EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO 2024

WED-AM-3-6262

Personalized Federated Domain-Incremental Learning based on Adaptive Knowledge Matching

Yichen Li[1], Wenchao Xu[2], Haozhao Wang[1], Yining Qi[1], Jingcai Guo[2], Ruixuan Li[1]

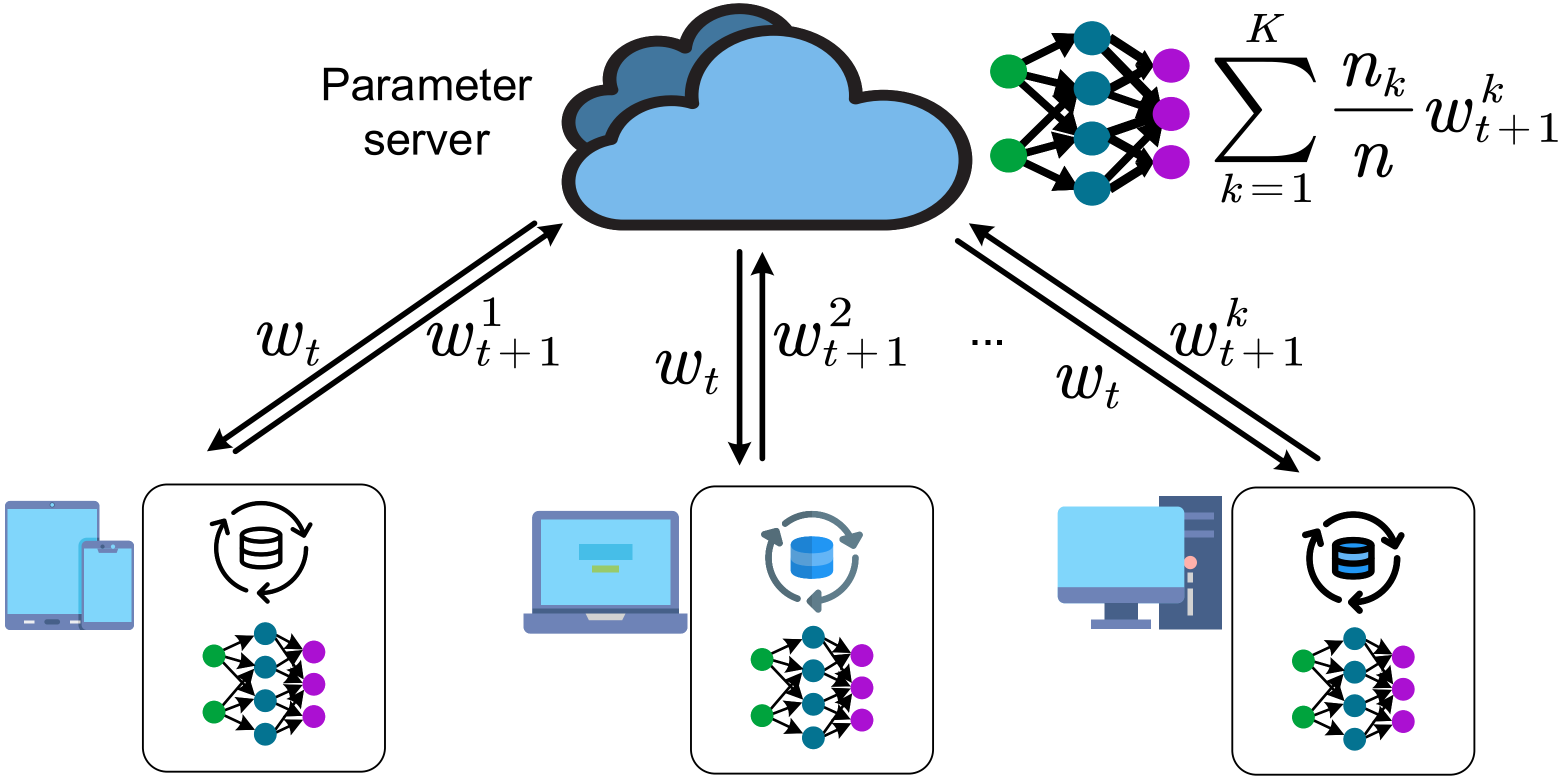
[1]Huazhong University of Science and Technology, Wuhan, China

