Architecting Multi-Scale Neural Networks For Life Science Discovery

Enabling 21st century science with Dell EMC

Presenter: Andrew Underwood & Alex Filby

intel DELL EMC

HELPING OUR CUSTOMERS TO SOLVE THE WORLD'S BIGGEST CHALLENGES

The fight against cancer

Natural disasters

Development

Work directly with Dell EMC HPC and AI experts

The Dell EMC HPC and AI Innovation Lab develops technologies and solutions, generates best practices and know-how, and assists customers in their quest to answer complex questions and make the right solution design decisions.

Typical HPC and AI Innovation Lab projects

While the list of potential projects is virtually limitless, some common projects include:

- Technology comparison
- System parameter sweeps
- GPU test comparison
- Efficiency tuning
- HPC network evaluation
- HPC storage system optimization
- Proof of concept studies
- Vertical Solutions

"Our lab is staffed by engineers with advanced degrees and many years of industry experience in domains such as mechanical engineering and bioinformatics. We also have engineers with computer science backgrounds, providing expertise in file systems, interconnects and HPC management tools."

—Onur Celebioglu, HPC Engineering Director and head of the HPC and AI Innovation Lab, Dell EMC

World-class infrastructure in the Innovation Lab

13K ft.^{2} lab, 1,300+ servers, ~10PB storage dedicated to HPC in collaboration with the community

Zenith

- TOP500-class system based on Intel Scalable Systems Framework (OPA, KNL, Xeon, OpenHPC)
- 424 nodes dual Intel Xeon Gold processors, Omni-Path fabric.
- +160 Intel Xeon Phi (KNL) servers.
- **Over 1 PF combined performance!**
- #265 on Top500 June 2018, 1.86 PF theoretical peak
- Lustre, Isilon H600, Isilon F800 and NSS storage
- Liquid cooled and air cooled

Rattler

- Research/development system with Mellanox, NVIDIA and Bright Computing
- 88 nodes with EDR InfiniBand and Intel Xeon Gold processors
- 32x PowerEdge C4140 nodes with 4x NVIDIA GPUs

Other systems

- 32 node AMD cluster, storage solutions, etc.

