MAFS: Modality-Aware Federated Semi-Supervised Learning with Selective Data Sharing Specified by Individual Clients

Yi-Chen Li (calvin0205calvin0205@gmail.com)

Network and Multimedia Systems Lab
Department of Computer Science
National Tsing Hua University

NMSL@NTHU
Networking and Multimedia Systems Lab

國立清華大學
NATIONAL TSING HUA UNIVERSITY

Outline

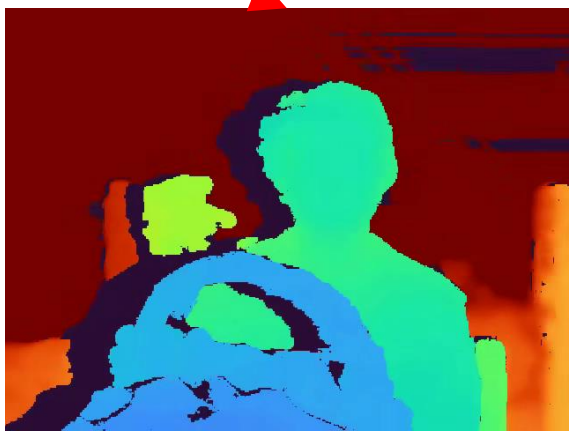
- Introduction
- Related Work
- Modality-Aware Federated Semi-Supervised Learning (MAFS)
- Multimodal Applications
- Experiment Setup
- Evaluations
- Conclusion & Future Works

Multi-Modal Sensors

- We often use **machine learning models** in our daily lives to help us finish some certain tasks.
- In order to train these models, we need to deploy **sensors** to collect training data.

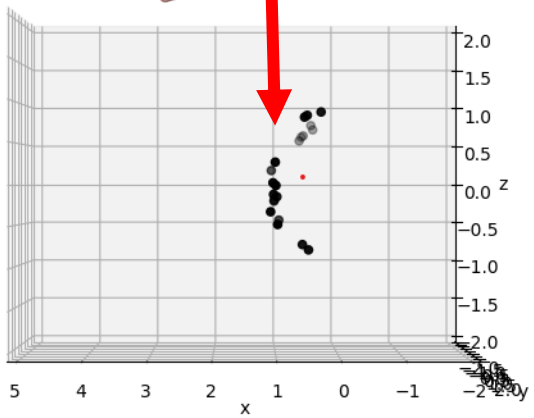
Intel Realsense

RGB image



IR image

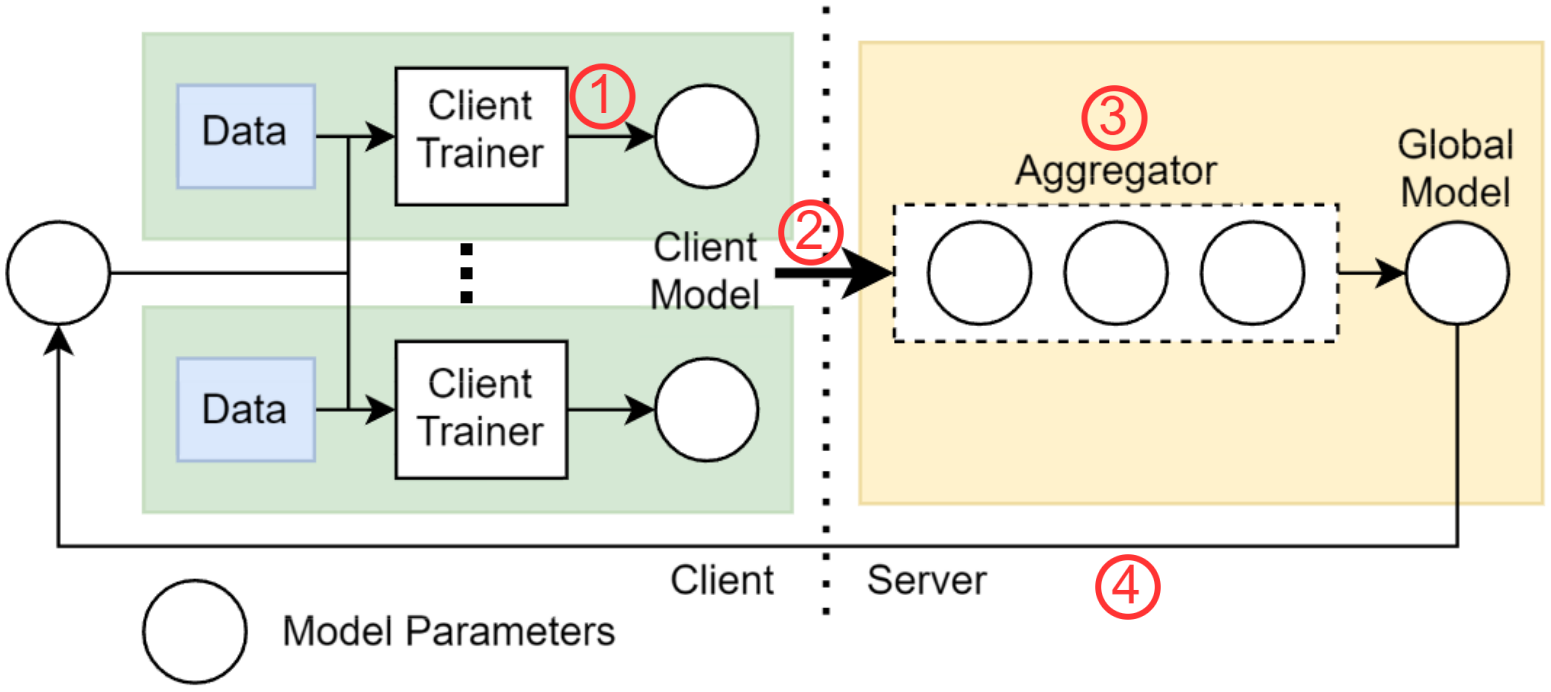
mmWave radar



point clouds

Federated Learning

- As **privacy** becomes increasingly important to everyone, how the collected sensor data is handled has become a very important issue.



① Train the client model.
② Upload client models.
③ Aggregate different client models.
④ Send back the global model.

Problem Statement

- **Labeled data scarcity** in Federated Learning poses challenges in training robust and generalizable models.
- Without enough labeled examples, a model might **overfit to the limited data** it has encountered. This leads to **poor performance on unseen data**.

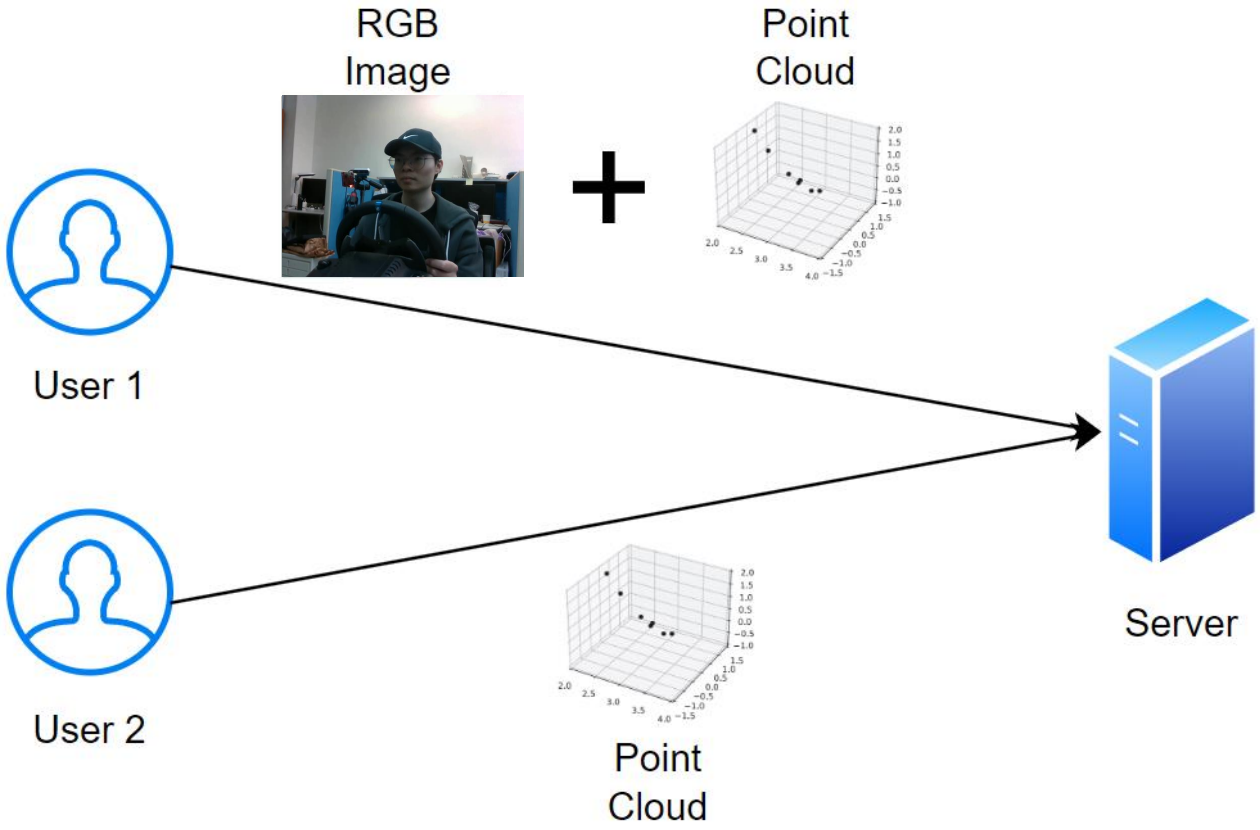
Objective

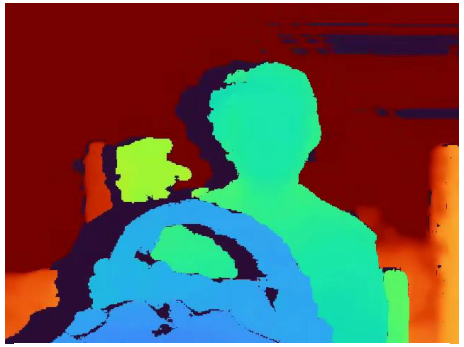
- Use **unlabeled data** to increase the accuracy of the client model and improve its robustness.

