MocoSFL: Enabling Cross-Client Collaborative Self-Supervised Learning

Rebecca Salganik

July 2024

UNIVERSITY of ROCHESTER

Agenda

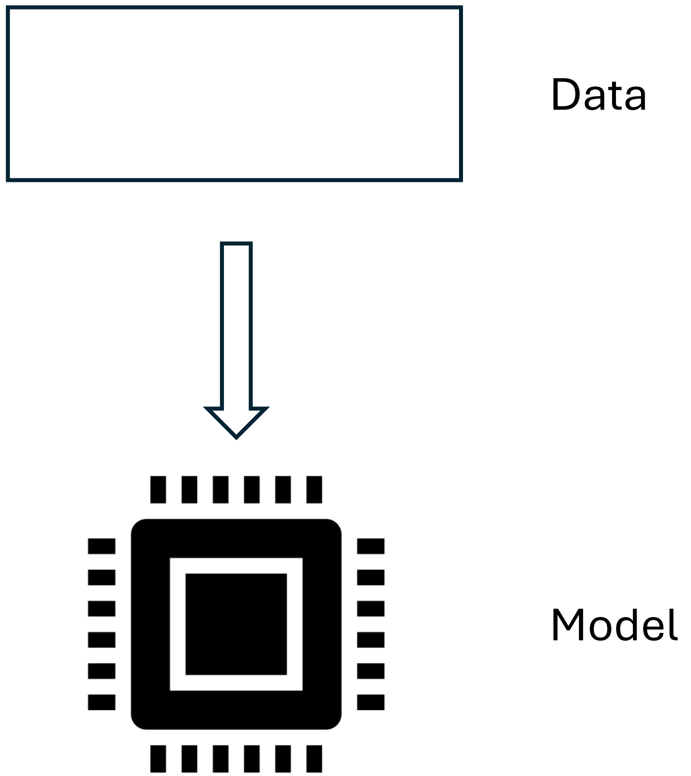
- Motivation
- Background
  - Federated Learning (FL)
  - Self Supervised Learning (SSL)
- Problem Definition
- Methodology
- Experimental Findings
- Conclusion

Motivation

- Users want to train personal, local models on their own data without sharing this data with other users → Federated Learning (FL)
- Some scenarios have scarce or non-existent labels → Self-Supervised Learning (SSL)
- Combining FL with SSL is the focus of this work

