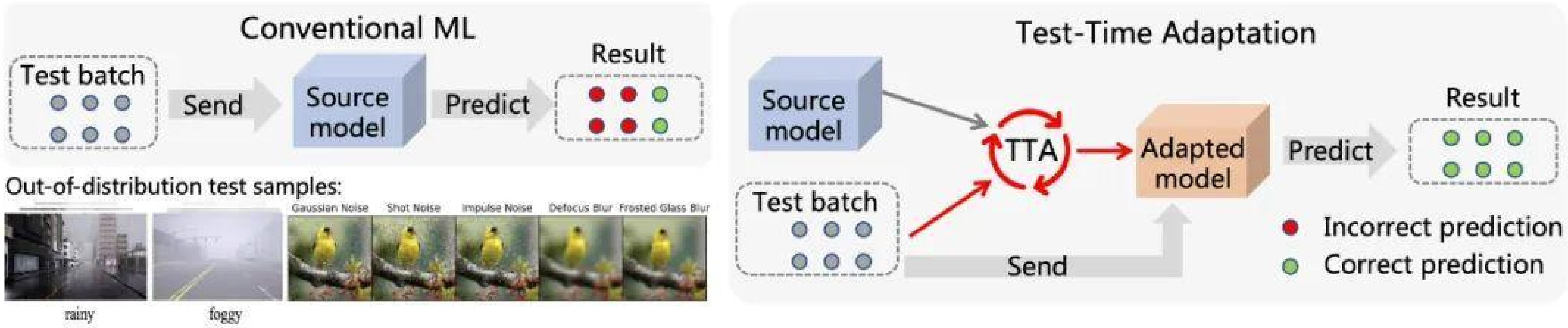
南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Adaptive Test-Time Personalization for Federated Learning

Wenxuan Bao[1*], Tianxin Wei[1*], Haohan Wang[1], Jingrui He[1],
[1]University of Illinois Urbana-Champaign
{wbao4,twei10,haohanw,jingrui}@illinois.edu

NeurIPS 2023

Test-Time Adaptation



Setting	Source data	Target data	Training loss	Testing loss	Offline	Online
Fine-tuning	×	x^t, y^t	$\mathcal{L}(x^t, y^t)$	--	√	×
Unsupervised domain adaptation	x^s, y^s	x^t	$\mathcal{L}(x^s, y^s) + \mathcal{L}(x^s, x^t)$	--	√	×
Test-time training [1]	x^s, y^s	x^t	$\mathcal{L}(x^s, y^s) + \mathcal{L}(x^s)$	$\mathcal{L}(x^t)$	×	√
Fully test-time adaptation [2]	×	x^t	×	$\mathcal{L}(x^t)$	×	√

Background

- 1) Federated Learning (FL) Core Challenge
- Clients exhibit distinct **distribution shifts**
- **Need:** Models must adapt to each client's unique data distribution for good performance

- 2) Limitations of Existing Methods

Personalized FL

- Relies on **labeled data** from test clients for personalization
- Infeasible for real-world scenarios

Test-Time Adaptation (TTA)

- Assumes training data from a **single source domain**
- Designed for **specific distribution shifts**
- Predefines adaptive modules — **ineffective for mismatched shifts**

Background

- 3) **Novel Setting: Test-Time Personalized FL (TTPFL)**
- **Training Phase:** Server trains a global model using source clients.
- **Test Phase:** Target clients adapt the global model locally with unlabeled data.
- **Goal:** Address complex distribution shifts for cross-device FL with unlabeled target clients.

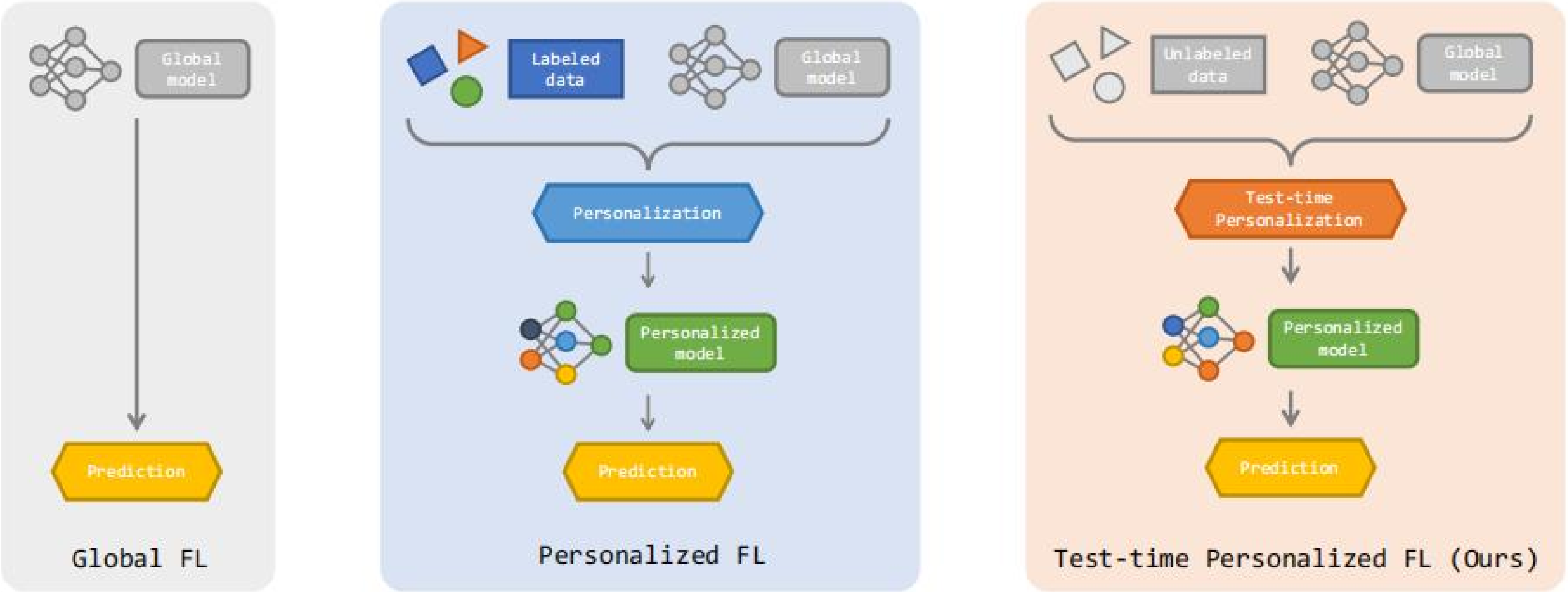


Figure 1: Comparison between the testing phase of GFL, PFL, and TTPFL. TTPFL enables model personalization without requiring labeled data.

