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

Wushi Dong[1], Murat Keceli[2], Rafael Vescovi[2], Hanyu Li[1], Corey Adams[2], Elise Jennings[2], Samuel Flender[2], Thomas Uram[2], Venkatram Vishwanath[2], Nicola Ferrier[2], Bobby Kasthuri[1,2], and Peter Littlewood[1,2].

[1]University of Chicago
[2]Argonne National Laboratory

THE UNIVERSITY OF CHICAGO Argonne NATIONAL LABORATORY

Contents

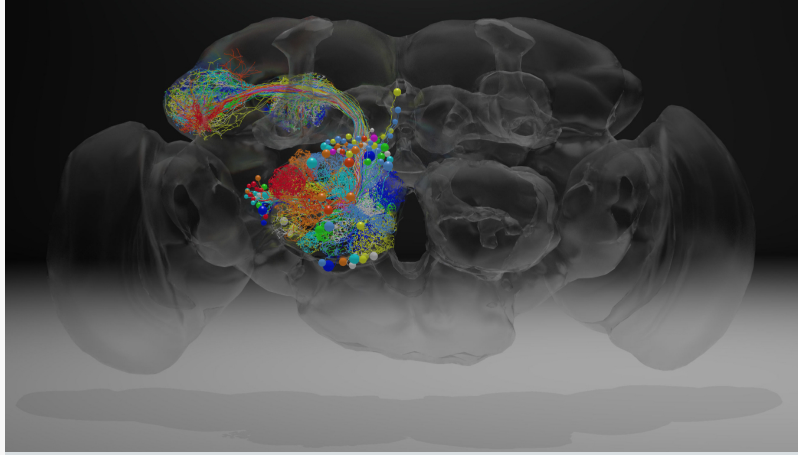
- Background
- Methods
- Results
- Summary

Contents

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

THE UNIVERSITY OF CHICAGO Argonne NATIONAL LABORATORY

Background



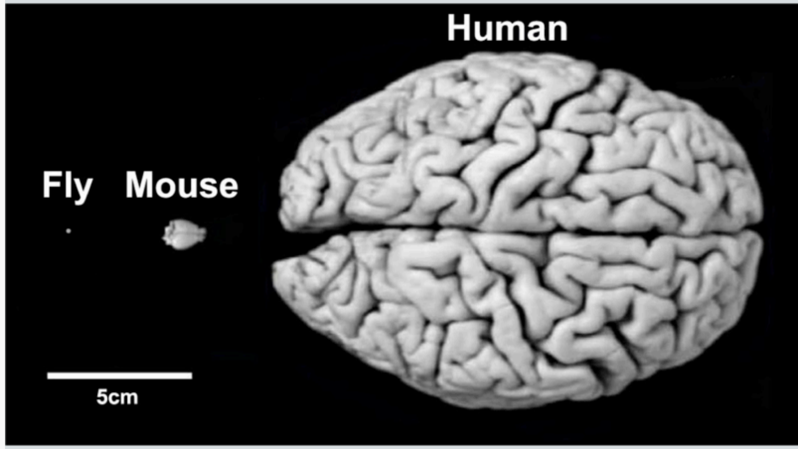
A 100,000-neuron *Drosophila* (Fruit fly) brain has been imaged at synaptic resolution.

Zheng, Zhihao, et al. Cell 174.3 (2018)

Brain mapping

- Reconstruction of neuronal wiring diagrams
- Aid in understanding of brain function
- Data density: 1 mm^3 ~ 10^{15} pixels ~ Petabytes (PB) ~ 100 million annotation hours
- Brain scales

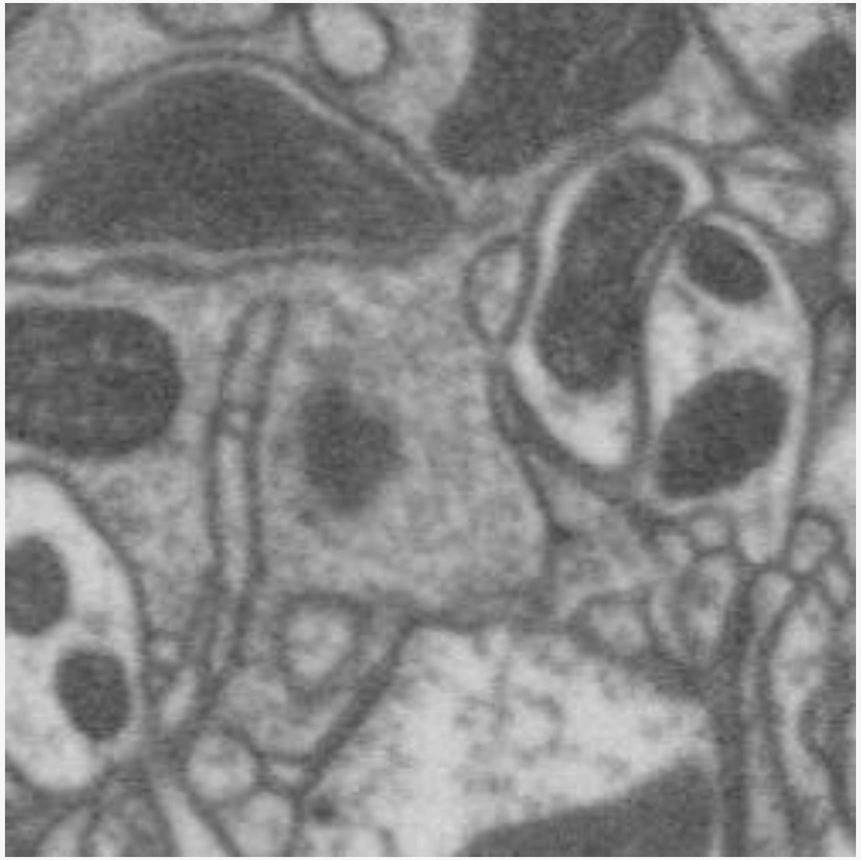
Species	# of neurons	Volume (mm^3)	Data size (PB)	Completion time
C. elegans	~10^2	\	\	1986
Drosophila	~10^5	~10^{-1}	~10^{-1}	2018
Mouse	~10^8	~10^2	~10^2	*2025?*
Human	~10^{11}	~10^6	~10^6	*???*



Background

Scanning Electron-Microscopy (SEM) Imaging Data

- Fib-25 dataset
  - *Drosophila* optical lobe
  - Focused Ion Beam (FIB) SEM
  - At a resolution of 8 × 8 × 8 nm
  - 52 × 53 × 65 μm^3 total volume
    - Training: 520^3-voxels subvolume
    - Testing: 250^3-voxels subvolume