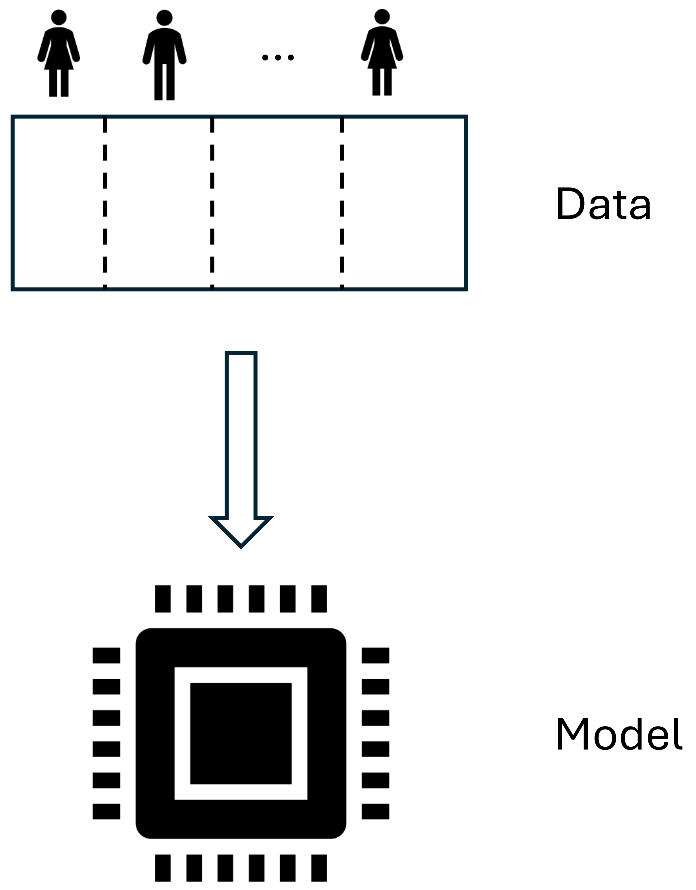
Federated Learning

- In the classic ML setting, a single dataset is fed to a single model



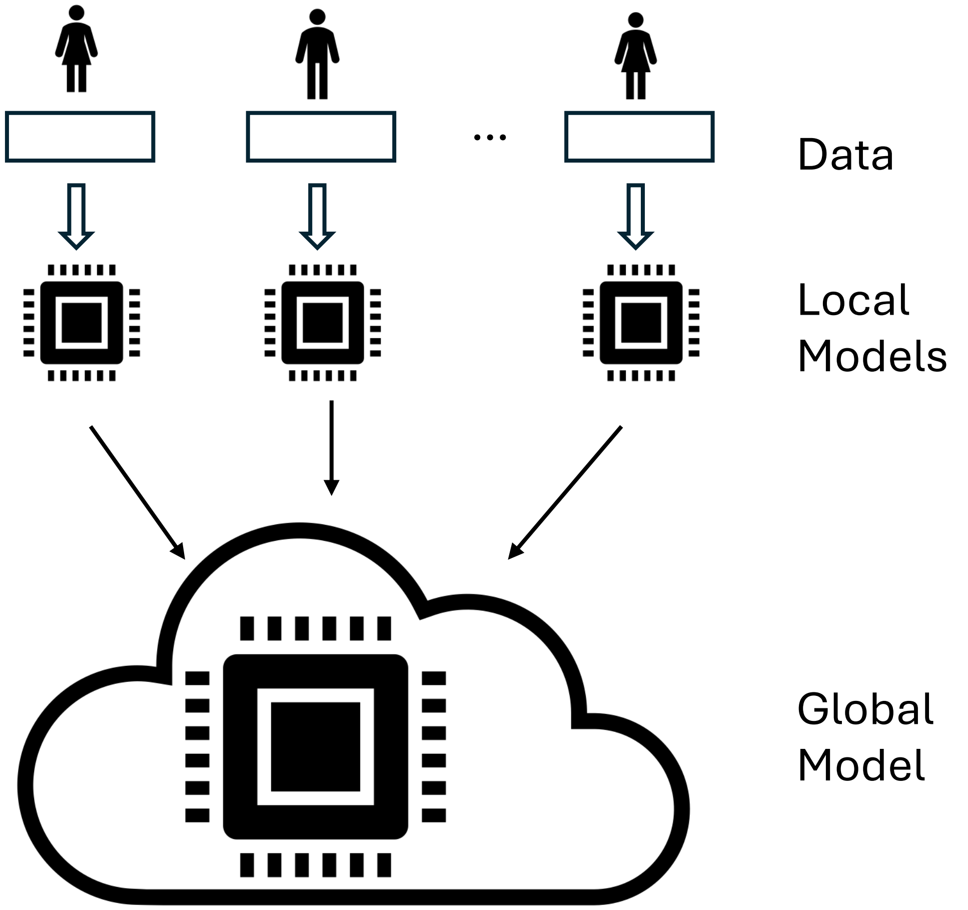
Federated Learning

- In the classic setting, a single dataset is fed to a single model
- However, this requires all the data to be pooled in a single collection
- This is not compatible with private industrial settings where users don't wish to share their data



Federated Learning

- Federated Learning (FL) allows users to keep their data private
- Ex: **FedAvg**[1] **--** users train a local version of their model and updates occur by sharing weights with a global model



1 McMahan et al. (2017). Communication efficient learning of deep networks from decentralized data. PMLR

Self Supervised Learning

- In the classic setting, each data point is associated with a label

Jaguar

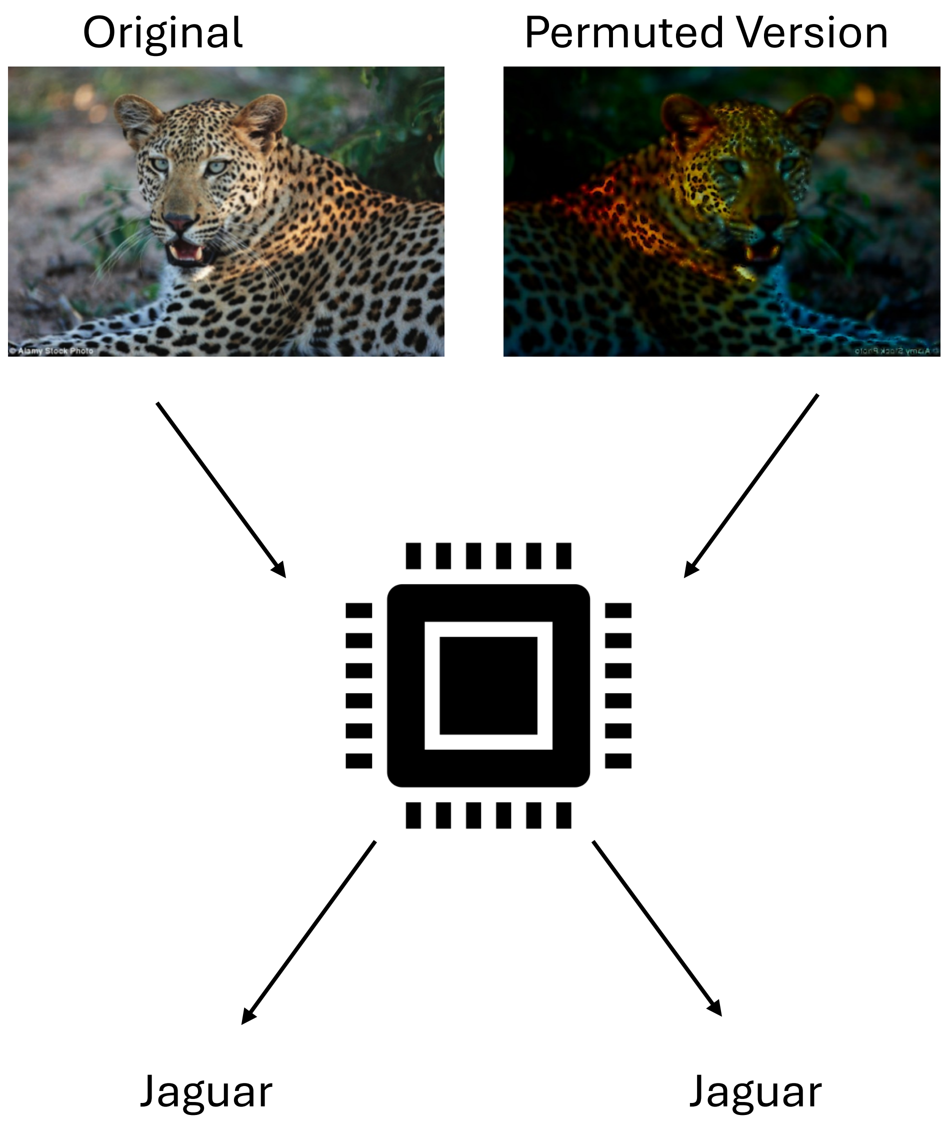
Self Supervised Learning

- In the classic setting, each data point is associated with a label
- But labeling is expensive and some scenarios have scarce or non-existent labeling
- Self supervised learning (SSL) aims to avoid reliance on labels

?