intel DELL EMC

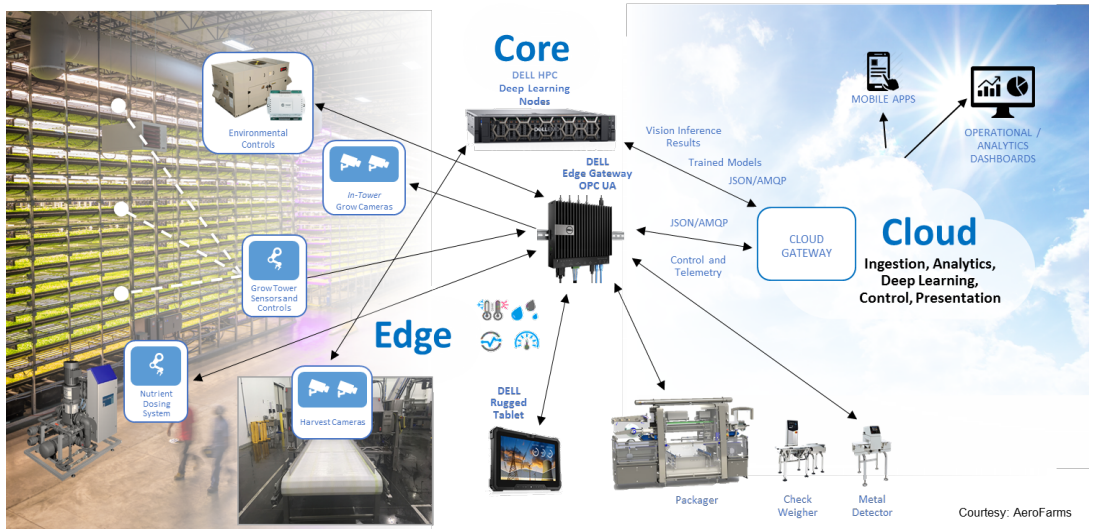
A Simple Goal

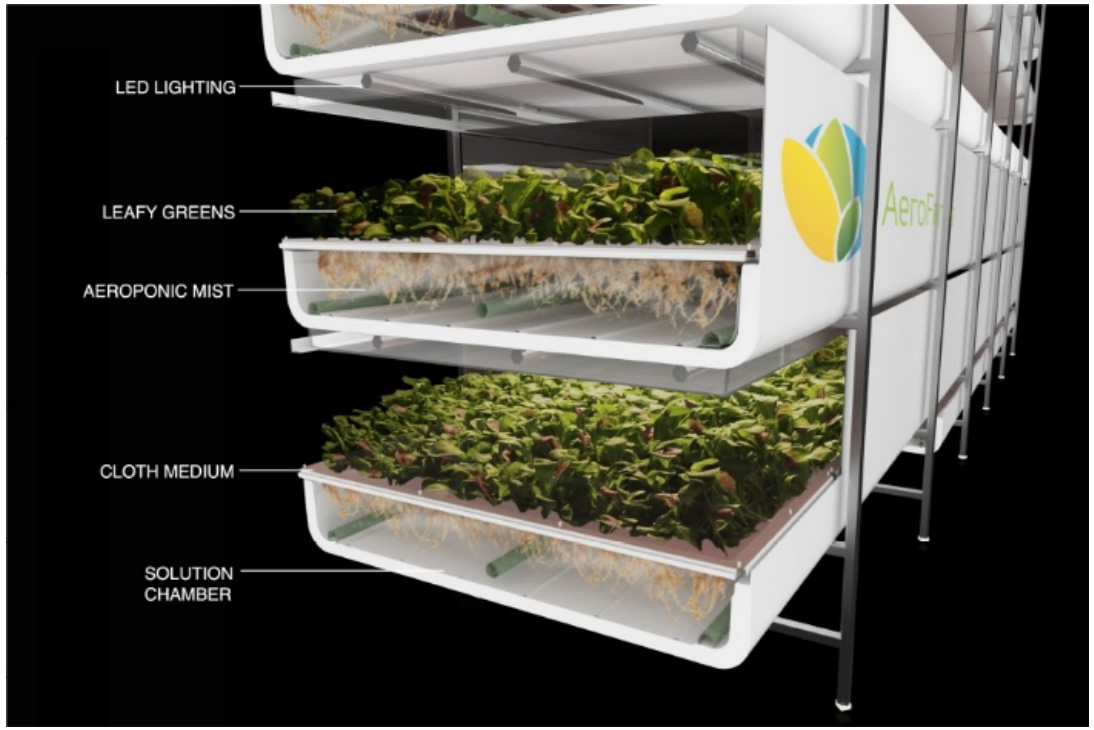
- Maximize performance on a single node
- Scale out performance on multiple nodes
- Enable new science with Xeon & TensorFlow (and more!)

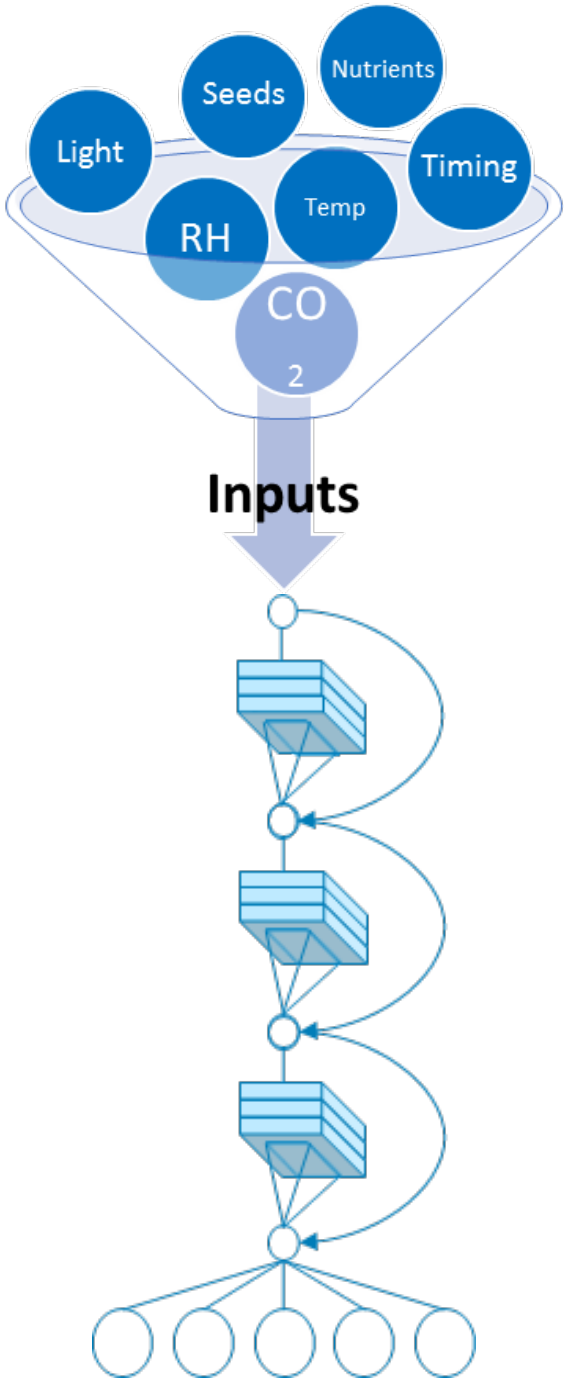
Plants x Data Science:
Phenomics and the Future of Indoor Agriculture

Democratising food and farming


AeroFarms®
Farming Locally. Globally.
Science goes digital. Farming goes vertical.
intel
DELL EMC