[2]The Hong Kong Polytechnic University, Hongkong, China

Federated Incremental Learning

I. Background
II. Motivations
III. Methodology
IV. Experimental Results
V. Conclusion

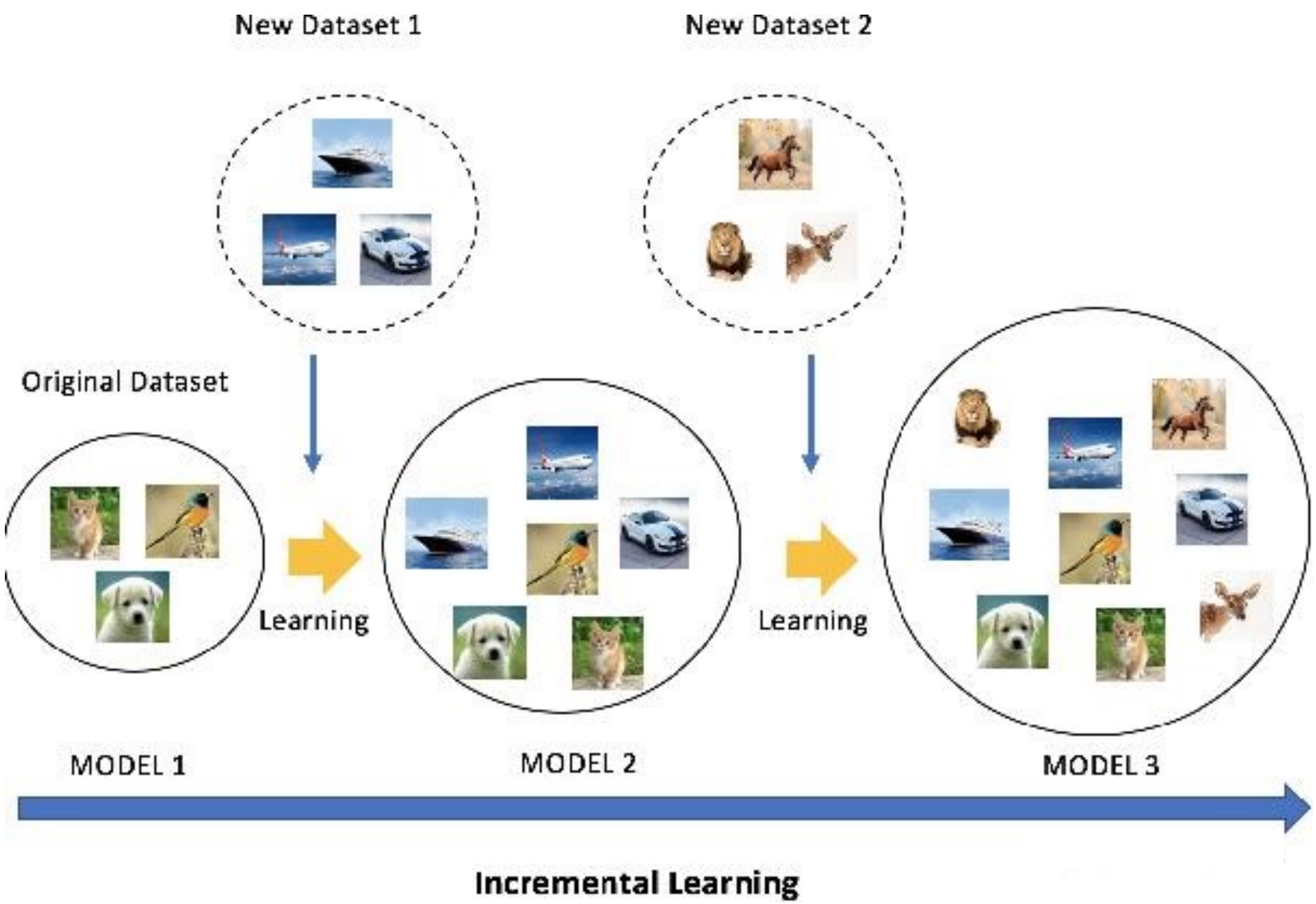
Background: Federated Learning



FedAvg: Global model is obtained by computing the average of parameters of multiple local models

