Enabling Scalable and Efficient Deep Learning on Supercomputers

Zhao Zhang

Texas Advanced Computing Center

TACC

Motivating Applications

- Neural image resolution enhancement with super resolution generative adversarial network.
  - In collaboration with Salk Institute
  - ~600 GB neural image dataset
  - TensorLayer + TensorFlow + Horovod
- Plasma reactor disruption prediction with fusion recurrent neural network.
  - In collaboration with ICES, UT Austin
  - ~1.7 TB text data
  - TensorFlow + MPI4Py

Motivating Applications

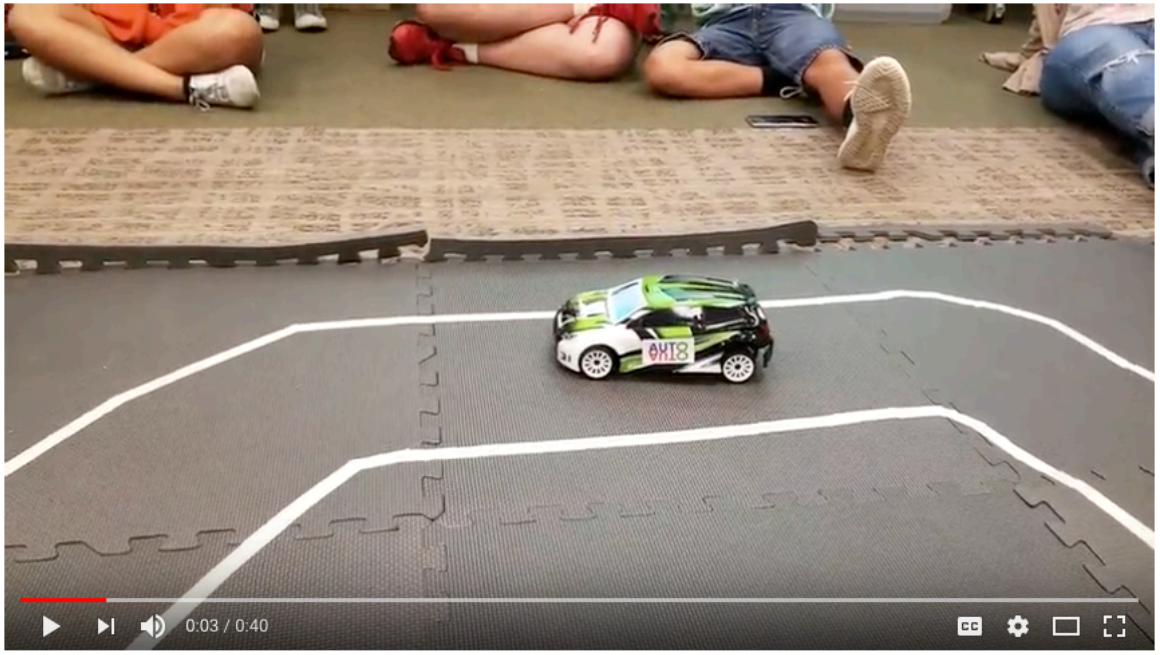
- MRI image analysis for multiple sclerosis patient management (Ponnada Narayana and et al., Texas Medical Center and UT Tyler)
- Deep Natural Language Understanding with Probabilistic Logic and Distributional Similarity (Katrin Erk, UT Austin)
- Functional Mapping between Geophysical Parameters and the Observable States (Alex Sun, BEG)
- Cancer Drug Treatment Prediction (Beibei Huang, MD Anderson)
- Deep Learning Driven Detector Design (Sofia Vallescorsa, CERN)
- Hyper-parameter Tuning with Black-box Optimization (Sanjay Shakkott, UT Austin)
- Power Control in Mars Rover (Chris Mattmann, NASA)
- Geological Image Analysis (David Walling, TACC)
- Ancient Biographical Text Analysis, (David Walling, TACC)
- Austin Traffic Analysis (Weijia Xu, TACC)

