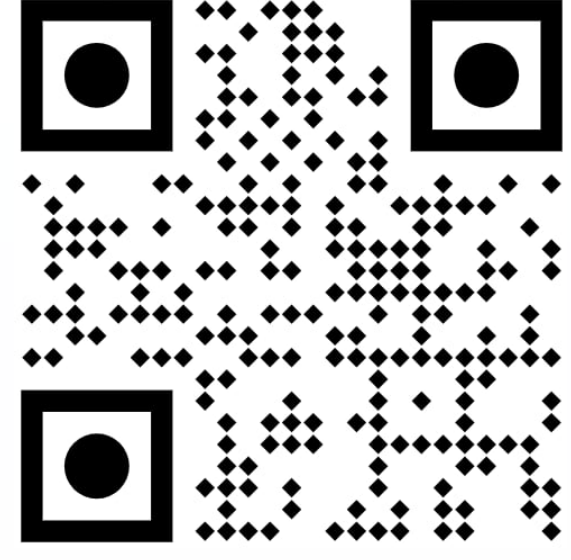


Computation and Communication Efficient Federated Unlearning via On-server Gradient Conflict Mitigation and Expression

Minh-Duong Nguyen¹, Senura Hansaja Wanasekara², Le-Tuan Nguyen¹, Ken-Tye Yong², Quoc-Viet Pham³, Nguyen H. Tran², Dung D. Le¹

¹VinUniversity, ²University of Sydney, ³Trinity College Dublin

GitHub →



1 Motivation & Problem

Federated Unlearning (FUL) removes the influence of selected clients from a trained federated model.

- ▶ Retraining is faithful but **computationally and communicatively expensive**.
- ▶ Approximate unlearning is faster, but often **sacrifices faithfulness** or retained performance.



Goal: forget client-specific knowledge while preserving retain-client knowledge, without accessing raw client data on the server.