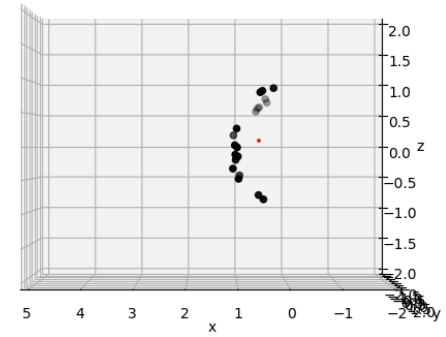
Challenges

- How to **get** the unlabeled data?
- How to **utilize** the unlabeled data?
- How to solve the **missing modality** issue?

Solution 1: Selective Data Sharing



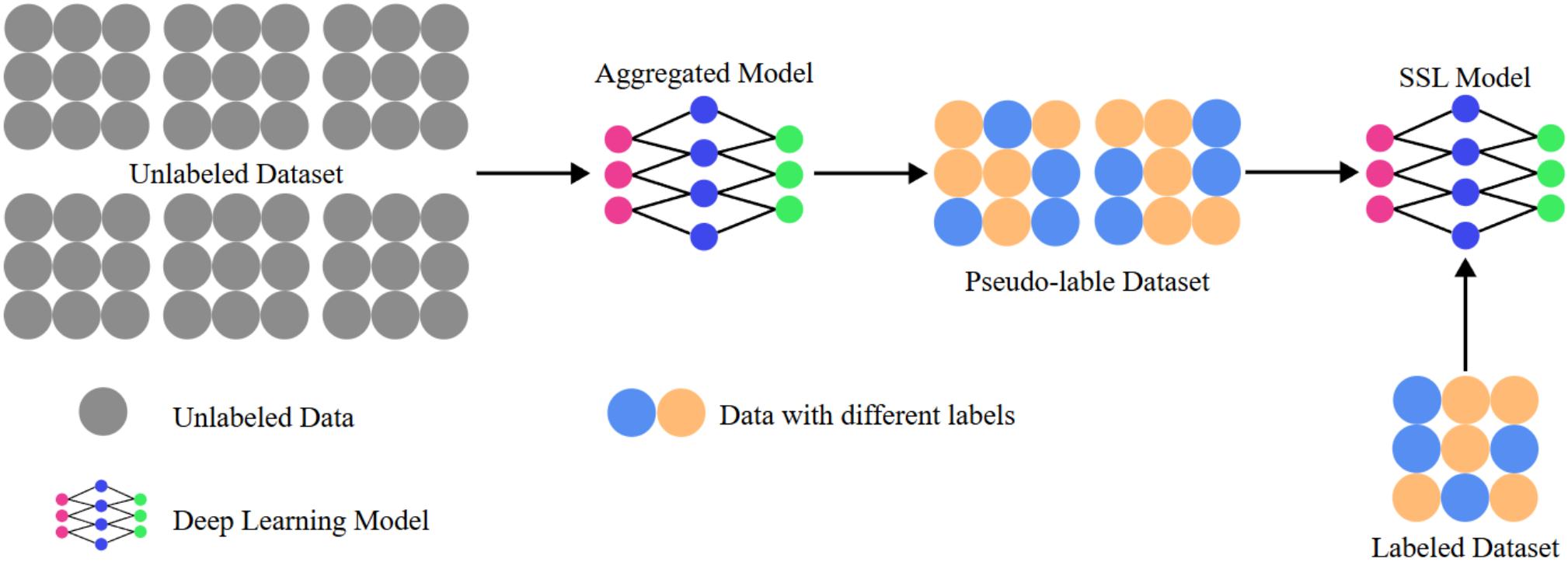




- Users can decide **which types of data to share** based on their own privacy considerations.

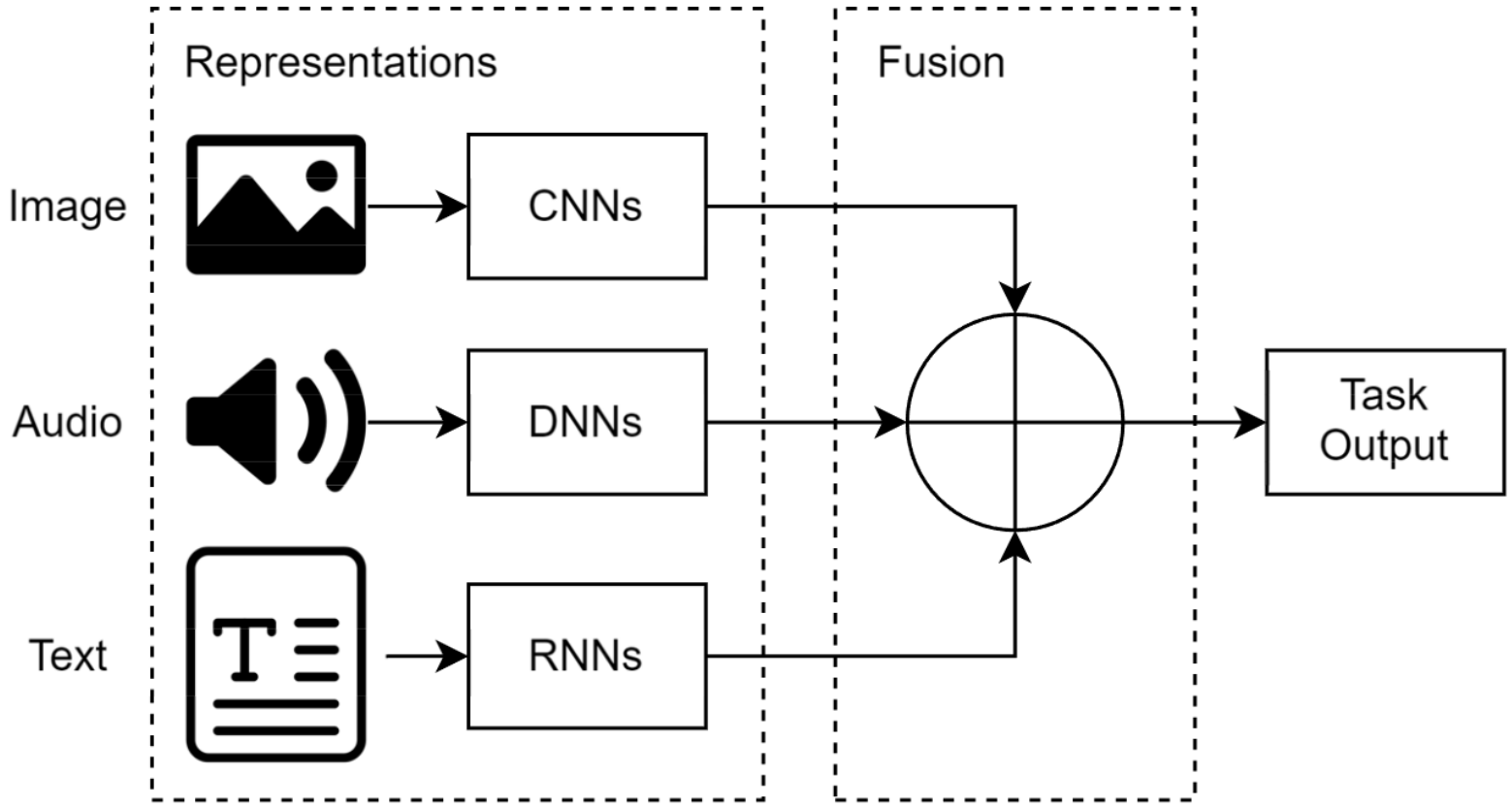
Solution 2: Semi-Supervised Learning

- The **knowledge** contained in the **unlabeled data** can significantly improve model performance.
- We need to effectively obtain **correct labels** for unlabeled data.



Solution 3: Multimodal Representation Learning

- We use **different networks** to process **different modalities**.
- We fill the missing modality input with **zeros**.



Outline

- Introduction
- Related Work
- Modality-Aware Federated Semi-Supervised Learning (MAFS)
- Multimodal Applications
- Experiment Setup
- Evaluations
- Conclusion & Future Works

Related Work

- **Multimodal Federated Learning (MFL)**
  - Modality-specific feature extraction [AAAI'22, SIGIR'21]

The features obtained after extraction contain less information than the raw data.

