Scaling Distributed Training of Flood-Filling Networks on HPC Infrastructure for Brain Mapping

Wushi Dong
University of Chicago

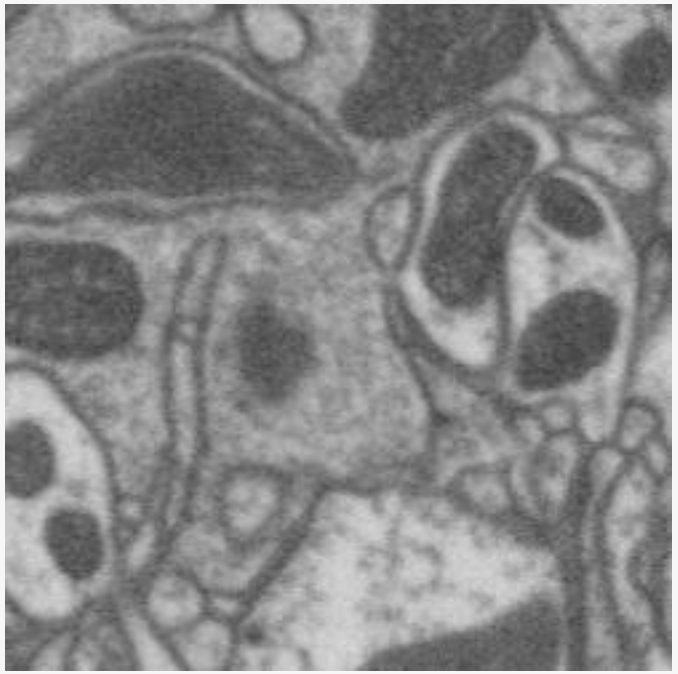
Contents

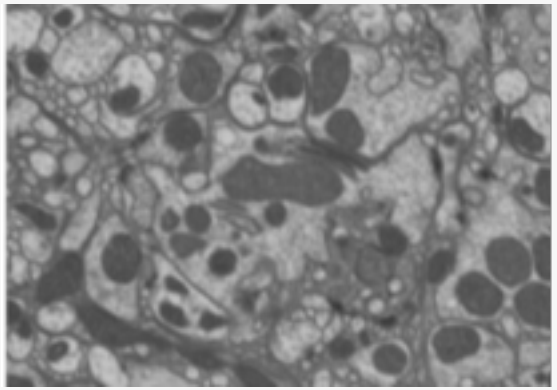
Contents

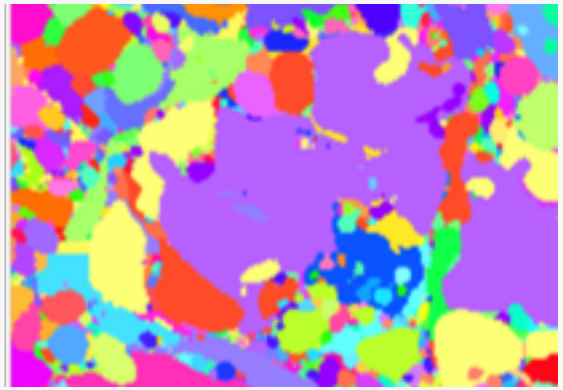
- Background
- Methods
- Results
- Summary