Motivating Applications

- MRI image analysis for multiple sclerosis patient management (Ponnada Narayana and et al., Texas Medical Center and UT Tyler)
- Deep Natural Language Understanding with Probabilistic Logic and Distributional Similarity (Katrin Erk, UT Austin)
- Functional Mapping between Geophysical Parameters and the Observable States (Alex Sun, BEG)
- Cancer Drug Treatment Prediction (Beibei Huang, MD Anderson)
- Deep Learning Driven Detector Design (Sofia Vallescorsa, CERN)
- Hyper-parameter Tuning with Black-box Optimization (Sanjay Shakkott, UT Austin)
- Power Control in Mars Rover (Chris Mattmann, NASA)
- Geological Image Analysis (David Walling, TACC)
- Ancient Biographical Text Analysis, (David Walling, TACC)
- Austin Traffic Analysis (Weijia Xu, TACC)

Deep Learning Activities

- Machine Learning Summer Institute
- PetroBras 12-week Deep Learning Class
- Code @ TACC



Observed Problems

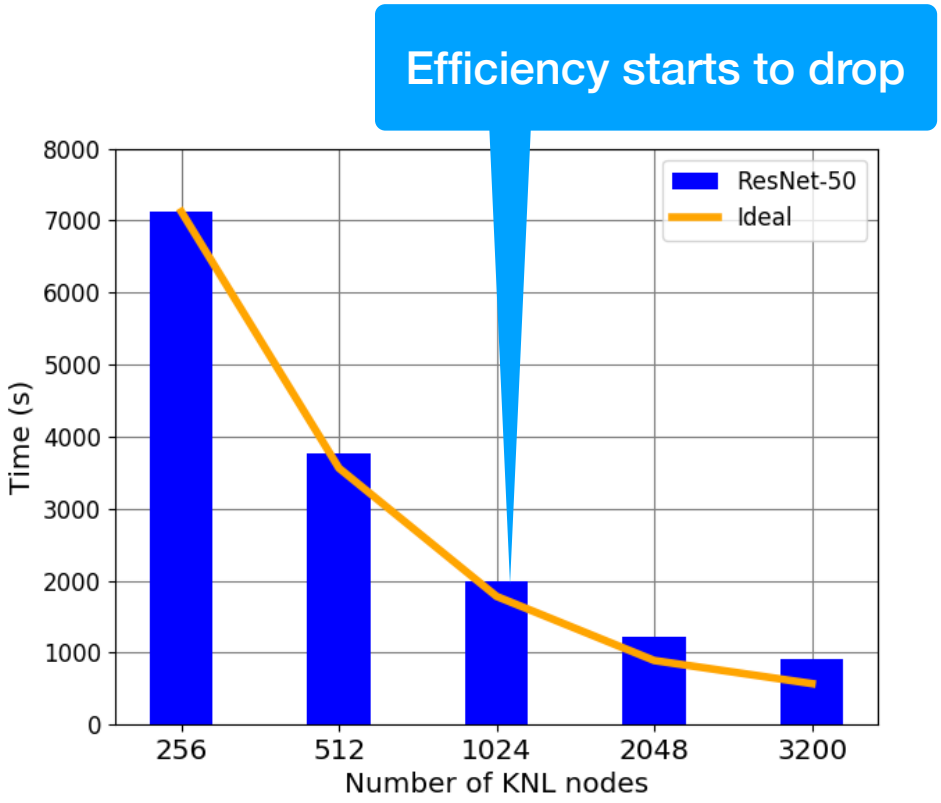
- Large batch sizes result in degraded test accuracy
- I/O results in file system slowdown or unresponsiveness