Self Supervised Learning

- The most popular form of SSL is contrastive learning
- Contrastive learning treats each individual data point as an independent class and trains a model to recognize it irrespective of permutations[1]



1 Chen et al. (2020). A simple framework for contrastive learning of visual representations. PMLR.

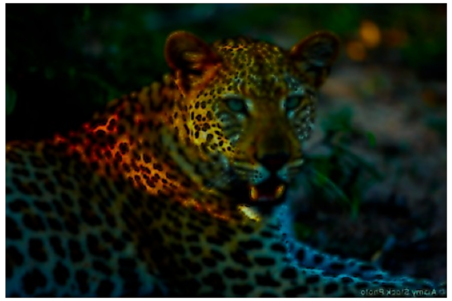
Self Supervised Learning

- The loss for contrastive learning can be formulated as:

$$\ell_{Q,K,N} = -\log \frac{exp(Q \cdot K^{+}/\tau)}{exp(Q \cdot K^{+}/\tau) + \sum_{N \in M} exp(Q \cdot N/\tau)}$$

where Q is a query key, K^{+} is a positive key, and N is a negative key

Q



K^{+}

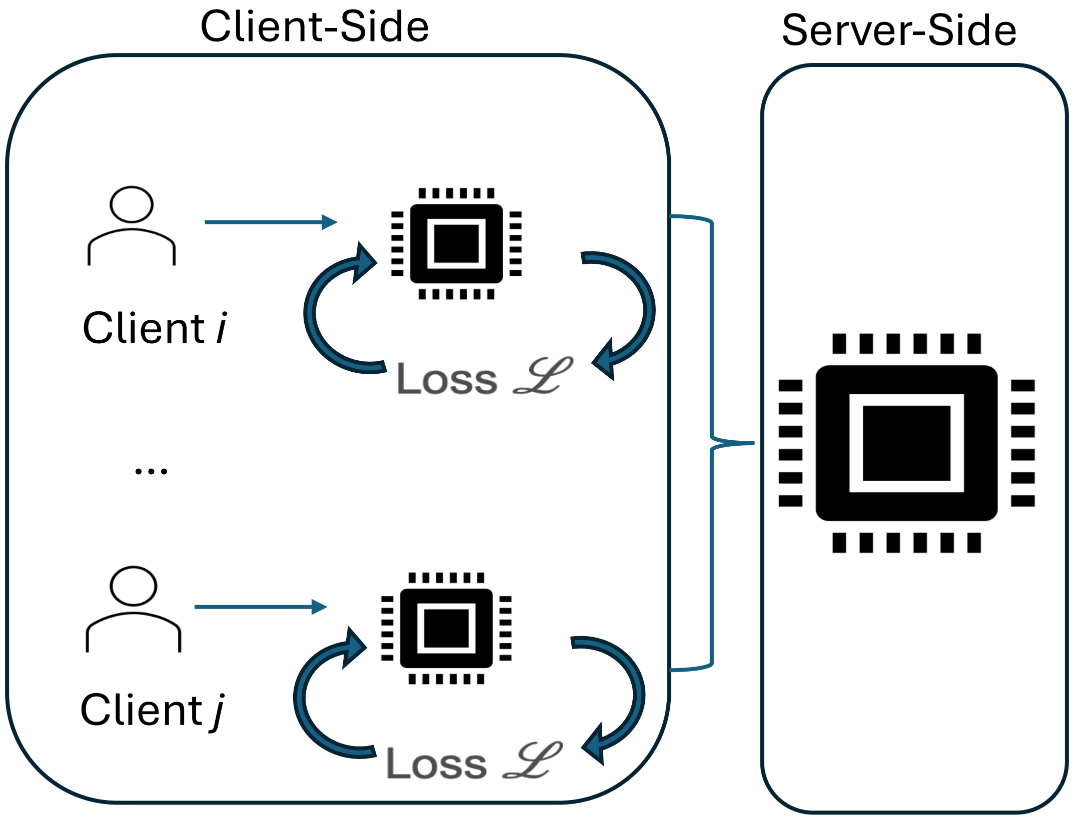
N

Problem Definition

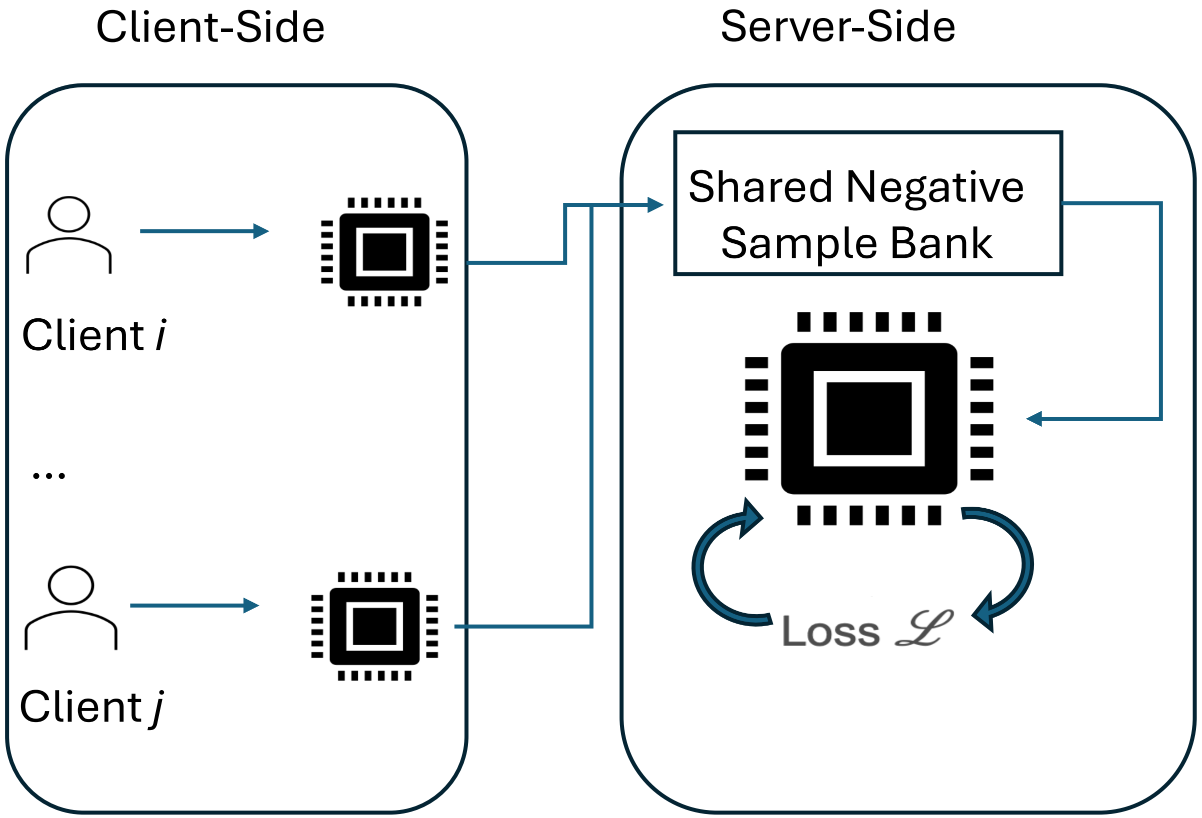
- The purpose of this work is to define a method for collaborative learning – ie combining FL and SSL for image classification.
- SOTA methods for combining FL + SSL have two key drawbacks:
  1. Large memory requirements
     1. SSL requires a models with a large number of parameters, otherwise accuracy degrades
  2. Low accuracy with large client base
     1. SSL requires a lot of data per client in order to maintain bank of “hard” negative examples