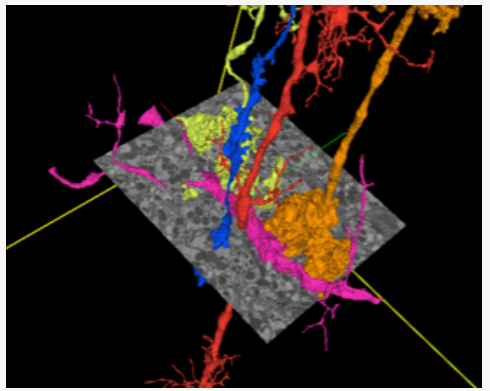
Background

- Brain mapping



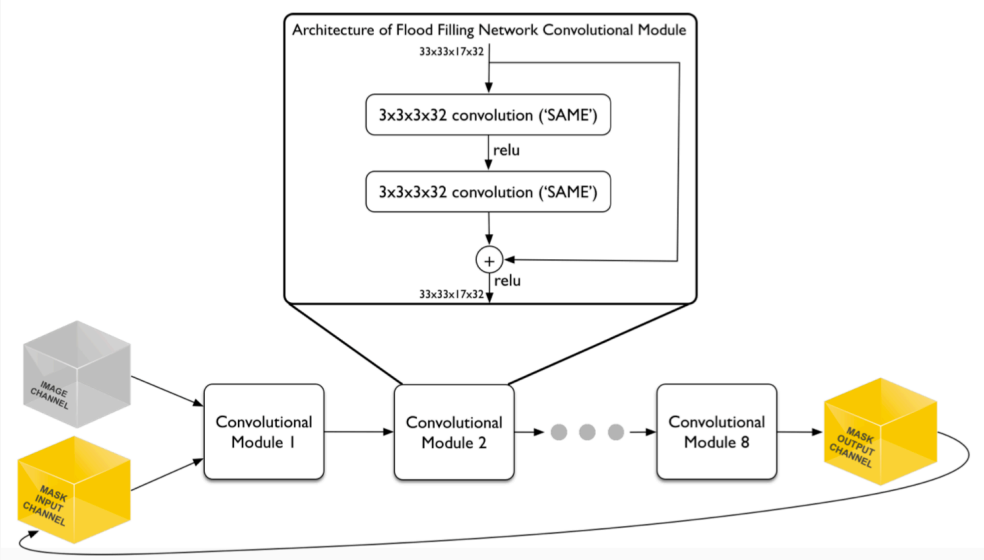


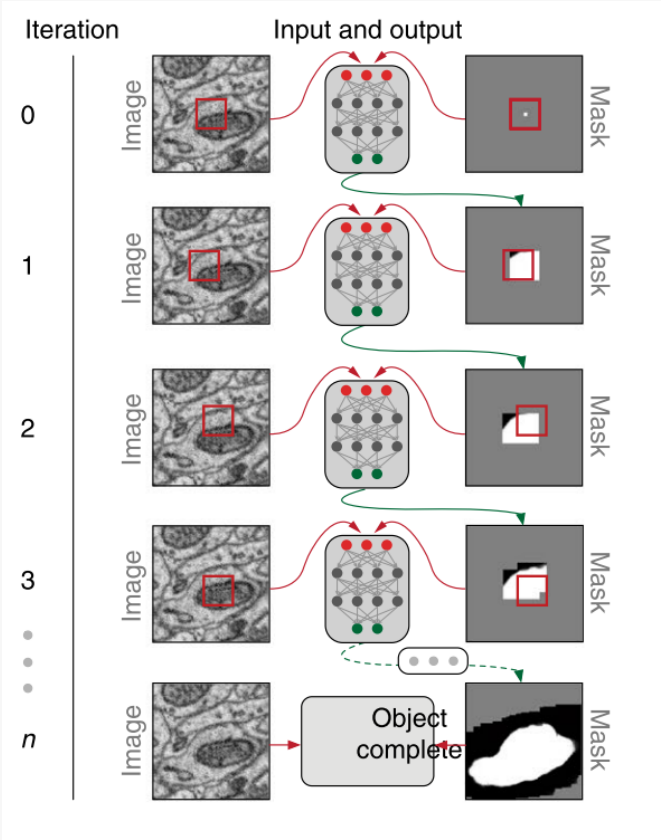


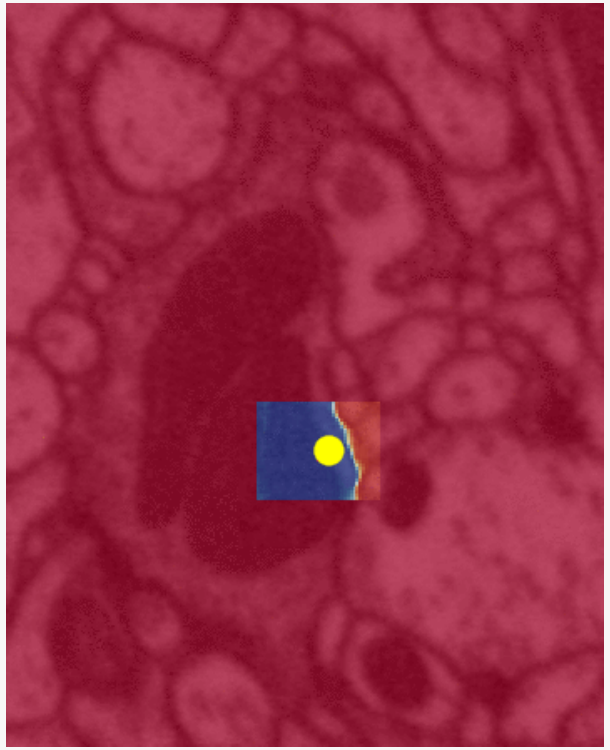


Background

- Brain mapping
- Flood-Filling Networks (FFN)