Motivation

- **Core Conflict in TTPFL**
- Existing methods fail to handle diverse distribution shifts in FL with unlabeled test clients.

- **Gap 1:** TTA Ignores FL's Multi-Source Nature
- In FL, data comes from **multiple source clients** with inherent distribution shifts.
- Overlooking inter-source domain relationships degrades generalization to target clients.

- **Gap 2:** TTA **Lacks Flexibility** for Diverse Shifts
- TTA predefines adaptive modules
- **Performance Trade-off:** A module effective for one shift harms performance on another

- **Motivation for ATP**
- Adapts to **multiple distribution shifts**
- Learns optimal **adaptation rates** from FL's multi-source clients
- Enables effective unlabeled personalization in TTPFL

Motivation

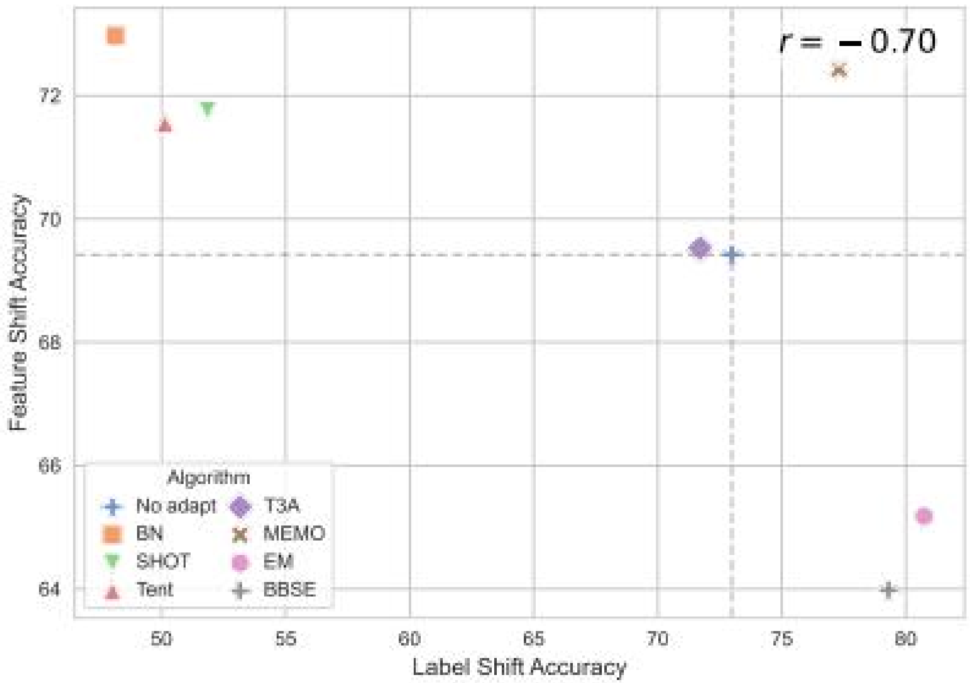


Figure 2: Performance trade-off of existing TTA methods under two distribution shifts.

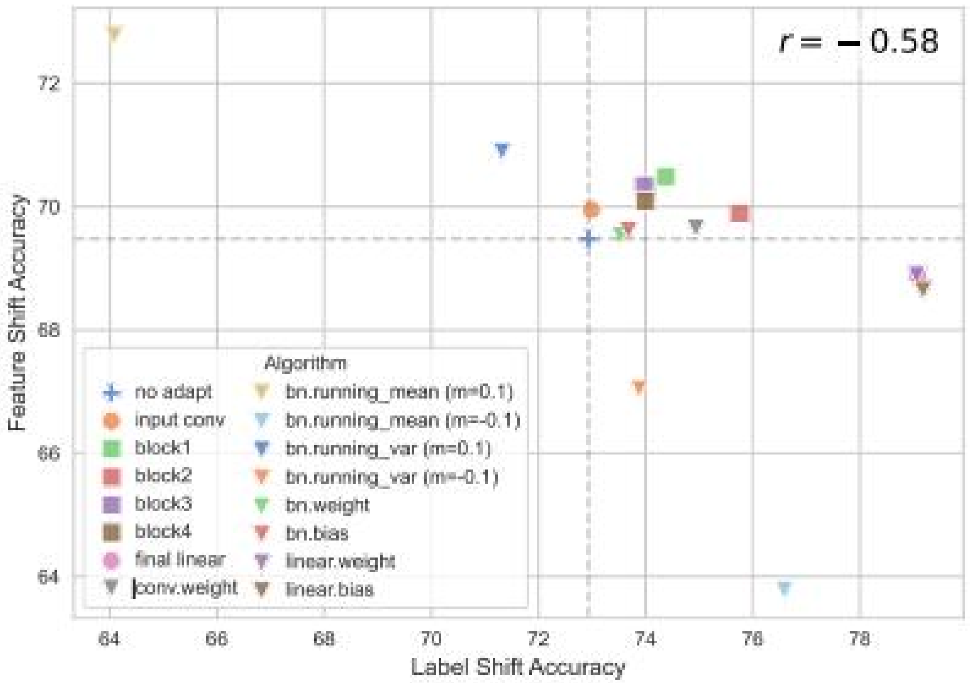


Figure 3: Performance trade-off of entropy minimization when adapting different modules.