- **Federated Semi-Supervised Learning (FSSL)**
  - Client-side FSSL [arxiv'20]

The SSL model trained by each client will be biased towards its own data.

  - Server-side FSSL [TMC'23]

Only applicable to unimodal datasets, and requires all raw data for training.

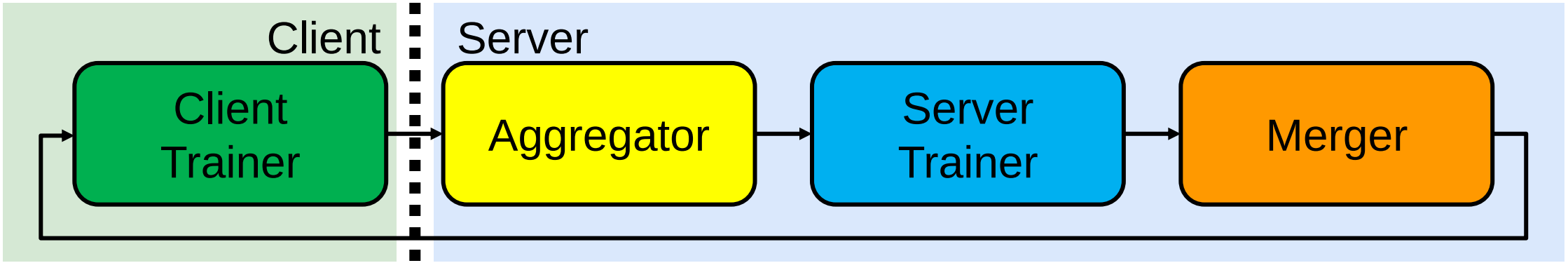
- **Modality-Aware Selective Data Sharing**
  - HPFL [TOMM'24]

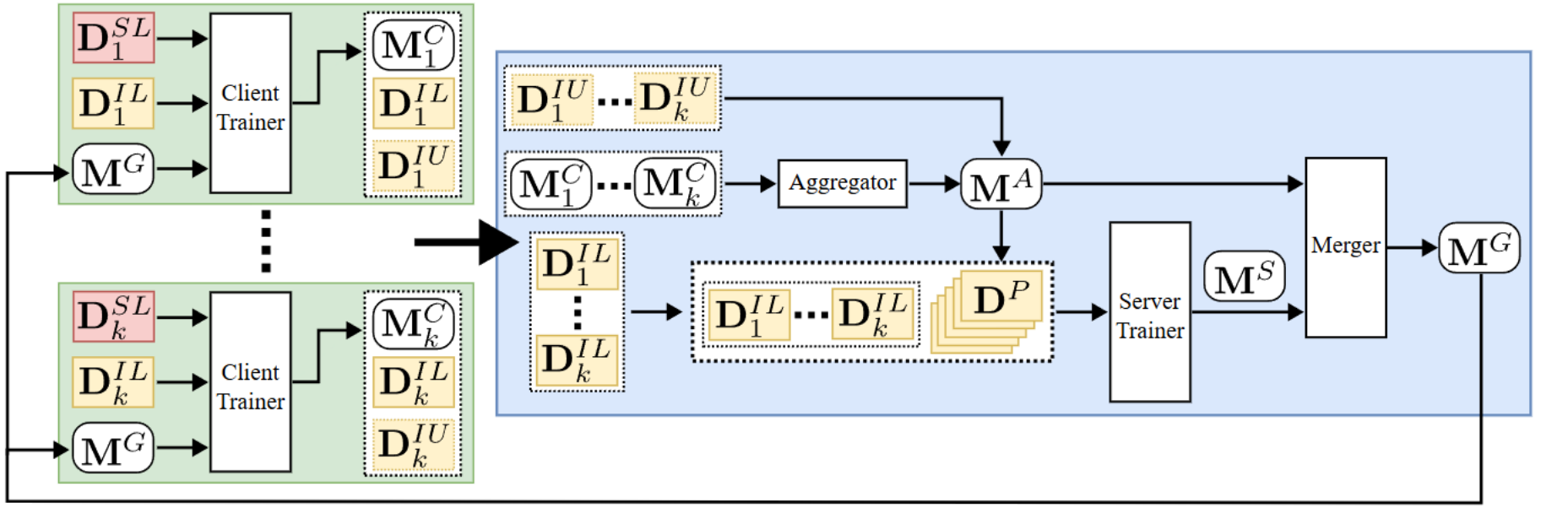
This paradigm cannot perform well on labeled data scarcity problem.

Outline

- Introduction
- Related Work
- Modality-Aware Federated Semi-Supervised Learning (MAFS)
- Multimodal Applications
- Experiment Setup
- Evaluations
- Conclusion & Future Works

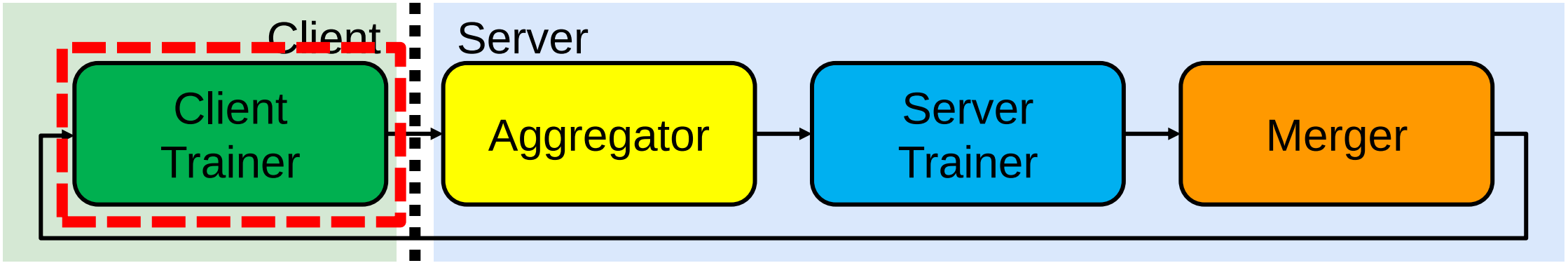
MAFS Paradigm Workflow



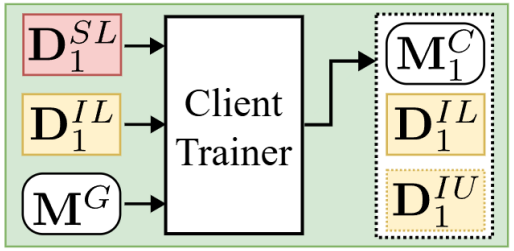


$\mathbf{D}^{SL}$ Sensitive Labeled Data
$\mathbf{D}^{IU}$ Insensitive Unlabeled Data
$\mathbf{M}^{C}$ Client Model Parameter
$\mathbf{M}^{A}$ Aggregated Model Parameter
$\mathbf{D}^{IL}$ Insensitive Labeled Data
$\mathbf{M}^{G}$ Global Model Parameter
$\mathbf{M}^{S}$ Semi-Supervised Learning Model Parameter

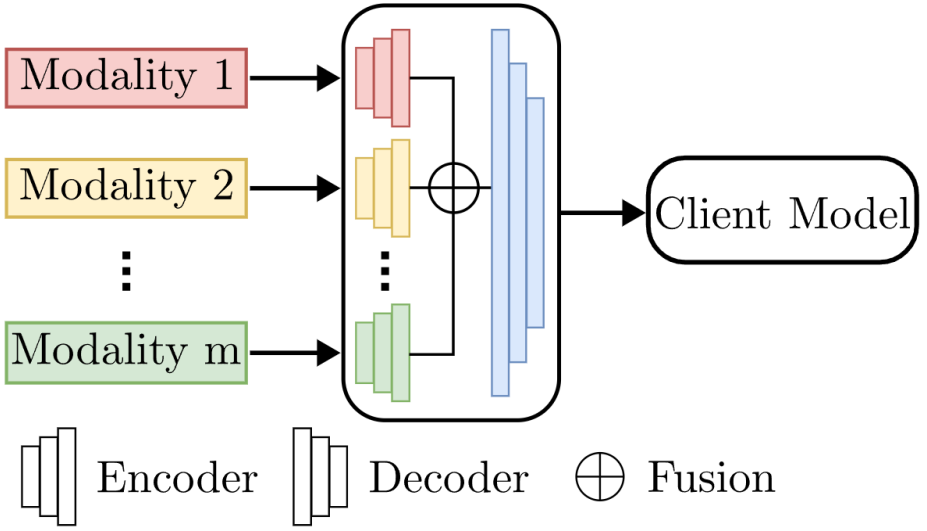
Client Trainer



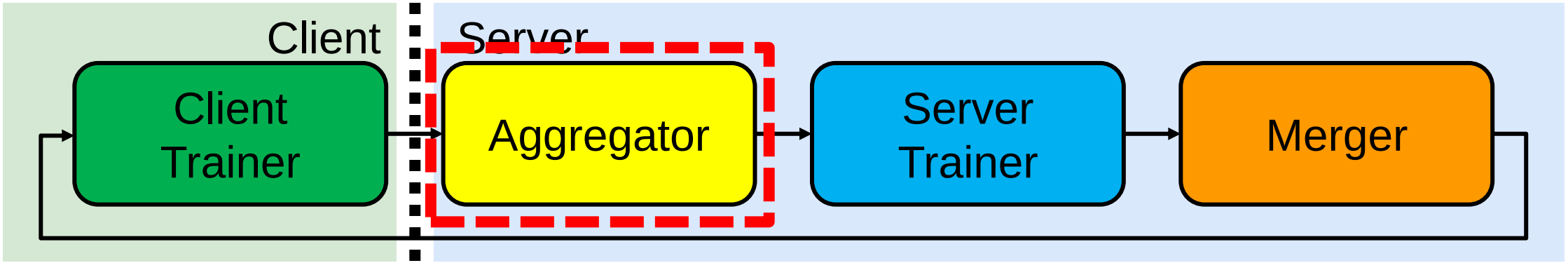
- Methodology
  1. Use an **encoder** to convert raw data into **feature vectors**.
  2. **Fuse feature vectors** of different modalities through mid-level fusion.
  3. Feed the fused vector into a **decoder** to obtain the output.