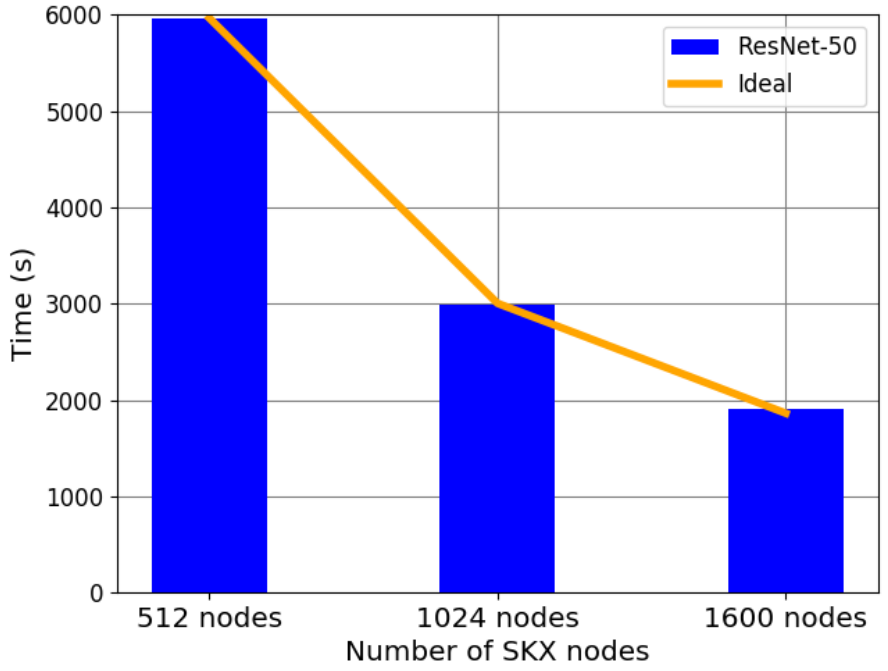
Train with Large Batch Size

- Layer-wise Adaptive Rate Scaling
  - Intuition: learning rate should be adjusted according to the norm of the weights in each layer
- Result
  - 90-epoch ResNet-50 training finished in 20 mins on 2,048 KNL with 74.9% top-1 accuracy

Train with Large Batch Size

- Using batch size of 32K on KNL and SKX nodes





You, Yang, Zhao Zhang, Cho-Jui Hsieh, James Demmel, and Kurt Keutzer. "ImageNet training in minutes." In Proceedings of the 47th International Conference on Parallel Processing, p. 1. ACM, 2018. Best paper

Train with Large Batch Size

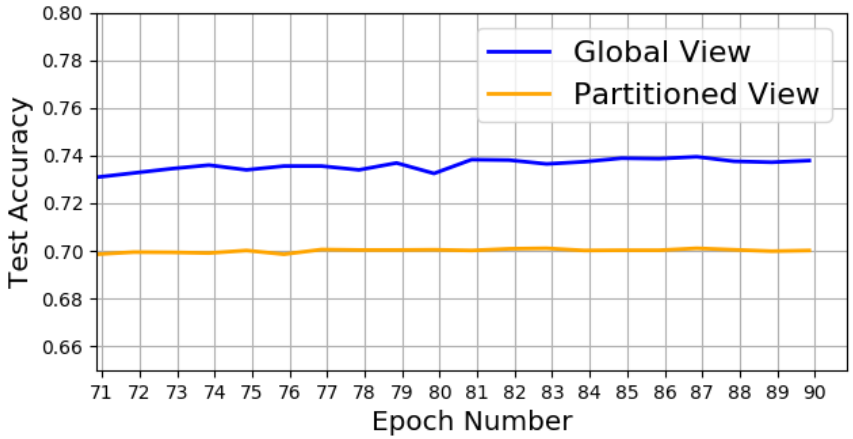
- Follow-on Work
  - 6.6 minutes using 2,048 Tesla P40 GPUs from researchers in Tencent and Hong Kong Baptist University
    - Half-precision for forward computation and back propagation, single precision for LARS
  - 224 seconds using 2,176 Tesla V100 GPUs from Sony researchers
    - 2D-Torus optimization with LARS

Deep Learning I/O

- Problem Statement
  - DL's long lasting, repeated, high volume, and highly concurrent file access can easily saturate the metadata and data service of traditional shared file system.
  - ResNet-50 with Kera, TensorFlow, and Horovod on 16 nodes, each with 4 GPUs
    - 128K stat() and readdir() operations with 64-way concurrent access
    - 117M stat(), open(), close() operations with 256-way concurrent access
    - ~180M read() operations with same concurrency
    - ~ 8 hour duration

