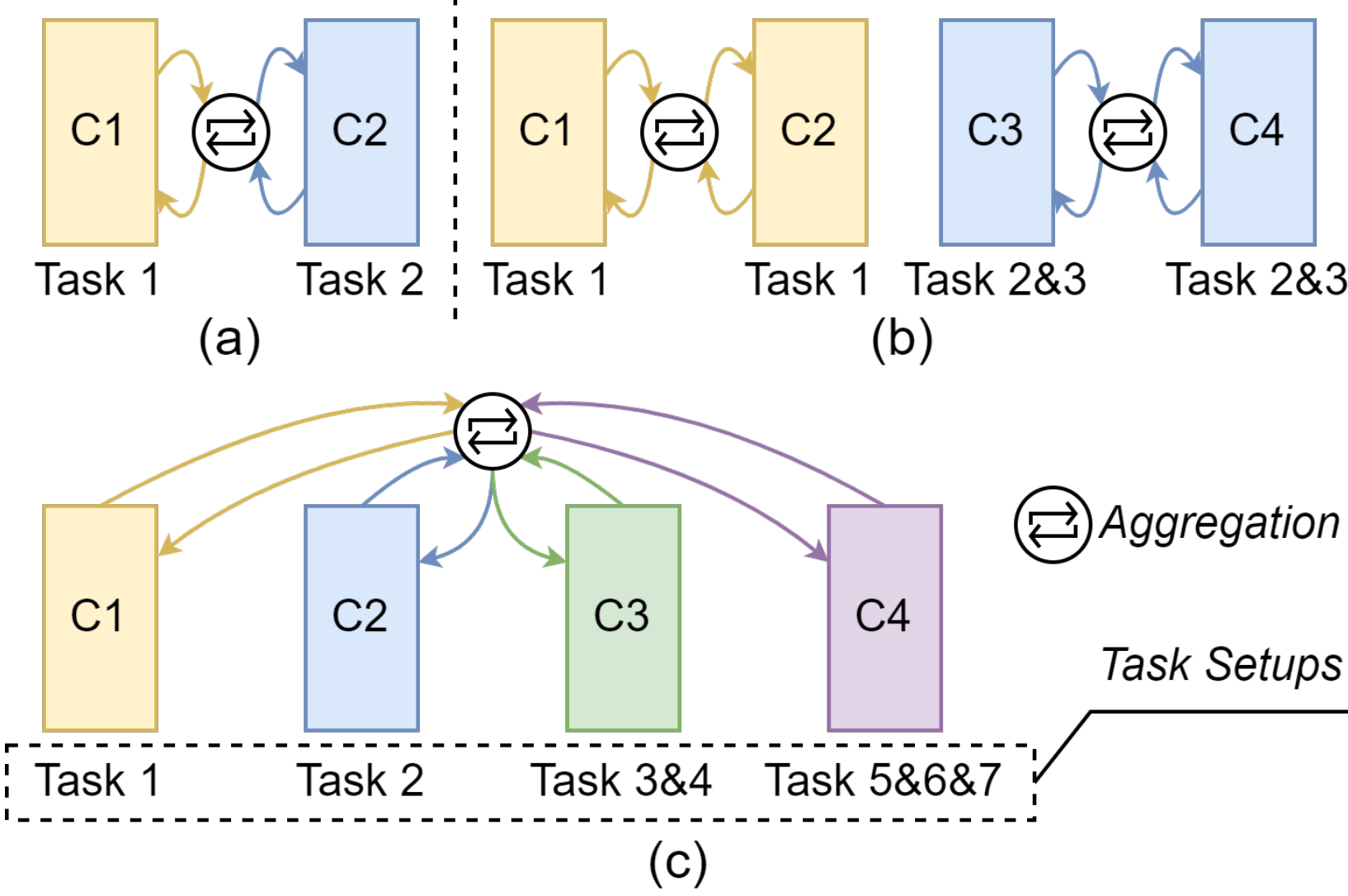


Motivation

Federated Multi-Task Learning (FMTL)

Conduct Multi-Task Learning in federated learning scenario



Limitations of Prior Works

- (a) Each client is dedicated to a single task
- (b) Clients are grouped w/ identical task setup