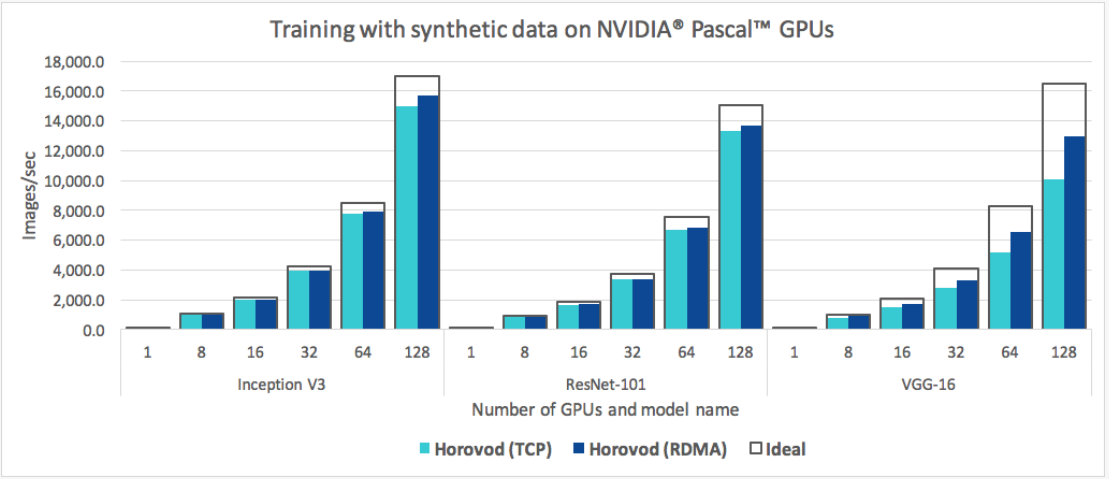


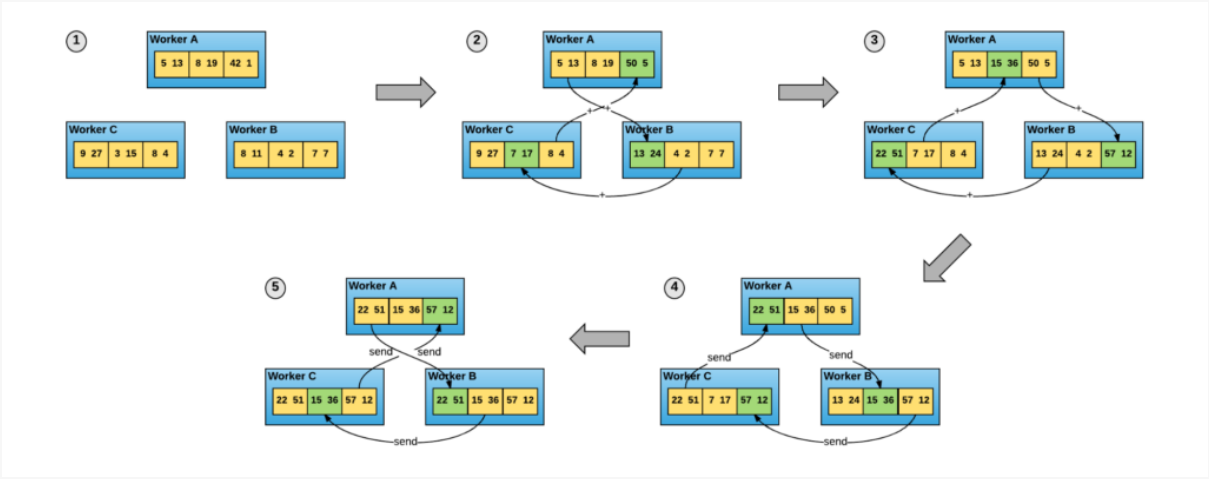


Contents

Methods

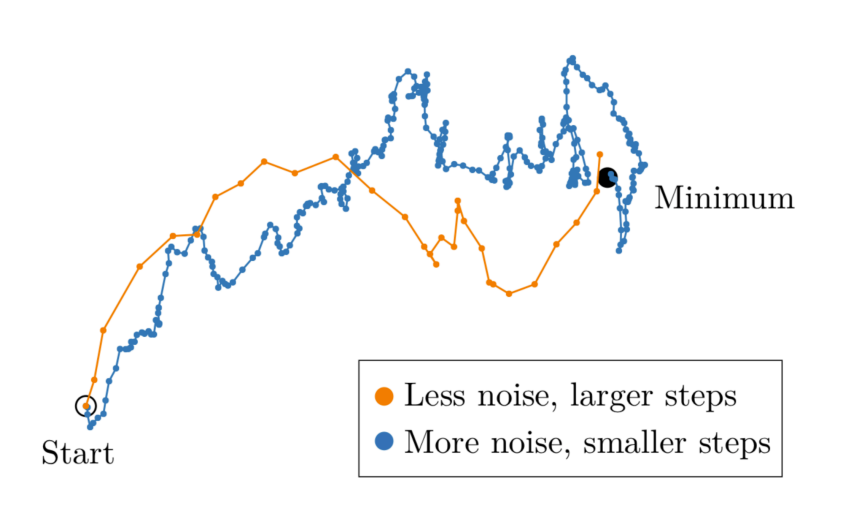