95%
Less
Water
390x
More
Productivity
Pesticide
Free
Produce
intel
DELL EMC

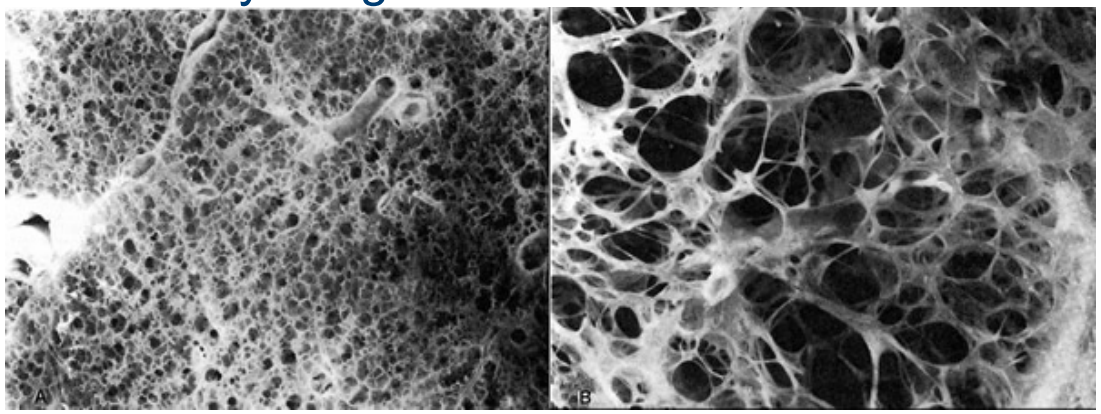
Building a Digital Radiologist, using Dell EMC Deep Learning Techniques

The Importance of Early Detection

- Emphysema is estimated to affect more than 3 million people in the U.S.1, and more than 65 million people worldwide2.
  - Severe emphysema (types 3 / 4) are life threatening
    - Early detection is important to try to halt progression
- Pneumonia affects more than 1 million people each year in the U.S.3, and more than 450 million4 each year worldwide.
  - 1.4 million deaths per year worldwide
    - Treatable with early detection

Healthy Lung

Severe Emphysema



https://www.ctsnet.org/article/airway-bypass-stenting-severe-emphysema

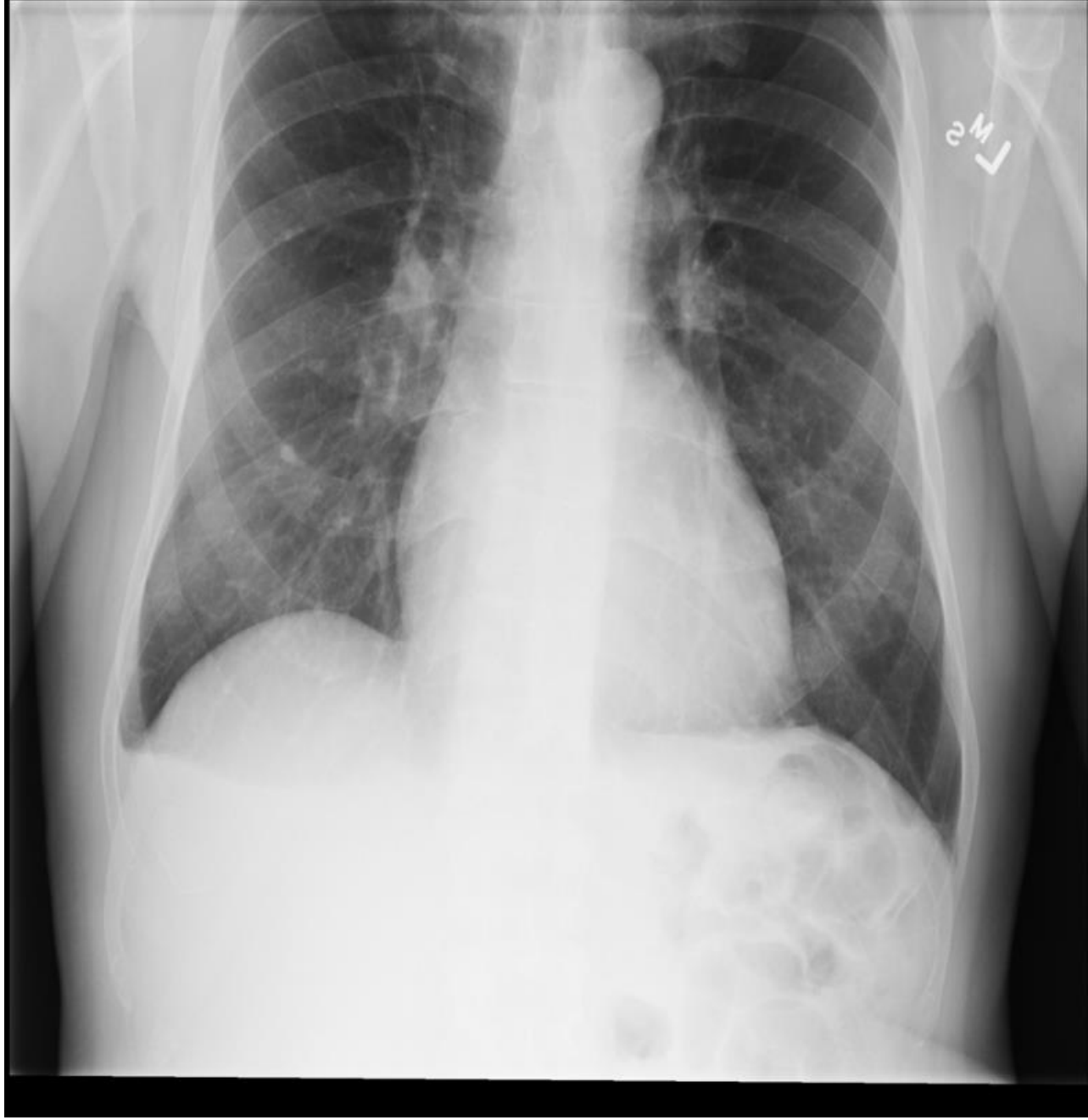
1. www.emphysemafoundation.org/index.php/the-lung/copd-emphysema
2. http://www.who.int/respiratory/copd/burden/en/
3. https://www.cdc.gov/features/pneumonia/index.html
4. https://doi.org/10.1016%2FS0140-6736%2810%2961459-6