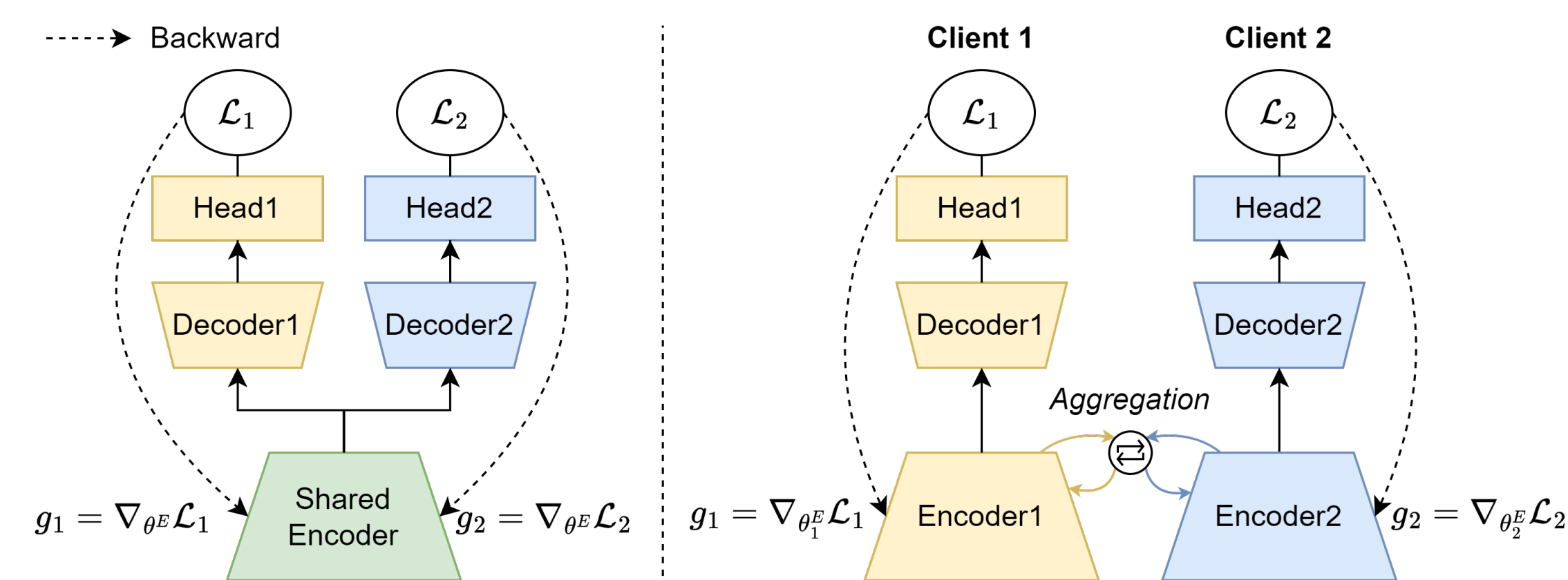
Hetero-Client Federated Multi-Task Learning

- ✓ A novel setting: each client w/ unique task setup
- ✓ Flexible and broaden applicability