- Synchronous VS Asynchronous training

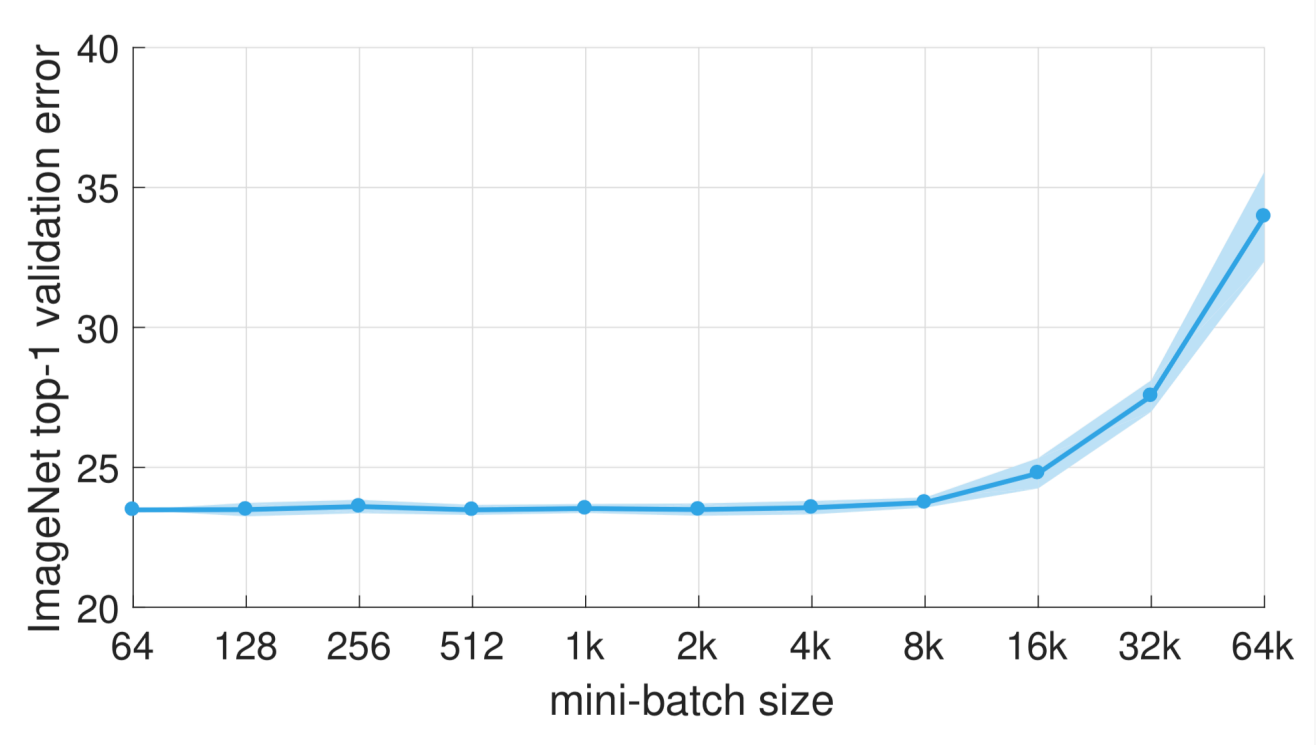




Methods

- Synchronous VS Asynchronous training
- Large-batch training



