CSE 847 - Federated Semi-Supervised Learning in Image Classification

Bao Hoang and Manh Tran

1. Introduction

- **Fully labeled datasets are often limited** in real-world scenarios.

  => Semi-supervised learning (SSL) can utilize unlabeled data to improve model performance.

- **Privacy constraints** prevent data sharing between clients to central server.

  => Federated learning algorithm is needed for model training without sharing data, thus preserving data privacy.

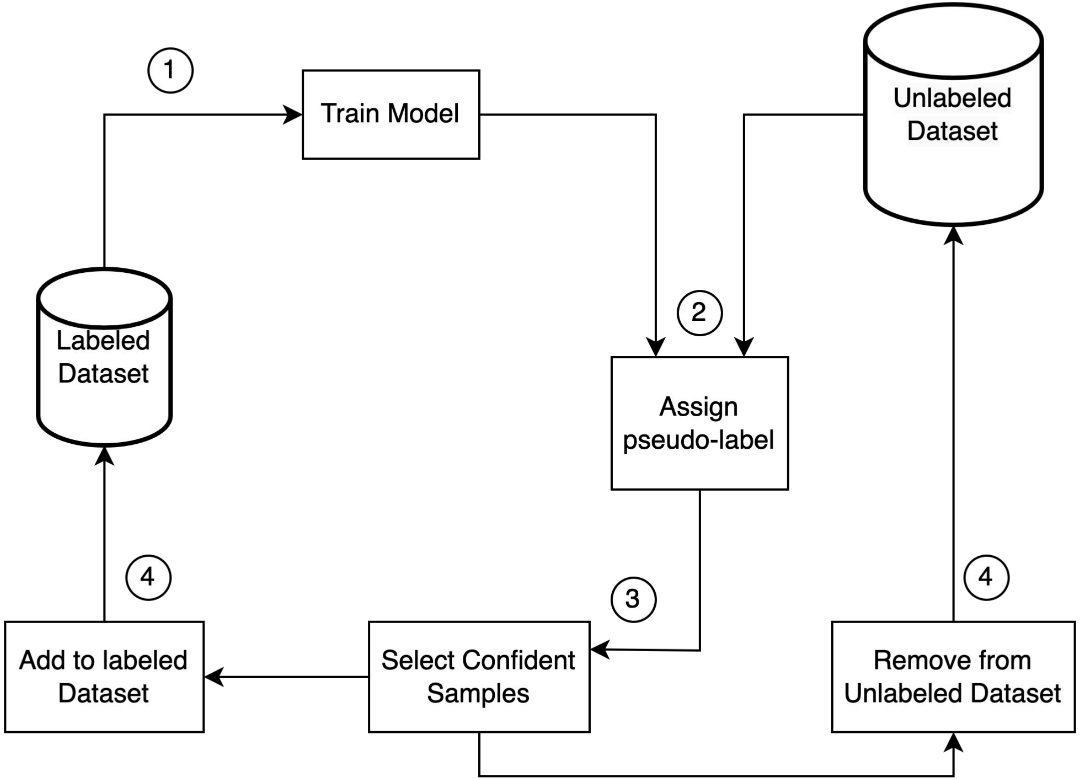
- This project focuses on investigate semi-supervised and also adapting them to the federated learning setting.