Method——ATP: adaptive test-time personalization

- **Modules: Fine-grained model components — each has an adaptation rate $\alpha^{[l]}$**

- BN layers split into running mean/variance/weight/bias

- **Update Direction:** Unsupervised shift-aware direction for adaptation

Update trainable parameters $\boldsymbol{h}_k^{[l]} = -\nabla_{\boldsymbol{w}^{[l]}} \ell_H(f(\boldsymbol{X}_k; \boldsymbol{w}_G))$

$$\ell_H(\hat{\boldsymbol{Y}}) = \frac{1}{B} \sum_{b=1}^{B} (-\sum_c \hat{y}_{b,c} \log \hat{y}_{b,c})$$

Update running statistics $\boldsymbol{h}_k^{[l]} = \hat{\boldsymbol{w}}_k^{[l]} - \boldsymbol{w}_G^{[l]}$

$$\boldsymbol{w}_k^{[l]} \leftarrow \boldsymbol{w}_G^{[l]} + \alpha^{[l]} \boldsymbol{h}_k^{[l]}$$

- **Module Update:** $\boldsymbol{w}_k^{[l]} \leftarrow \boldsymbol{w}_G^{[l]} + \alpha^{[l]} \boldsymbol{h}_k^{[l]}$ $\quad \boldsymbol{w}_k \leftarrow \boldsymbol{w}_G + (\boldsymbol{A}\boldsymbol{\alpha}) \odot \boldsymbol{h}_k$

Method——ATP: adaptive test-time personalization

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

- **Training Phase**

Algorithm 1 ATP Training

```
ServerTrain(w_G, α_G^0 = 0)
  Broadcast w_G to all source clients
  for communication round t = 1 to T do
      S^t ← (random set of C source clients)
      for source client S_i ∈ S^t in parallel do
          α_i^t ← ClientTrain(S_i, α_G^{t-1})
      α_G^t = (1/C) Σ_{S_i ∈ S^t} α_i^t
  return α_G^T

ClientTrain(S_i, α)        # Run on source client S_i
  for local epoch e = 1 to E do
      B^{S_i} ← (split D^{S_i} into K^{S_i} batches of size B)
      for batch k = 1 to K^{S_i} do
          (X_k^{S_i}, Y_k^{S_i}) ← (k-th labeled batch in B^{S_i})
          Estimate update direction h_k^{S_i} with unla-
          beled X_k^{S_i} according to Eq. (4) and (5)
          w_k^{S_i} ← w_G + (Aα) ⊙ h_k^{S_i}
          α ← α − η ∇_α ℓ_CE(f(X_j^{S_i}; w_k^{S_i}), Y_k^{S_i})
  return α
```

- **Testing phase**

ATP-batch for test-time batch adaptation/
ATP-online for online test-time adaptation

Algorithm 2 ATP Testing

```
ClientTest(T_j, w_G, α)    # Run on target client T_j
  B^{T_j} ← (split X^{T_j} into K^{T_j} batches of size B)
  h_history ← 0            # Cumulative moving average
  for batch k = 1 to K^{T_j} do
      Estimate update direction h_k^{T_j} with unlabeled
      X_k^{T_j} according to Eq. (4) and (5)
      if TTBA then
          w_k^{T_j} ← w_G + (Aα) ⊙ h_k^{T_j}
      else if OTTA then
          h_history ← ((k−1)/k) h_history + (1/k) h_k^{T_j}
          w_k^{T_j} ← w_G + (Aα) ⊙ h_history
      Make prediction: Ŷ_k^{T_j} = f(X_k^{T_j}; w_k^{T_j})
```

Communication Cost: $2TD \rightarrow D + 2Td$

Experiments

RQ1: Can ATP handle different distribution shift?

Table 1: Accuracy (mean ± s.d. %) on target clients under various distribution shifts on CIFAR-10

Method	Feature shift	Label shift	Hybrid shift	Avg. Rank
No adaptation	69.42 ± 0.13	72.98 ± 0.24	63.68 ± 0.24	7.7
BN-Adapt	73.52 ± 0.22	54.54 ± 0.10	50.42 ± 0.39	7.0
SHOT	71.76 ± 0.17	48.13 ± 0.18	44.68 ± 0.32	9.3
Tent	71.76 ± 0.09	50.13 ± 0.21	46.05 ± 0.26	8.3
T3A	69.53 ± 0.08	71.70 ± 0.32	62.17 ± 0.17	8.0
MEMO	72.43 ± 0.22	77.30 ± 0.15	68.07 ± 0.28	4.3
EM	65.18 ± 0.12	80.73 ± 0.18	69.85 ± 0.43	5.0
BBSE	63.98 ± 0.17	79.30 ± 0.17	67.96 ± 0.43	6.7
Surgical	69.85 ± 0.22	76.00 ± 0.17	66.94 ± 0.43	6.3
ATP-batch	73.68 ± 0.10	79.90 ± 0.22	73.05 ± 0.35	2.3
ATP-online	**74.06 ± 0.18**	**81.96 ± 0.14**	**75.37 ± 0.22**	**1.0**

Table 2: Accuracy (mean ± s.d. %) on target clients under hybrid shift on Digits-5 and PACS