Previous Works

FL-SSL[1]



MocoSFL



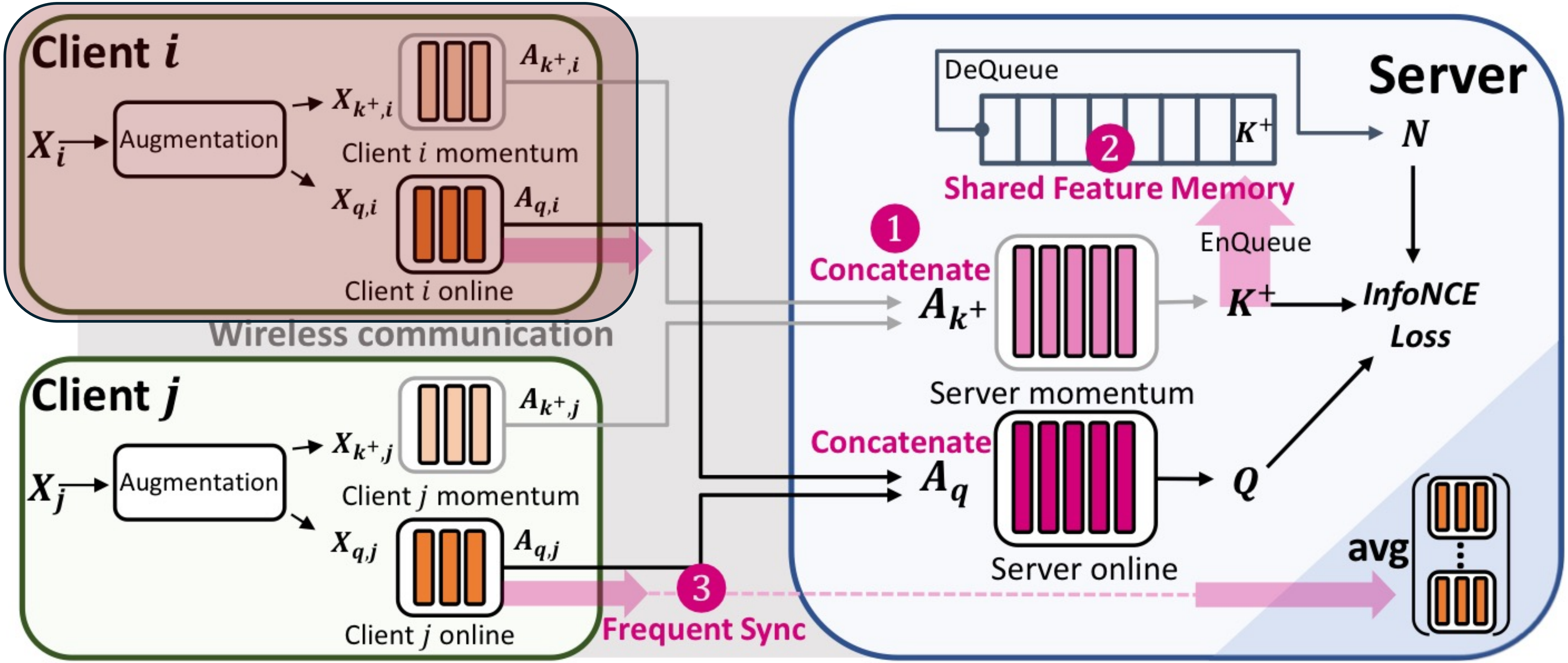
1 Zhuang et al. (2022). Divergence-aware federated self-supervised learning. ICLR.

MocoSFL – Overview

- Three key contributions:
  1. Latent vectors sent by all clients are concatenated before being processed by the server-side model
  2. Model uses a shared feature memory which is updated by positive keys contributed by all clients in each step of training
  3. Non-IID performance is improved by using higher synchronization frequency