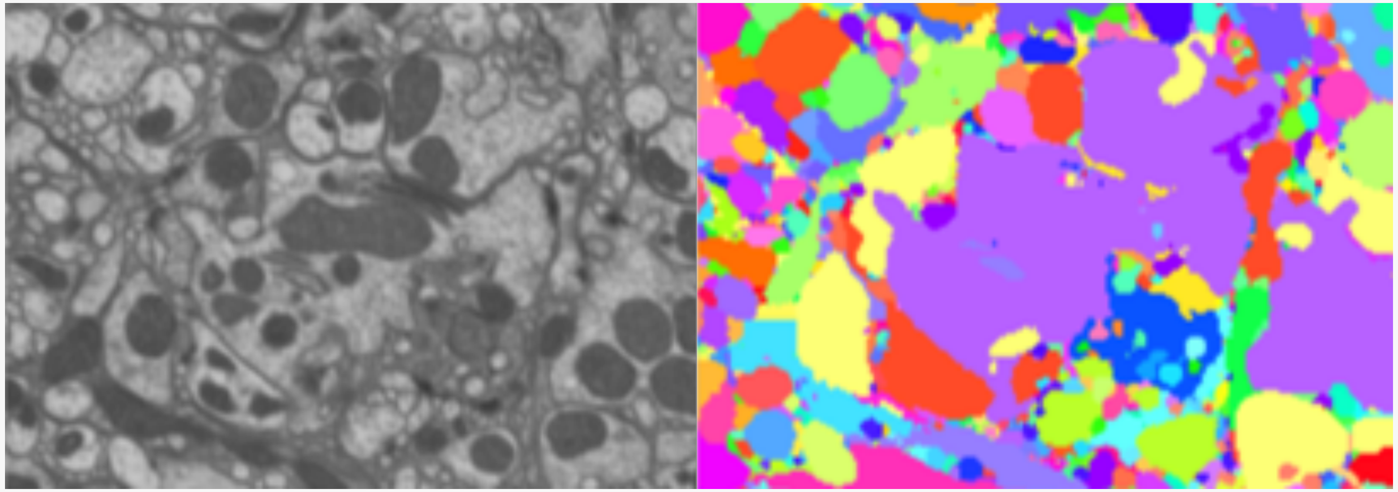


Images of cellular structure in consecutive slices of Fib-25 dataset.

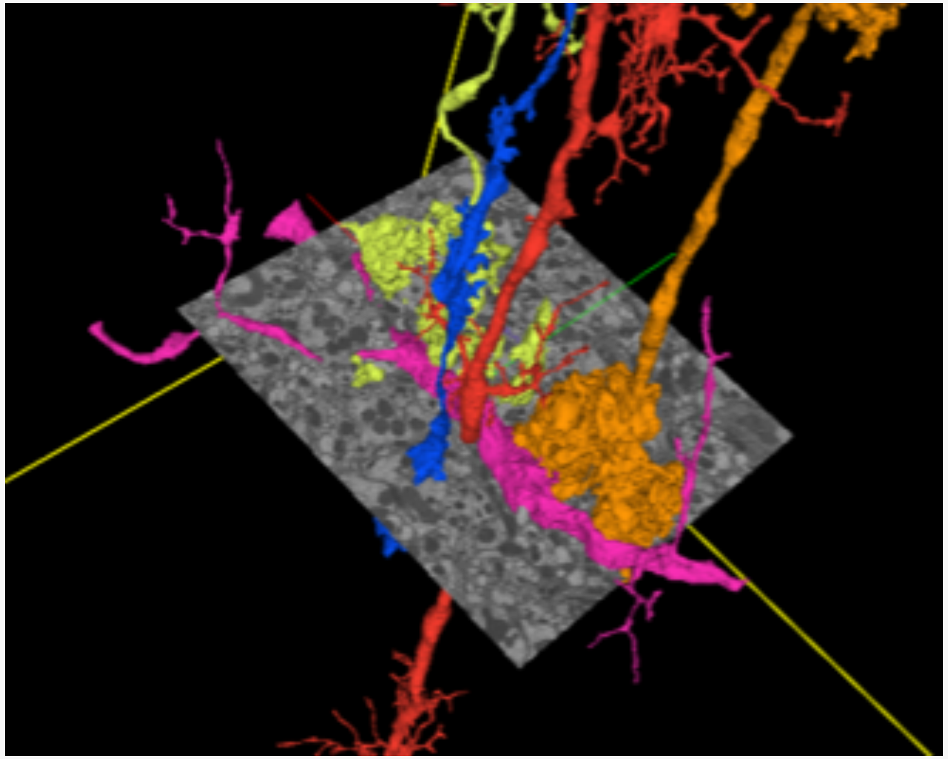
Takemura, Shin-ya, et al. Proceedings of the National Academy of Sciences 112.44 (2015)

Background

3D volumetric segmentation



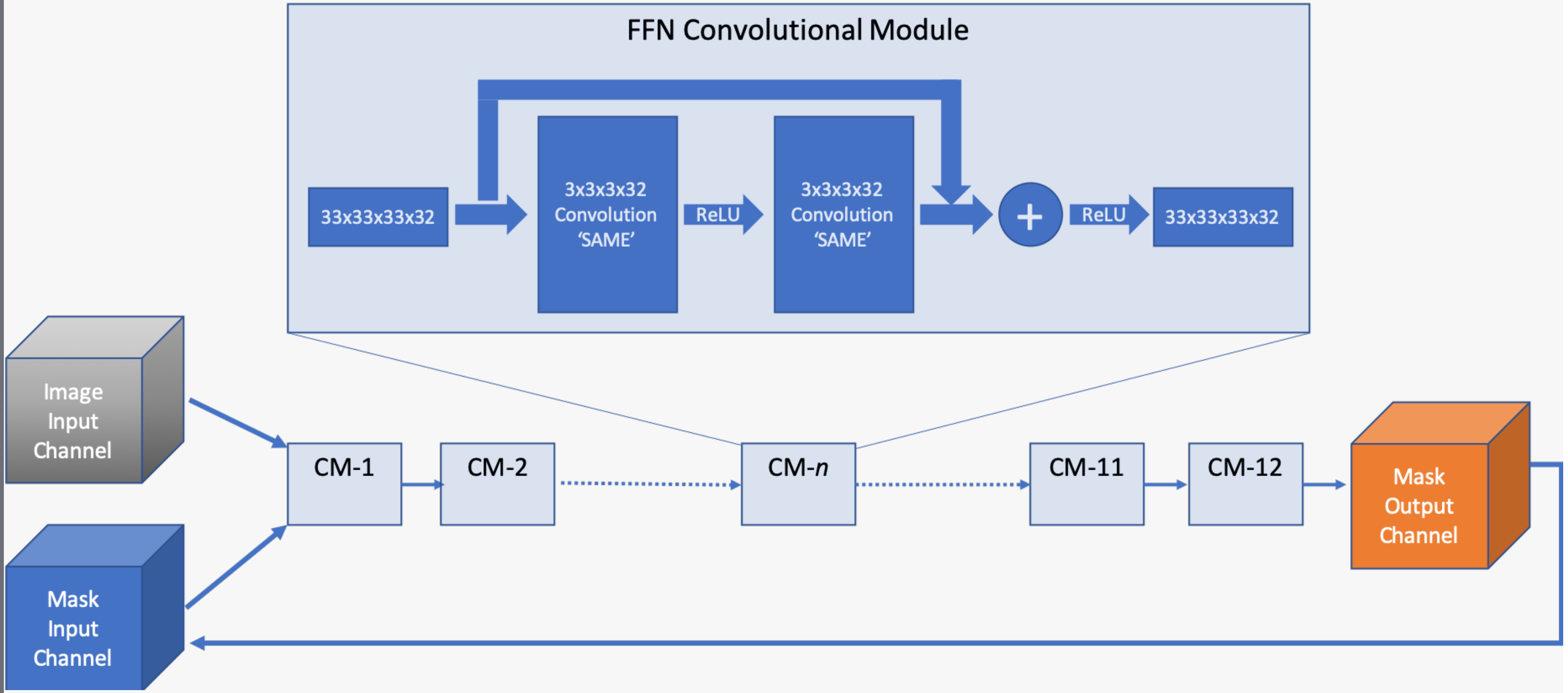
Raw imaging data (left) and manually annotated image (left) showing cell boundaries.