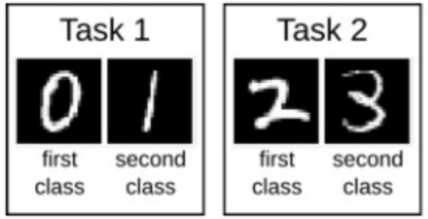
Background: Incremental Learning

Illustration of Continual Learning/Incremental Learning/Lifelong Learning



Three Typical Scenarios

- **Class-Incremental Learning**



$$P(Y^1) \neq P(Y^2)$$

- **Domain-Incremental Learning**

$$P(X^1) \neq P(X^2)$$

- **Task-Incremental Learning**

$$P(Y^1) \neq P(Y^2), P(X^1) \neq P(X^2), |Y^1| \neq |Y^2|$$

Motivations

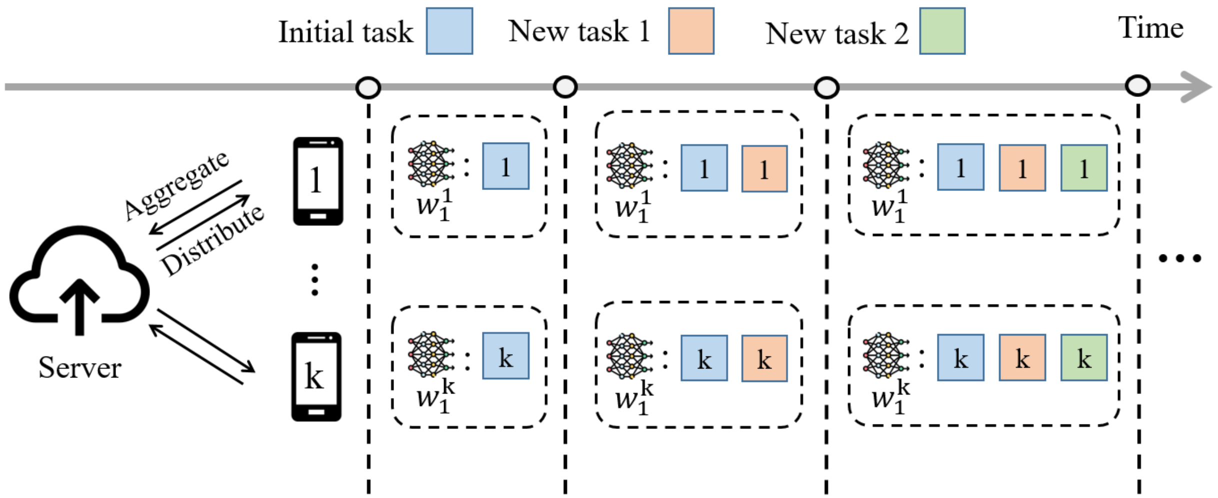
◆ **Catastrophic Forgetting**: clients are difficult to learn new data while retaining previous information.

➢ especially when data is non-identically and independently distributed (Non-IID) across clients.

◆ **Domain Shift**: existing FIL methods only focus on the label information in class-incremental tasks.

➢ fail to work with domain shifts between incremental tasks.

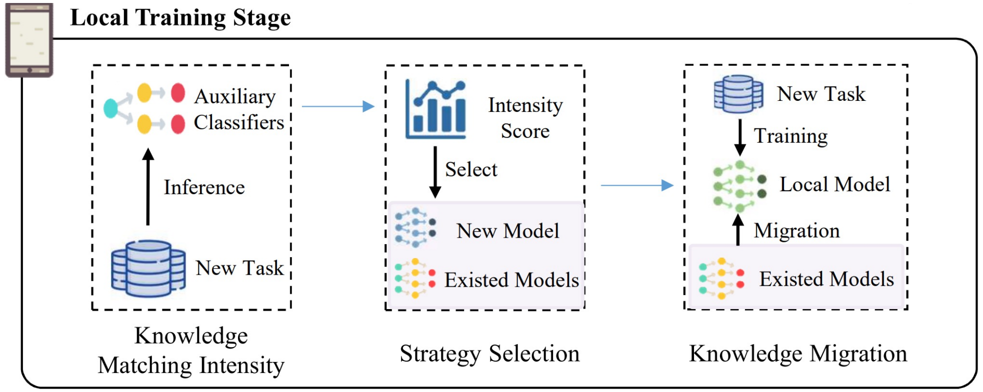
		Spatial heterogeneity			
		No changes (IID data)	Changes in the input space throughout clients $P_i(x) \neq P_j(x)$	Changes in the behaviour throughout clients $P_i(y\|x) \neq P_j(y\|x)$	Changes in the input space and behaviour throughout clients
Temporal heterogeneity	No changes (IID data)	OUT OF SCOPE	[35–38, 44] [55, 61–81]	[39–42] [99–101]	[45] [50–52]
	Changes in the input space over time, $P^t(x) \neq P^{t+k}(x)$ (Virtual Concept Drift)	[8, 117–122] [131–137]	NOT ADDRESSED SO FAR		
	Changes in the behaviour over time, $P^t(y\|x) \neq P^{t+k}(y\|x)$ (Real Concept Drift)	[123–125] [138–144]			
	Changes in the input space and behaviour over time (Total Concept Drift)	NO SPECIFIC ALGORITHMS			



