Label Hungry? Not Anymore: Towards Federated Semi-supervised Learning

Deepthi Hegde, Geoffrey Li, Ishna Kaul, Roshan Ramkeesoon
University of Washington

Outline

Problem Formulation

Related Work

Methodology

Results

Conclusion

Outline

Problem Formulation

Related Work

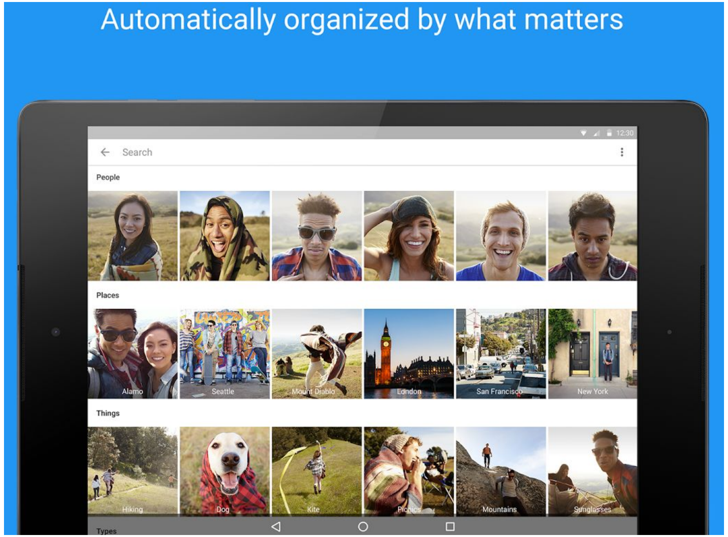
Methodology

Results

Conclusion

Challenge for FL

- While users generate a wealth of rich data, **labeled data** is scarce on client devices
  - ~1.2 trillion photos taken on mobile devices in 2017 [1]
- Limits usability of FL in domains where deep learning has shown high performance, such as image classification



[1] Richter, Felix. “Infographic: Smartphones Cause Photography Boom.” *Statista Infographics*, 31 Aug. 2017, www.statista.com/chart/10913/number-of-photos-taken-worldwide/.

Problem Formulation

- Apply semi-supervised learning in federated setting to leverage unlabelled data
- EMNIST federated digit recognition dataset
- Study model performance while varying the ratio of labelled to unlabeled data

Outline

Problem Formulation

Related Work

Methodology

Results

Conclusion

Related Work

- **Federated learning** [2]:
  - Privacy-preserving ML training on a large corpus of decentralized data residing on devices like mobile phones.
- **Semi-supervised learning:**
  - Combines small amount of labelled data with large amounts of unlabelled data during training [3].
- **Representation Learning:**
  - Supervision is derived from the data itself using auxiliary tasks
    1. Autoencoder network
    2. Image rotations [4]

[2] McMahan, et al. "Communication-efficient learning of deep networks from decentralized data." arXiv preprint arXiv:1602.05629 (2016).
[3] Chapelle, et al. *Semi-supervised learning*. Cambridge, Mass.: MIT Press. ISBN 978-0-262-03358-9
[4] Gidaris, et al. "Unsupervised representation learning by predicting image rotations." arXiv preprint arXiv:1803.07728 (2018).

Outline

Problem Formulation

Related Work

Methodology

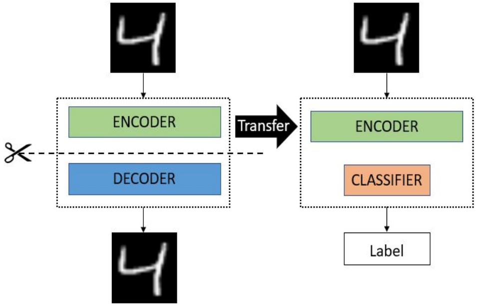
Results

Conclusion

Methodology

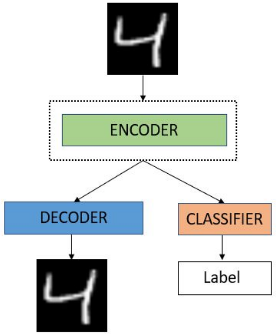
Self-supervised pre-training (SSPT)

- Train autoencoder in federated computation
- Transfer encoder layer weights to classifier
- Train classifier in second federated computation