2. Outline

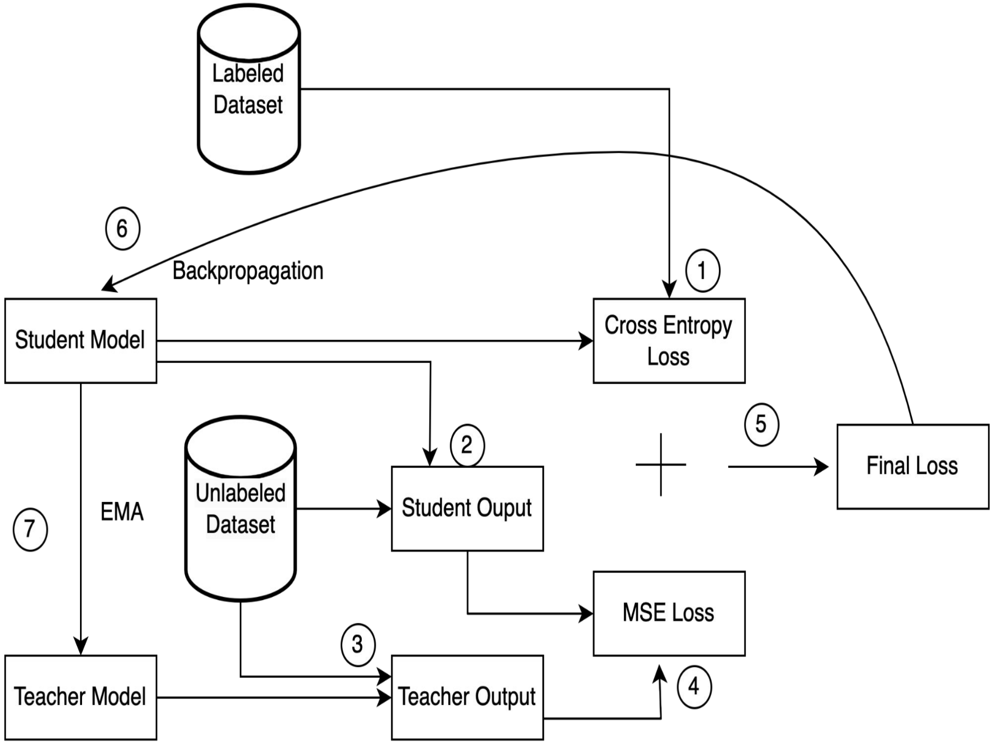
- In this project, we investigate 4 semi-supervised methods:
  1. Self-training
  2. Mean Teachers
  3. FixMatch
  4. MixMatch
- We also implement Federated Average as federated algorithm.
- We evaluate performance using 3 computer vision architectures on 3 datasets.

3. Self-training

- In each iteration, a supervised model is trained on the labeled data (1), then it is used to generate pseudo-labels for the unlabeled data (2).
- The most confident pseudo-labeled samples are added to the labeled dataset (3, 4), which is then used for training in the subsequent iteration.