D^{SL}: Sensitive Labeled Data
D^{IL} : Insensitive Labeled Data
M^G : Global Model
M^C : Client Model



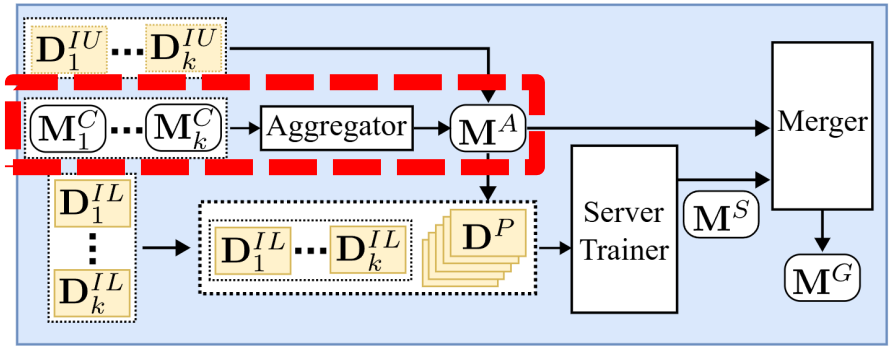
Aggregator



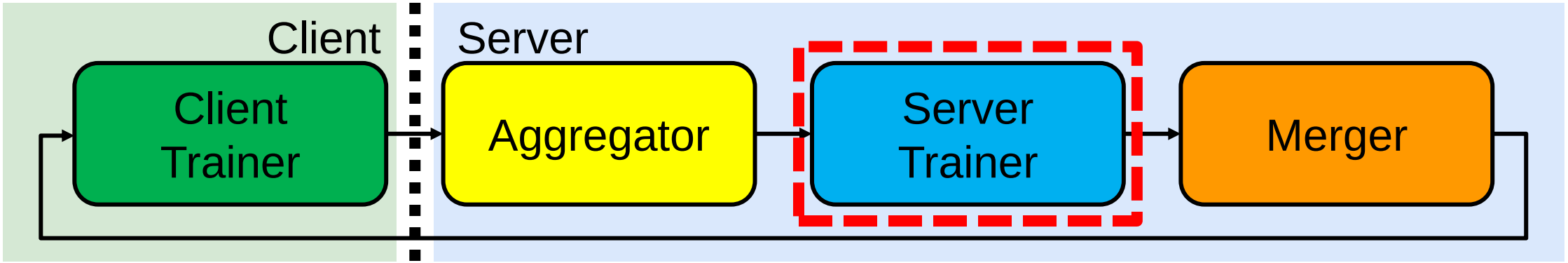
- We **aggregate all the client models** sent from different clients through the server aggregator, which is the same as the FL workflow.
- The default aggregator use Fed Avg to generate the aggregated model.

$$\mathbf{M}_t^A = \frac{1}{K}\sum_{i=1}^{K}\mathbf{M}_{i,t}^C$$

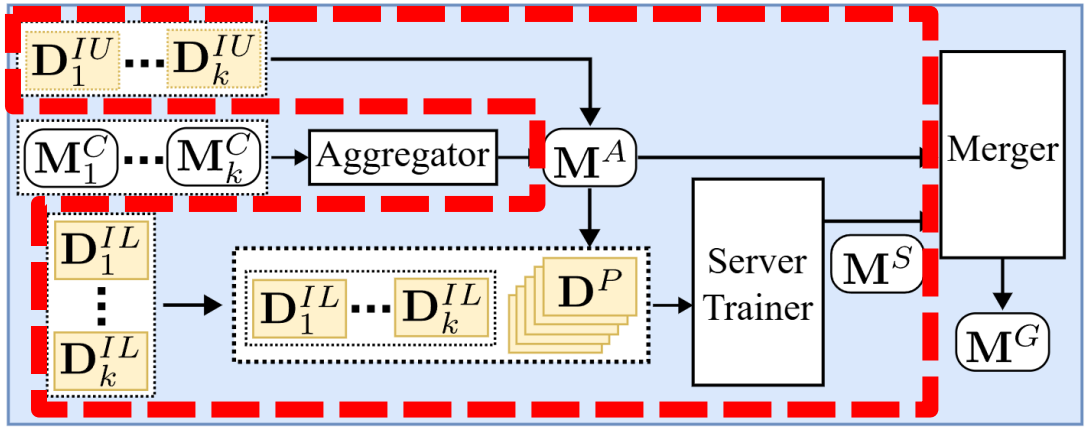
- MAFS can be generalized for different FL algorithms.



Server Trainer

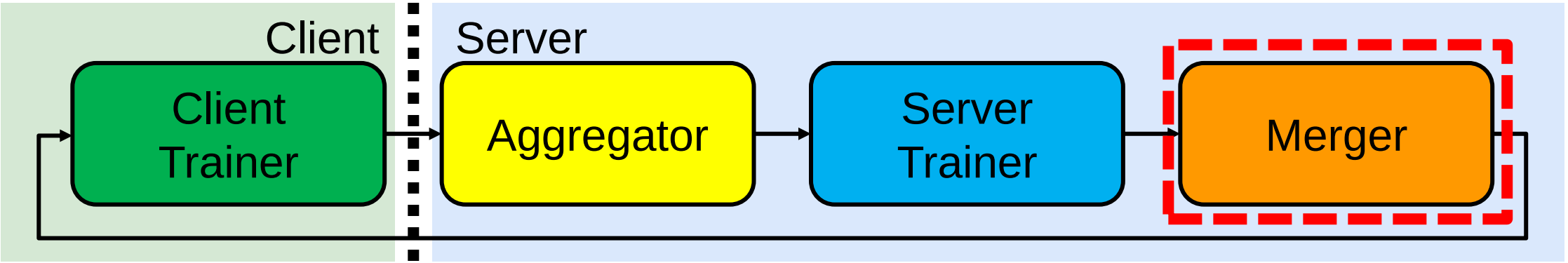


- Methodology
  1. Use the aggregated model to **pseudo-label** the **insensitive data** and generate the **pseudo-label dataset**.
  2. We use system parameter **τ** as the **pseudo-labeling threshold**.
  3. Train using both the **labeled dataset** and the **pseudo-label dataset** to generate the **semi-supervised model**.

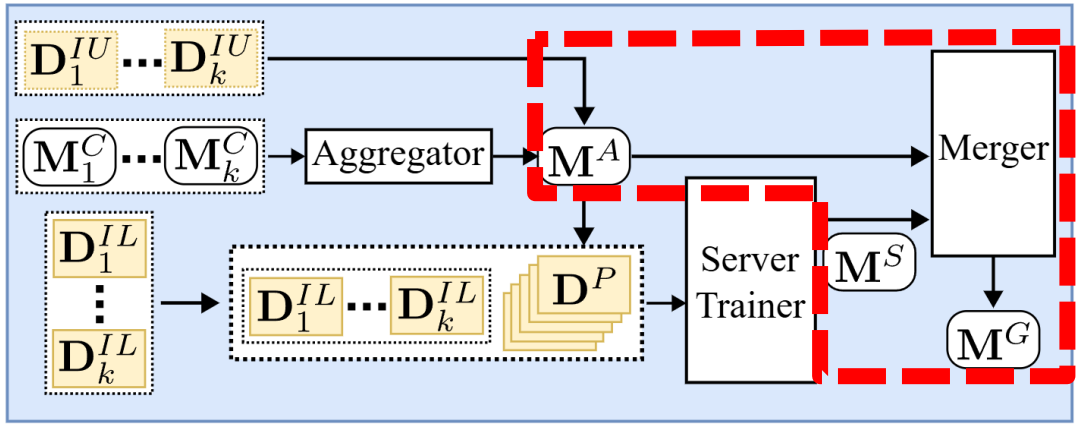


D^{IU} : Insensitive Unlabeled Data
D^{IL} : Insensitive Labeled Data
D^P : Pseudo-label Data
M^A : Aggregated Model
M^S : Semi-Supervised Learning Model

Merger



- The **aggregated model** represents the knowledge learning from **labeled data**, while the **semi-supervised learning model** represents **pseudo-label data**.
- We use the **weighted sum** and set the system parameter **α** to balance the proportion between the aggregated model and the semi-supervised learning model.