Federated Domain-Incremental learning (FDIL)

Methodology: pFedDIL

- How to discern the similar knowledge between domain-incremental tasks?
- How to transfer the shared knowledge?



Step 1. Knowledge Matching

$$\widetilde{\rho}^k = \frac{1}{N^{t+1}} \sum_{i=1}^{N^{t+1}} f(x_i^{t+1}; \widetilde{\theta}^k). \quad \underline{\text{auxiliary classifier } \widetilde{\theta}^k = [\theta_1^k, \theta_2^k, \ldots, \theta_d^k]}$$

Step 2. Strategy Selection

$$w_{t+1}^k = OPT \begin{cases} \hat{w} & if\ max(\widetilde{\rho}_k) < \lambda \\ \widetilde{w}^k[m] & else\ max(\widetilde{\rho}_k) = \rho_k^m \geq \lambda\,. \end{cases} \quad \underline{\widetilde{w}^k = [w_1^k, w_2^k, \ldots, w_d^k]}$$

Step 3. Knowledge Migration

$$\min_{w_{t+1}^k} \mathcal{L}_{Local}^k(w_{t+1}^k) = \mathcal{L}_{(t+1)}^k(w_{t+1}^k) + \mathcal{L}_{KM}^k(w_{t+1}^k),$$

where $\mathcal{L}_{KM}^k(w_{t+1}^k) = \widetilde{\rho}_k \cdot ||w_{t+1}^k - \widetilde{w}^k||^2 = \sum_{i=1}^{d} \rho^i \cdot ||w_{t+1}^k - w_i^k||^2.$

Algorithm 1: pFedDIL

Input : T: the communication round; K: client number;
C: the fraction of active client in each round;
$\{\mathcal{T}_{(t)}\}_{t=1}^n$: the distributed dataset with n tasks;
w: the parameter of the target classification model;
θ: the parameter of the auxiliary classifier.

```
Initialize the parameter w and θ;
for t = 1 to T do
    S_t ← (random set of ⌈C · K⌉ clients); for each selected client k ∈ S_t
        receives w_t from the server;
        calculate knowledge matching intensity ρ̃^k with (2);
        select learning strategy w^k_{t+1} with (4);
        set local models w^k_{t+1} and θ^k_{t+1};
        for e = 1 to E do
            update θ^k_{t+1} with (3);
            update w^k_{t+1} through adaptive knowledge migration with (5);
        end
        pushes w^k_{t+1} to the server.
    end
    w_t ← ServerAggregation({w^k}_{k∈S_t})
end
```