4. Mean-Teachers

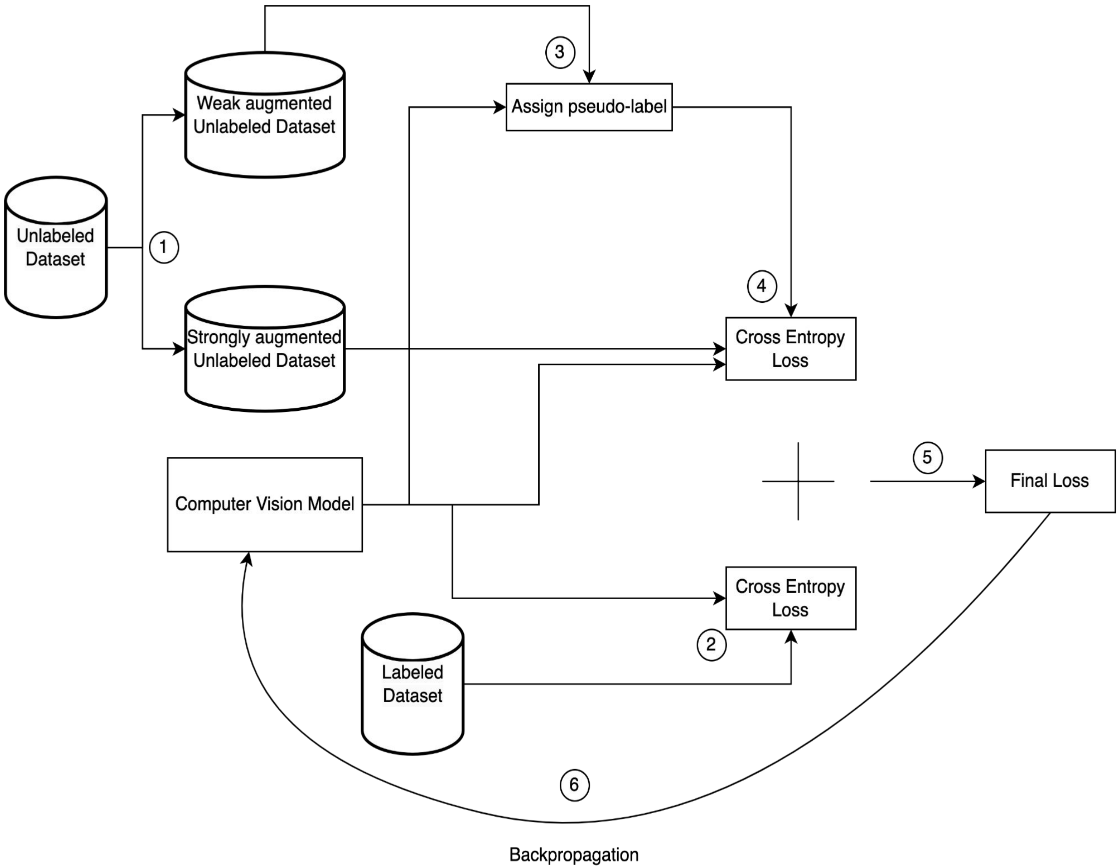


- The Mean Teacher algorithm train a student and teacher model.
- The student model minimizes a classification loss on labeled data (1) and a consistency loss aligning its outputs with the teacher model's on unlabeled data (2, 3, 4).
- The teacher model's weights are updated via an Exponential Moving Average (EMA) of the student's (7):

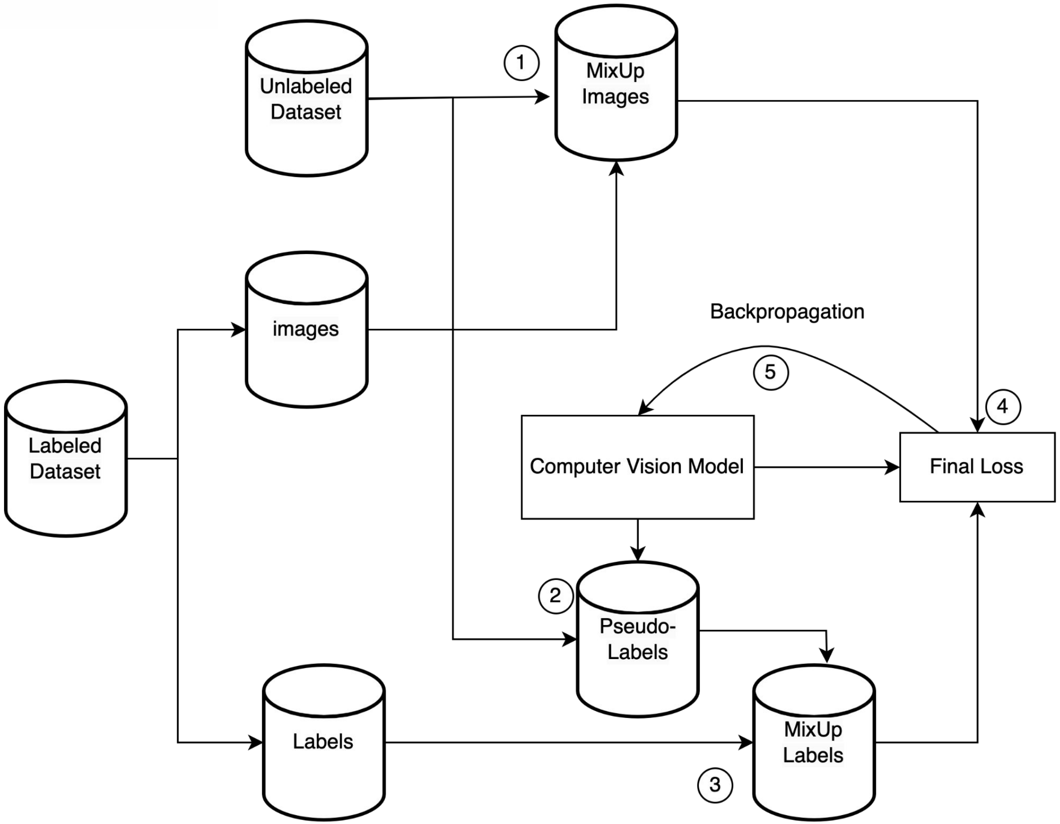
$$\theta_t \leftarrow \alpha\theta_t + (1-\alpha)\theta_s$$

5. FixMatch

- FixMatch uses two types of augmentations: "weak" and "strong," for unlabeled data (1).
- FixMatch generates pseudo-labels from "weak" augmentations (3), then calculates the cross-entropy loss with "strong" augmentations (4).
- It also combines this with the cross-entropy loss of the labeled dataset (2) to compute the final loss (5).



