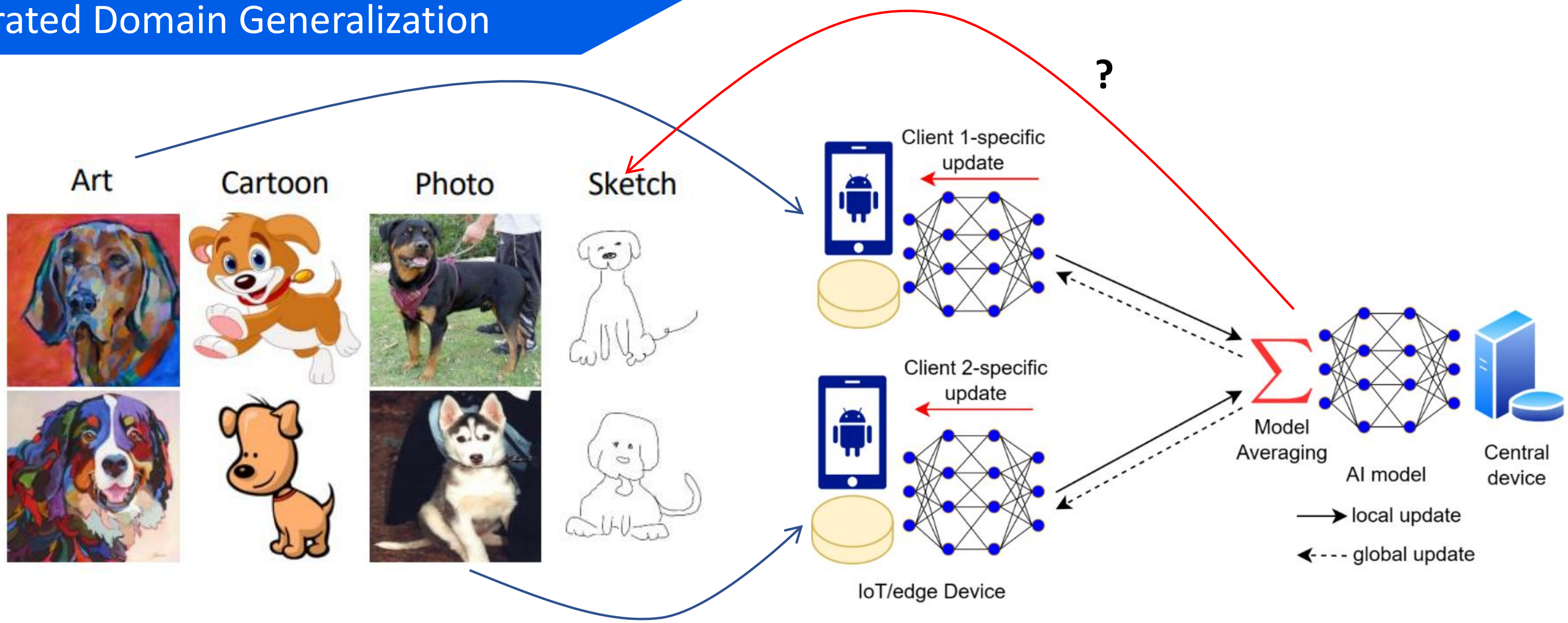


- Data collected from different domains (e.g., Art, Photo, Sketch), client models trained on different domains ➔ heterogeneous distributions.

[*] PACS dataset, https://paperswithcode.com/dataset/pacs

Overview

Federated Domain Generalization



- Data collected from different domains (e.g., Art, Photo, Sketch), client models trained on different domains ➔ heterogeneous distributions.
- Federated Domain Generalization (FDG) focuses on the domain generalization problem in FL, aiming to ensure that performance on test domains is as good as performance on training domains.