Experiments - Performance Overview

Datasets: Digit10、Office31、DomainNet

Baselines： FedAvg、FedProx、FedCIL、DANN+FL、Source-Only、Sharing

Test Accuracy & Ablation Study

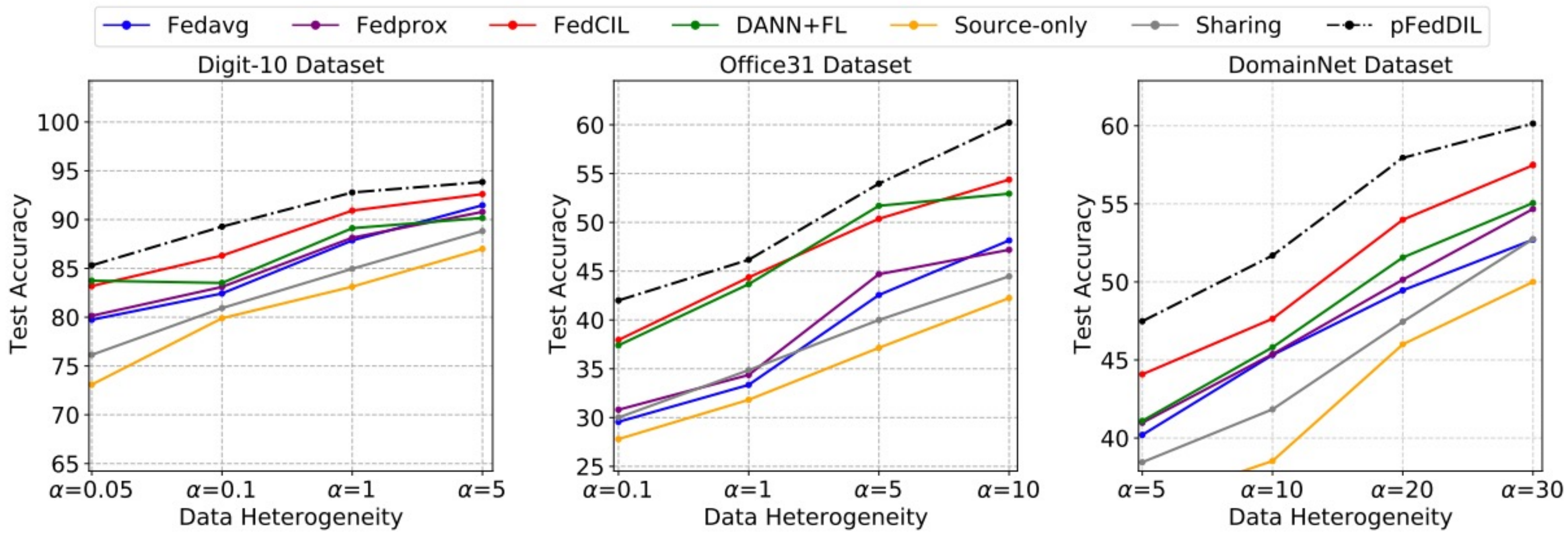
Method	**Digit-10** ($\lambda = 0.5$)					
	MNIST →	EMNIST →	USPS →	SVHN	Avg	Δ(↑)
FedAvg [24]	92.82	84.02	81.02	71.87	82.43	6.85↑
FedProx [19]	93.01	84.68	78.30	76.42	83.10	6.18↑
FedCIL [28]	94.64	87.52	82.16	80.92	86.31	2.97↑
DANN [10] + FL	**96.07**	86.71	79.11	72.14	83.51	5.77↑
Source-Only	92.82	82.15	75.53	69.06	79.89	9.39↑
Sharing	92.67	82.91	76.17	71.96	80.93	8.35↑
pFedDIL -w/o Migration	92.82	87.09	84.00	81.89	86.45	2.83↑
pFedDIL -w/o Sharing	92.79	90.13	**87.59**	86.19	89.18	0.11↑
pFedDIL -w/o Correlation	92.82	89.06	84.72	84.30	87.73	1.55↑
pFedDIL	92.82	**90.39**	87.14	**86.75**	**89.28**	/
Upper Bound -Disjoint	92.82	90.33	93.41	87.97	91.13	1.85↓

Method	**Office-31** ($\lambda = 0.8$)				
	Amazon →	Dlsr →	Webcam	Avg	Δ(↑)
FedAvg [24]	46.53	24.17	29.34	33.35	12.82↑
FedProx [19]	45.19	25.68	32.23	34.37	11.80↑
FedCIL [28]	49.38	39.65	44.04	44.36	1.81↑
DANN [10] + FL	**51.97**	35.96	43.08	43.67	2.50↑
Source-Only	46.53	20.61	28.32	31.82	14.35↑
Sharing	46.11	24.23	34.25	34.86	11.31↑
pFedDIL -w/o Migration	46.53	40.17	44.55	43.75	2.42↑
pFedDIL -w/o Sharing	46.63	42.90	47.03	45.52	0.65↑
pFedDIL -w/o Correlation	46.53	41.66	47.19	45.13	1.04↑
pFedDIL	46.53	**43.87**	**48.12**	**46.17**	/
Upper Bound -Disjoint	46.53	50.17	56.21	50.97	4.80↓