6. MixMatch



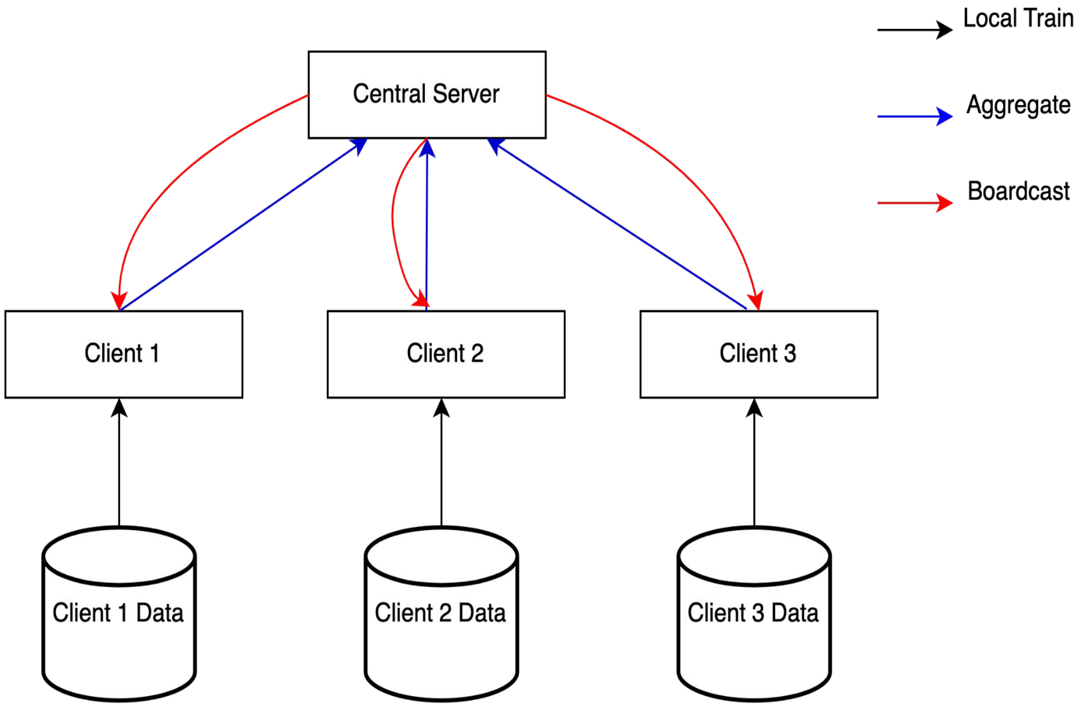
- MixMatch utilizes the MixUp formula to generate MixUp images (1) and MixUp labels (3).

$$\hat{x} = \lambda x_i + (1 - \lambda) x_j$$

$$\hat{y} = \lambda y_i + (1 - \lambda) y_j$$

- MixUp labels are generated using pseudo-labels (2) of unlabeled images and the labels of labeled images.
- Finally, the loss is computed using the MixUp images and labels (4).

7. Federated Average

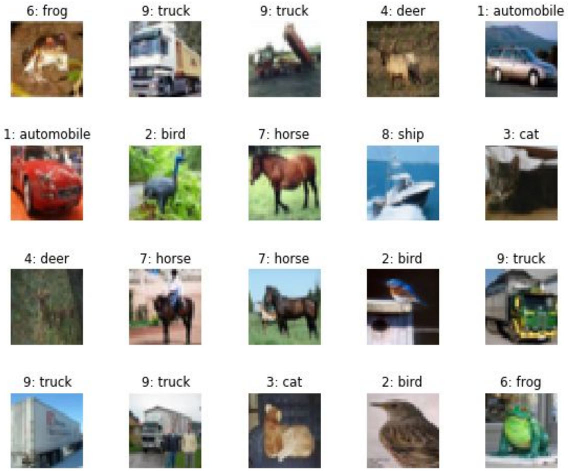


FedAvg has three main steps:

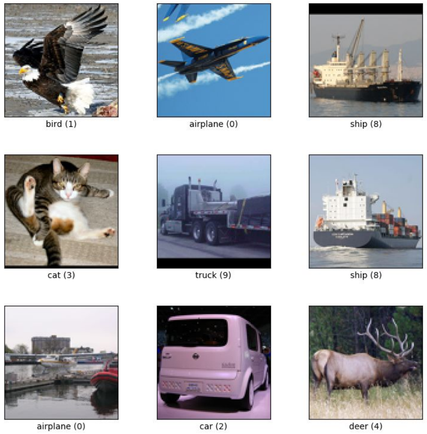
1. Each client trains locally using their own data.
2. Then, the clients send their model parameters to the central server, which aggregates the parameters .
3. Finally, the server broadcasts the aggregated model back to the clients, and the process continues from step 1.