CheXNet

Developed at Stanford University, CheXNet is a model for identifying thoracic pathologies from the NIH ChestXray14 dataset

- DenseNet121 topology
  - Pretrained on ImageNet
- Dataset contains 121K images
  - Multicategory / Multilabel
  - Unbalanced

https://stanfordmlgroup.github.io/projects/chexnet/



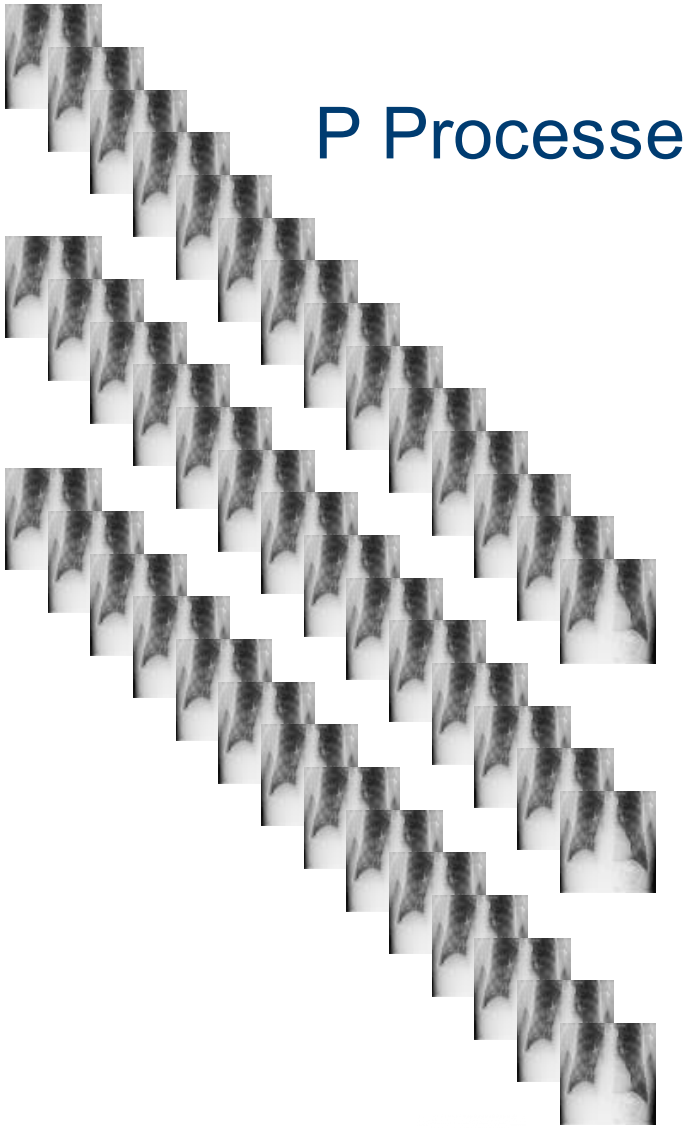
Parallelizing CheXNet

Parallelizing CheXNet

To Parallelize:

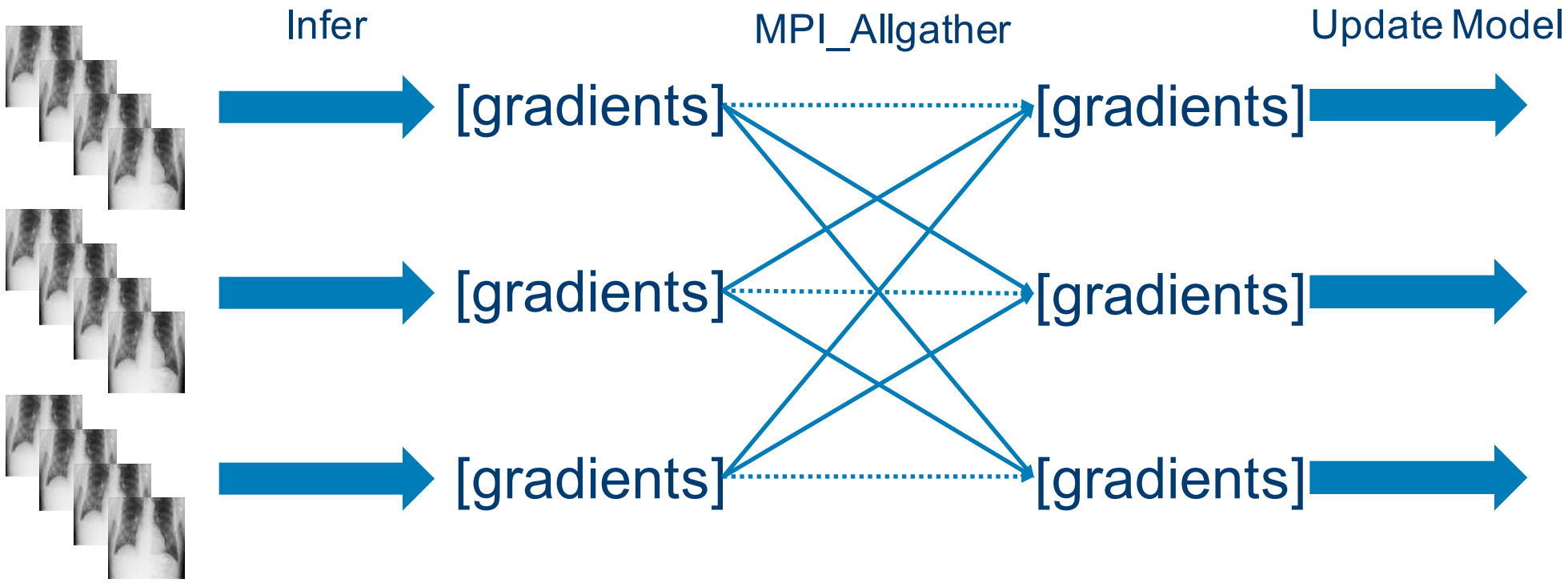
1. Each Process Has Access to Entire Data Set
2. Each Process Independently Shuffles Data Set
    a. Independent Random Seeds
3. Each Process Trains on the First N/P Images
4. Repeat Steps 2 + 3 Every Epoch

Training Data Set
(N Images)