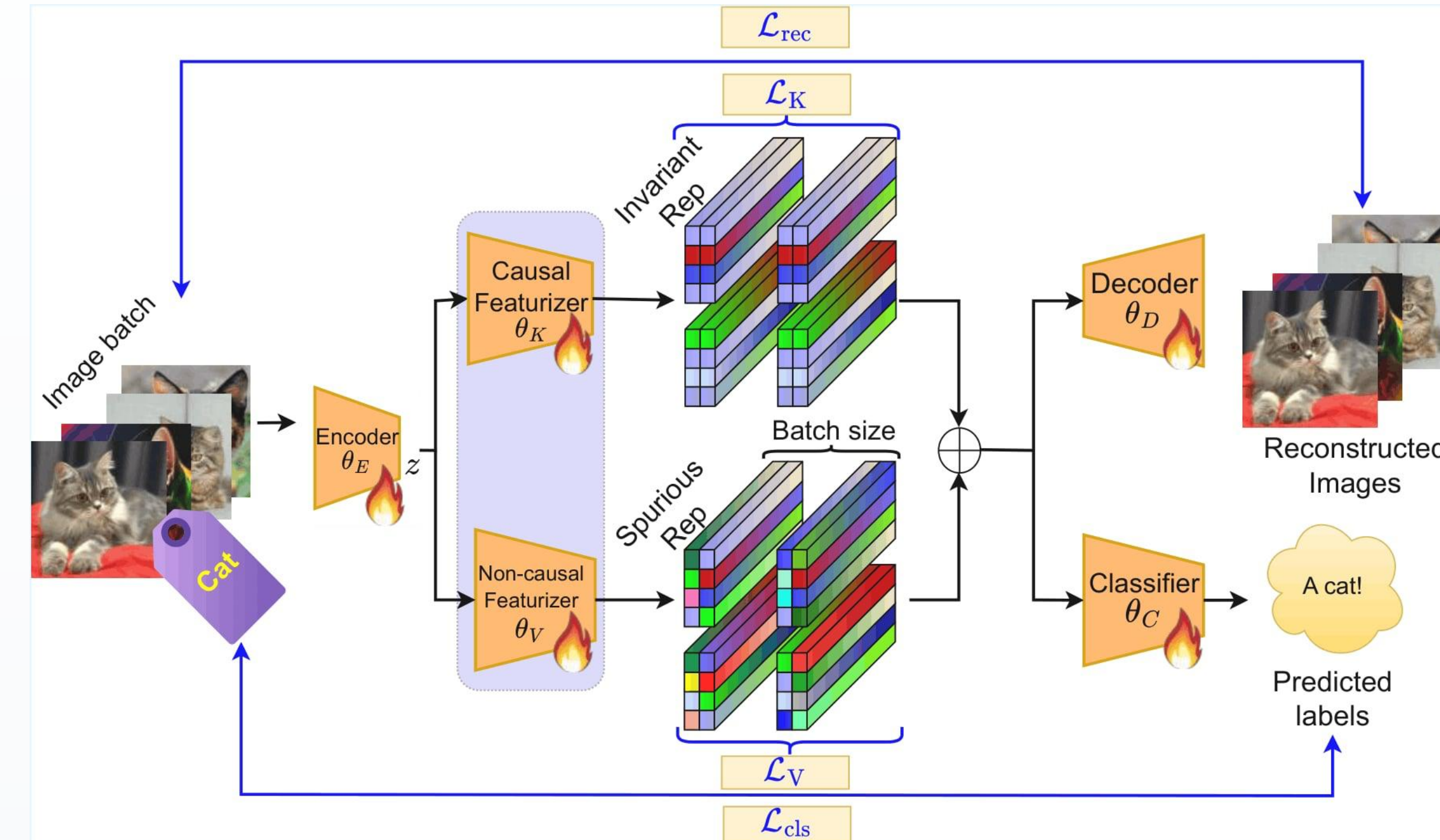
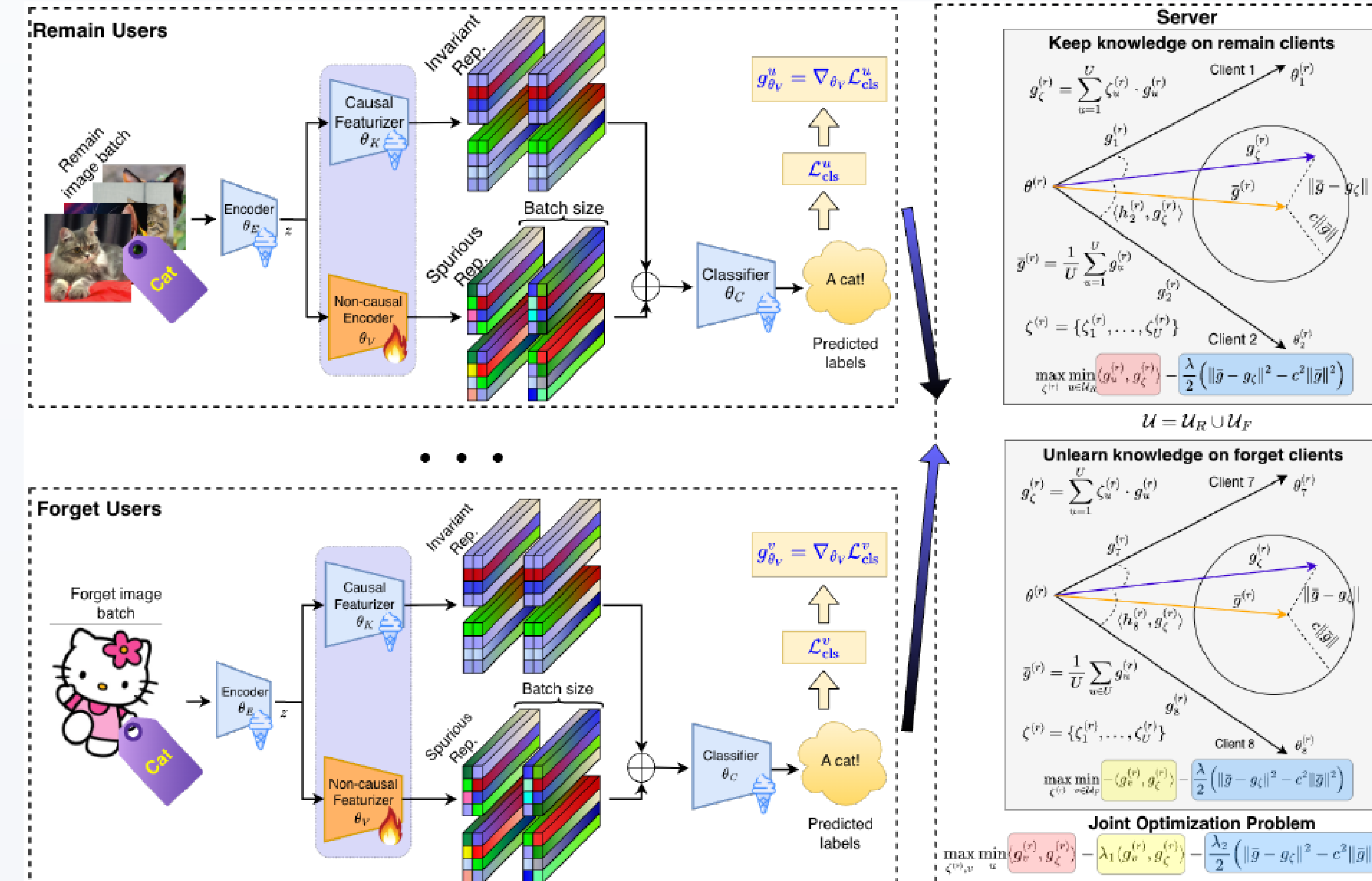
2 Key Contributions