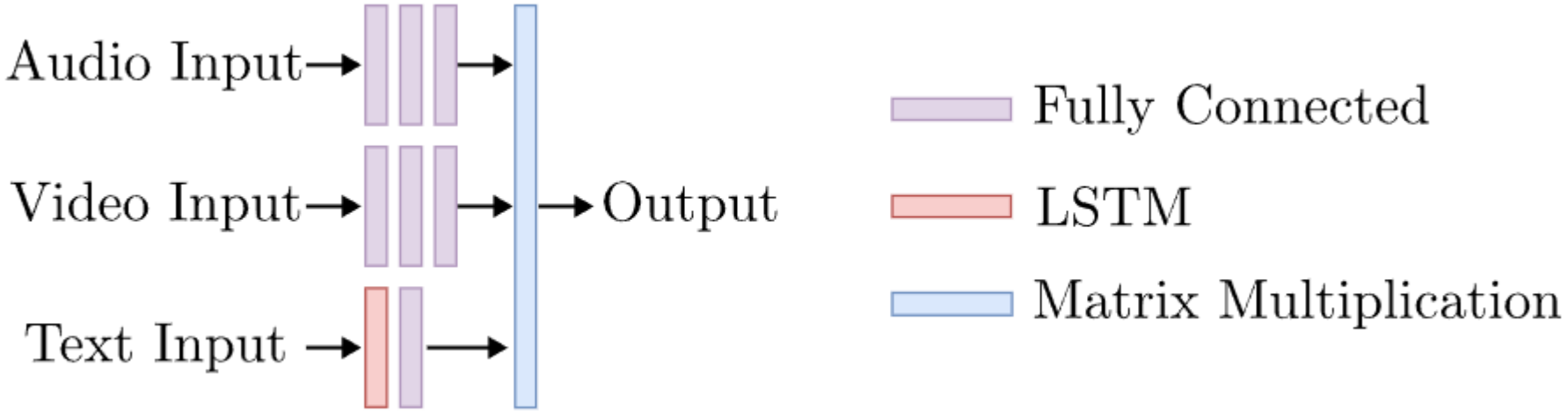


Outline

- Introduction
- Related Work
- Modality-Aware Federated Semi-Supervised Learning (MAFS)
- Multimodal Applications
- Experiment Setup
- Evaluations
- Conclusion & Future Works

Emotion Recognition (ER)

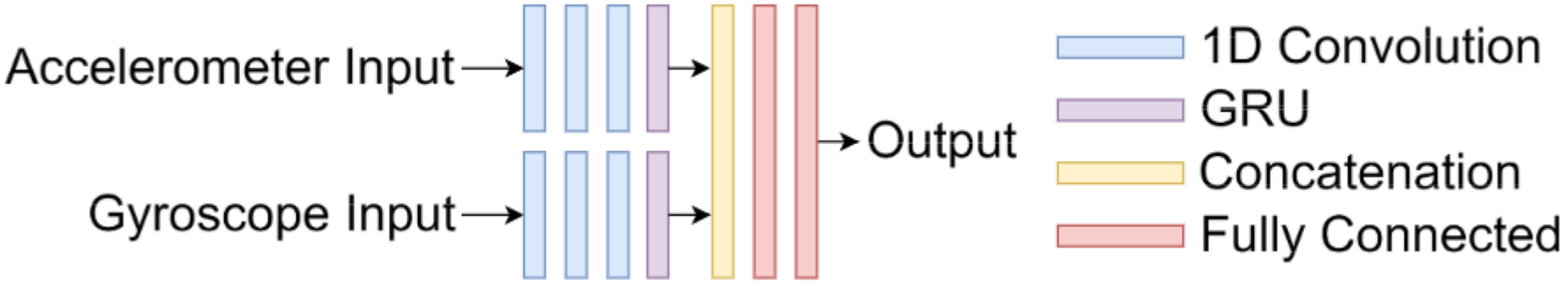
- We use the IEMOCAP [1] dataset for the ER application.
- We adopted the Low-rank-Multimodal-Fusion [2] approach as our neural network structure.



[1] Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette Chang, Sungbok Lee, and Shrikanth Narayanan. 2008. IEMOCAP: Interactive emotional dyadic motion capture database. Language Resources and Evaluation 42, 4 (2008).
[2] Zhun Liu, Ying Shen, Varun Bharadhwaj Lakshminarasimhan, Paul Liang, Amir Zadeh, and Louis-Philippe Morency. 2018. Efficient low-rank multimodal fusion with modality-specific factors. arXiv preprint arXiv:1806.00064 (2018).

Human Activity Recognition (HAR)

- We use the KU-HAR [1] dataset for the HAR application.
- We refer to FedMultimodal's [2] neural network structure for training.



[1] Niloy Sikder and Abdullah-Al Nahid. 2021. KU-HAR: An open dataset for heterogeneous human activity recognition. Pattern Recognition Letters 146 (2021), 46–54.
[2] Tiantian Feng, Digbalay Bose, Tuo Zhang, Rajat Hebbar, Anil Ramakrishna, Rahul Gupta, Mi Zhang, Salman Avestimehr, and Shrikanth Narayanan. 2023. Fedmultimodal: A benchmark for multimodal federated learning. In Proc. of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 4035–4045.

Outline

- Introduction
- Related Work
- Modality-Aware Federated Semi-Supervised Learning (MAFS)
- Multimodal Applications
- Experiment Setup
- Evaluations
- Conclusion & Future Works

Datasets and Data Partition

- **ER task**
  - IEMOCAP contains 4453 **triplets of audio, video, and text data**.
  - We split IEMOCAP into 3515 training (**80%**) and 938 testing (**20%**) samples.
  - We set different **Dirichlet parameters** to control the amount of data distributed to individual clients.

- **HAR task**
  - KU-HAR contains **accelerometer and gyroscope** data.
  - We use a subset of KU-HAR from **65 users** and **8 actions**.
  - We divided 65 users into **63 for training, 1 for validation, and 1 for testing**.
  - We perform **5-fold cross-validation**.

Hyperparameters

Hyperparameter	ER Task	HAR Task
Rounds	100	200
Client Epochs per Round	3	1
Server Epochs per Round	10	10
Batch Size	16	16
Loss Function	Cross-Entropy	NLLLoss
Learning Rate	$\eta_t = 0.003 \times 0.965^{t-1}$	$\eta_t = 0.001$
Client Optimizer	Adam	SGD
Client Weight Decay	0.002	0.0005
Server Optimizer	SGD	SGD
Server Momentum	0.9	0.9
Server Weight Decay	0.0005	0.0005
Regularization Parameter (FedProx)	0.001	0.001

System Parameter

- **ER task**
  - No. client $c \in \{8, 16, 32\}$
  - Pseudo-label threshold $\tau \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$
  - Merger weight $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$
  - Dirichlet Distribution Parameter $\in \{0.1, 1, 10\}$
- **HAR task**
  - Pseudo-label threshold $\tau \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$
  - Merger weight $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$
- **SOTAs**
  - FedAvg [1], FedProx [2], FedOpt [3]

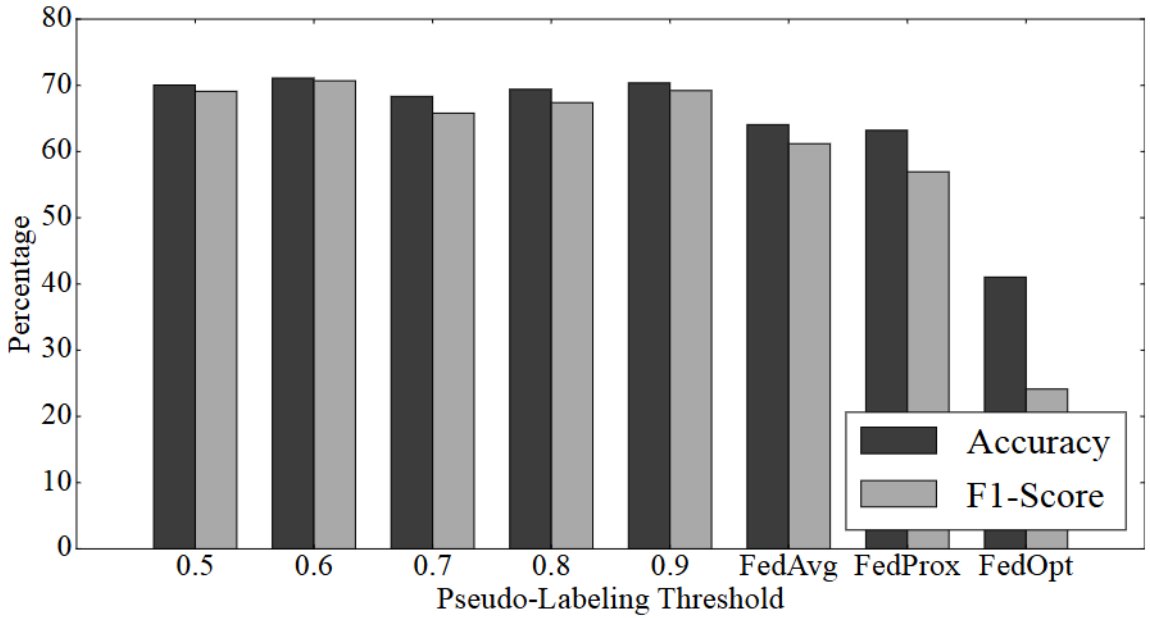
[1] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In Proc. of PMLR International Conference on Artificial Intelligence and Statistics (AISTATS). 1273–1282.
[2] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. 2020. Federated optimization in heterogeneous networks. Proc. of Machine learning and systems 2 (2020), 429–450.
[3] Sashank Reddi, Zachary Charles, Manzil Zaheer, Zachary Garrett, Keith Rush, Jakub Konečn `y, Sanjiv Kumar, and Brendan McMahan. 2020. Adaptive federated optimization. arXiv preprint arXiv:2003.00295 (2020).