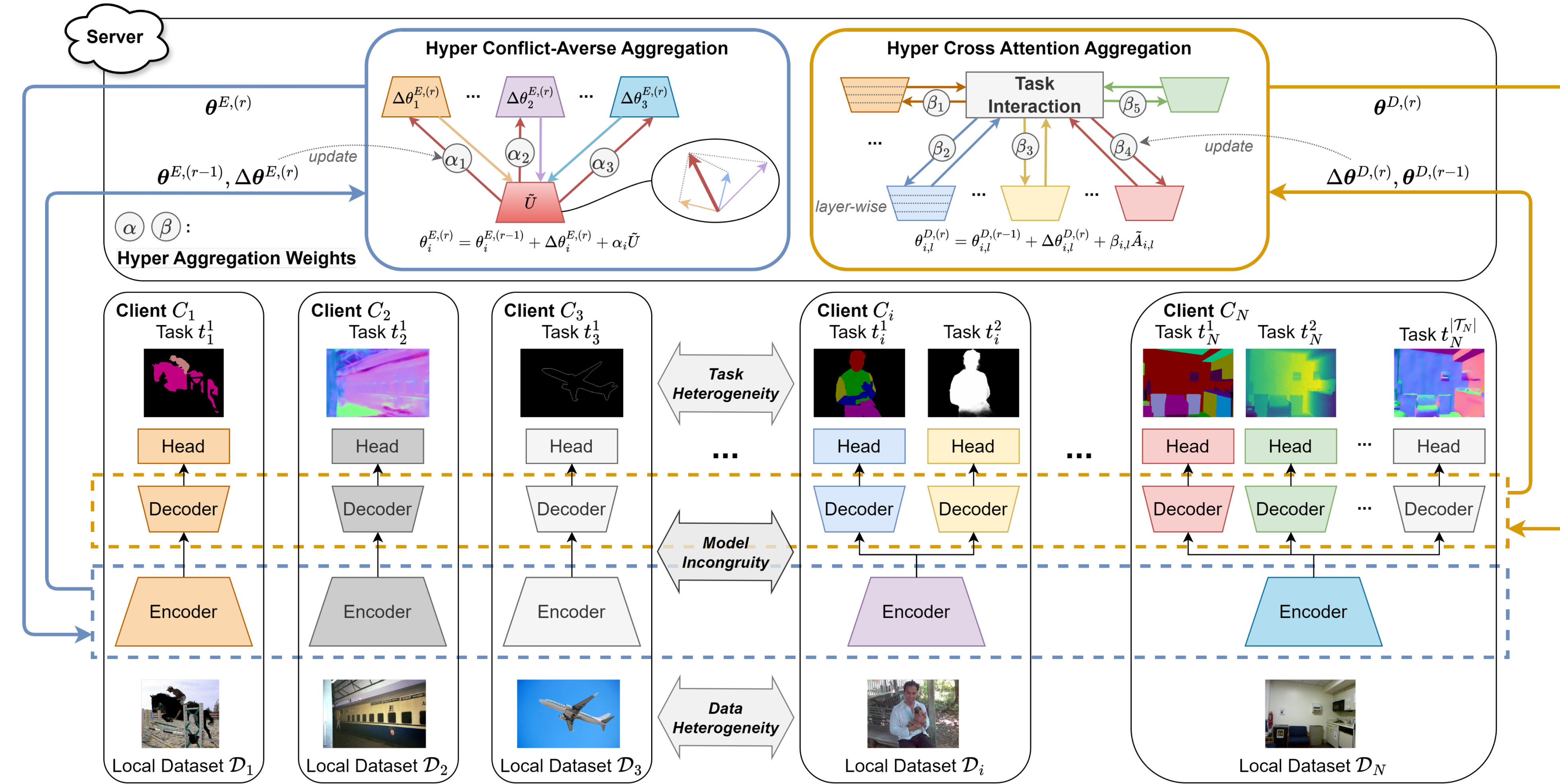
Challenges

- Data heterogeneity** clients w/ distinct private data domains
- Task heterogeneity** clients w/ distinct dense prediction tasks
- Model incongruity** disparity in multi-task model structure

Theorem: Difference in optimizing MTL and FL



FedHCA² Framework



Hyper Conflict-Averse Aggregation

- ✓ Alleviate update conflicts from heterogeneity
- ✓ Maintain shared knowledge

Conflict in $\Delta\theta_i^E$ and $\Delta\theta_j^E$ measured by $\langle \Delta\theta_i^E, \Delta\theta_j^E \rangle$
 \uparrow inner product $\approx \downarrow$ conflict

Solve for optimal aggregated update \tilde{U} :

$$\max_{\tilde{U}} \min_i \langle \Delta\theta_i^E, \tilde{U} \rangle \quad \text{s.t.} \quad \|\tilde{U} - \Delta\bar{\theta}^E\| \leq c \|\Delta\bar{\theta}^E\|$$

Compute personalized update:

$$\theta_i^{E,(r)} = \theta_i^{E,(r-1)} + \Delta\theta_i^{E,(r)} + \alpha_i \tilde{U}$$