Semi-supervised joint optimisation (SSJO)

- Single federated computation
- All examples generate unsupervised loss
- Labeled examples generate additional supervised loss



Experimental Setup

- **Mask ratio**: the fraction of data for which labels are removed
  - Varied over {0.0, 0.8, 0.9, 0.95, 0.98, 0.99}

Models:

- **FSL:** Fully Supervised Learning (baseline)
- **SSPT:** Self-Supervised Pre-Trained model
- **SSJO:** Semi-Supervised Jointly Optimized model

Architecture:

- **Classification head:** (Input-size: 784) → (embedding dense-layer: 256) → (Output-size: 10)
- **Reconstruction head:** (Input-size: 784) → (embedding dense-layer: 256) → (Output-size: 784)

Federated Setting

Batch Size	20
Client Learning Rate	1e-3
Client Momentum	0.99
Epochs/Client	10
Clients/Round	100
Communication Rounds	100

Central Setting

Batch Size	20
Learning Rate	1e-4
Momentum	0.99
Epochs	10

Outline

Problem Formulation

Related Work

Methodology

Results

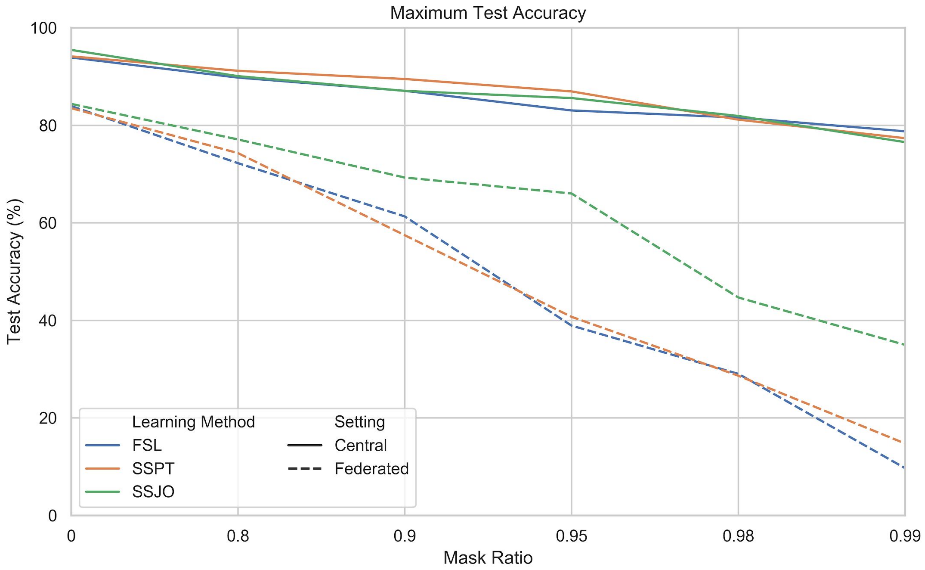
Conclusion

Results

- Central Setting:
  - similar results across three methods
  - Marginal increase in test accuracy for SSJO and SSPT methods over FSL

- Federated Setting:
  - SSJO outperforms SSPT and FSL by greater margin as mask ratios increase

- As expected, lower test accuracy for Federated vs. Central
  - Possibility for higher federated accuracy with refined hyperparameter search and additional communication rounds