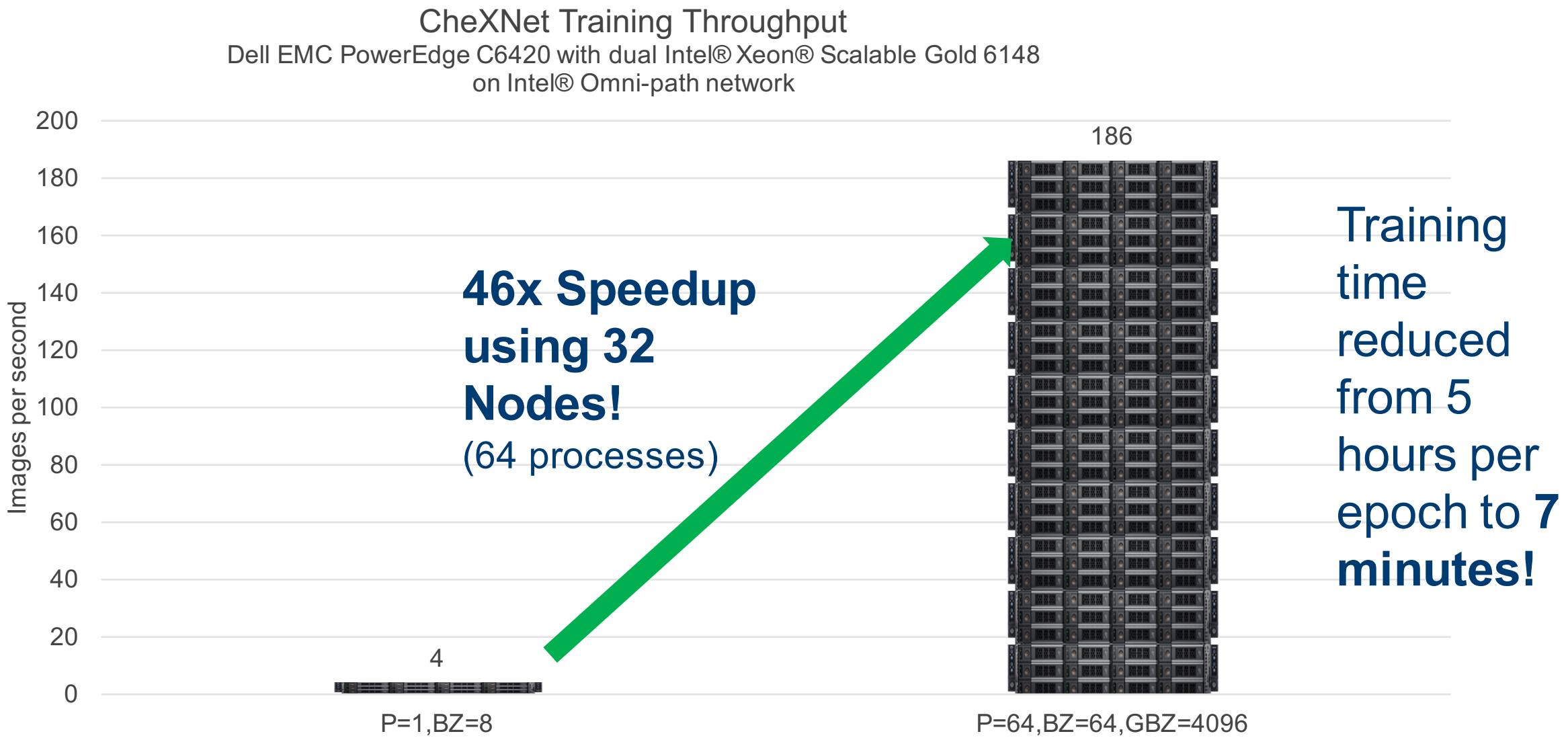


P Processes

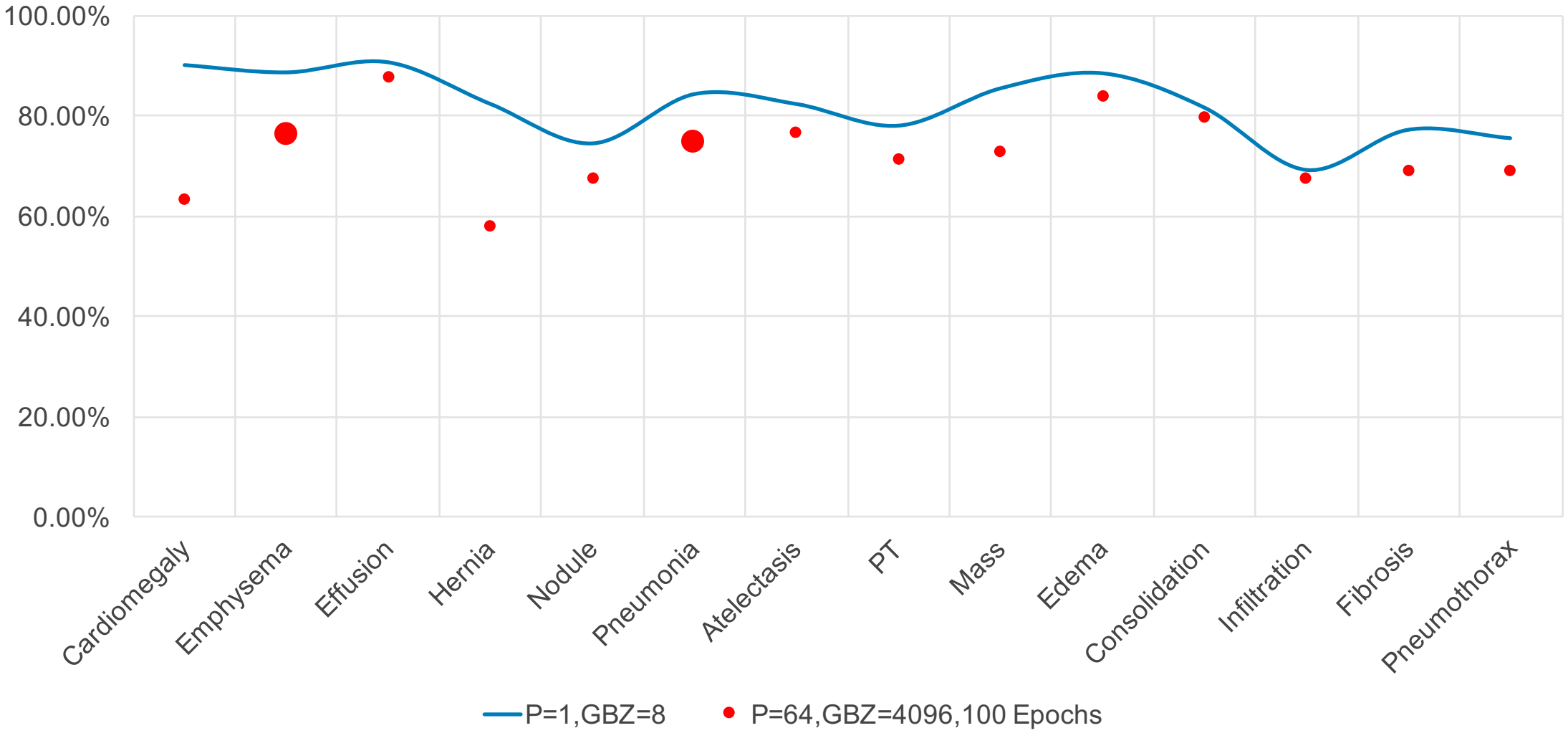
Parallelizing CheXNet



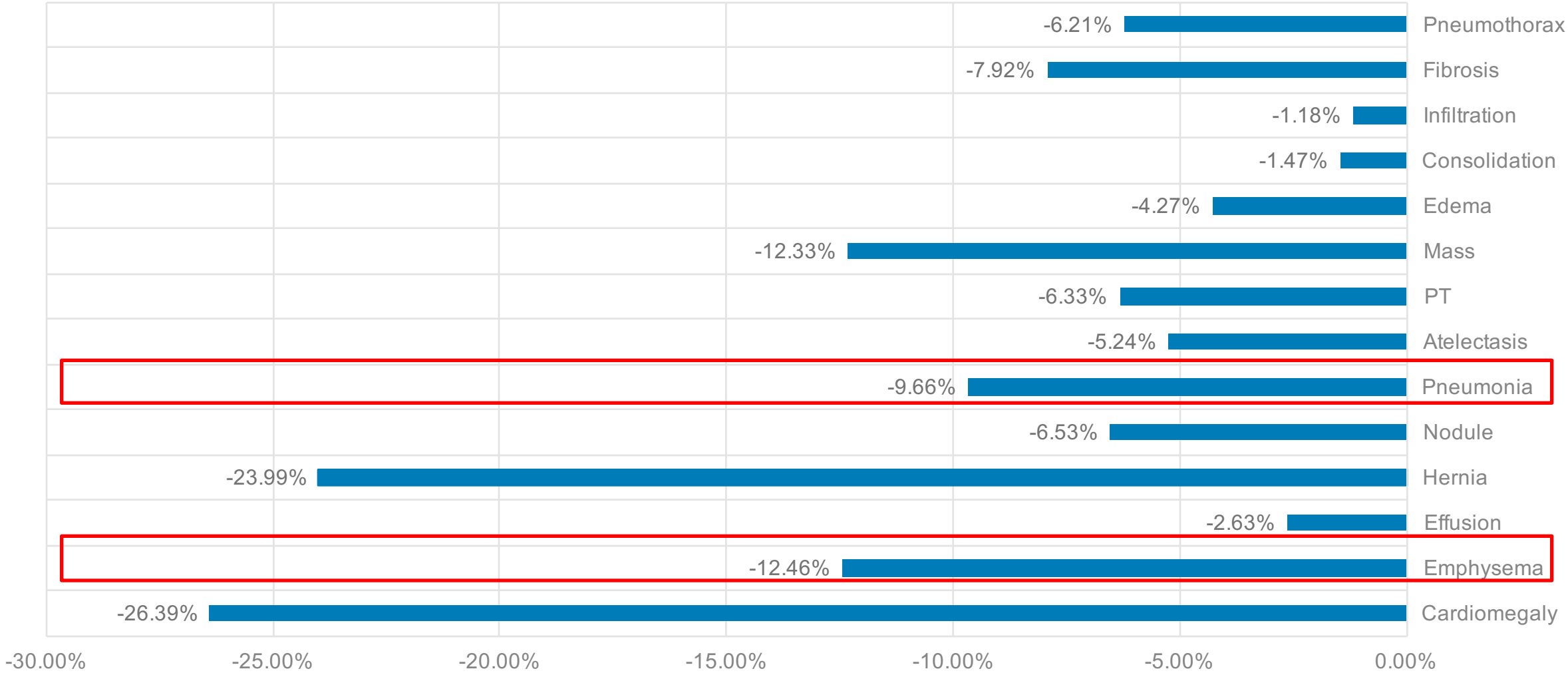
CheXNet – Parallel Speedup