[*] PACS dataset, https://paperswithcode.com/dataset/pacs

Revisiting Domain Generalization (DG)

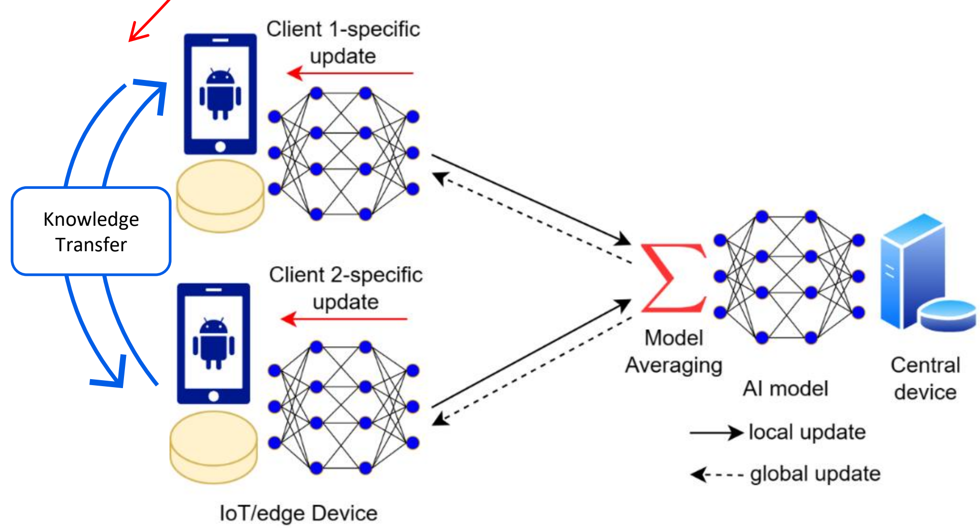
$$\overbrace{\mathcal{E}(\theta;\mathcal{D}_{\mathcal{T}})}^{\text{Generalization gap on an unseen domain } \mathcal{D}_{\mathcal{T}}} \leq \sum_{u\in\mathcal{U}_{\mathcal{S}}} \rho_u \left(\underbrace{\mathcal{E}(\theta;\mathcal{D}_u)}_{\text{Local risk}} + \overbrace{\sum_{v\in\mathcal{U}_{\mathcal{T}}}^{v\neq u} d_{\mathcal{H}}(\mathcal{D}_u,\mathcal{D}_v)}^{\text{Domain Divergence}} + \mathcal{O}\left[\sqrt{\overbrace{\frac{\log M + \log 1 : \delta}{2N_u}}^{\text{Gap due to finite sampling}}}\right]\right) \underbrace{+\ \zeta^*}_{\text{Optimal joint risk}}$$

Revisiting Domain Generalization (DG)

$$\overbrace{\mathcal{E}(\theta;\mathcal{D}_{\mathcal{T}})}^{\text{Generalization gap on an unseen domain } \mathcal{D}_{\mathcal{T}}} \leq \sum_{u\in\mathcal{U}_S} \rho_u \left(\underbrace{\mathcal{E}(\theta;\mathcal{D}_u)}_{\text{Local risk}} + \overbrace{\sum_{v\in\mathcal{U}_T}^{v\neq u} d_{\mathcal{H}}(\mathcal{D}_u,\mathcal{D}_v)}^{\text{Domain Divergence}} + \mathcal{O}\left[\overbrace{\sqrt{\frac{\log M + \log 1:\delta}{2N_u}}}^{\text{Gap due to finite sampling}} \right] \right) + \underbrace{\zeta^*}_{\text{Optimal joint risk}}$$

Issues of Conventional DG approaches to FDG

- Reduce the domain divergence, i.e., minimizing the distance between domains' data.