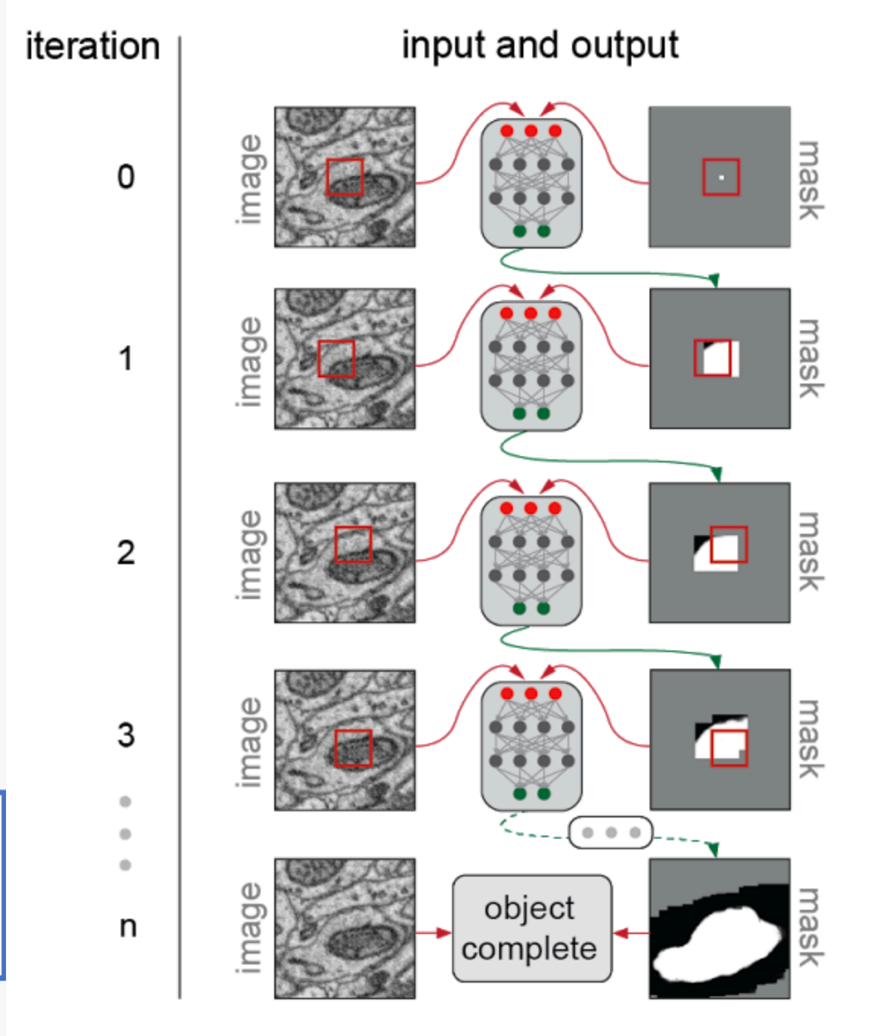


3D reconstruction of selected neurons from ~200 slices.

Background

Flood-Filling Networks (FFN)