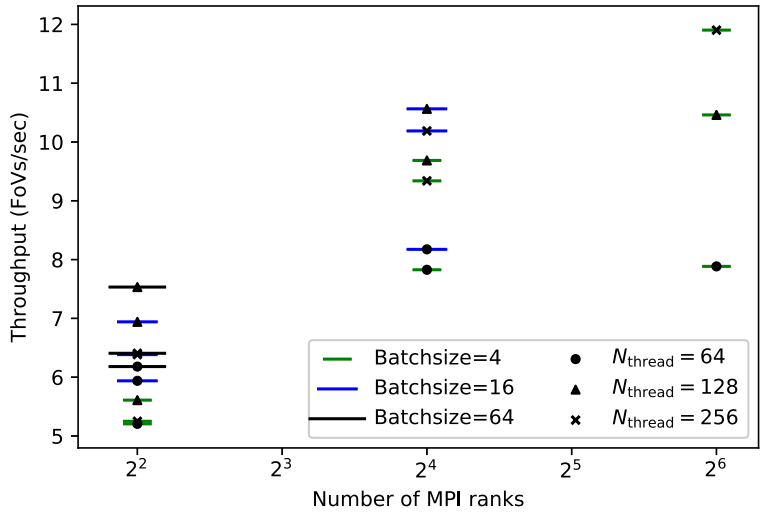


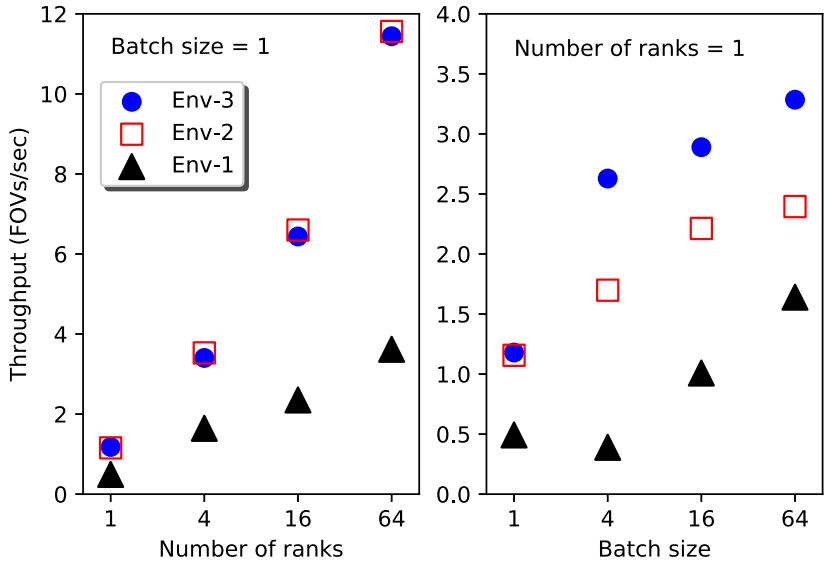
Contents

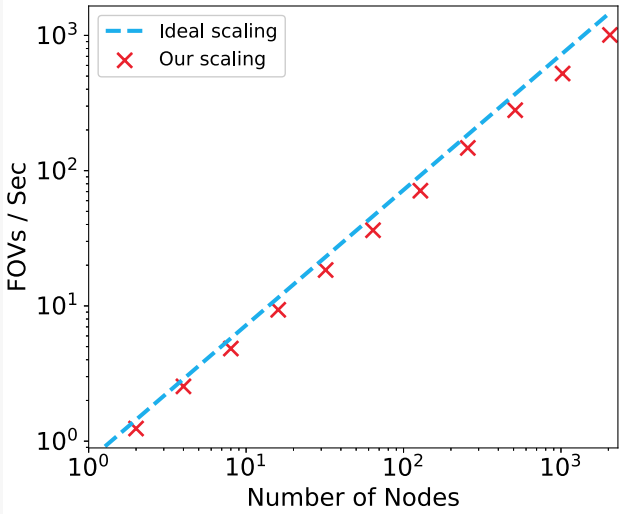
- Background
- Methods
- **Results**
- Summary