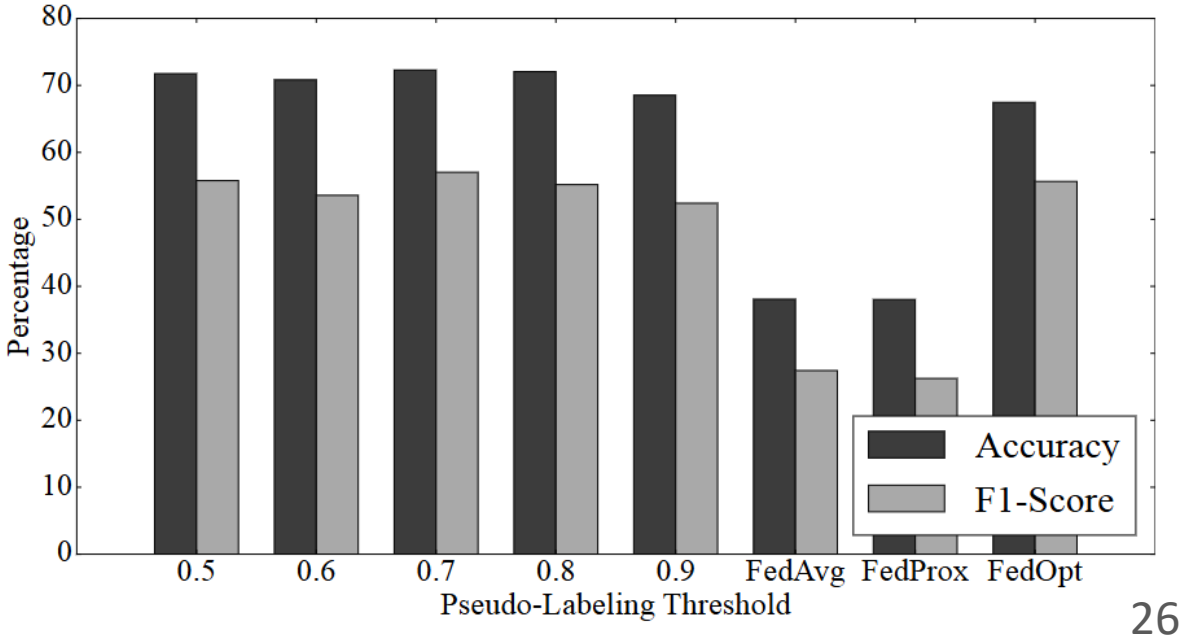
Outline

- Introduction
- Related Work
- Modality-Aware Federated Semi-Supervised Learning (MAFS)
- Multimodal Applications
- Experiment Setup
- Evaluations
- Conclusion & Future Works

Impact of Pseudo-Labeling Threshold τ

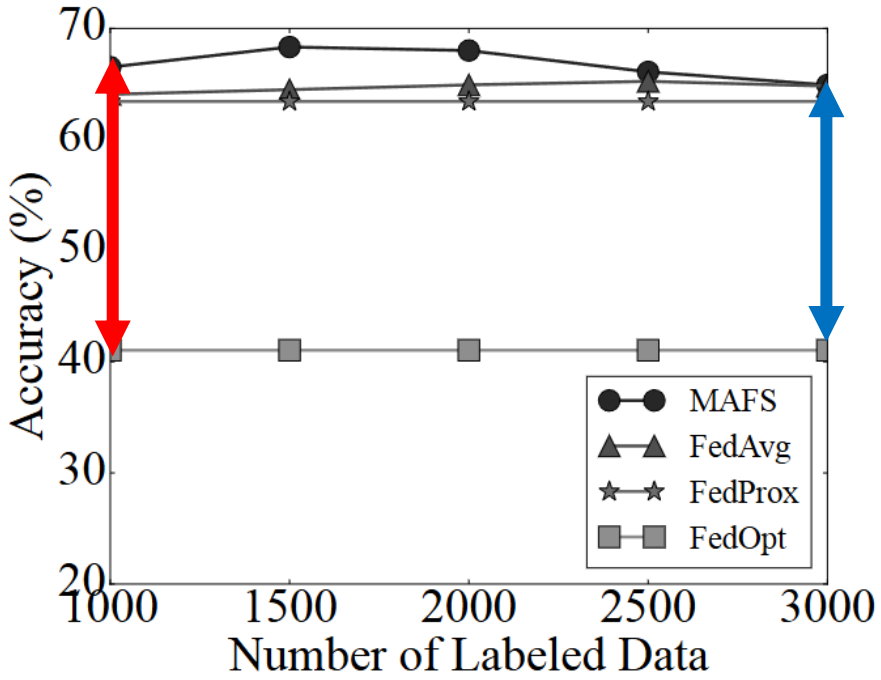
- For the **ER** task, using a threshold τ of **0.6** during pseudo-labeling resulted in the largest improvements in accuracy (6.94%) and F1-score (9.49%).
- For the **HAR** task, τ does not affect the accuracy and F1-score much.
- We recommend using **0.6** as the default τ value for both tasks.

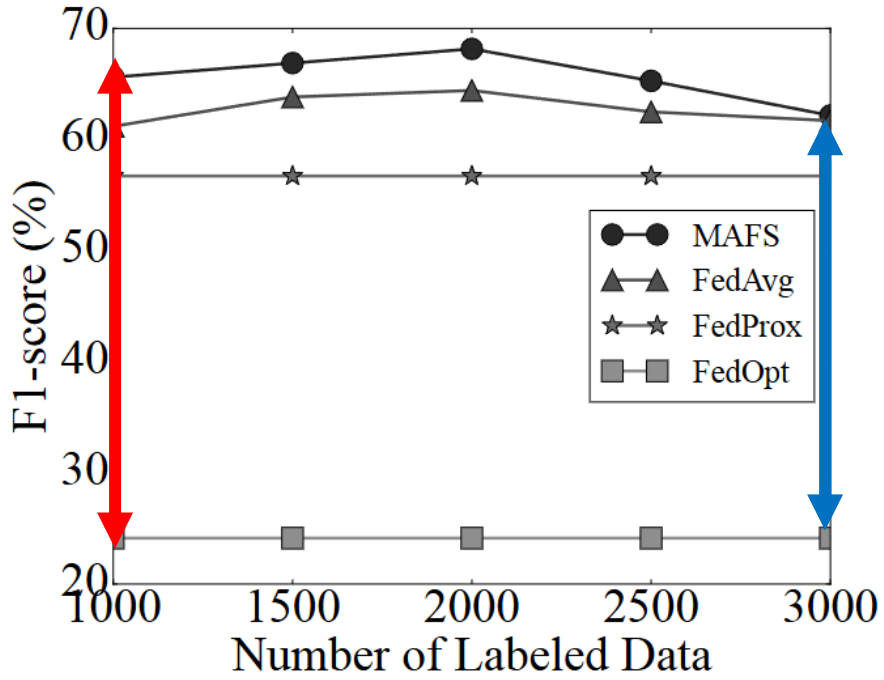




Impact of Labeled Data Proportion in ER Task

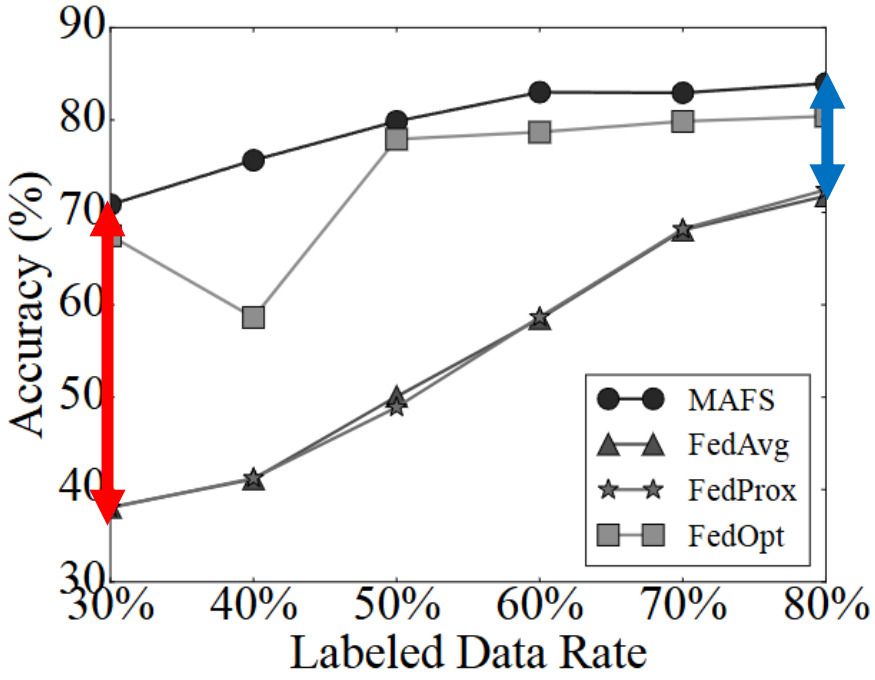
- Training solely with labeled data leads to a **decline in both accuracy and F1-score** as the quantity of labeled data decrease.
- **MAFS enhances model performance** while increasing the amount of training data.