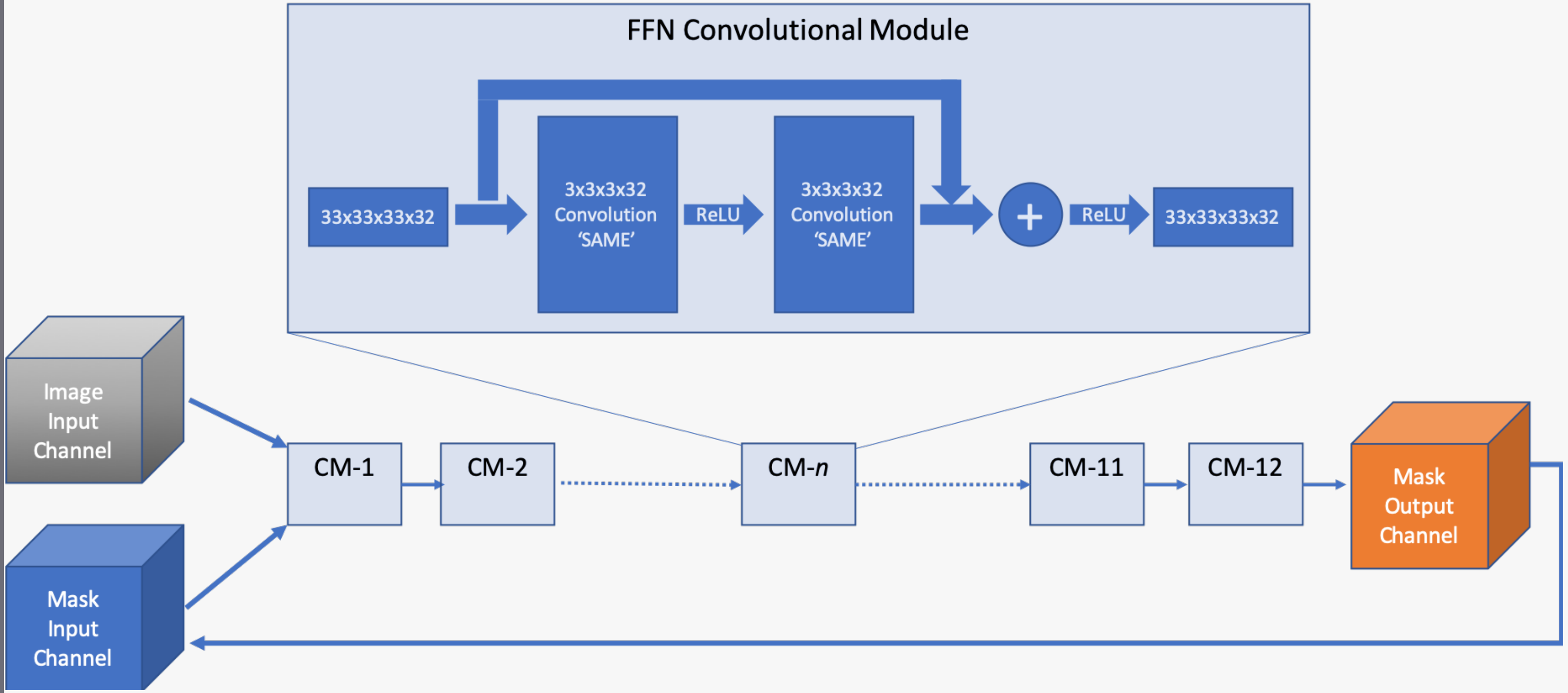


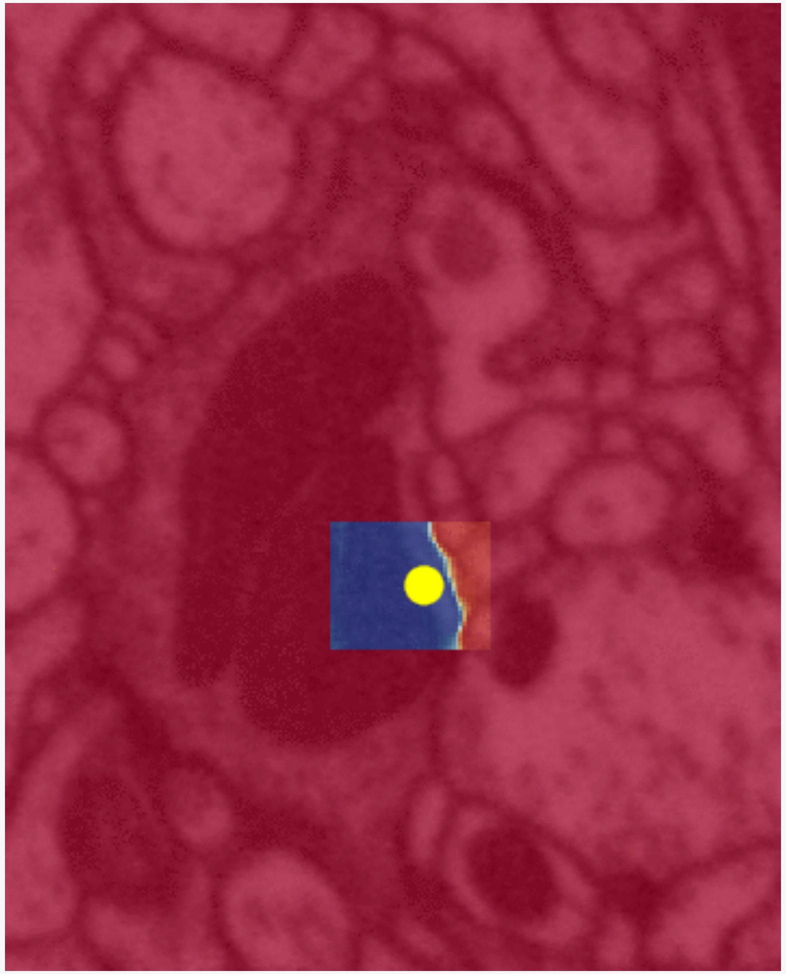


Januszewski, Michał, et al. Nature methods 15.8 (2018)

Background

Flood-Filling Networks (FFN)





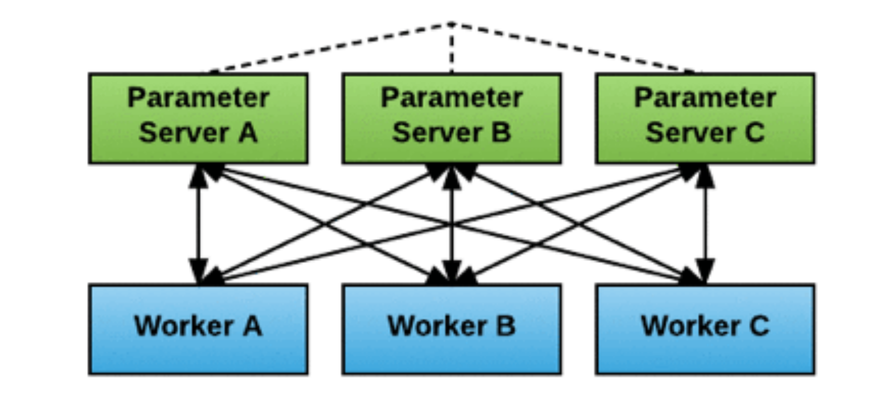
Januszewski, Michał, et al. Nature methods 15.8 (2018)