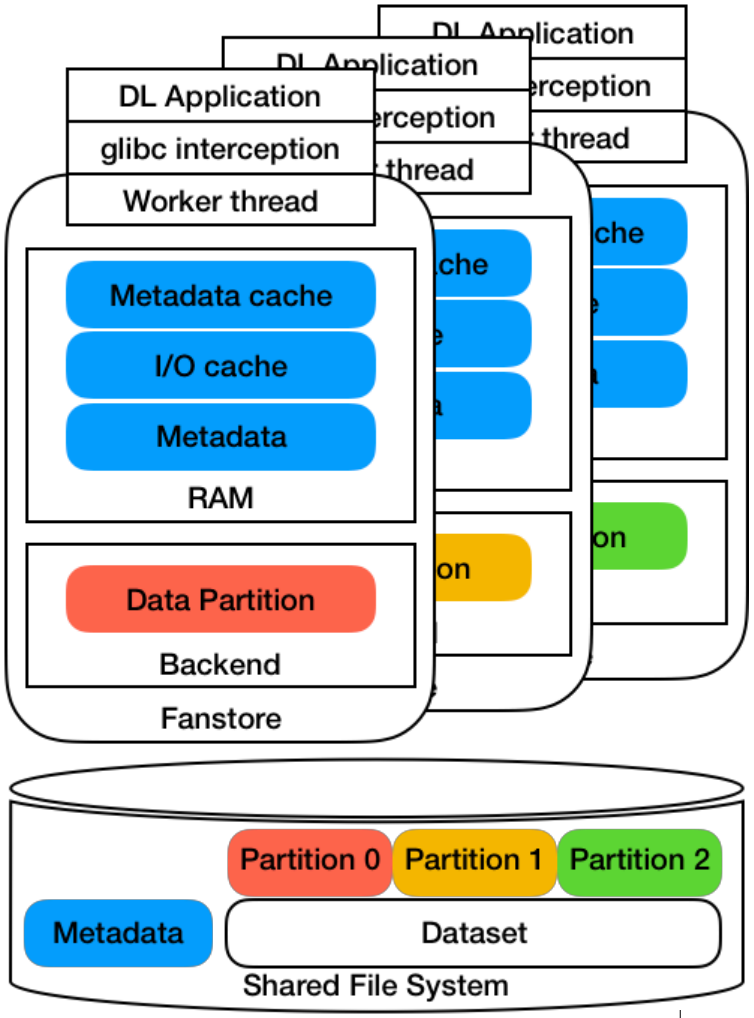
I/O Requirements

- Global namespace
- POSIX compliant interface
- High metadata availability
- High data availability
- Relaxed writing consistency
- Higher utilization of local storage space