Results

- Throughput optimization

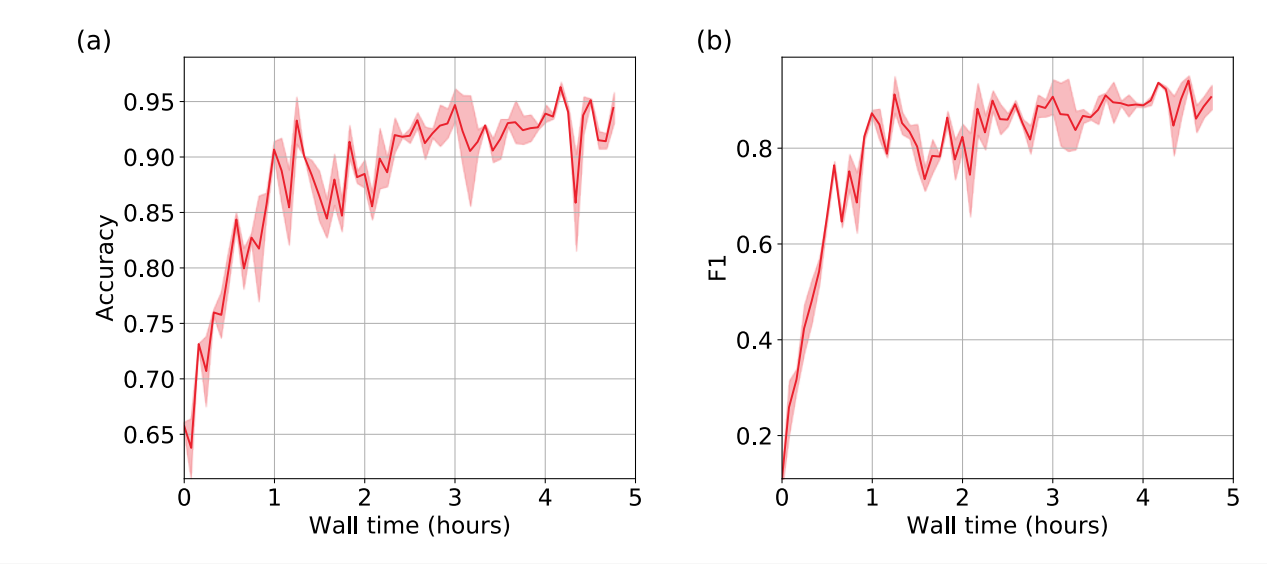


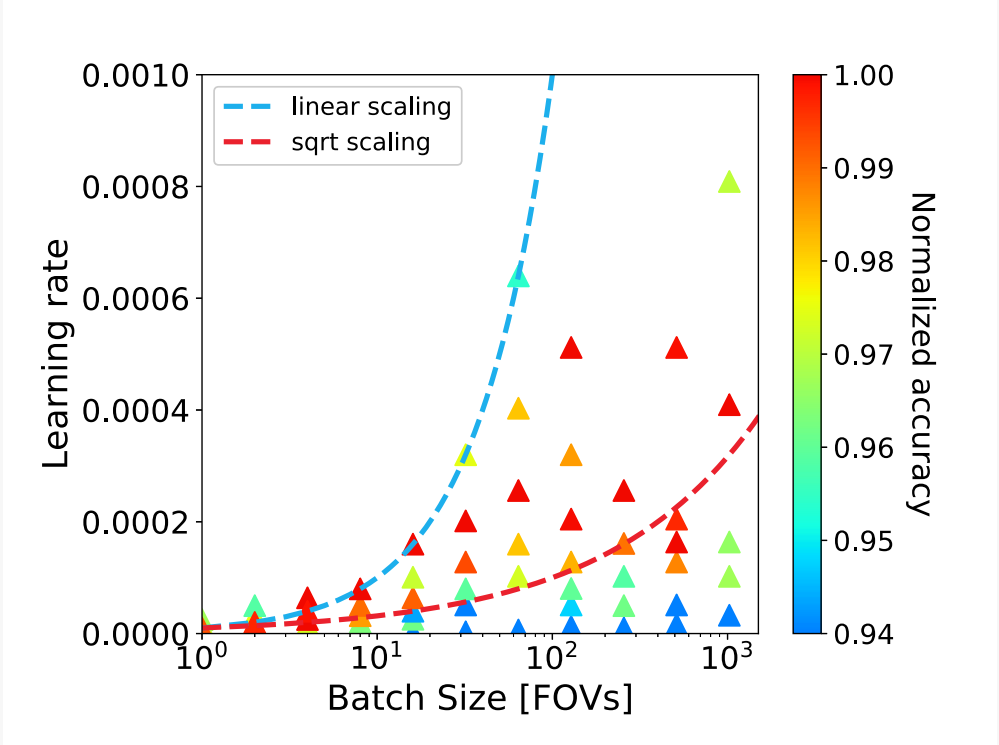




Results

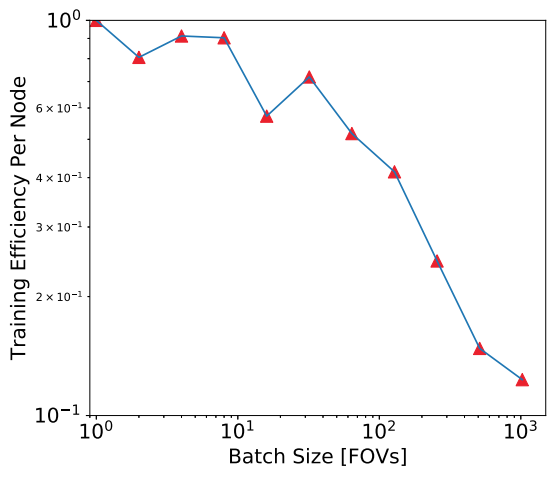
- Throughput optimization
- Optimal learning rate scaling