Contents

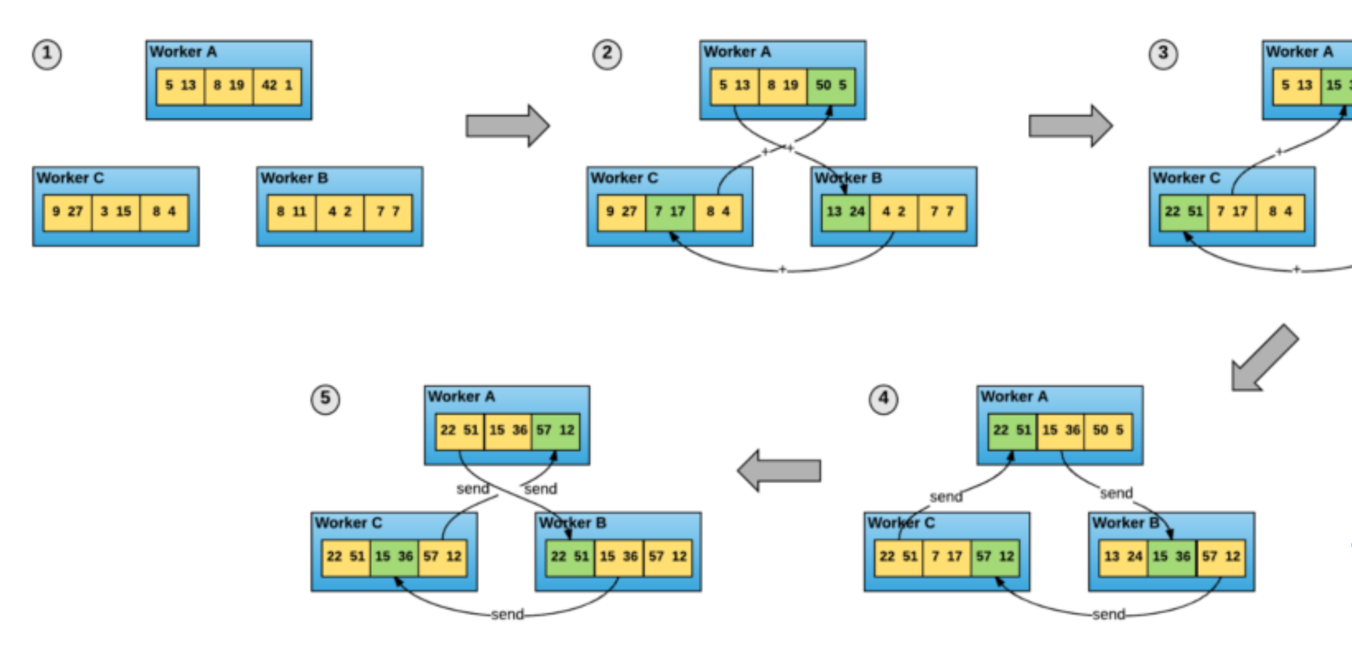
Methods

Asynchronous VS Synchronous training

Asynchronous training



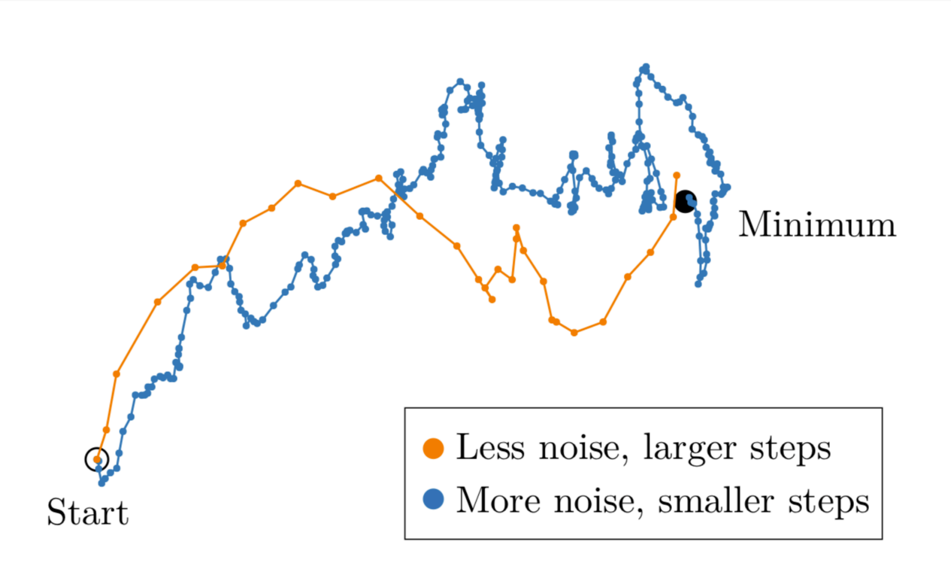
Synchronous training

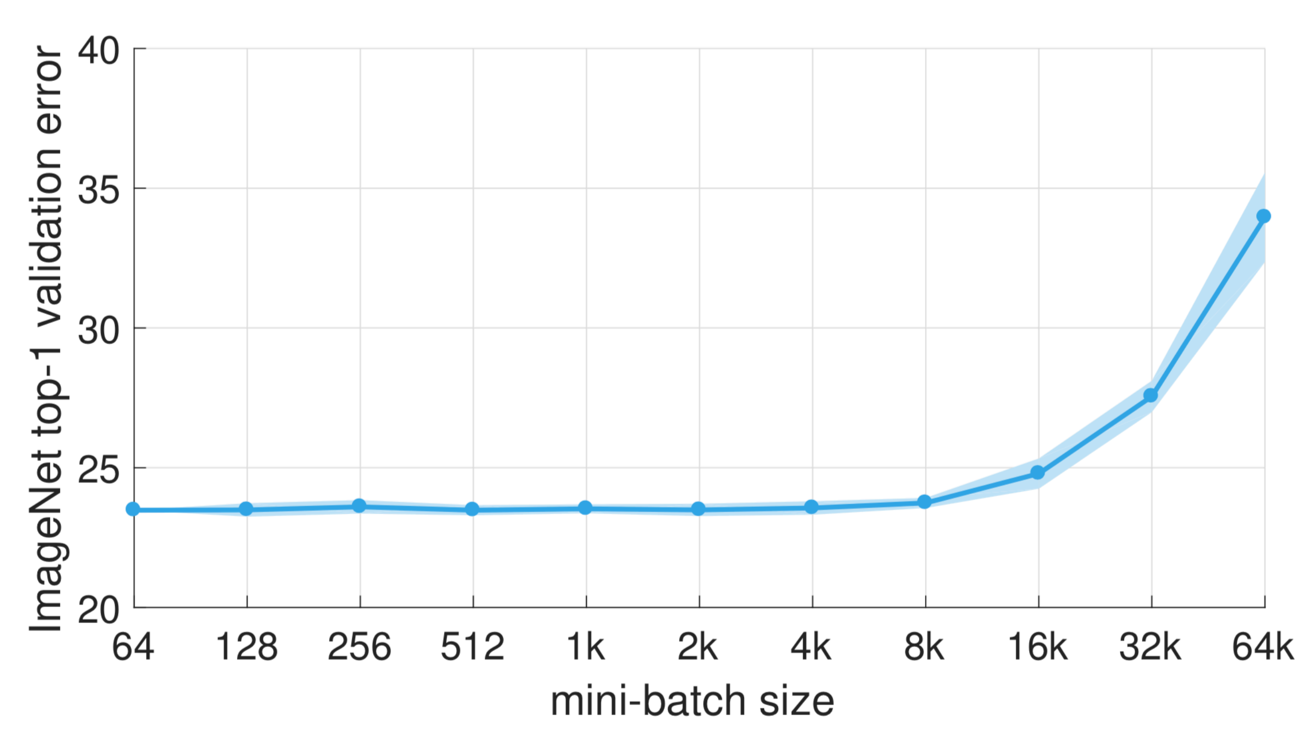


Sergeev, Alexander, and Mike Del Balso. arXiv preprint arXiv:1802.05799 (2018)

Methods

Large-batch training





Goyal, Priya, et al. arXiv preprint arXiv:1706.02677 (2017)