Method	Digits-5					PACS			
	MNIST	SVHN	USPS	SynthDigits	MNIST-M	Art	Cartoon	Photo	Sketch
No adaptation	95.47 ± 0.22	52.28 ± 1.45	89.62 ± 0.44	79.75 ± 0.69	55.62 ± 0.80	71.57 ± 1.16	74.71 ± 0.70	90.25 ± 0.75	74.20 ± 0.72
BN-Adapt	94.90 ± 0.29	57.57 ± 0.53	89.51 ± 0.39	75.34 ± 0.48	59.68 ± 0.44	73.55 ± 0.51	71.54 ± 0.55	92.07 ± 0.26	70.92 ± 0.53
SHOT	94.69 ± 0.31	57.91 ± 0.23	89.55 ± 0.69	76.43 ± 0.34	60.19 ± 0.69	69.32 ± 0.67	67.77 ± 0.40	86.97 ± 0.60	59.40 ± 0.91
Tent	95.48 ± 0.29	60.67 ± 0.49	91.65 ± 0.61	78.56 ± 0.45	62.49 ± 0.73	71.59 ± 0.71	71.03 ± 0.97	88.06 ± 0.24	63.15 ± 1.10
T3A	94.63 ± 0.61	49.90 ± 1.10	88.46 ± 0.75	75.47 ± 1.14	51.25 ± 1.55	72.15 ± 0.72	75.02 ± 0.78	91.51 ± 0.62	70.14 ± 1.21
MEMO	95.92 ± 0.19	52.85 ± 1.09	89.84 ± 0.44	80.12 ± 0.90	55.48 ± 1.13	71.47 ± 1.29	75.57 ± 0.98	90.65 ± 0.90	76.30 ± 0.65
EM	96.64 ± 0.31	57.21 ± 1.65	92.29 ± 0.32	85.69 ± 0.46	62.08 ± 0.60	73.96 ± 1.85	78.91 ± 0.92	92.30 ± 0.92	80.82 ± 1.52
BBSE	94.47 ± 0.58	57.26 ± 1.47	91.34 ± 0.39	85.54 ± 0.46	61.59 ± 0.91	74.33 ± 1.78	78.69 ± 1.00	91.82 ± 0.68	80.15 ± 1.42
Surgical	97.35 ± 0.13	59.93 ± 2.01	94.19 ± 0.40	86.06 ± 0.44	65.87 ± 0.78	74.59 ± 2.69	77.48 ± 0.64	92.34 ± 0.78	80.90 ± 3.42
ATP-batch	97.81 ± 0.27	62.18 ± 1.71	95.41 ± 0.26	87.91 ± 0.45	69.98 ± 1.96	82.92 ± 0.96	**79.64 ± 0.75**	95.40 ± 0.41	82.28 ± 1.57
ATP-online	**97.81 ± 0.23**	**62.64 ± 1.92**	**95.56 ± 0.23**	**88.33 ± 0.47**	**70.78 ± 2.36**	**83.51 ± 0.84**	79.46 ± 0.77	**95.52 ± 0.40**	**82.80 ± 1.69**

Experiments

Table 5: ATP with different model architectures, accuracy (mean $\pm$ s.d. %) on target clients

Method	Shallow CNN on CIFAR-10				ResNet-50 on CIFAR-100			
	Feature shift	Label shift	Hybrid shift	Avg. Rank	Feature shift	Label shift	Hybrid shift	Avg. Rank
No adaptation	64.39 ± 0.18	69.33 ± 0.37	61.99 ± 0.47	7.3	45.31 ± 0.30	51.63 ± 0.15	40.01 ± 0.17	7.3
BN-Adapt	66.46 ± 0.22	54.99 ± 0.38	50.40 ± 0.43	7.0	47.75 ± 0.29	34.85 ± 0.26	30.31 ± 0.09	7.3
SHOT	65.60 ± 0.18	49.98 ± 0.29	45.95 ± 0.47	9.0	45.42 ± 0.30	31.06 ± 0.32	27.44 ± 0.14	9.3
Tent	65.61 ± 0.24	50.12 ± 0.25	45.91 ± 0.49	8.7	45.91 ± 0.46	31.34 ± 0.11	27.93 ± 0.31	8.3
T3A	64.31 ± 0.27	66.96 ± 0.43	59.65 ± 0.58	8.3	45.31 ± 0.30	51.42 ± 0.15	39.89 ± 0.20	7.7
MEMO	65.89 ± 0.31	71.95 ± 0.25	64.17 ± 0.47	5.3	<u>48.42 ± 0.14</u>	55.19 ± 0.28	42.53 ± 0.20	3.7
EM	61.74 ± 0.25	<u>76.28 ± 0.29</u>	67.54 ± 0.41	5.0	43.00 ± 0.31	<u>59.34 ± 0.15</u>	44.82 ± 0.27	5.0
BBSE	56.92 ± 0.53	75.99 ± 0.44	66.64 ± 0.53	6.3	37.26 ± 0.64	56.97 ± 0.20	40.09 ± 0.51	7.0
Surgical	64.45 ± 0.12	73.75 ± 0.42	65.67 ± 0.44	5.7	45.18 ± 0.38	54.83 ± 0.26	42.50 ± 0.33	6.7
ATP-batch	<u>66.90 ± 0.05</u>	76.23 ± 0.32	<u>68.88 ± 0.35</u>	<u>2.3</u>	48.35 ± 0.45	58.06 ± 0.53	<u>46.82 ± 0.32</u>	<u>2.7</u>
ATP-online	**67.13 ± 0.17**	**78.56 ± 0.32**	**71.52 ± 0.51**	**1.0**	**49.08 ± 0.26**	**61.86 ± 0.25**	**49.51 ± 0.23**	**1.0**

Experiments

RQ2: Does ATP learn adaptation rates specific to distribution shift?