Hyper Cross Attention Aggregation

- ✓ Model cross-task relationships w/ attention
- ✓ Fine-grained w/ layer-wise strategy

$$V_l = [\Delta\theta_{1,l}^D, \dots, \Delta\theta_{K,l}^D]^T$$

$$\tilde{A}_{i,l} = \text{Softmax}(\Delta\theta_{i,l}^D V_l^T / \sqrt{d}) V_l$$

Hyper Aggregation Weights

- ✓ Assess importance of aggregated parameters
- ✓ High-level guidance in harmonizing local update and collaborative update

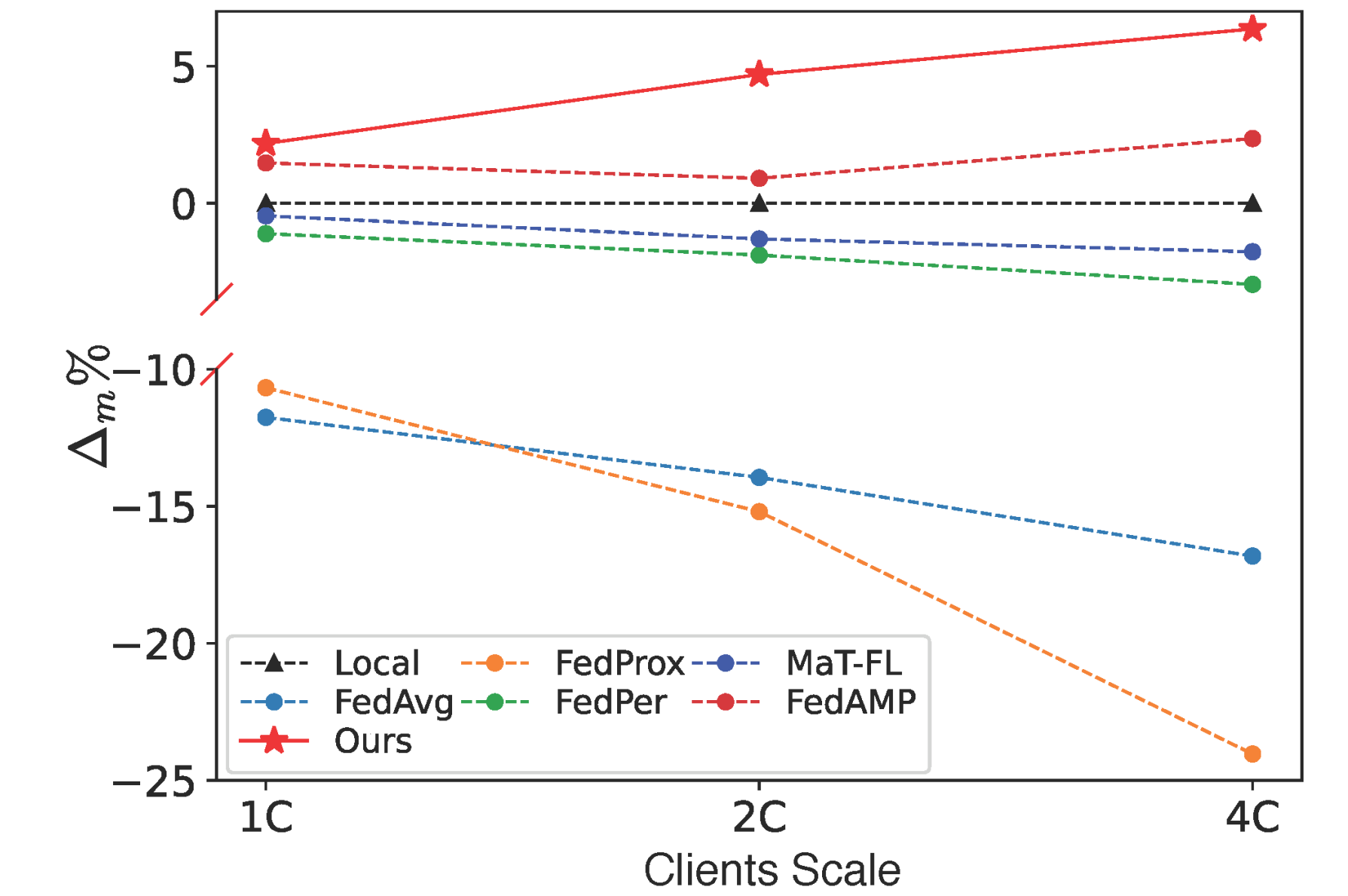
$$\theta_i^{(r)} = \theta_i^{(r-1)} + \Delta\theta_i^{(r)} + \psi_i \tilde{\theta}_i$$