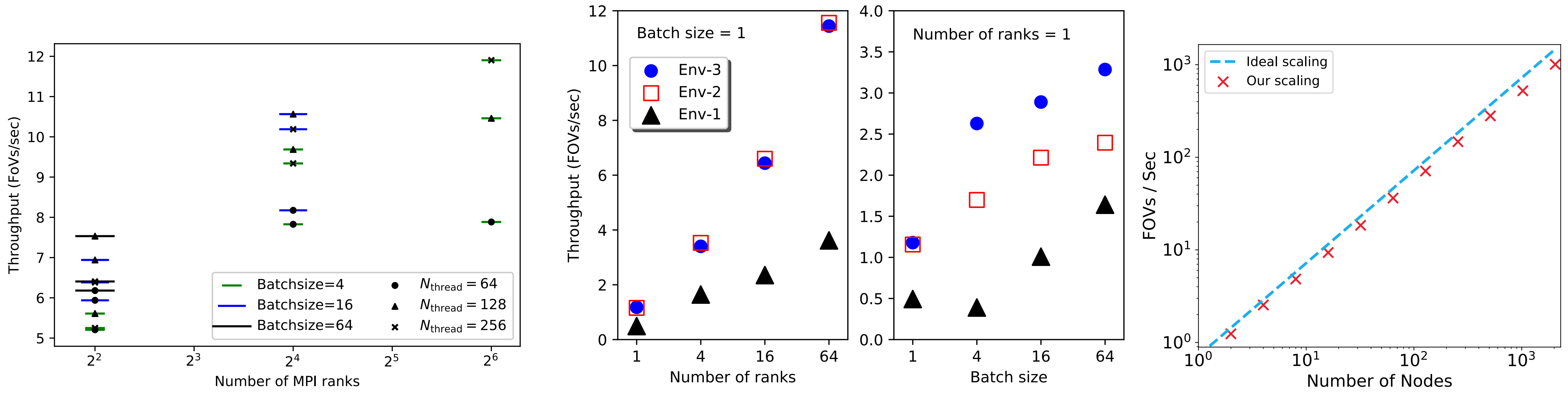
Contents

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

Results

Throughput optimization

- Theta supercomputer (KNL-based system)



Results

Evaluation metrics

Adapted Rand Error (ARE)

True Positive (TP)

False Positive (FP)

True Negative (TN)

False Negative (FN)

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

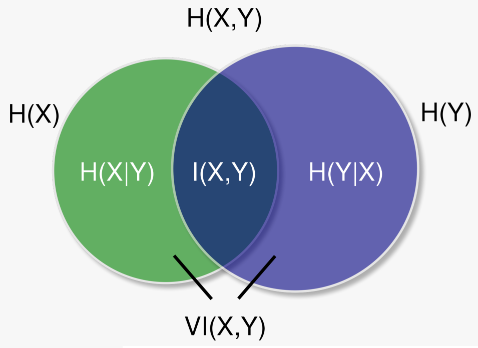
$$\text{precision} = \frac{TP}{TP + FP}$$

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

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

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

Variation of Information (VOI)



H(X) denotes the entropy of X

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

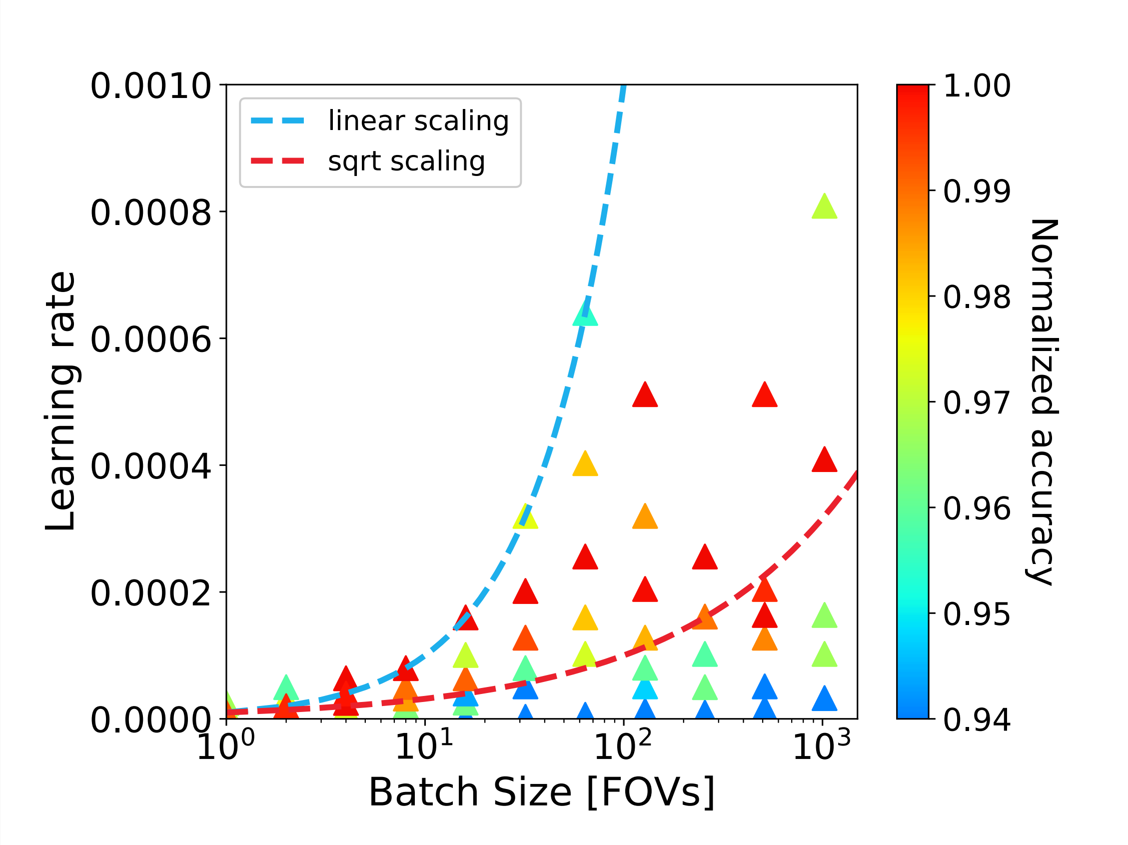
$$\text{VOI}_{\text{merge}} = H(Y|X)$$

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

Arganda-Carreras, Ignacio, et al. Frontiers in neuroanatomy 9 (2015)