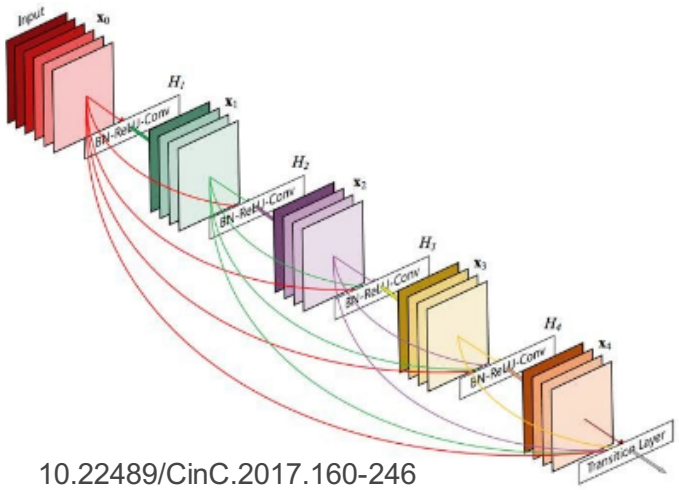
Parallelizing CheXNet - Accuracy



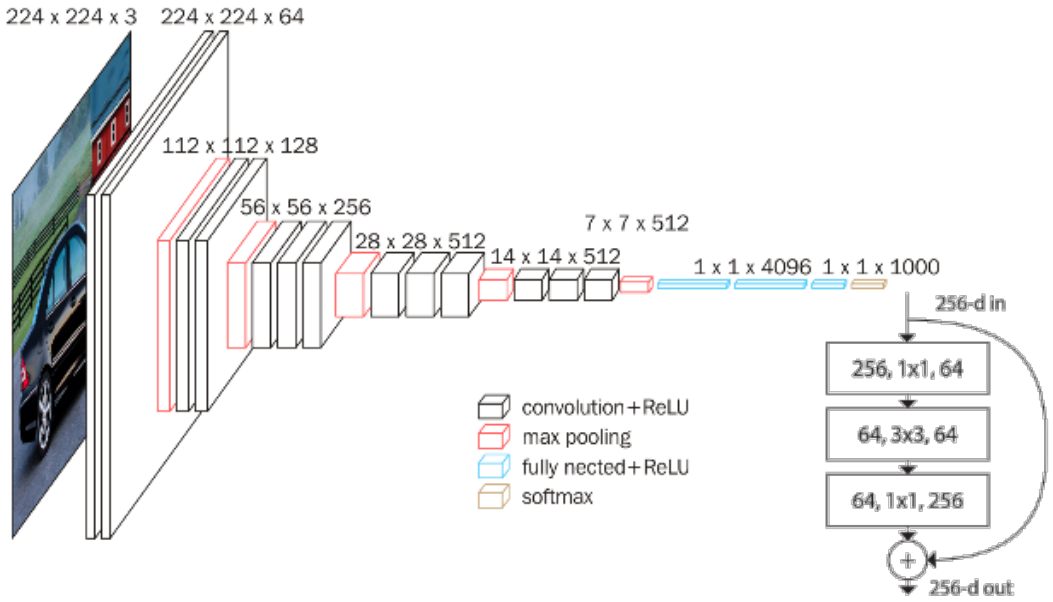
Parallelizing CheXNet – Accuracy Relative to single-process



Can We Do Better?



10.22489/CinC.2017.160-246