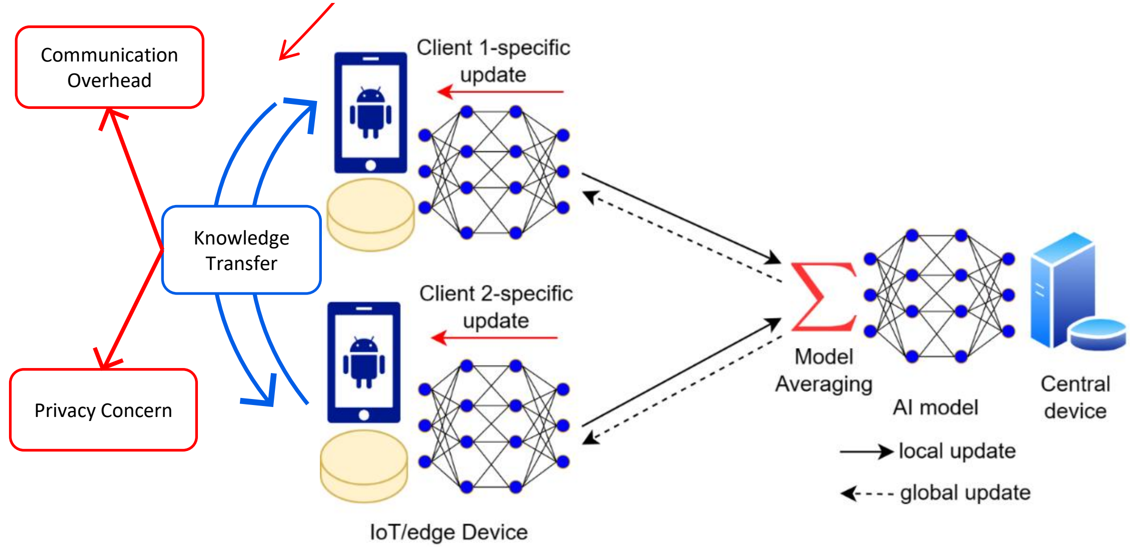


Revisiting Domain Generalization (DG)

$$\overbrace{\mathcal{E}(\theta;\mathcal{D}_{\mathcal{T}})}^{\text{Generalization gap on an unseen domain } \mathcal{D}_{\mathcal{T}}} \le \sum_{u\in\mathcal{U}_S} \rho_u \left(\underbrace{\mathcal{E}(\theta;\mathcal{D}_u)}_{\text{Local risk}} + \overbrace{\sum_{v\in\mathcal{U}_{\mathcal{T}}}^{v\neq u} d_{\mathcal{H}}(\mathcal{D}_u,\mathcal{D}_v)}^{\text{Domain Divergence}} + \mathcal{O}\left[\overbrace{\sqrt{\frac{\log M + \log 1:\delta}{2N_u}}}^{\text{Gap due to finite sampling}}\right]\right) + \underbrace{\zeta^*}_{\text{Optimal joint risk}}$$

Issues of Conventional DG approaches to FDG

- Reduce the domain divergence, i.e., minimizing the distance between domains' data.
- Require data accessibility among users → communication overhead, privacy protection.

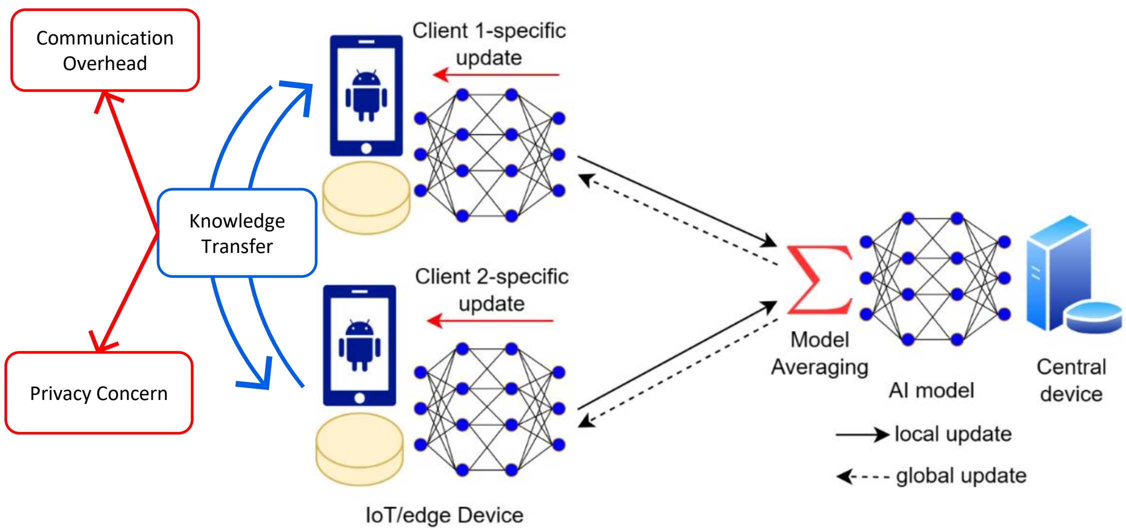


Revisiting Domain Generalization (DG)