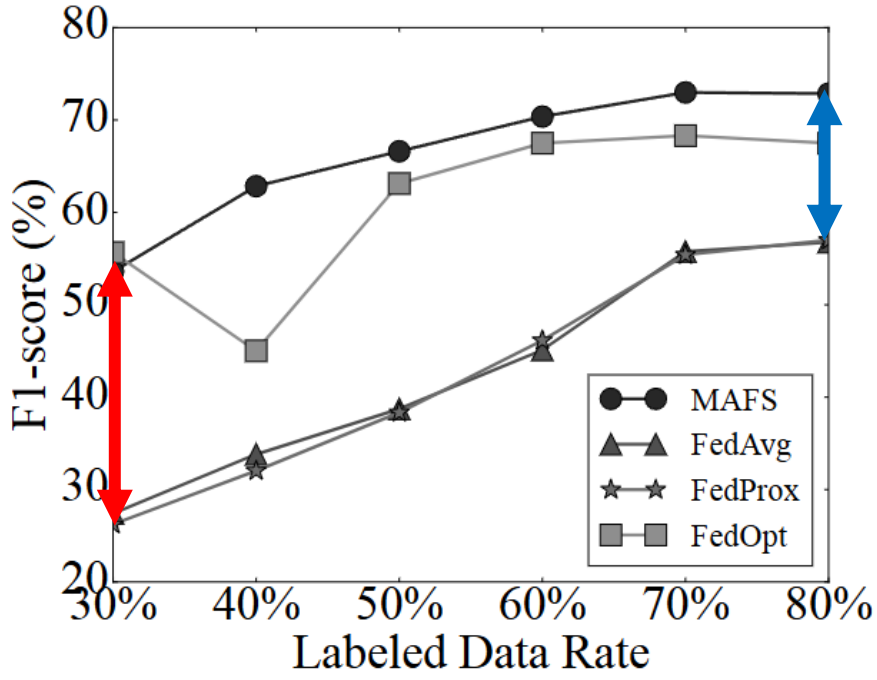




Impact of Labeled Data Proportion in HAR Task

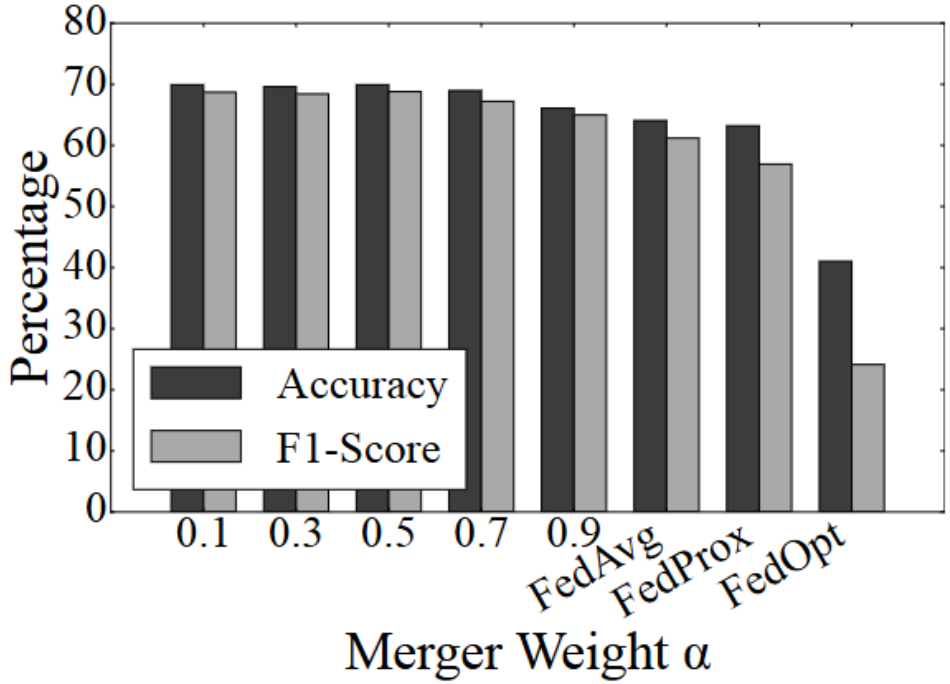
- **MAFS performs the best** among all compared methods, regardless of whether the labeled data rate is high or low.
- **The lower the labeled data rate, the greater the improvement** in model performance by MAFS.

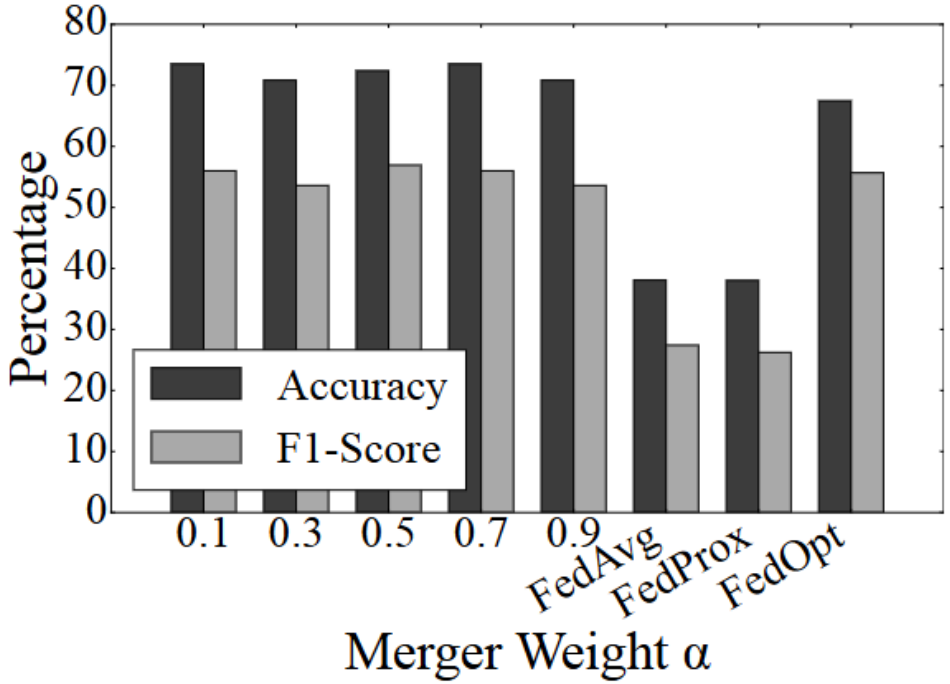




Impact of Merger Weight α

- A gradual **decrease in the α value** correlates with **increases in both model accuracy and F1-score.**
- We recommend setting the α value to **0.1**.





Impact of the Dirichlet Parameter

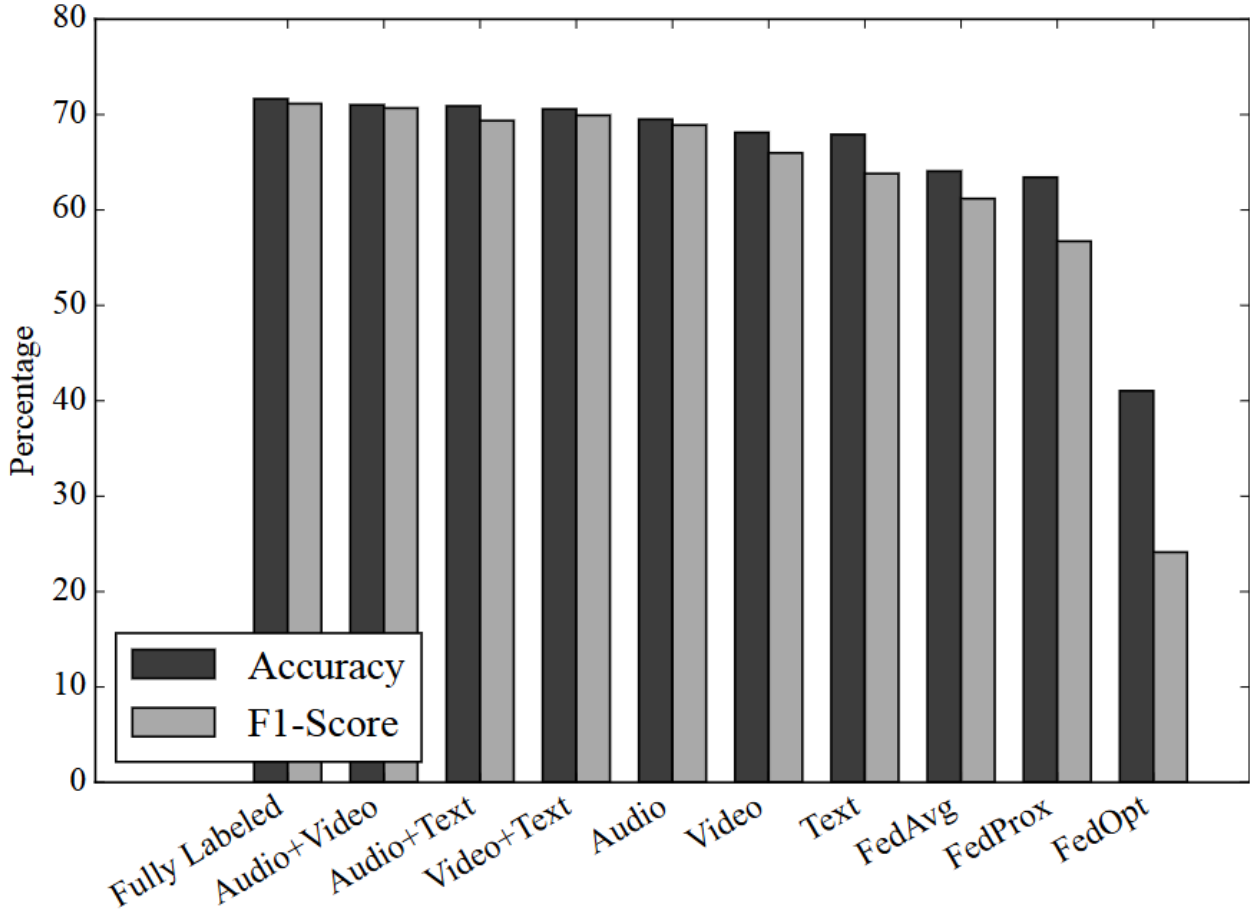
- MAFS leads to **very little fluctuations** in the accuracy and F1-score **regardless of the non-i.i.d. severity**.
- MAFS consistently **outperforms FedAvg** by a large margin
- MAFS **increases the amount of data available** for training the SSL model even in situations with considerable non-i.i.d. distribution.