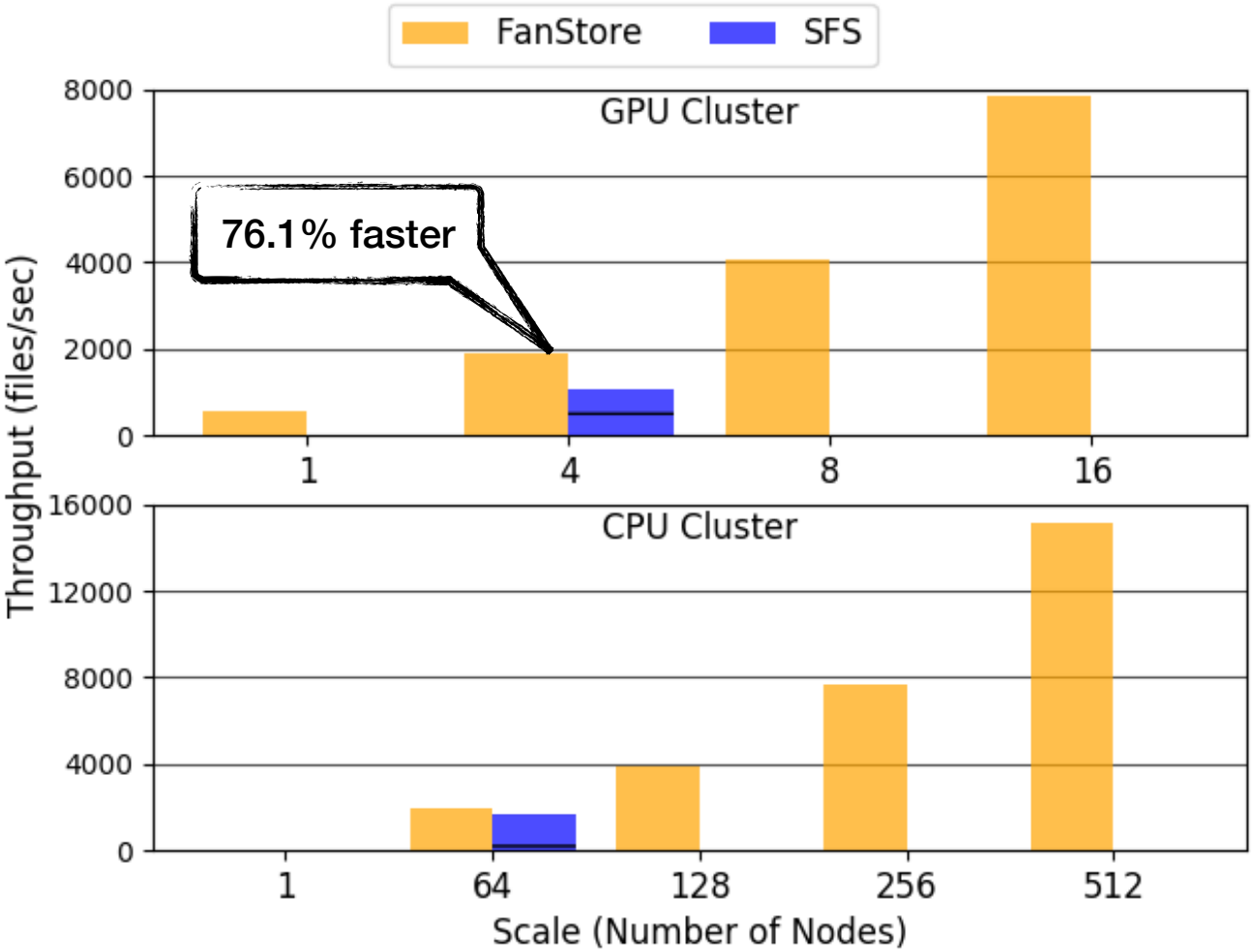


FanStore Design

- FanStore is a transient runtime file system that optimizes I/O for distributed DL training.
- Data is partitioned (optionally compressed) and spread across local storage space
- File access functions are intercepted and handled in user space
- Remote file access is in the form of MPI round-trip message