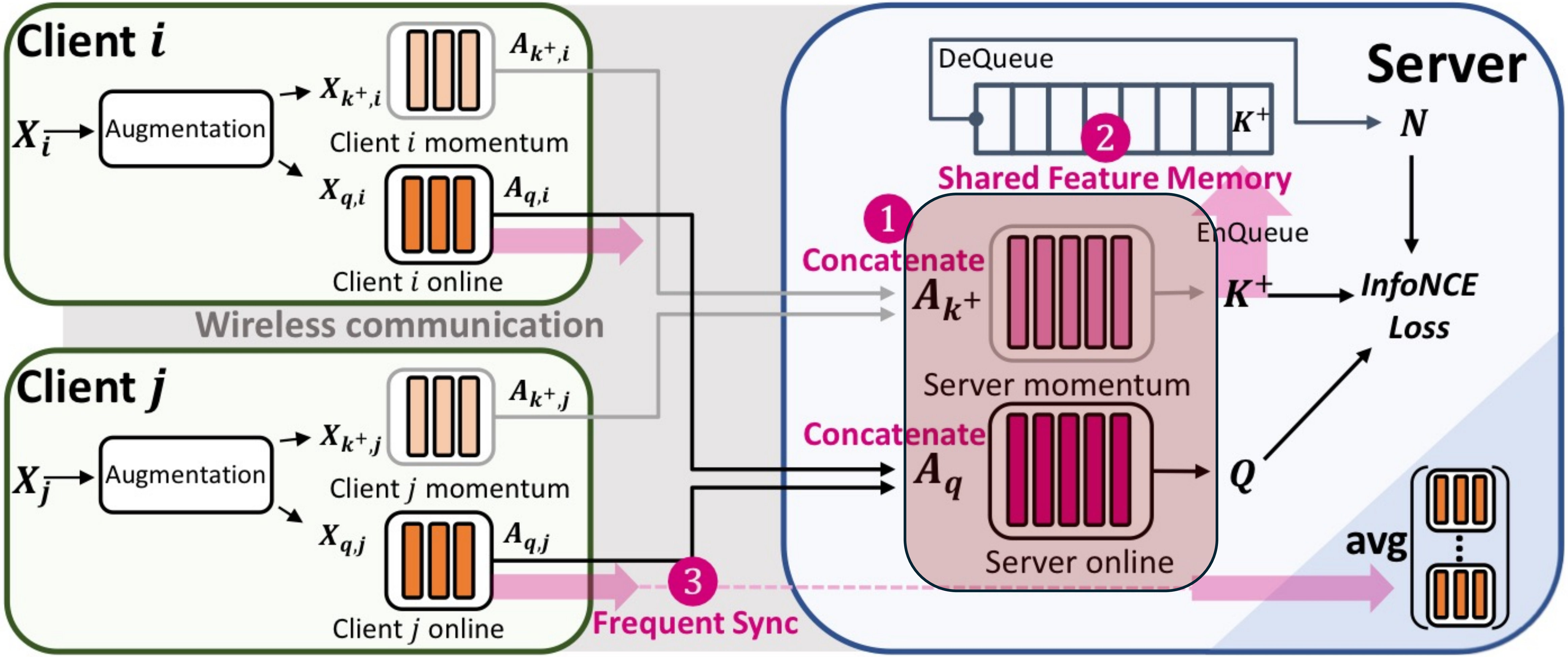
MocoSFL

- Each client has a local model
- Given input X_i, the local model performs augmentation to generate a query $X_{q,i}$ and positive key $X_{k^+,i}$



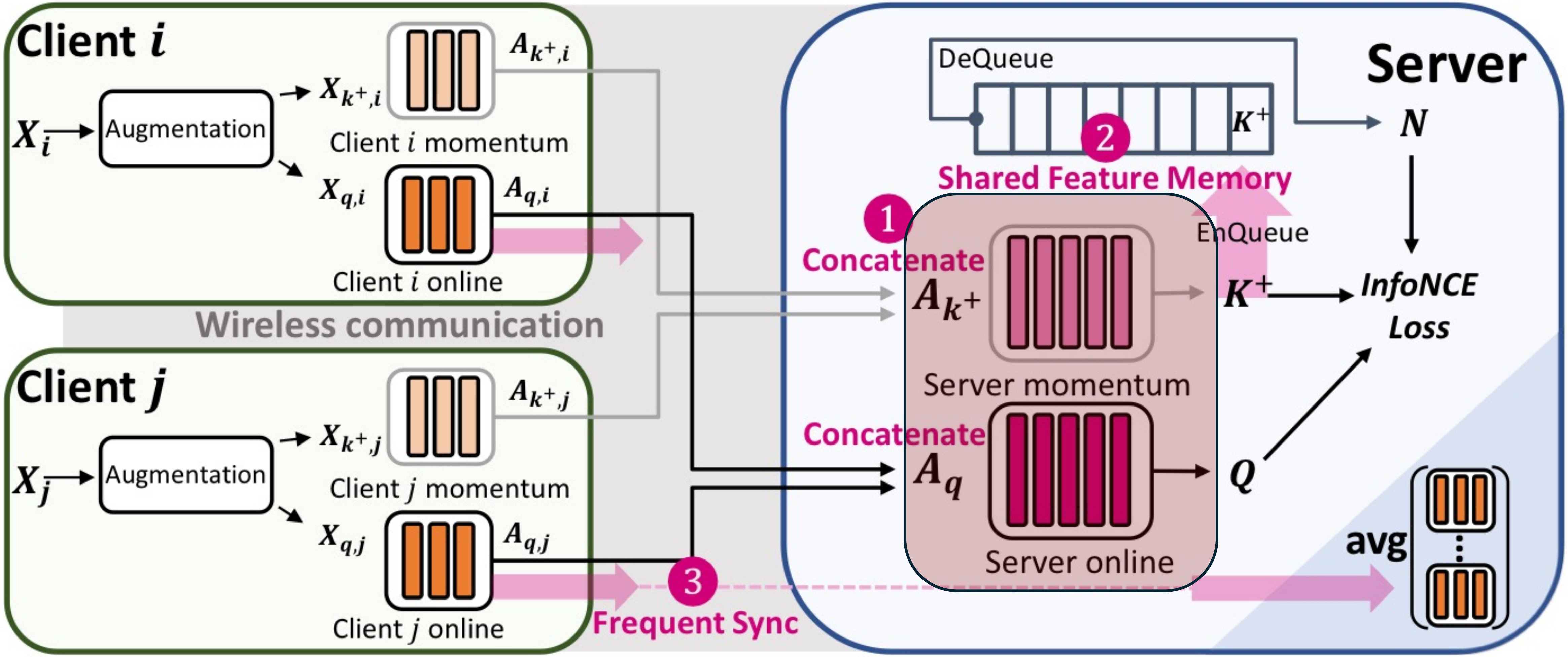
MocoSFL

- Given input X_i, the local model performs augmentation to generate a query $X_{q,i}$ and positive key $X_{k^+,i}$
- These are concatenated over all the clients into A_q and A_{k^+}