$$\overbrace{\mathcal{E}(\theta;\mathcal{D}_{\mathcal{T}})}^{\text{Generalization gap on an unseen domain } \mathcal{D}_{\mathcal{T}}} \leq \sum_{u\in\mathcal{U}_S} \rho_u \left(\underbrace{\mathcal{E}(\theta;\mathcal{D}_u)}_{\text{Local risk}} + \overbrace{\sum_{v\in\mathcal{U}_{\mathcal{T}}}^{v\neq u} d_{\mathcal{H}}(\mathcal{D}_u,\mathcal{D}_v)}^{\text{Domain Divergence}} + \mathcal{O}\left[\overbrace{\sqrt{\frac{\log M + \log 1\!:\delta}{2N_u}}}^{\text{Gap due to finite sampling}} \right] \right) + \underbrace{\zeta^*}_{\text{Optimal joint risk}}$$

Issues of Conventional DG approaches to FDG

- Reduce the domain divergence, i.e., minimizing the distance between domains' data.
- Require data accessibility among users → communication overhead, privacy protection.



How to make the models trained on heterogeneous network domains scalable and generalizable to new/unseen domains in the distributed network settings?

FDG via On-Server Optimization

On-Server Optimization

$$\arg\min \sum_{u\in\mathcal{U}_S} \rho_u \left(\overbrace{\mathcal{E}(\theta;\mathcal{D}_u)}^{\text{FL Loss}} + \overbrace{\sum_{v\in\mathcal{U}_S}^{v\neq u} d_{\mathcal{H}}\big(\underbrace{g(\theta;\mathcal{D}_u)\,,g(\theta;\mathcal{D}_v)}_{\text{Local Gradients}}\big)}^{\text{Domain Divergence}} \right)$$

- Exploit ***local gradient trajectories*** at the server to estimate the ***domain divergence***.
- Identify an optimal aggregated gradient for achieving ***domain invariance***
  → no need for data exchange.

FDG via On-Server Optimization

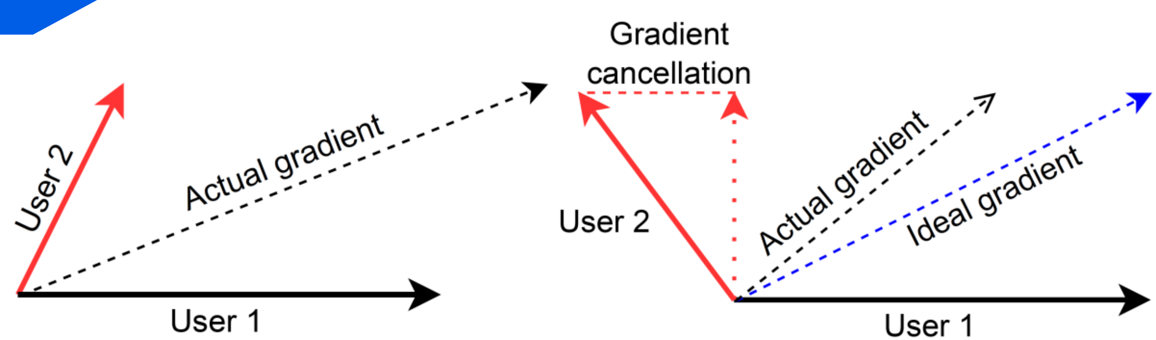
On-Server Optimization

$$\arg\min \sum_{u\in\mathcal{U}_S} \rho_u \left(\overbrace{\mathcal{E}(\theta;\mathcal{D}_u)}^{\text{FL Loss}} + \overbrace{\sum_{v\in\mathcal{U}_S}^{v\neq u} d_{\mathcal{H}}\big(\underbrace{g(\theta;\mathcal{D}_u)\,,g(\theta;\mathcal{D}_v)}_{\text{Local Gradients}}\big)}^{\text{Domain Divergence}} \right)$$

- Exploit ***local gradient trajectories*** at the server to estimate the ***domain divergence***.
- Identify an optimal aggregated gradient for achieving ***domain invariance***
  → no need for data exchange.

Gradient Matching Motivation

- Gradient Conflict between clients indicate the domain divergence.
- Naïve averaging of vanilla FL lead to negative transfer.