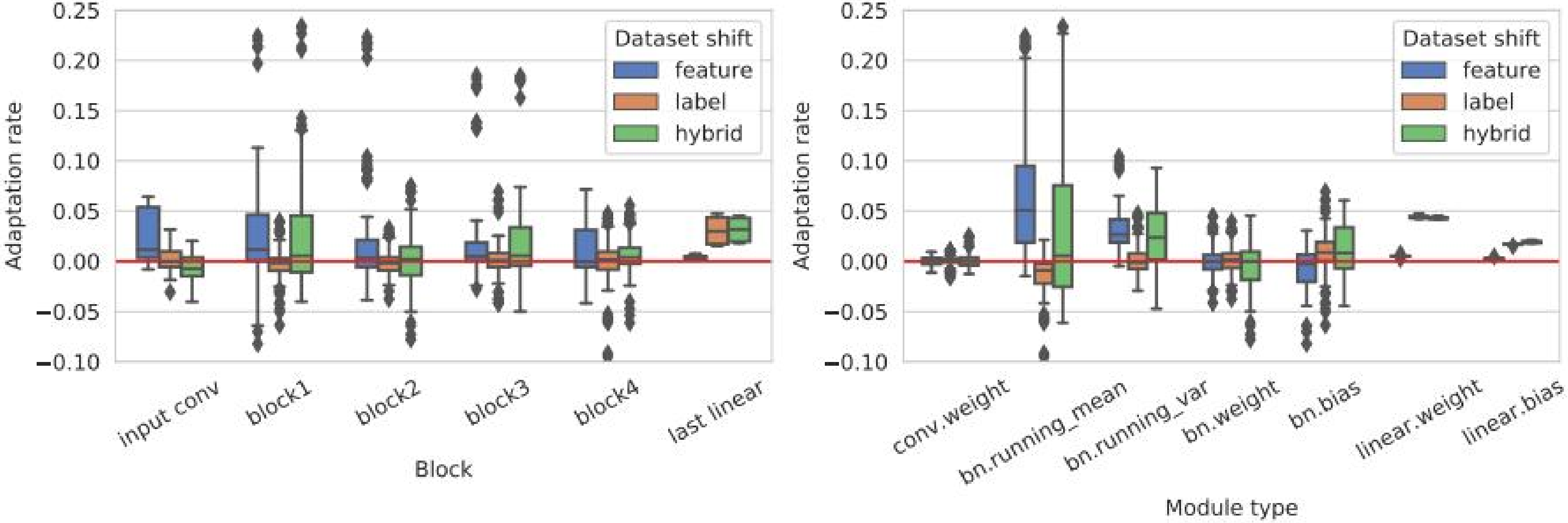


Figure 4: Adaptation rates learned by ATP with different distribution shifts on CIFAR-10

Experiments

RQ2: Does ATP learn adaptation rates specific to distribution shift?

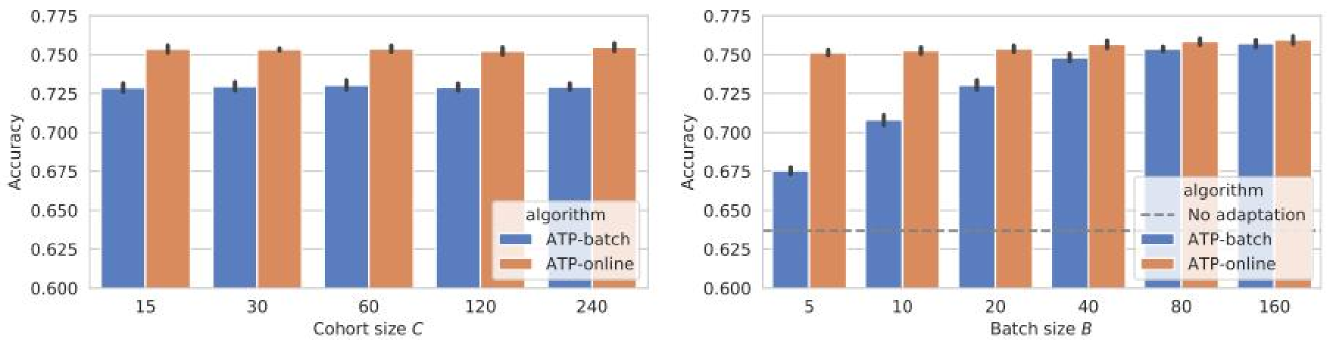


Figure 5: Effect of cohort size and batch size

Table 3: Train and test adaptation rates with different distribution shifts, accuracy (mean $\pm$ s.d. %)

Train	Test		
	Feature shift	Label shift	Hybrid shift
No adaptation	69.42 ± 0.13	72.98 ± 0.24	63.68 ± 0.24
Feature shift	$\mathbf{73.68 \pm 0.10}$	65.05 ± 1.82	60.64 ± 1.43
Label shift	67.99 ± 0.28	$\mathbf{79.90 \pm 0.22}$	69.50 ± 0.52
Hybrid shift	72.69 ± 0.14	78.92 ± 0.34	$\mathbf{73.05 \pm 0.35}$

Table 7: Accuracy (%), ATP enhances different global models under hybrid shift on CIFAR-10

Method	FedAvg	FedProx ($\mu = 0.01$)	q-FFL ($q = 1$)
No adaptation	63.68 ± 0.24	63.77 ± 0.25	63.87 ± 0.23
ATP-batch	73.05 ± 0.35	72.95 ± 0.33	73.15 ± 0.21
ATP-online	75.37 ± 0.22	75.51 ± 0.19	75.79 ± 0.15

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

FedCTTA: A Collaborative Approach to Continual Test-Time Adaptation in Federated Learning

Rakibul Hasan Rajib*, Md Akil Raihan Iftee*, Mir Sazzat Hossain*, A. K. M. Mahbubur Rahman*, Sajib Mistry†, M Ashraful Amin* and Amin Ahsan Ali*
*Center for Computational & Data Sciences, Independent University, Bangladesh
†Curtin University
rakibul@iub.edu.bd, iftee1807002@gmail.com, {sazzat, akmmrahman}@iub.edu.bd, Sajib.Mistry@curtin.edu.au, {aminmdashraful, aminali}@iub.edu.bd

IJCNN 2025

Motivation

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

- Test data on FL clients is typically temporally **streaming** and **non-stationary.** Models need to continuously adapt to dynamically evolving distributions during the test phase.

- Existing **FL-TTA** methods:
- **a) Computationally Intensive Methods (FedICON)**
  - requires extensive contrastive learning
  - computationally intensive for resource-constrained clients

- **b) Single-Client Adaptation Methods (ATP)**
  - assumes static test-time distributions
  - lack inter-client knowledge sharing

c) Collaborative Methods with Privacy/Scalability Issues(FedTSA)
  - based on sharing local feature means
  - requires server-side learning (6.2 millions parameters)
  - memory bank to store local feature means

Method——FedCTTA

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

- **Problem Setting**

$C = \{C_1, C_2, \ldots, C_N\}$ each receiving a data stream $D_t^{(i)}$

Objective: enable continuous adaptation of client models θ_i to dynamic distributions, while avoiding catastrophic forgetting and ensuring generalizability

- **Local Test-time Adaptation**
- TTA-grad

$$H(p) = -\sum_{k=1}^{K} p_k \log(p_k) \quad L_{\text{ent}} = \frac{1}{|\mathcal{D}_t^{(i)}|} \sum_{x \in \mathcal{D}_t^{(i)}} H(f_{\theta_i}(x)).$$