$$\nabla_{\psi_i} \mathcal{L}_i = (\nabla_{\psi_i} \theta_i^{(r)})^T \nabla_{\theta_i^{(r)}} \mathcal{L}_i = (\tilde{\theta}_i)^T \nabla_{\theta_i^{(r)}} \mathcal{L}_i$$

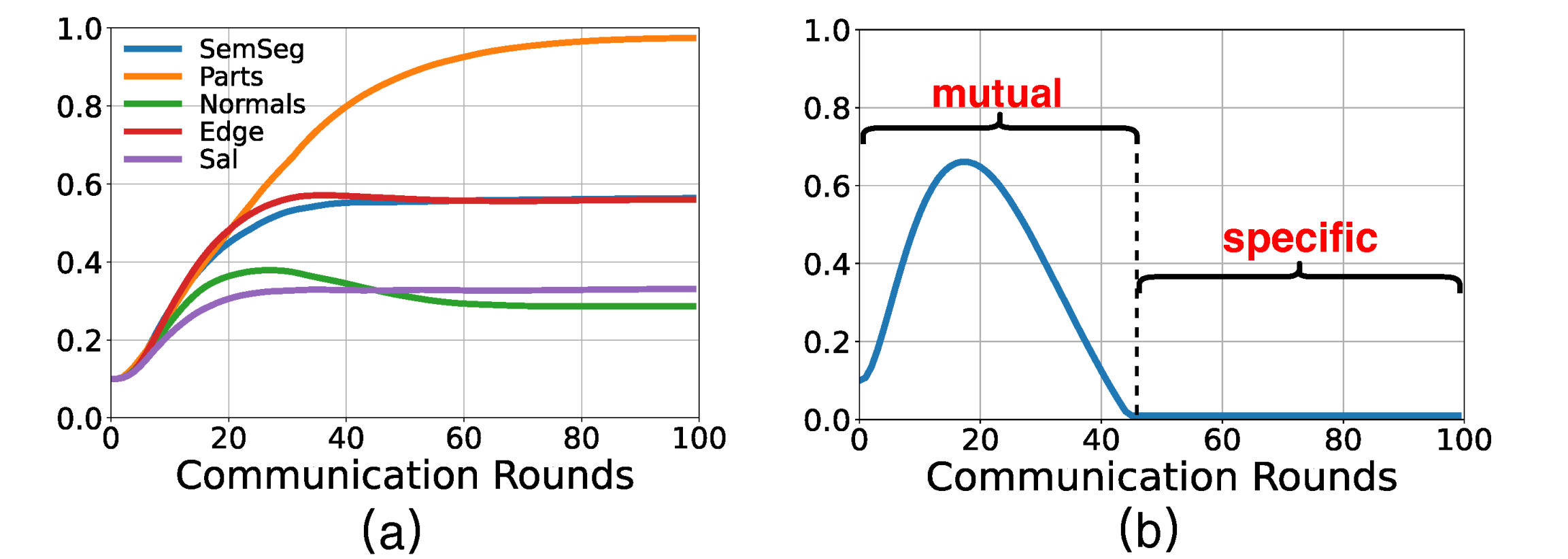
Experiments

HC-FMTL Benchmark

- 5 ST client (PASCAL-Context) + 1 MT client (NYUD-v2)
- 4 ST client (NYUD-v2) + 1 MT client (PASCAL-Context)



Hyper Aggregation Weights on Encoder



Hyper Aggregation Weights on Decoder