8. Experimentals Setting

- For the CIFAR-10 and Cat-Dog datasets, we split the training dataset into 10% labeled and 90% unlabeled data.
- For the STL-10 dataset, the labeled and unlabeled sets are already predefined.



CIFAR-10



STL-10

Cat Dog

9. Centralized Setting Results

Table 1. Accuracy of Semi-Supervised Learning Methods on Different Datasets With Different Model Architectures in Centralized Setting

Methodology	CIFAR-10	STL-10	Cat And Dogs
Simple CNN			
Baseline	67.94%	68.14%	70.14%
Golden Baseline	82.01%	–	81.76%
Self-training	69.14%	69.38%	72.61%
Mean teacher	68.71%	70.59%	72.27%
FixMatch	70.70%	69.71%	72.81%
MixMatch	**72.35%**	**71.75%**	**74.34%**
ResNet-18			
Baseline	68.53%	75.55%	71.38%
Golden Baseline	85.84%	–	88.83%
Self-training	70.45%	77.61%	72.52%
Mean teacher	70.64%	74.38%	67.08%
FixMatch	75.12%	80.20%	76.91%
MixMatch	**79.33%**	**80.86%**	**81.30%**
DenseNet-121			
Baseline	67.00%	75.70%	73.46%
Golden Baseline	87.56%	–	88.38%
Self-training	70.09%	77.05%	74.35%
Mean teacher	70.23%	75.52%	70.79%
FixMatch	75.95%	77.37%	73.15%
MixMatch	**79.92%**	**80.32%**	**81.90%**

- Semi-supervised methods **outperform standard baseline**, showing the importance of unlabeled data.
- FixMatch and MixMatch outperform other semi-supervised methods due to **strong data augmentation and consistency regularization.**
- Semi-supervised methods **are weaker than the golden baseline**, emphasizing the need for high-quality labeled data.

10. Decentralized Setting Results

- Performance across all methods, models, and datasets **in decentralized setting** lower than **centralized setting**, highlighting the challenge of training model without centralized data.

- MixMatch and Self-training **maintain effectiveness in decentralized settings**, while FixMatch and Mean Teacher decrease significantly.

Table 2. Accuracy of Semi-Supervised Learning Methods on Different Datasets With Different Model Architectures In Decentralized Setting

Methodology	CIFAR-10	STL-10	Cat And Dogs
Simple CNN			
Baseline	61.41%	62.00%	69.85%
Golden Baseline	76.84%	–	80.82%
Self-training	68.24%	**69.42%**	**74.59%**
Mean teacher	62.23%	61.98%	68.51%
FixMatch	65.15%	63.04%	71.91%
MixMatch	**69.66%**	68.50%	72.12%
ResNet-18			
Baseline	62.80%	71.04%	68.51%
Golden Baseline	82.89%	–	85.91%
Self-training	70.51%	**76.62%**	74.84%
Mean teacher	61.23%	68.94%	59.71%
FixMatch	70.04%	75.85%	67.47%
MixMatch	**74.01%**	73.56%	**80.20%**
DenseNet-121			
Baseline	61.56%	72.09%	70.54%
Golden Baseline	84.32%	–	86.90%
Self-training	68.56%	**76.00%**	73.55%
Mean teacher	61.48%	69.73%	66.68%
FixMatch	70.70%	73.23%	67.42%
MixMatch	**75.81%**	74.85%	**80.37%**

11. Conclusion

- Semi-supervised methods are effective in settings where labeled images are scarce but there are many unlabeled images.
- Although some semi-supervised methods are advanced, they still cannot compare to fully-labeled baselines, indicating that labeled datasets continue to play a significant role.
- Decentralized settings are challenging for training effective machine learning models compared to centralized settings due to the constraints on data sharing.

Thank you for listening