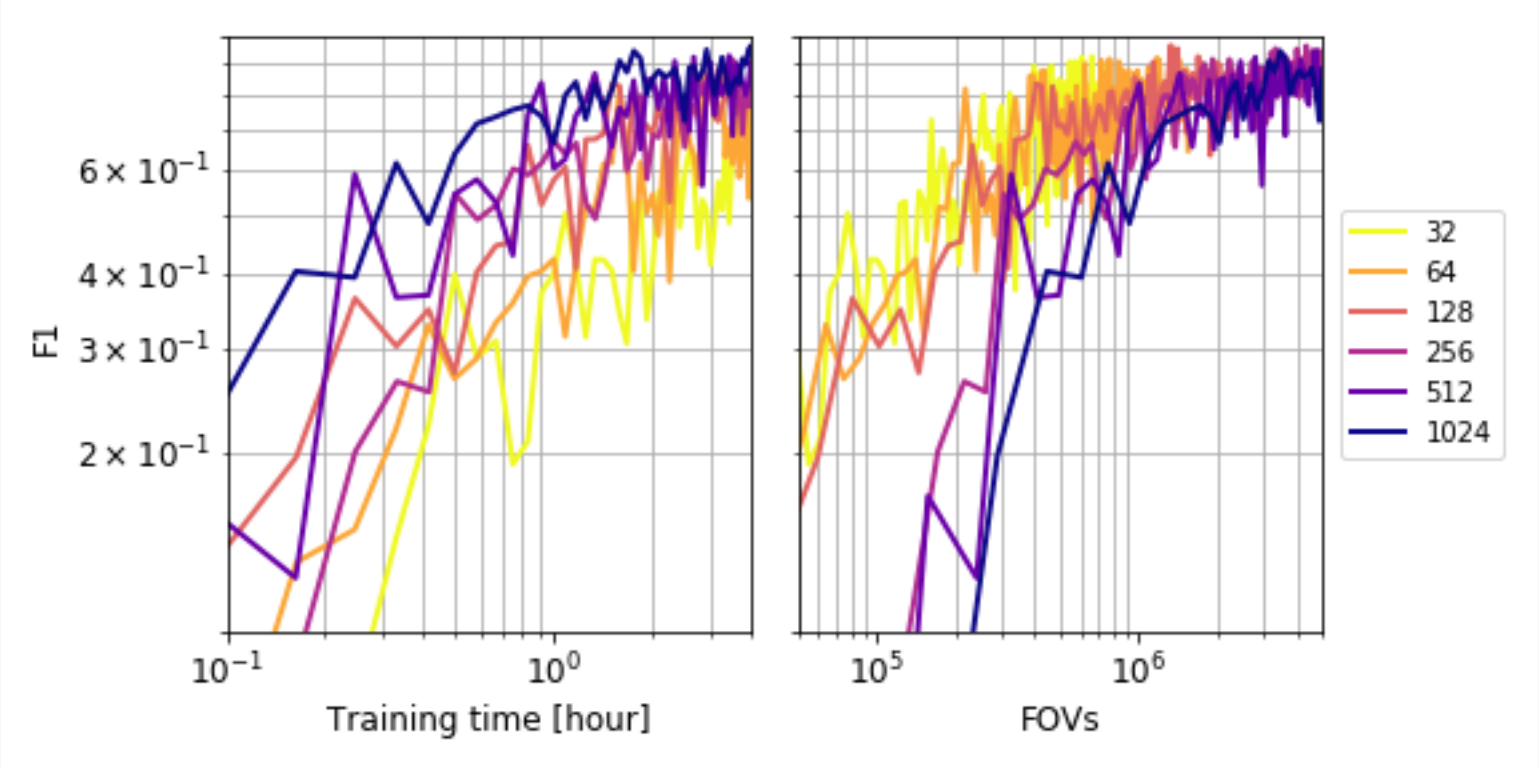




Results

- Throughput optimization
- Optimal learning rate scaling
- Effect of batch sizes





Results

- Throughput optimization
- Optimal learning rate scaling
- Effect of batch sizes
- Training evaluation

True Positive (TP)

False Positive (FP)

True Negative (TN)

False Negative (FN)

Adapted Rand Error (ARE)

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$\text{F1} = \frac{2 \times \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{ARE} = 1 - \text{F1}$$

Variation of Information (VOI)