Results

Optimal learning rate scaling

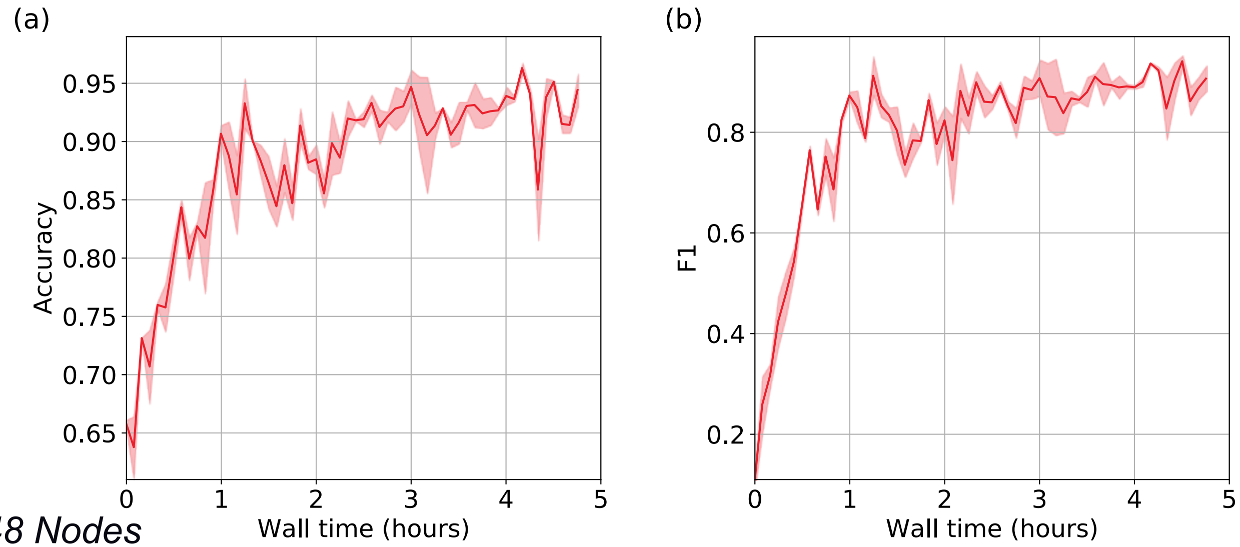


** Normalized accuracy =*

$$\frac{\textit{Accuracy reached after a certain number of steps}}{\textit{Max accuracy reached within the same batch size}}$$

Results

Training results



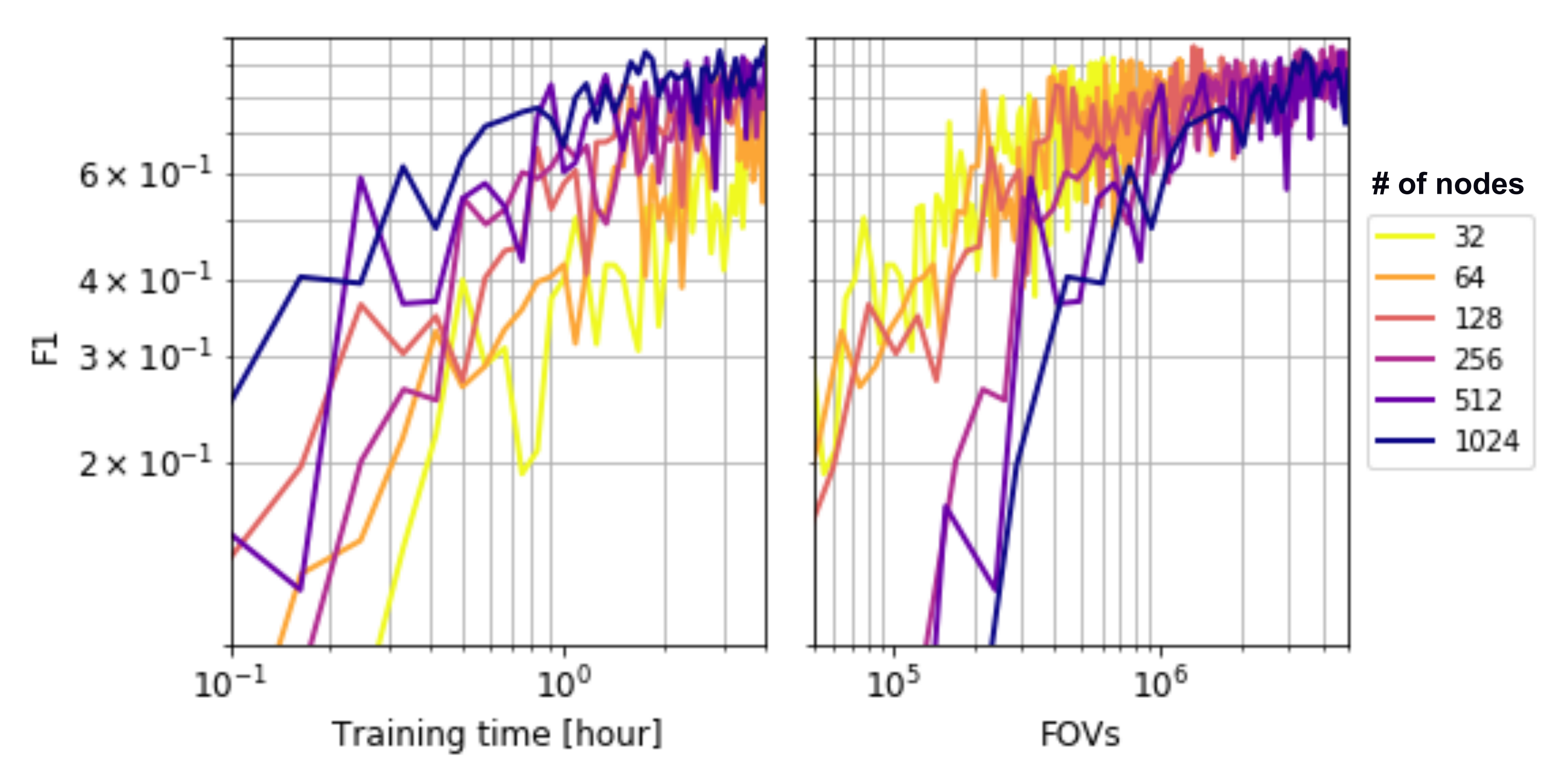
2048 Nodes

Enables exploration of the hyperparameter space to further improve training performance and to test different model architectures.