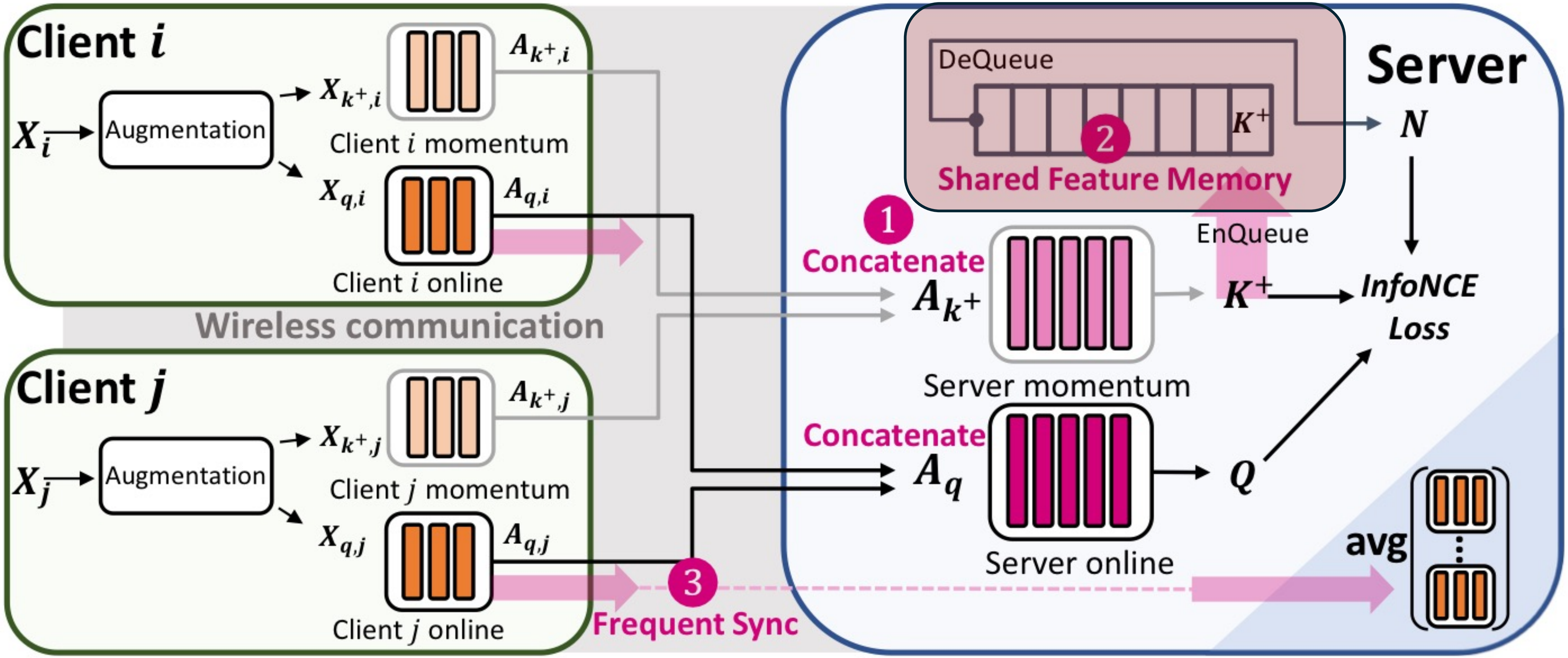
MocoSFL

- **Concatenation** reduces the hardware resource requirements on the local model
- Local models can train using **micro-batches** to reduce their memory consumption without degrading overall accuracy