- TTA-bn

$$\mu_i^{\text{new}} = (1-\alpha)\mu_i^{\text{old}} + \alpha \cdot \mathbb{E}_{x \sim \mathcal{D}_t^{(i)}}[x],$$

$$\sigma_i^{2,\text{new}} = (1-\alpha)\sigma_i^{2,\text{old}} + \alpha \cdot \text{Var}(x \sim \mathcal{D}_t^{(i)}),$$

- **Similarity-Aware Aggregation**
- functional similarity

$$D_{\text{noise}} = \{z_i\}_{i=1}^{M}$$

$$\mu_i = \frac{1}{M} \sum_{k=1}^{M} f_{\theta_i}(z_k), \quad \mu_j = \frac{1}{M} \sum_{k=1}^{M} f_{\theta_j}(z_k).$$

$$D_{ij} = -\|\mu_i - \mu_j\|_2,$$

$$\theta_i^{\text{new}} = \sum_{j=1}^{K} \frac{\exp(D_{ij})}{\sum_{k=1}^{K} \exp(D_{ik})} \theta_j,$$

Method——FedCTTA

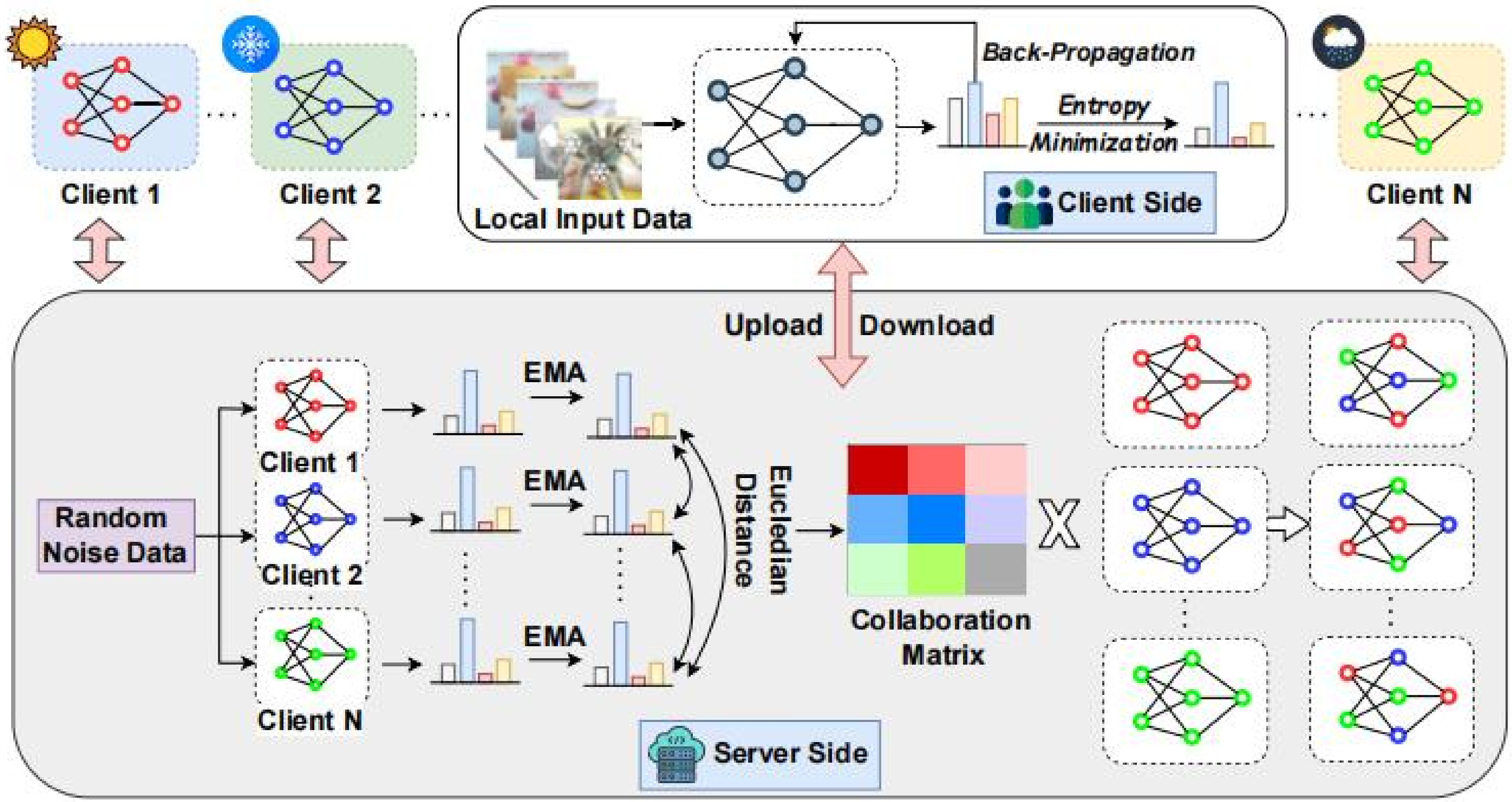


Fig. 1: Illustration of the FedCTTA framework. The server aggregates models received from all the clients based on functional similarity, then distributes the personalized aggregated models back to the clients for the next round. This process continues iteratively to adapt the models collaboratively across all clients. The figure depicts both client-to-server communication (model updates) and server-to-client communication (aggregated model distribution).

Experiments

南京航空航天大學 NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

TABLE I: Performance comparison of various federated learning methods with our proposed FedCTTA on CIFAR10-C and CIFAR100-C datasets. We evaluate all methods under two TTA setups: TTA-Grad, where all model parameters are updated during adaptation, and TTA-BN, where only batch normalization layers are updated.