$$\text{VOI}_{\text{split}} = H(X|Y)$$

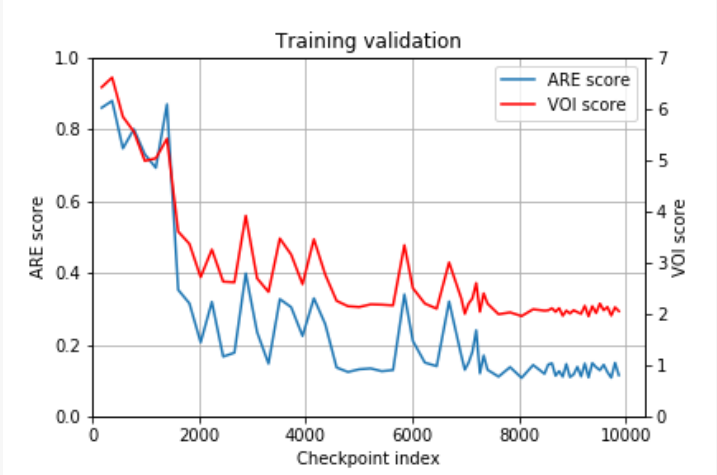
$$\text{VOI}_{\text{merge}} = H(Y|X)$$

$$\text{VOI} = \text{VOI}_{\text{split}} + \text{VOI}_{\text{merge}}$$

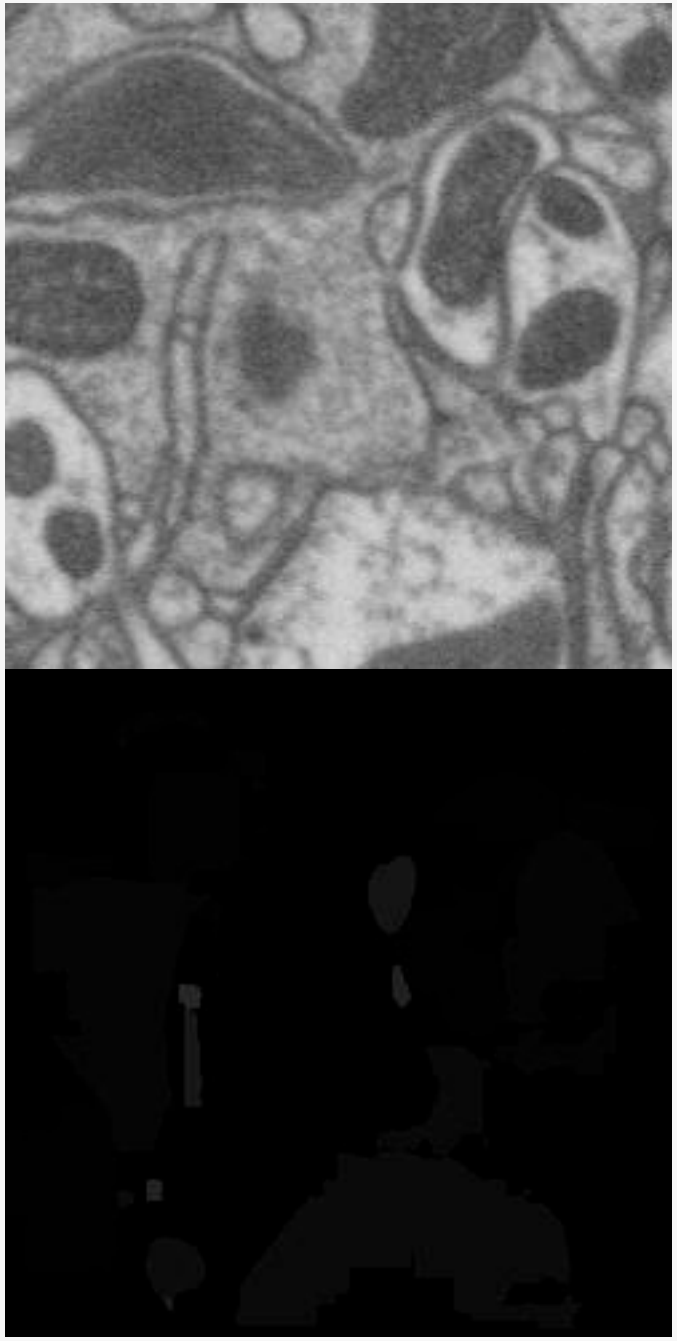
H(X) denotes the entropy of X

Results

- Throughput optimization
- Optimal learning rate scaling
- Effect of batch sizes
- Training evaluation



Segmentation	ARE	VOI
MALA	N/A	1.1470
CELIS-MC	N/A	1.1208
Original FFN paper	0.0973	3.2085
Ours	0.1074	1.9518



Contents

Summary

- Implemented data-parallel synchronous training of FFN and scaled it up to 2048 KNL nodes on Theta
- Reduced FFN training time needed to reach similar levels of evaluation quality
- Showed the tradeoff between compute-efficiency and time-efficiency
- First step towards a complete computational pipeline to produce large-scale brain maps
- Take-home message: Efficient training on HPC requires efficient usage of large training batches.

Acknowledgement

Murat Keceli, Rafael Vescovi, Hanyu Li, Corey Adams, Elise Jennings, Samuel Flender, Thomas Uram, Venkatram Vishwanath, Nicola Ferrier, Bobby Kasthuri, and Peter Littlewood.