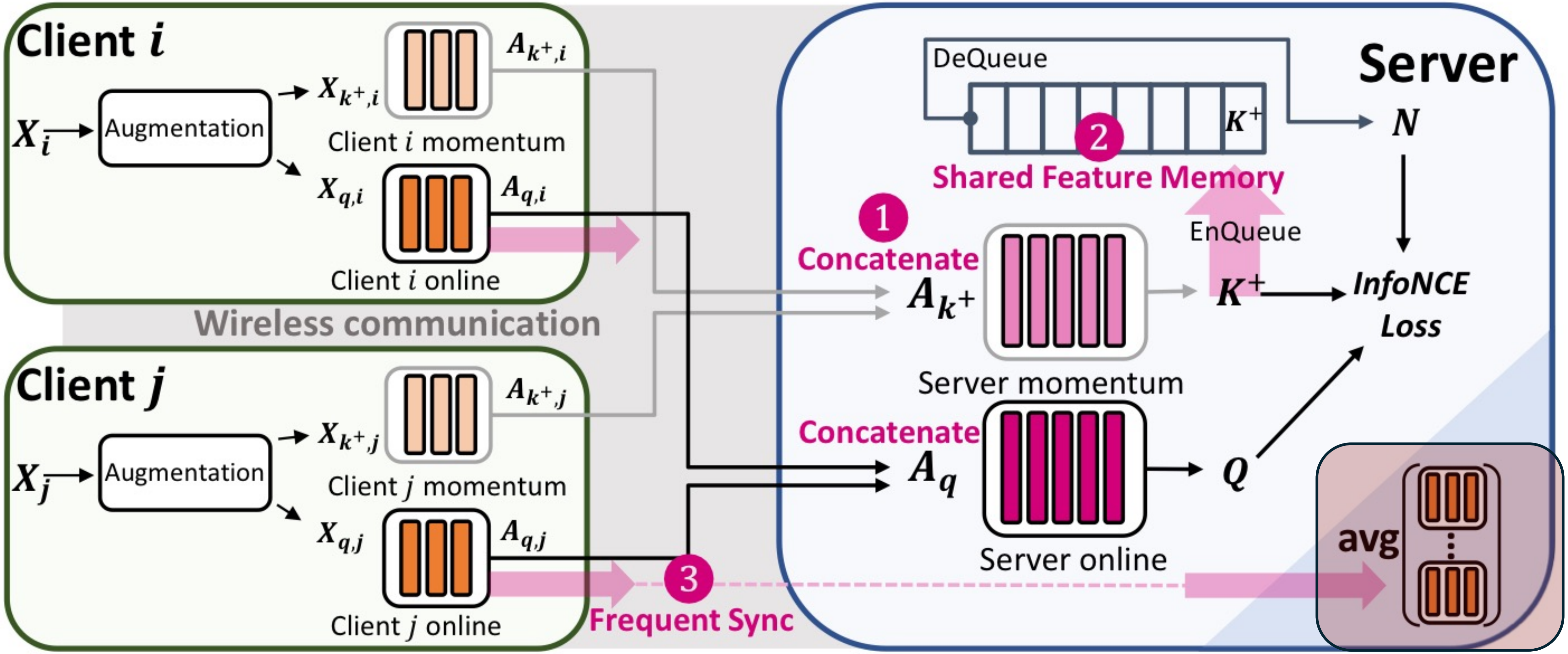
MocoSFL

- Positive keys from previous batches become negative keys for the next batch
- Keeping shared feature memory on server mitigates the large data requirements for SSL
- See Eq. 2 for a bound on the hardness of each new query



MocoSFL

- Since local models are lightweight, their weights require less overhead
- Local models can be synched after every batch, lowering divergence and improving generalization to non-IID setting

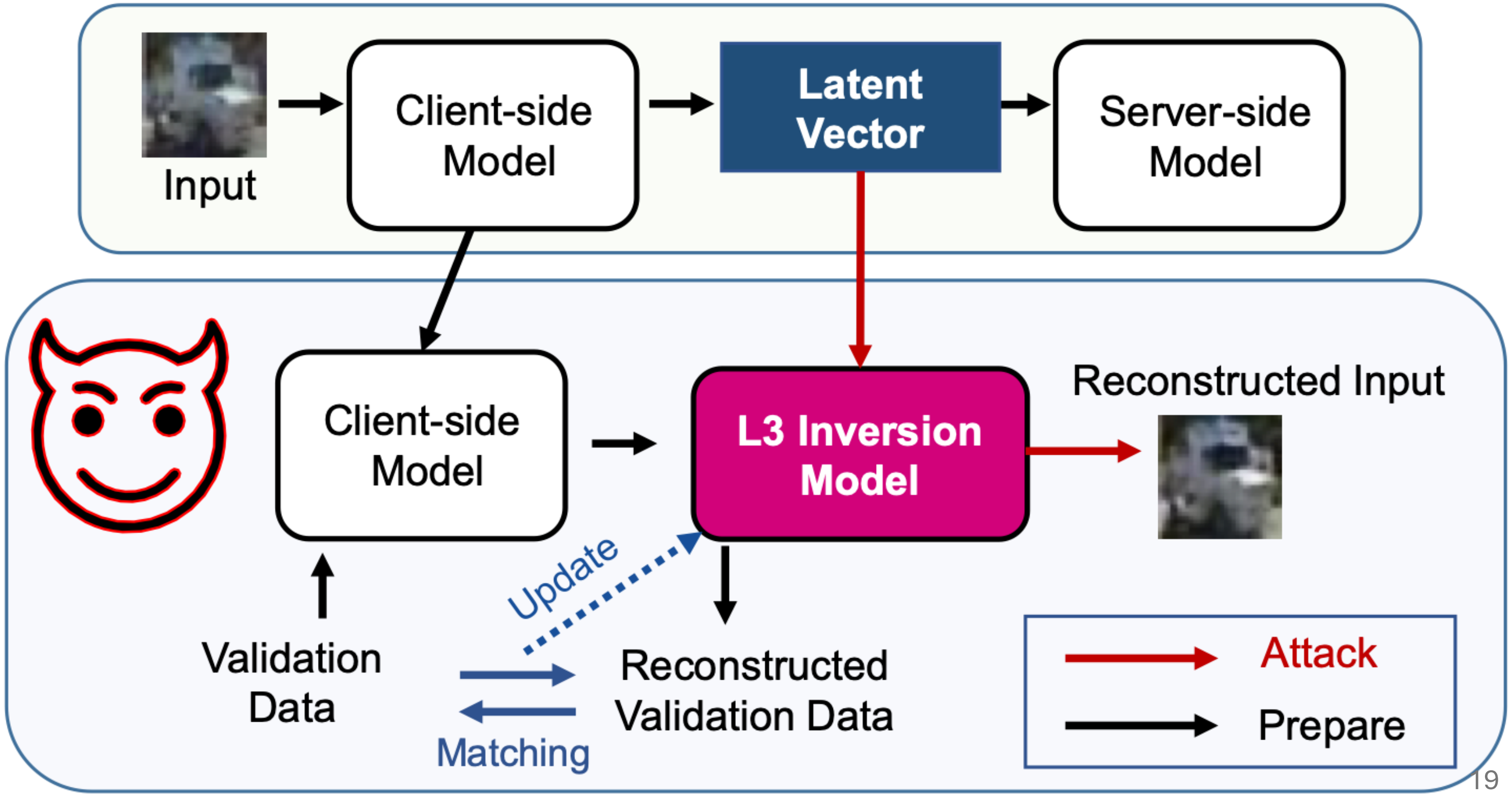


MocoSFL – Privacy Concerns

- Concatenation of latent vectors creates vulnerability to Model Inversion Attacks (MIA)

Client's View:

Server's View: (Honest-but-curious)