Method	NIID				IID			
	CIFAR10-C		CIFAR100-C		CIFAR10-C		CIFAR100-C	
	TTA-grad	TTA-bn	TTA-grad	TTA-bn	TTA-grad	TTA-bn	TTA-grad	TTA-bn
No-Adapt	58.47±0.19	58.61±0.17	30.22±0.12	30.22±0.12	58.64±0.22	58.55±0.21	30.22±0.12	30.22±0.12
Local	63.82±0.31	64.65±0.29	52.85±0.32	55.99±0.34	63.96±0.33	64.79±0.31	52.94±0.31	56.05±0.34
FedAvg	61.15±0.24	61.45±0.23	51.63±0.17	57.13±0.43	66.12±0.26	67.41±0.27	62.54±0.31	63.96±0.31
FedAvg+FT	63.82±0.27	61.45±0.23	47.83±0.58	57.13±0.43	63.79±0.30	67.41±0.27	61.72±0.59	63.96±0.31
FedProx	61.68±0.22	61.45±0.23	53.00±0.38	57.13±0.43	66.12±0.24	67.41±0.27	62.33±0.67	63.96±0.31
FedAvgM	61.50±0.25	61.37±0.19	52.31±0.46	57.13±0.43	63.60±0.28	67.41±0.27	54.66±0.27	63.96±0.31
MOON	61.58±0.23	61.45±0.23	54.26±0.27	57.13±0.43	66.05±0.25	67.41±0.27	62.40±0.23	63.96±0.31
pFedSD	61.31±0.21	61.45±0.23	53.33±0.37	57.13±0.43	66.14±0.26	67.41±0.27	62.32±0.33	63.96±0.31
pFedGraph	62.38±0.26	64.21±0.25	57.01±0.38	58.73±0.38	66.10±0.29	64.42±0.28	62.48±0.30	58.75±0.63
LDAWA	61.85±0.23	61.45±0.23	53.61±0.33	57.13±0.43	65.92±0.26	67.41±0.27	62.37±0.41	63.96±0.31
FedTSA	63.39±0.27	66.19±0.26	58.03±0.38	62.93±0.29	66.29±0.28	67.51±0.27	62.62±0.36	63.70±0.34
FedCTTA	**66.23±0.28**	**66.50±0.27**	**64.81±0.29**	**63.39±0.28**	**66.64±0.29**	**67.78±0.28**	**64.15±0.28**	**64.52±0.28**

	spatial heterogeneity	temporal heterogeneity
NIID	SH_t=0.2	TH_i=0.02
IID	$SH_t = 0.05$	TH_i=0.02

Experiments

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

TABLE II: Detailed performance comparison under spatial IID and temporal heterogeneity using the TTA-bn method.

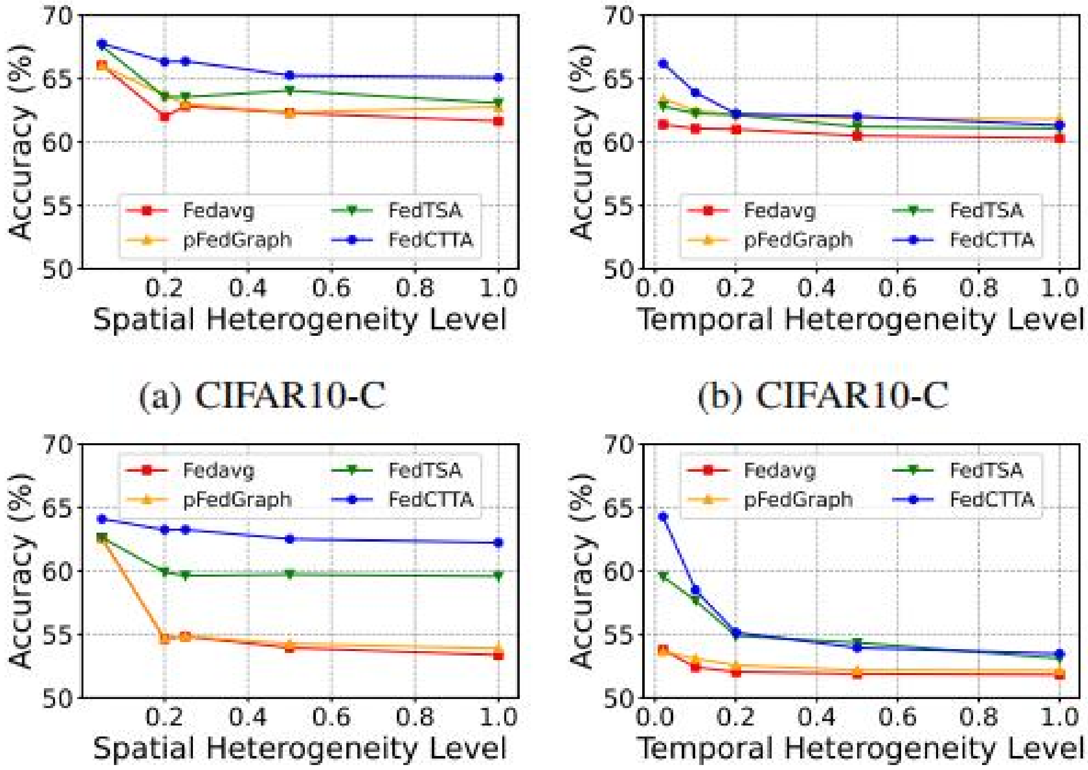
Datasets	Method	Gaussian	Shot	Impulse	Defocus	Glass	Motion	Zoom	Snow	Frost	Fog	Brightness	Contrast	Elastic	Pixelate	Jpeg	Mean
	Time	t →															
CIFAR10-C	Source	37.30	38.44	26.08	28.99	33.92	27.44	30.21	34.53	32.89	10.63	36.46	23.53	37.51	41.43	43.70	30.54
	Local	69.76	69.81	63.21	68.69	61.40	63.92	66.58	67.14	67.44	55.35	71.16	39.68	64.93	70.25	71.58	64.72
	FedAvg	72.28	72.46	66.31	71.08	64.22	66.45	69.61	70.03	69.34	58.32	73.60	42.84	68.34	72.92	74.55	67.49
	pFedGraph	68.40	69.45	62.88	68.35	61.48	63.82	65.72	66.70	66.57	55.56	70.37	40.52	65.14	69.28	70.96	64.34
	FedTSA	72.56	72.72	66.06	71.13	64.25	66.35	69.39	70.10	69.66	58.74	73.61	42.61	68.47	72.98	74.69	67.55
	FedCTTA	73.21	73.04	66.27	71.91	64.64	66.32	69.58	70.24	70.38	57.57	74.20	41.93	68.16	73.72	75.00	67.74
CIFAR100-C	Source	14.05	16.64	34.76	41.60	19.35	38.15	43.32	36.48	27.63	20.96	54.91	17.24	35.02	11.45	41.73	30.22
	Local	51.59	53.02	50.26	65.13	50.77	63.23	65.07	58.10	58.17	51.10	66.70	61.25	56.67	59.98	51.57	57.51
	FedAvg	57.33	58.60	56.74	69.07	57.51	68.94	70.92	64.29	64.19	57.44	72.40	67.36	63.70	66.33	57.56	63.50
	pFedGraph	52.05	53.42	50.48	65.60	51.19	63.61	65.49	58.39	58.64	51.39	67.08	61.64	57.14	60.46	52.06	57.91
	FedTSA	57.56	58.75	57.23	69.73	56.27	69.18	71.05	64.33	64.60	56.44	73.10	67.77	63.30	66.58	58.21	63.61
	FedCTTA	56.90	59.58	57.06	72.12	58.47	70.04	71.84	65.32	65.66	58.27	74.34	68.41	64.29	67.32	58.97	64.57