1. Propose **FOUL** (Federated On-server UnLearning), a two-stage framework for client-wise FUL.
2. **Learning-to-Unlearn (L2U)**: disentangle features into causal / domain-invariant and non-causal / domain-specific parts.
3. **On-server unlearning**: update only the non-causal branch via gradient matching, using gradients from both retain and forget clients.
4. Introduce **Domain-wise FUL benchmark** setting based on domain-generalization datasets.

3 Methods and Materials

3.1 Learning to unlearn

- Disentangle the feature extractor into **causal** and **non-causal** subnetworks.
- Causal features are label-discriminative and domain-invariant.
- Non-causal features capture domain-specific information.



3.2 Unlearning Phase

- On-server unlearning updates only the **non-causal branch** via gradient matching using gradients from both retain and forget clients.
- Pull retain gradients closer to **reduce conflict** and preserve retained knowledge.
- Push forget gradients away to **induce conflict** and remove forgotten influence.

4 Results

- **Datasets:** PACS, VLCS, OfficeHome, TerraIncognita.
- **Backbones:** ResNet-18 on PACS/VLCS/OfficeHome; ResNet-50 on TerraIncognita.
- **20 IID clients** total, 5 clients per domain.
- Forget set = one domain; retain set = remaining domains.
- **Metrics:** FA↓, RA↑, TA↑, MIA↓, comm. cost, comp. cost, T2F↑.

Group	Methods	PACS				TerraIncognita			
		FA (↓)	RA (↑)	TA (↑)	MIA (↓)	FA (↓)	RA (↑)	TA (↑)	MIA (↓)
Retrain	Retrain	70.51	82.84	77.45	50.02	30.64	42.41	38.94	50.71
	FATS	74.45 (+3.94)	80.91 (−1.93)	75.98 (−1.47)	55.72 (+5.70)	33.07 (+2.43)	40.96 (−1.45)	37.34 (−1.60)	60.81 (+10.10)
	NoT	73.24 (+2.73)	79.25 (−3.59)	75.28 (−2.17)	59.13 (+9.11)	32.26 (+1.62)	38.32 (−4.09)	36.08 (−2.86)	62.90 (+12.19)
Approx. Unlearn.	FedCDP	77.37 (+6.86)	78.13 (−4.71)	76.41 (−1.04)	72.26 (+22.24)	34.56 (+3.92)	38.79 (−3.62)	36.75 (−2.19)	81.91 (+31.20)
	FedRecovery	76.48 (+5.97)	76.97 (−5.87)	74.81 (−2.64)	75.24 (+25.22)	32.11 (+1.47)	37.49 (−4.92)	35.59 (−3.35)	84.03 (+33.32)
	FedOSD	72.89 (+2.38)	80.15 (−2.69)	75.49 (−1.96)	58.84 (+8.82)	32.63 (+1.99)	40.77 (−1.64)	36.76 (−2.18)	60.54 (+9.83)
	FFMU	73.31 (+2.80)	78.27 (−4.57)	74.14 (−3.31)	60.62 (+10.60)	33.76 (+3.12)	39.39 (−3.02)	36.04 (−2.90)	65.85 (+15.14)
	FUSED	75.94 (+5.43)	79.34 (−3.50)	76.86 (−0.59)	58.72 (+8.70)	33.64 (+3.00)	38.73 (−3.68)	36.74 (−2.20)	62.03 (+11.32)
FOUL	MoDE	72.53 (+2.02)	79.01 (−3.83)	75.79 (−1.66)	59.79 (+9.77)	32.17 (+1.53)	38.04 (−4.37)	35.74 (−3.20)	63.47 (+12.76)
	(1)	69.53 (−0.98)	93.11 (+14.55)	77.14 (−0.31)	53.82 (+3.80)	27.97 (−2.67)	43.81 (+1.40)	38.16 (−0.78)	56.40 (+5.69)
	(1) + (2)	70.97 (+0.46)	92.33 (+14.49)	76.43 (−1.02)	51.93 (+1.91)	29.92 (−0.72)	42.13 (−0.28)	39.16 (+0.22)	57.11 (+6.40)