On-Server Gradient Matching

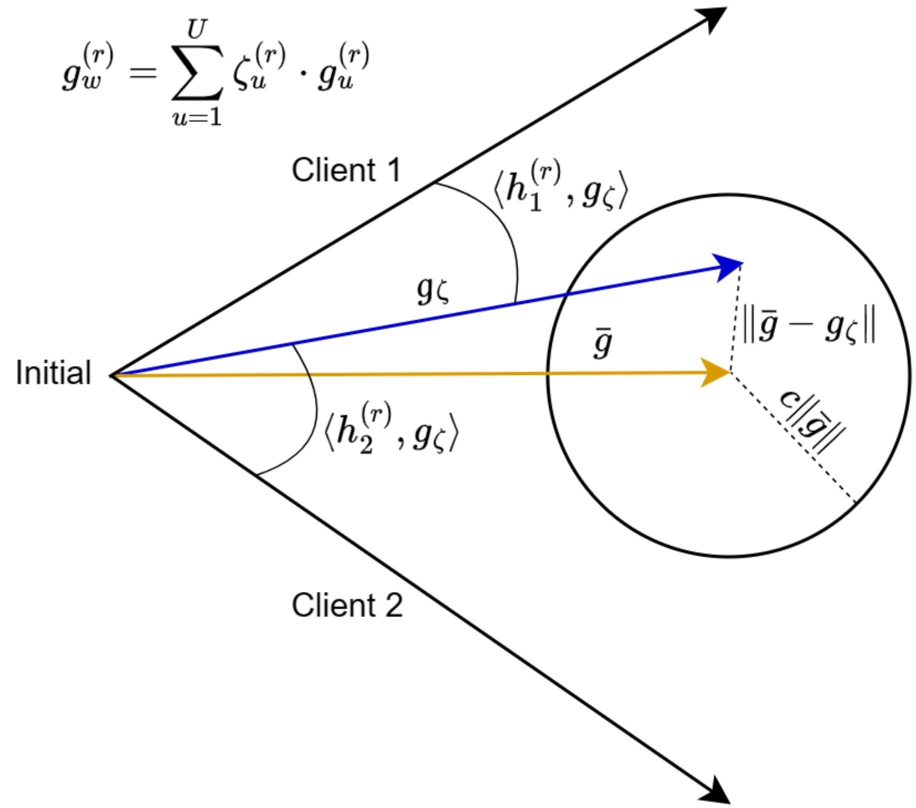
Gradient Matching

- Reduce the cosine similarity between updated trajectories of different clients.
- Reduce the negative transfer on distinct clients.
- Each distributed client must compete to obtain the best updated policy.

Search Space Limitation

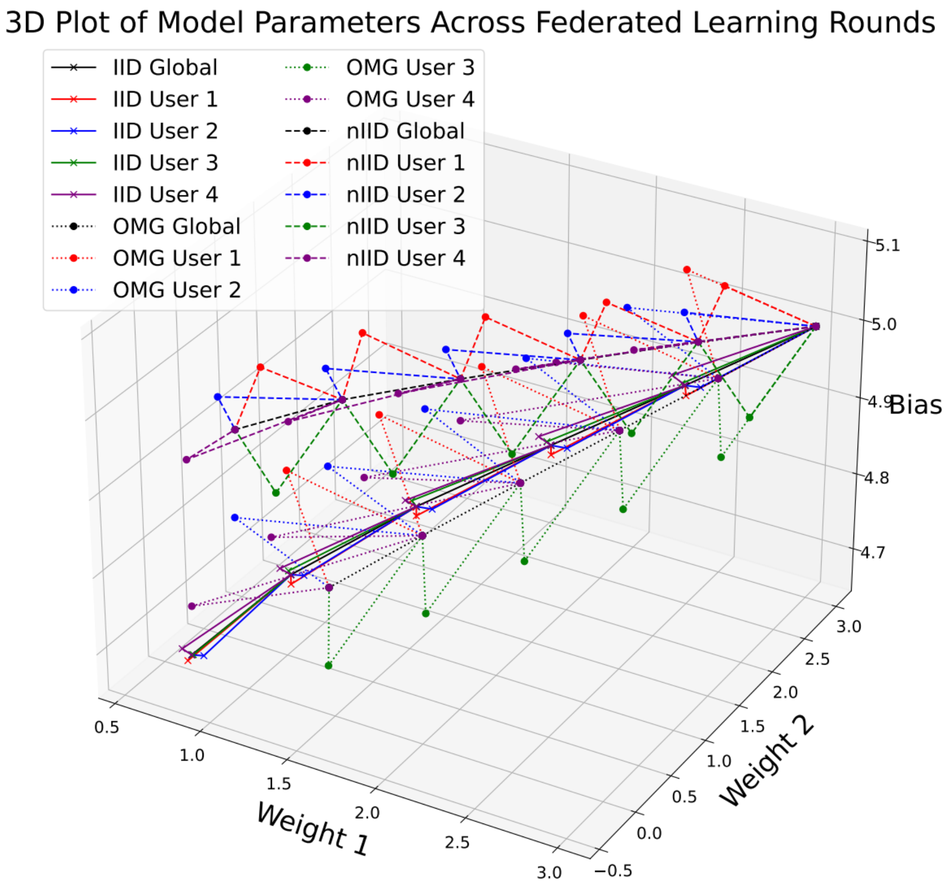
- Reduce computation resources required.
- Mitigate the risk of getting trapped in local minima.