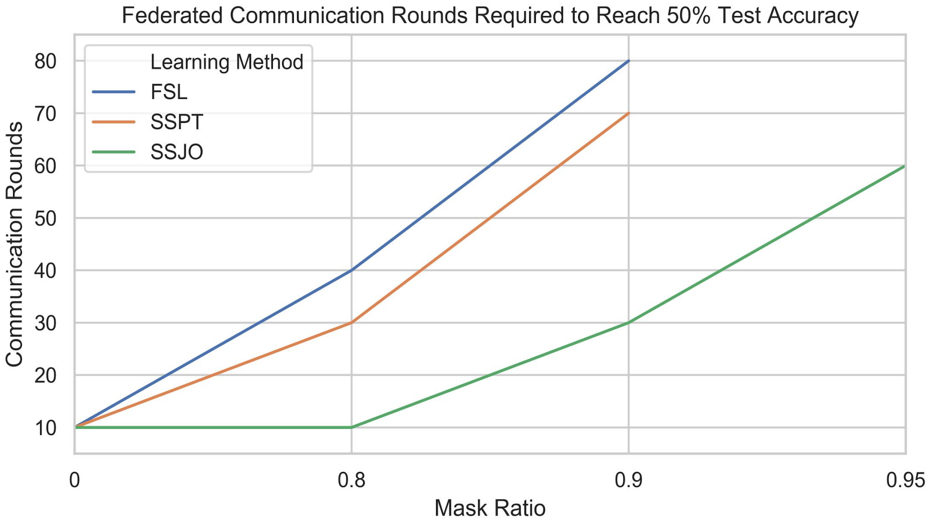
FSL: Fully Supervised Learning

SSPT: Self-supervised Pre-Training

SSJO: Semi-supervised Joint Optimization

Results

- FSL requires more communication rounds than SSPT and SSJO
- SSJO requires the least communication rounds



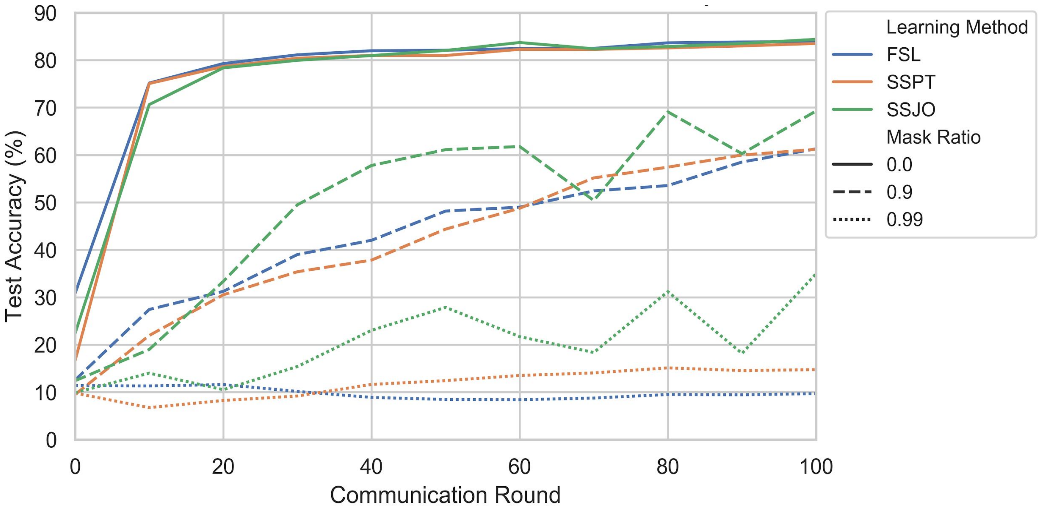
FSL: Fully Supervised Learning

SSPT: Self-supervised Pre-Training

SSJO: Semi-supervised Joint Optimization

Results

- SSJO test accuracy is unstable during training as mask ratio increases
- Model might be oscillating between local optima in the two heads



FSL: Fully Supervised Learning SSPT: Self-supervised Pre-Training SSJO: Semi-supervised Joint Optimization

Outline

Problem Formulation

Related Work

Methodology

Results

Conclusion

Conclusion

- In our experiments, SSL:
  - achieved similar (low mask-ratio) or higher (high mask-ratio) test accuracy to FSL
  - required fewer communications rounds to reach a threshold of 50% test accuracy

- Joint optimization: higher test accuracy but more unstable training

- Future work:
  - learning rate decay to stabilize SSJO training
  - the use of more specialized auxiliary tasks such as predicting rotations
  - the performance of these methods on richer datasets such as federated CIFAR 100

Acknowledgement

We would like to thank Peter Kairouz, Zachary Charles, Boqing Gong, Brendan McMahan, and other members of the Google Research team, Seattle, for fruitful discussions and feedback.

We would also like to thank Megan Hazen for her help and support throughout the project.


END OF
PRESENTATION!!

Questions?

Appendix: Rotation Model

- Convolution model on EMNIST
- SSPT only
- No major difference in performance
    - Dataset: EMNIST may not be ideal for rotation prediction task
    - Model architecture: may not be ideal for pretraining