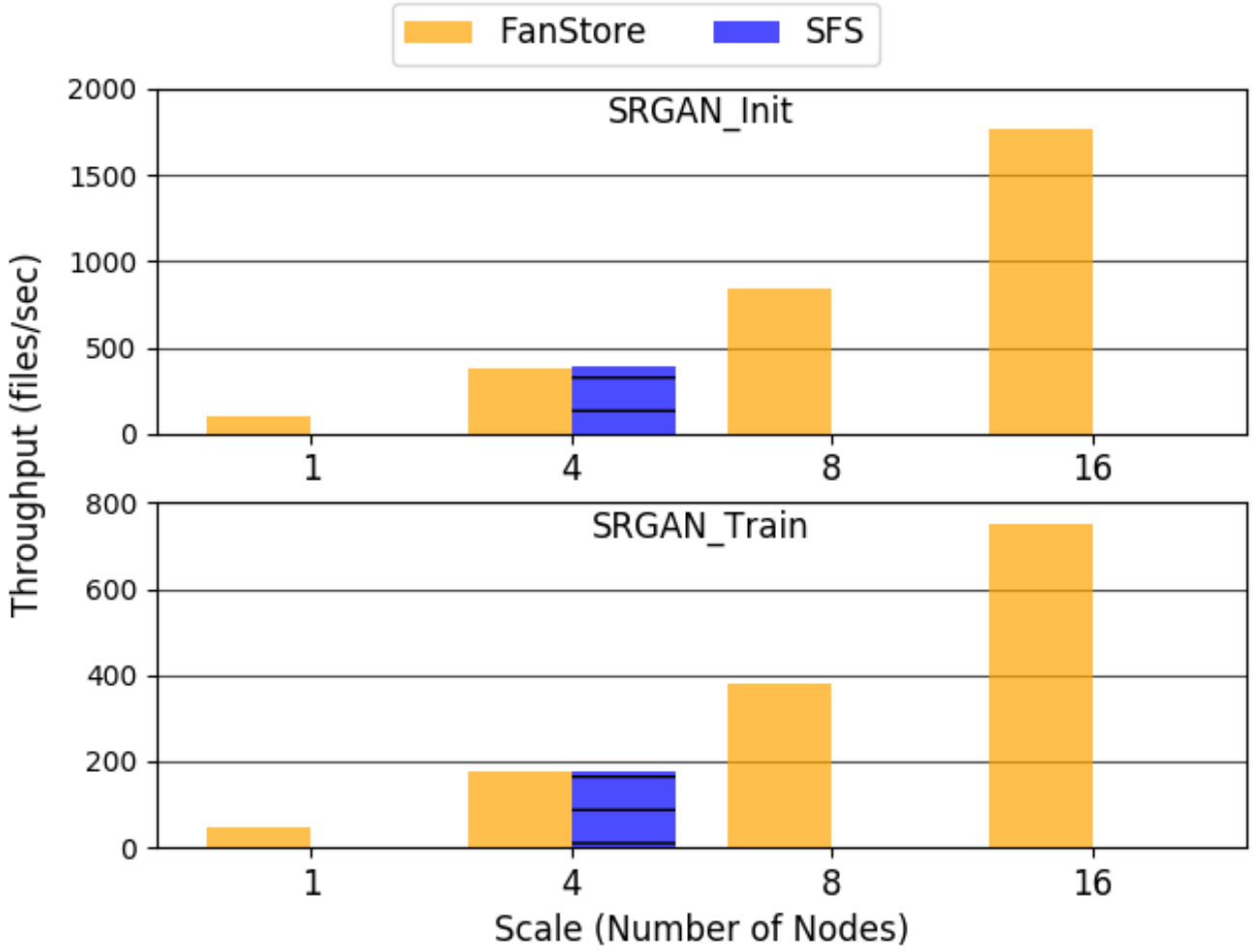
ResNet-50 Results



End with scalability

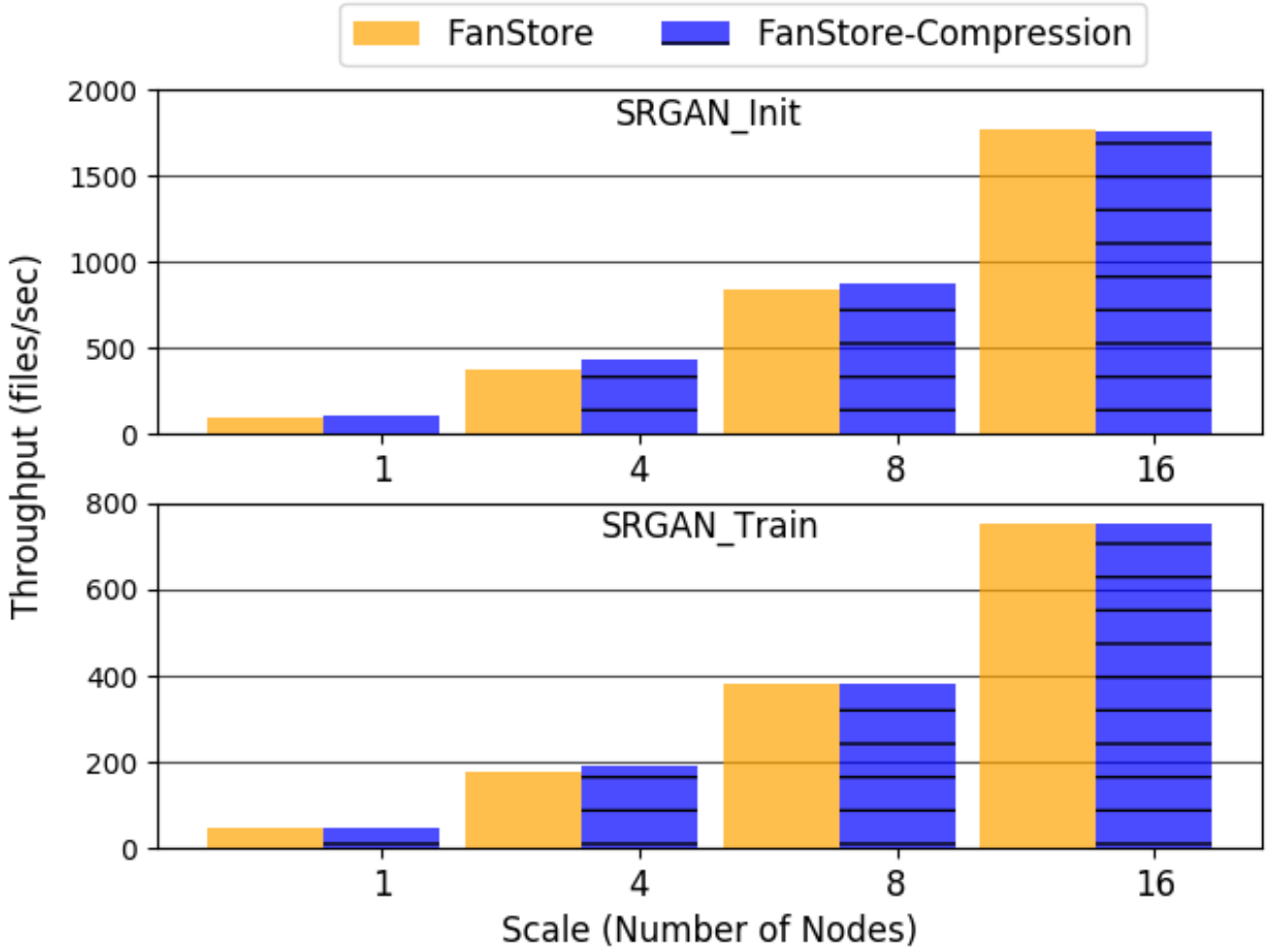
Zhao Zhang, Lei Huang, Uri Manor, Lingjing Fang, Gabriele Merlo, Craig Michoski, John Cazes, Niall Gaffney.
"FanStore: Enabling Scalable and Efficient I/O for Distributed Deep Learning".
Preprint: https://arxiv.org/abs/1809.10799

SRGAN Results



Zhao Zhang, Lei Huang, Uri Manor, Lingjing Fang, Gabriele Merlo, Craig Michoski, John Cazes, Niall Gaffney.
"FanStore: Enabling Scalable and Efficient I/O for Distributed Deep Learning".
Preprint: https://arxiv.org/abs/1809.10799

SRGAN Results with Compression



Zhao Zhang, Lei Huang, Uri Manor, Lingjing Fang, Gabriele Merlo, Craig Michoski, John Cazes, Niall Gaffney.
"FanStore: Enabling Scalable and Efficient I/O for Distributed Deep Learning".
Preprint: https://arxiv.org/abs/1809.10799

Conclusion

- We faced two major obstacles on our way to scalable deep learning: test-accuracy loss at scale and I/O
- With collaboration with UC Berkeley and UC Davis, we validate LARS algorithm's effectives on preserving test accuracy
- We designed and implemented FanStore to leverage the local node storage and interconnect to address the I/O issue at scale
- Real world applications such as SRGAN, FRNN, and ResNet-50 can scale linearly to hundreds of nodes with FanStore

Questions

- zzhang@tacc.utexas.edu