$$g_w^{(r)} = \sum_{u=1}^{U} \zeta_u^{(r)} \cdot g_u^{(r)}$$

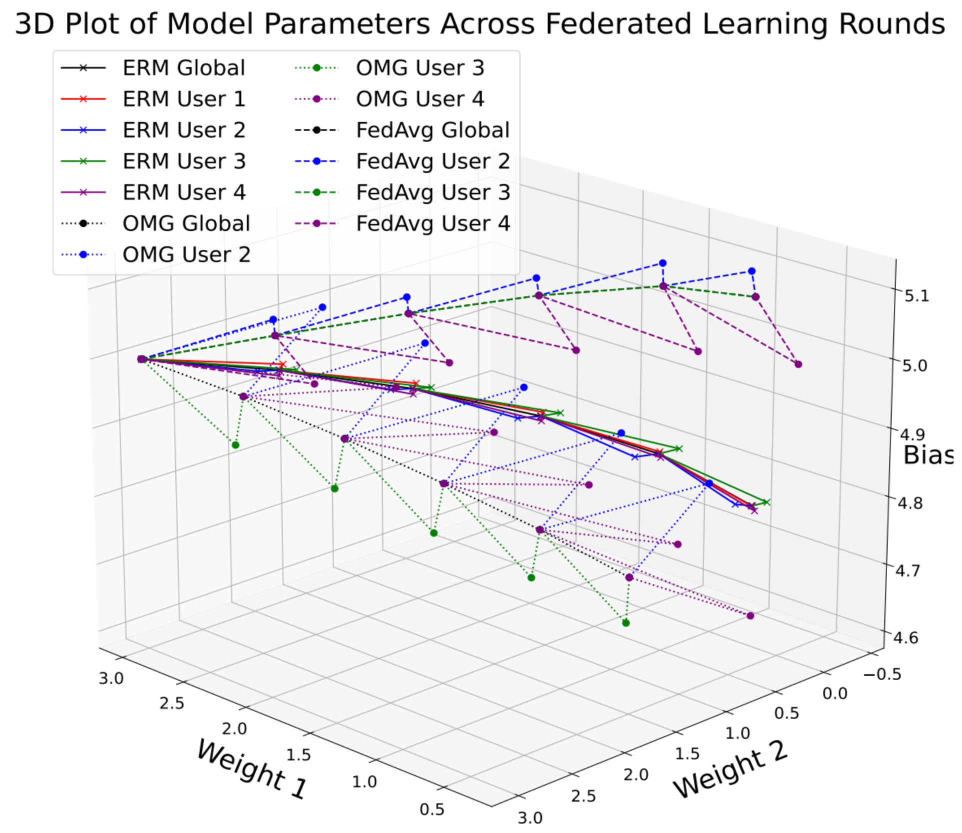


$$\underbrace{\max_{g_\zeta} \min_{u \in U} \langle h_u^{(r)}, g_\zeta \rangle}_{\text{Gradient matching}} - \underbrace{\frac{\lambda}{2}\left(\|\bar{g} - g_\zeta\| - c^2 \|\bar{g}\|^2 \right)}_{\text{Searching space limitation}}$$

Illustrative Toy Task



FL setting: all clients participated.