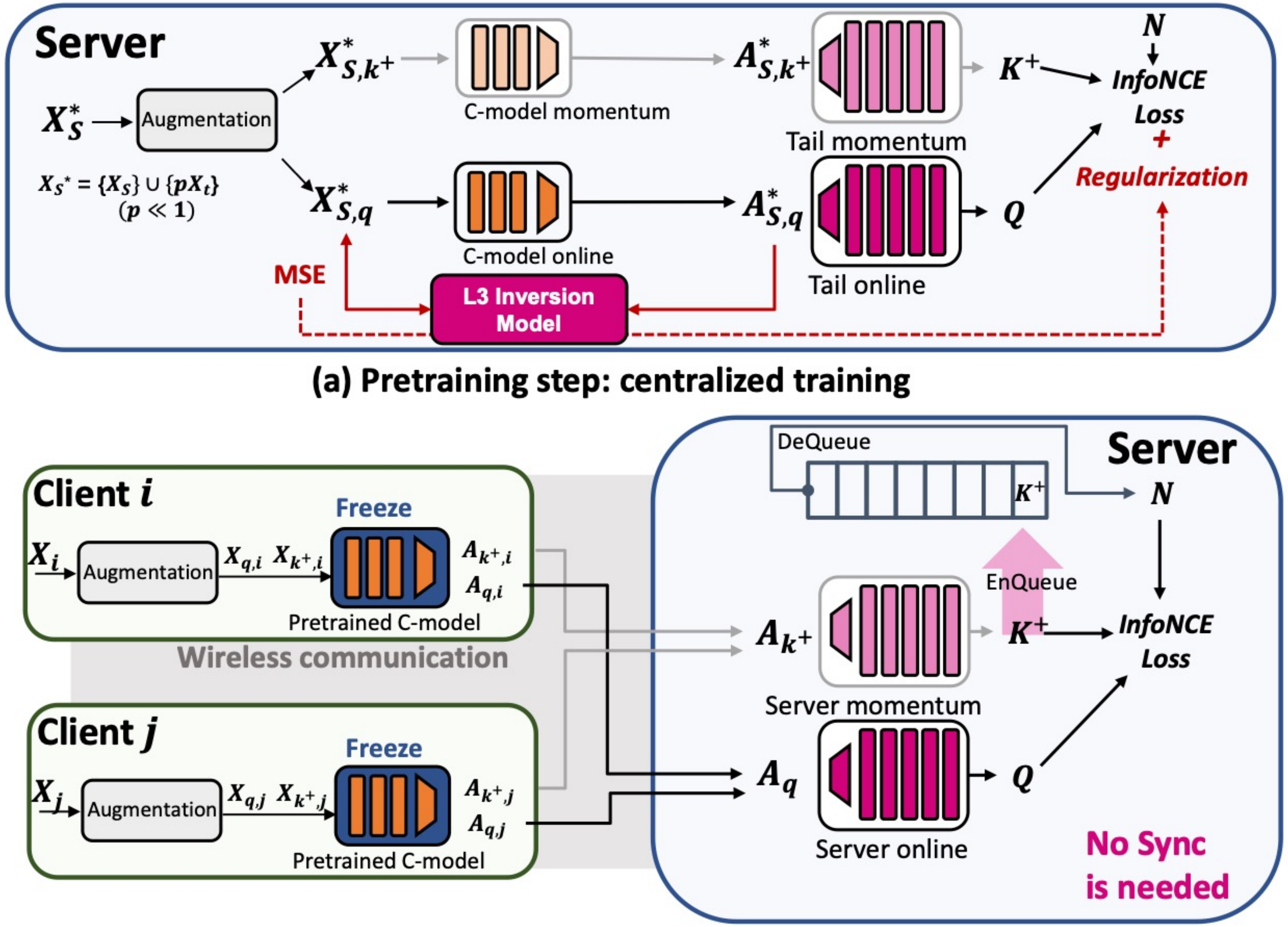


MocoSFL – Privacy Concerns

TAResSFL:

(a) Server model is pre-trained on subset of training data combined with out of domain data then transferred to clients

(b) Client side models are frozen



(a) Pretraining step: centralized training

(b) Transfer step: MocoSFL + TAResSFL

Experimental Results – Non-IID Performance

Method	CIFAR-10		CIFAR-100	
	$N_C = 5$	$N_C = 20$	$N_C = 5$	$N_C = 20$
FL-BYOL (Zhuang et al., 2022)	83.34	75.77	61.78	52.78
MocoSFL-1 (ours)	87.81	85.84	58.78	57.80
MocoSFL-3 (ours)	87.29	85.32	57.70	57.52

- In cases with high client count, MocoSFL has significant gains
- However, in cases with higher task complexity, FL-BYOL outperforms MocoSFL

Experimental Results – Client Count Scaling

Method	Dataset	IID			non-IID		
		$N_C = 100$	$N_C = 200$	$N_C = 1000$	$N_C = 100$	$N_C = 200$	$N_C = 1000$
MocoSFL-1	CIFAR-10	87.29	87.38	87.51	87.71	87.39	86.46
	CIFAR-100	58.91	59.15	58.85	59.22	58.90	56.75
	ImageNet-12	92.02	91.73	91.76	92.24	91.44	91.28
MocoSFL-3	CIFAR-10	87.29	87.15	87.25	87.10	85.22	84.75
	CIFAR-100	58.41	58.30	58.80	58.69	58.59	56.88
	ImageNet-12	92.08	92.24	92.02	92.60	91.83	91.28

- As the number of clients scales up, the performance remains relatively stable

Experimental Results – Privacy Evaluation

Method	Metric	Target Data		
		0.0%	**0.5%**	**1.0%**
MocoSFL-1	Accuracy (%) Attack MSE	81.14±0.47 0.039±0.005	80.78±1.34 0.033±0.014	79.96±2.96 0.039±0.002
MocoSFL-3	Accuracy (%) Attack MSE	81.19±2.32 0.045±0.003	80.51±1.49 0.035±0.003	**83.13±2.40** **0.039±0.002**

- Applying TAResSFL achieves good accuracy while having high MIA resistance
  - With TAResSFL the reconstructed pictures are harder to identify and more blurry
- Trade-off between layer cutoff for attack resistance and accuracy

Conclusion

- Goal:
  - Design a federated learning system that can be used for self-supervised learning on computer vision tasks
- Contributions:
  - MocoSFL, a novel FL-SSL model that uses a small client-side model, latent vector concatenation, and feature sharing
    - Addresses two major challenges in achieving high accuracy in FL-SSL schemes for cross client applications: (1) Large data requirement, (2) large hardware requirements
    - Addresses communication overhead and privacy issues inherent to SFL-based schemes