Method	**DomainNet** ($\lambda = 0.8$)							
	Clipart →	Infograph →	Painting →	Quickdraw →	Real →	Sketch	Avg	Δ(↑)
FedAvg [24]	48.43	37.18	45.80	45.32	44.91	50.25	45.32	6.38↑
FedProx [19]	47.39	38.43	44.31	47.96	42.38	51.77	45.37	6.32↑
FedCIL [28]	**52.14**	43.68	47.10	48.75	42.89	51.26	47.64	4.05↑
DANN [10] + FL	50.07	39.74	43.73	45.08	43.28	52.96	45.81	5.88↑
Source-Only	48.43	31.01	33.12	38.15	37.62	42.85	38.53	13.16↑
Sharing	48.43	36.03	39.58	41.80	40.41	44.76	41.84	9.86↑
pFedDIL -w/o Migration	48.43	43.83	50.05	49.44	46.10	54.90	48.79	2.90↑
pFedDIL -w/o Sharing	48.52	**47.37**	51.16	53.98	50.11	55.24	51.06	0.63↑
pFedDIL -w/o Correlation	48.43	46.71	51.09	52.34	48.07	56.95	50.60	1.09↑
pFedDIL	48.43	47.19	**52.48**	**54.07**	**50.36**	**57.61**	**51.69**	/
Upper Bound -Disjoint	48.43	52.33	54.21	52.89	53.62	58.06	53.26	1.57↓

Experiments - Performance Overview

Parameter Sensitivity & Communication Efficiency



a smaller α indicates higher data heterogeneity

Dataset		$\lambda = 0.0$	$\lambda = 0.2$	$\lambda = 0.5$	$\lambda = 0.8$	$\lambda = 1.0$
Digit-10	Avg ACC	82.43	85.02	**89.28**	90.77	93.94
	Model Size	22Mb	37Mb	**46Mb**	71Mb	88Mb
Office-31	Avg ACC	33.35	40.51	39.98	**46.17**	50.38
	Model Size	22Mb	42Mb	48Mb	**53Mb**	66Mb
DomainNet	Avg ACC	45.32	48.19	49.65	**51.69**	53.75
	Model Size	22Mb	73Mb	84Mb	**91Mb**	132Mb

Dataset	FedAvg		FedCIL		DANN+FL		pFedDIL	
	Rounds	Time	Rounds	Time	Rounds	Time	Rounds	Time
Digit10	480(r)	**32.14**(s)	492(r)	50.06(s)	553(r)	46.77(s)	**379**(r)	34.91(s)
Office31	463(r)	**44.96**(s)	482(r)	78.32(s)	488(r)	62.89(s)	**411**(r)	47.58(s)
DomainNet	835(r)	**92.22**(s)	904(r)	165.41(s)	861(r)	158.01(s)	**725**(r)	105.38(s)

Conclusions

In this paper, we seek to tackle the catastrophic forgetting in the FDIL scenario.

We propose a personalized federated domain-incremental learning approach based on adaptive knowledge matching, named pFedDIL. We leverage the auxiliary classifier to calculate the knowledge-matching intensity for the incremental task-learning strategy selection and knowledge migration. Furthermore, we propose sharing partial parameters between the target classification model and the auxiliary classifier to condense model parameters.

Our extensive experiments across various settings and baselines validate the effectiveness of pFedDIL, making it a robust solution for federated domain-incremental learning.

EUROPEAN CONFERENCE ON COMPUTER VISION

MILANO 2024

Thank You

Yichen Li Ph.D. candidate 2023 - present
School of Computer Science and Technology
Huazhong University of Science and Technology

Research Interest:
Federated Learning LLM Edge-Cloud
Recommendation System Software Engineering

Welcome to Collaborate!

[1] Towards Efficient Replay in Federated Incremental Learning, CVPR2024
[2] DaFKD: Domain-aware Federated Knowledge Distillation, CVPR2023
[3] FedCDA: Federated Learning with Cross-rounds Divergence-aware Aggregation, ICLR2024
[4] SR-FDIL: Synergistic Replay for Federated Domain-Incremental Learning, IEEE TPDS2024