Uneven ← → **Even**

	Dirichlet Para. 0.1		Dirichlet Para. 1		Dirichlet Para. 10	
Algo.	**Accuracy**	**F1-score**	**Accuracy**	**F1-score**	**Accuracy**	**F1-score**
MAFS	69.93%	68.70%	69.72%	68.48%	69.82%	69.69%
FedAvg	64.07%	61.19%	63.43%	55.38%	66.52%	62.61%

MAFS Encourages Users to Share More Modalities for Better Performance

- Sharing two modalities, like audio and text, results in good performance.
- When clients share only one modality, such as video, the performance gap increases to 3.73% and 7.33%.



MAFS Allows Privacy-conscious Clients to Selectively Share Fewer Modalities

	Fewer Samples				More Samples			
# of clients share two modalities	0	2	4	6	0	2	4	6
Accuracy	0%	+2.02%	+2.77%	+2.66%	0%	+2.45%	+3.41%	+2.45%
F1-score	0%	+3.32%	+4.25%	+3.10%	0%	+3.98%	+4.95%	+3.51%

Summary

- MAFS **outperforms SOTAs** under different labeled data rates and data distributions.
- MAFS allows and encourages clients to **selectively share more data modalities**, while more clients share more data modalities lead to better model performance.
- MAFS works well in two sample tasks under **our recommended system parameters**, while the same hyperparameter search strategy can be readily applied to other tasks.

Outline

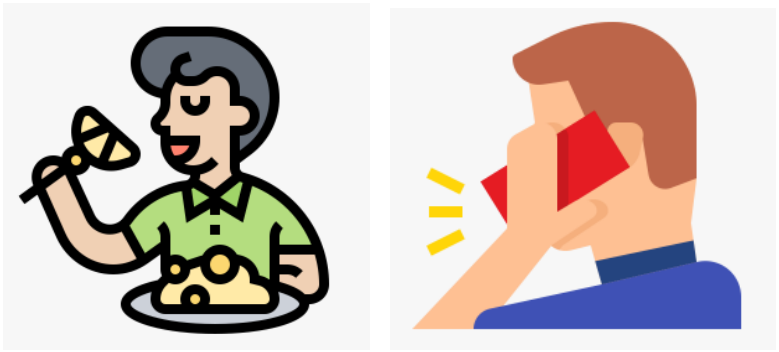
- Introduction
- Related Work
- Modality-Aware Federated Semi-Supervised Learning (MAFS)
- Multimodal Applications
- Experiment Setup
- Evaluations
- Conclusion & Future Works

Conclusion

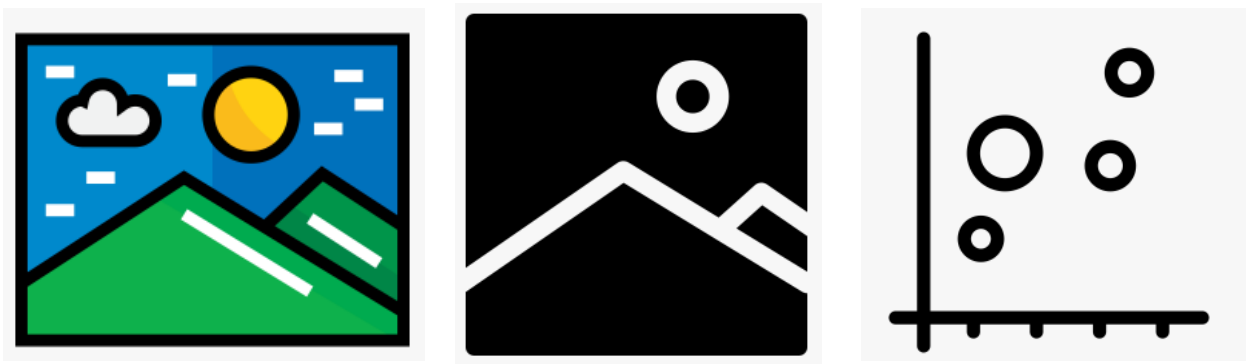
- MAFS paradigm collects **unlabeled insensitive data** from clients and uses **SSL pseudo-labeling** to generate usable data for server training.
- MAFS paradigm comes with a **modularized design** on FL clients and servers, allowing developers to readily augment FL neural network structures into MAFS-ied version.
- MAFS paradigm has been **applied to two sample classification problems** on Emotion Recognition (ER) and Human Activity Recognition (HAR) to demonstrate its practicality and efficiency

Future Works

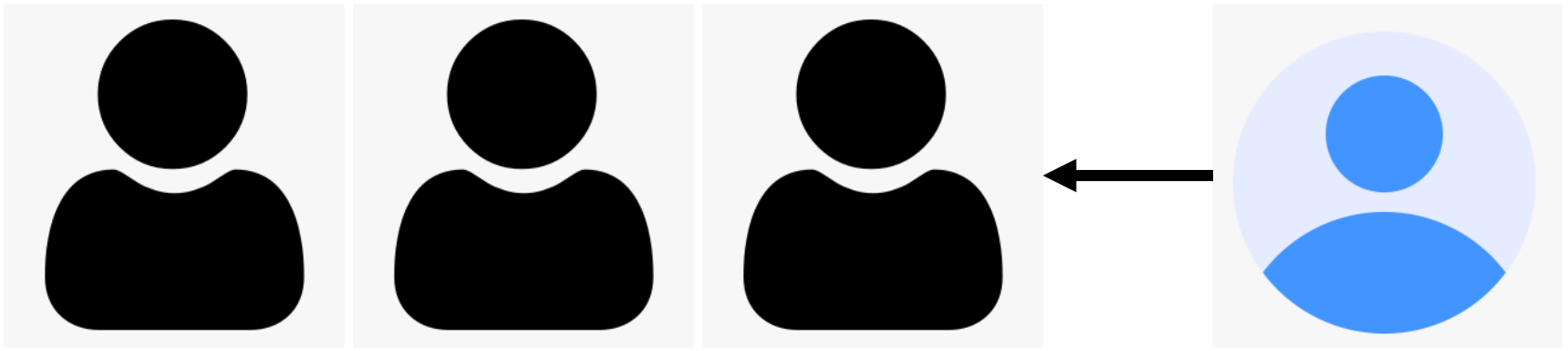
1



3



2



Thanks for listening

Special thanks for:
Chih-Fan Hsu, Inventec,
Chung-Chi Tsai, Qualcomm Technologies, Inc., USA,
Jian-Kai Wang, Qualcomm Technologies, Inc., Taiwan,
and all labmates.

Publications:

[1] G. Li, H. Chiang, **Y. Li**, S. Shirmohammadi, and C. Hsu, "A Driver Activity Dataset with Multiple RGB-D Cameras and mmWave Radars", in Proceeding of the 15th ACM Multimedia Systems Conference, 2024.
[2] C. Hsu, **Y. Li**, C. Tsai, J. Wang, and C. Hsu, "Federated Learning Using Multi-Modal Sensors with Heterogeneous Privacy Sensitivity Levels", ACM Transactions on Multimedia Computing, Communications, and Applications, 2024, Accepted.
[3] **Y. Li** , C. Hsu, C. Tsai, J. Wang, and C. Hsu, "MAFS: Modality-Aware Federated Semi-Supervised Learning with Selective Data Sharing Specified by Individual Clients", ACM Multimedia Asia, 2024, Under review.

Questions or comments?