THE UNIVERSITY OF CHICAGO Argonne NATIONAL LABORATORY

Results

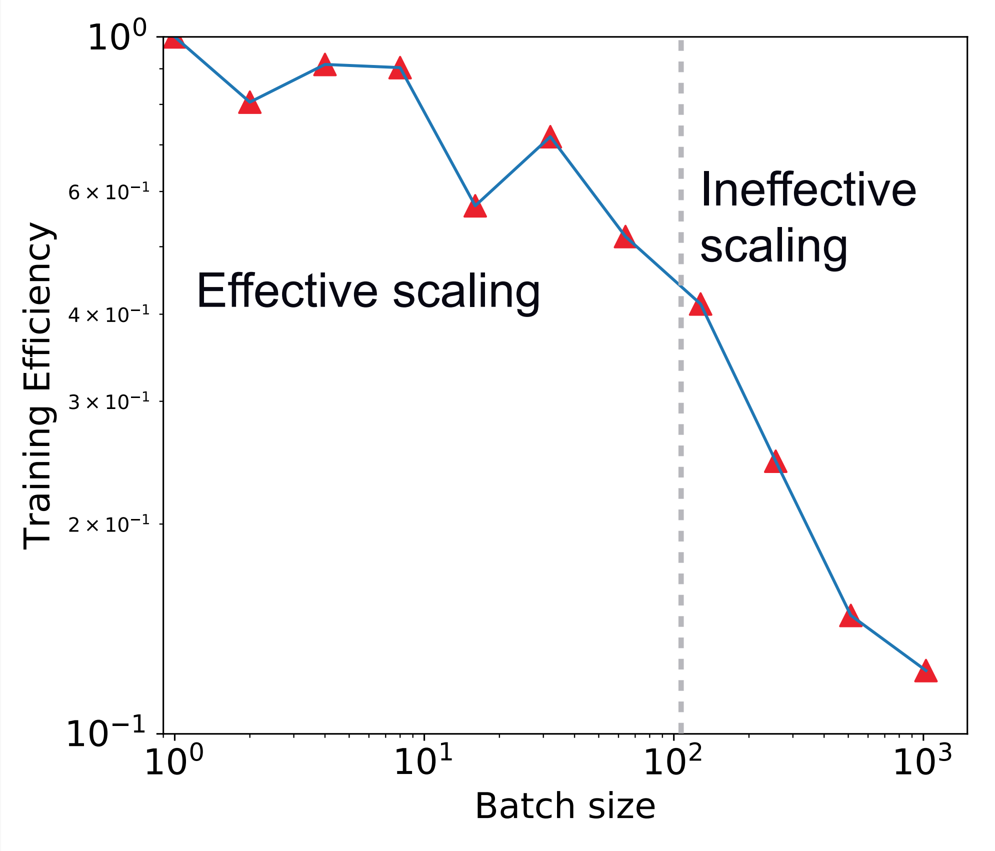
Compute-efficiency VS Time-efficiency



THE UNIVERSITY OF CHICAGO Argonne NATIONAL LABORATORY

Results

Effect of batch sizes on efficiency

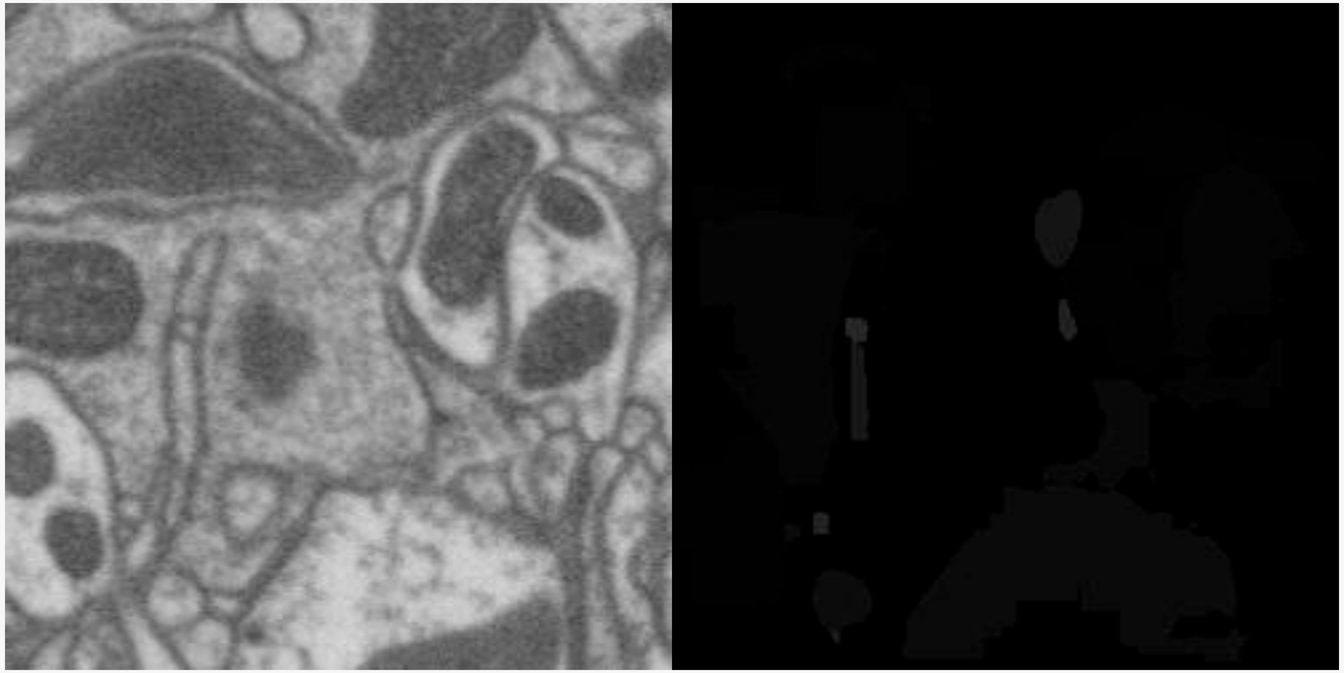


* *Training efficiency =*

(Time used to reach a specified F1 score)$^{-1}$

Results

Training evaluation



Fib-25 raw imaging data (left) and our volumetric segmentation results (left).

Segmentation	ARE	VOI
MALA	\	1.1470
CELIS-MC	\	1.1208
Original FFN paper	0.0973	3.2085 (Unagglomerated) —> 0.9375
Ours	0.1074	1.9518 (Unagglomerated)

Funke, Jan, et al. arXiv preprint arXiv:1709.02974 (2017)
Wolf, Steffen, et al. Proceedings of the IEEE International Conference on Computer Vision. 2017

Contents

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

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 good levels of evaluation quality
- Reduced training enables hyper-parameter optimization
  - Important for different data sets
- Showed the tradeoff between compute-efficiency and time-efficiency
- Take-home message: Efficient training on HPC requires efficient usage of large training batches.

Future works

- To implement automatic hyperparameter optimization to further improve training efficiency
- To test different model architectures

Acknowledgement

- University of Chicago Office of Research and National Laboratories and the Center for Data and Computing (CDAC)
- U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research, under contract number DE-AC02-06CH11357
- Argonne Leadership Computing Facility

Thank You!

For more information: arXiv 1905.06236

dongws@uchicago.edu