Experiments

TABLE III: The experimental setup and performance in the NIID scenario, where Clients 1–4, Clients 5–7, and Clients 8–10 share similar data distributions throughout the entire lifecycle, with a total of 10 clients.

Time	t →															
Client 1-4	Gaussian	Shot	Impulse	Defocus	Glass	Motion	Zoom	Snow	Frost	Fog	Brightness	Contrast	Elastic	Pixelate	Jpeg	Mean
Client 5-7	Glass	Motion	Zoom	Snow	Frost	Fog	Brightness	Contrast	Elastic	Pixelate	Jpeg	Gaussian	Shot	Impulse	Defocus	Mean
Client 8-10	Jpeg	Pixelate	Elastic	Contrast	Brightness	Fog	Frost	Snow	Zoom	Motion	Glass	Defocus	Impulse	Shot	Gaussian	Mean
No-Adapt	71.56	66.60	58.06	46.70	59.88	47.00	60.02	63.48	51.10	48.36	51.22	43.92	64.44	71.44	71.68	58.36
pFedGraph	68.54	66.06	61.72	54.42	66.40	57.04	67.60	54.36	65.54	61.98	69.18	50.86	65.46	67.56	67.66	62.95
FedTSA	68.70	67.30	63.02	56.58	66.48	57.36	68.32	57.26	66.38	62.08	69.18	52.72	66.58	68.20	69.76	63.99
FedCTTA	68.64	67.84	67.20	58.66	67.76	58.56	70.52	58.98	67.18	64.00	70.26	57.02	67.68	70.56	70.22	65.67

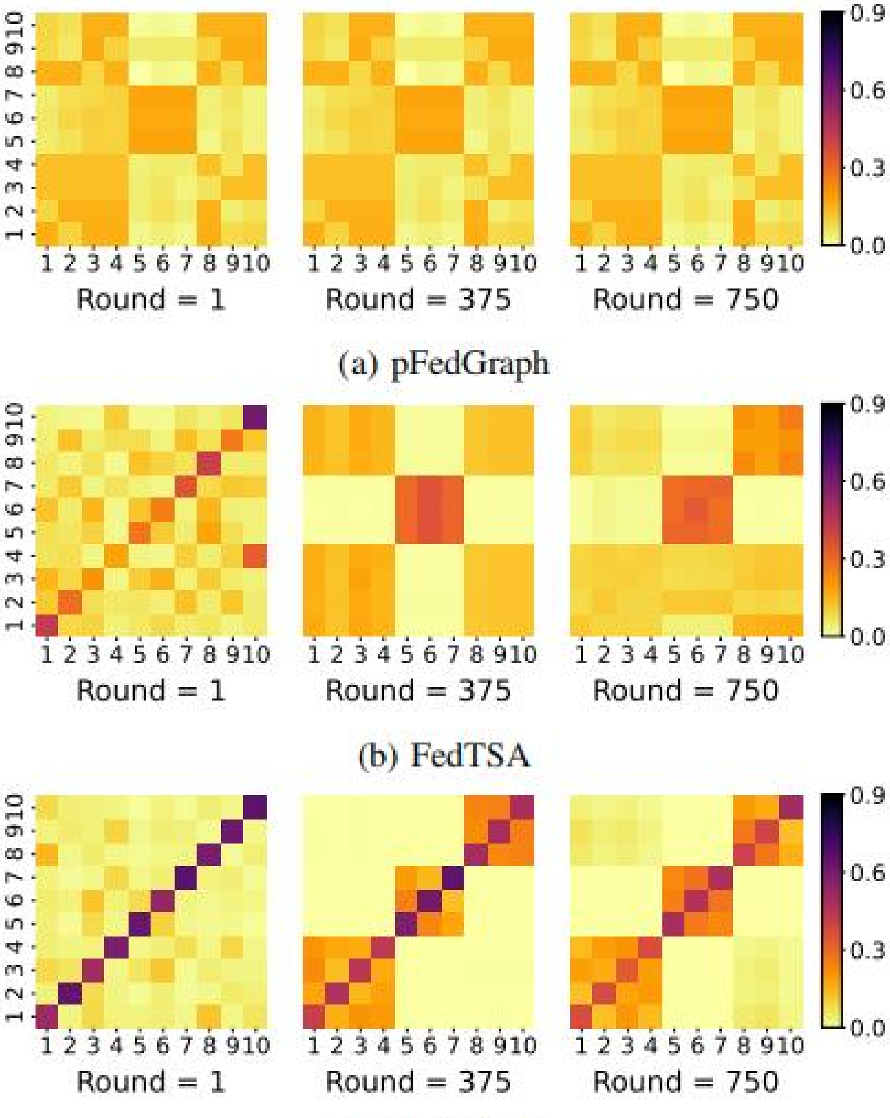
Experiments



(a) CIFAR10-C (b) CIFAR10-C

(c) CIFAR100-C (d) CIFAR100-C

Fig. 2: Comparison of accuracy on CIFAR-10C and CIFAR-100C under varying degrees of temporal and spatial heterogeneity.



(a) pFedGraph

(b) FedTSA

(c) FedCTTA

Thanks