FDG setting: 1 client being dropped out.

Experimental Evaluation

Algorithm	VLCS					PACS					OfficeHome				
	V	L	C	S	Avg	P	A	C	S	Avg	A	C	P	R	Avg
FedAvg	72.5 ± 0.8	61.1 ± 0.9	93.6 ± 1.0	65.4 ± 0.3	73.1	92.7 ± 0.6	77.2 ± 1.0	77.9 ± 0.5	81.0 ± 0.8	82.7	57.7 ± 0.9	48.3 ± 0.1	72.8 ± 0.2	75.3 ± 0.1	63.5
FedGA	74.4 ± 0.1	56.9 ± 1.0	94.3 ± 0.6	68.9 ± 0.9	73.4	93.9 ± 0.2	81.2 ± 0.7	76.7 ± 0.4	82.5 ± 0.1	83.5	58.5 ± 0.4	54.3 ± 0.6	73.3 ± 0.8	74.7 ± 1.0	65.2
FedSAM	74.5 ± 0.3	58.0 ± 0.4	92.9 ± 0.8	74.1 ± 0.7	74.8	91.2 ± 0.1	74.4 ± 0.9	77.7 ± 0.3	83.3 ± 0.2	81.6	55.3 ± 0.2	54.7 ± 0.4	73.5 ± 0.5	73.7 ± 0.7	64.3
FedIIR	76.1 ± 1.4	60.9 ± 0.2	96.3 ± 0.4	73.2 ± 0.8	76.6	94.2 ± 0.2	82.9 ± 0.8	75.8 ± 0.3	81.9 ± 0.8	83.7	57.1 ± 0.4	49.8 ± 0.6	74.2 ± 0.1	76.1 ± 0.1	64.4
FedSR	72.8 ± 0.3	62.3 ± 0.3	93.8 ± 0.5	74.4 ± 0.6	75.8	94.0 ± 0.6	82.8 ± 1.5	75.2 ± 0.5	81.7 ± 0.8	83.4	57.9 ± 0.2	50.3 ± 0.6	73.3 ± 0.1	75.5 ± 0.1	64.3
StableFDG	73.6 ± 0.1	59.2 ± 0.7	98.1 ± 0.2	70.2 ± 1.1	75.3	94.8 ± 0.1	83.0 ± 1.1	79.3 ± 0.2	79.7 ± 0.8	84.2	57.1 ± 0.3	57.9 ± 0.5	72.7 ± 0.6	72.1 ± 0.8	65.0
FedOMG	**82.3 ± 0.5**	**67.5 ± 0.4**	**99.3 ± 0.1**	**79.1 ± 0.5**	**82.0**	**98.0 ± 0.2**	**89.7 ± 0.4**	**81.4 ± 0.8**	**84.3 ± 0.5**	**88.4**	**65.4 ± 0.4**	**58.1 ± 0.3**	**77.5 ± 0.4**	**78.9 ± 0.5**	**70.0**
FedIIR+OMG	75.3 ± 1.3	64.0 ± 0.2	97.7 ± 0.1	72.8 ± 0.2	77.5	97.7 ± 0.1	83.0 ± 1.1	80.8 ± 0.2	79.3 ± 0.3	85.2	62.0 ± 0.3	52.8 ± 0.5	74.3 ± 0.6	76.9 ± 0.8	66.5
FedSAM+OMG	82.7 ± 0.7	69.4 ± 0.9	99.3 ± 0.3	78.5 ± 0.8	82.5	98.3 ± 0.1	88.9 ± 1.2	82.7 ± 0.3	85.5 ± 0.2	88.8	65.8 ± 0.2	58.9 ± 0.4	78.9 ± 0.5	79.3 ± 0.7	70.9
FedSR+OMG	73.6 ± 0.1	66.0 ± 0.3	94.8 ± 0.2	73.3 ± 3.3	76.9	97.2 ± 0.1	83.2 ± 1.1	79.8 ± 0.2	79.3 ± 3.3	84.8	61.7 ± 0.3	53.3 ± 0.5	73.6 ± 0.6	75.9 ± 0.8	66.1

Domain generalization performance on VLCS, PACS, and OfficeHome Datasets

THANK YOU