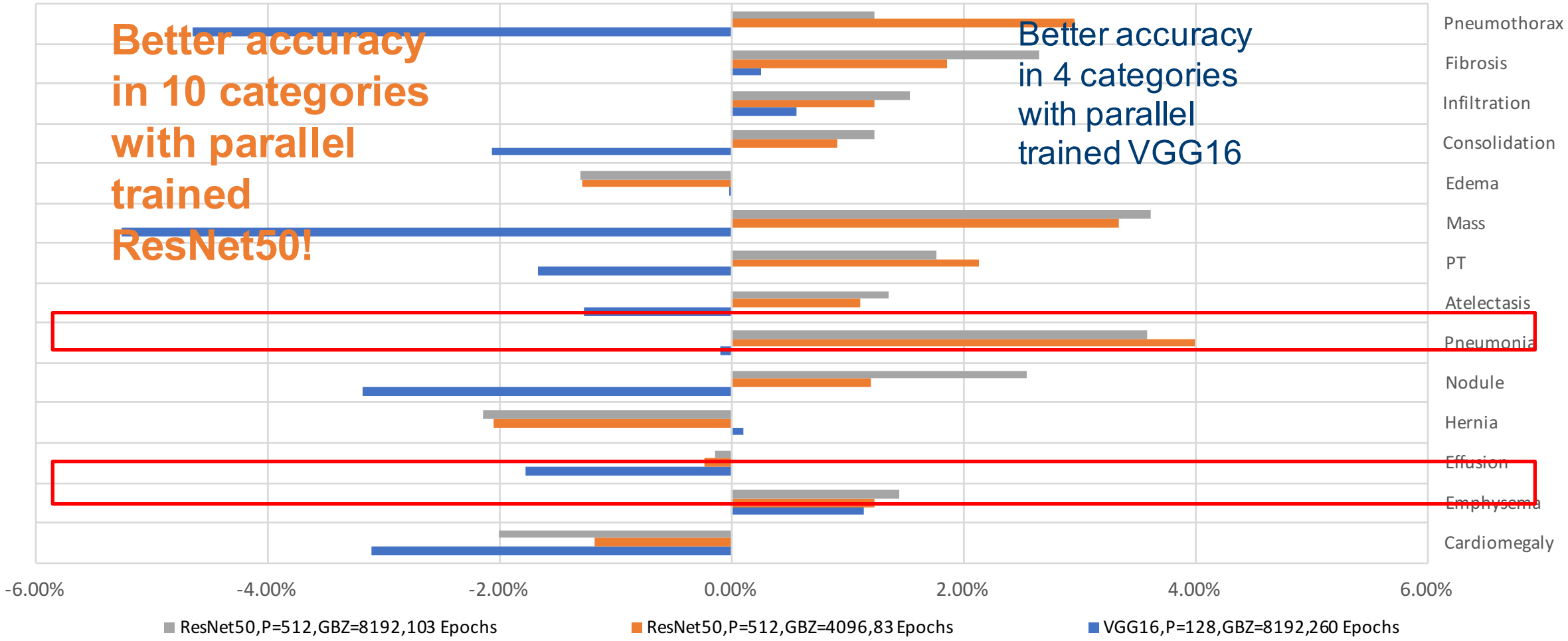
DenseNet is a relatively new topology, with lots of **batch normalization**

- Increasing batch size decreases accuracy, regardless of number of processes/nodes

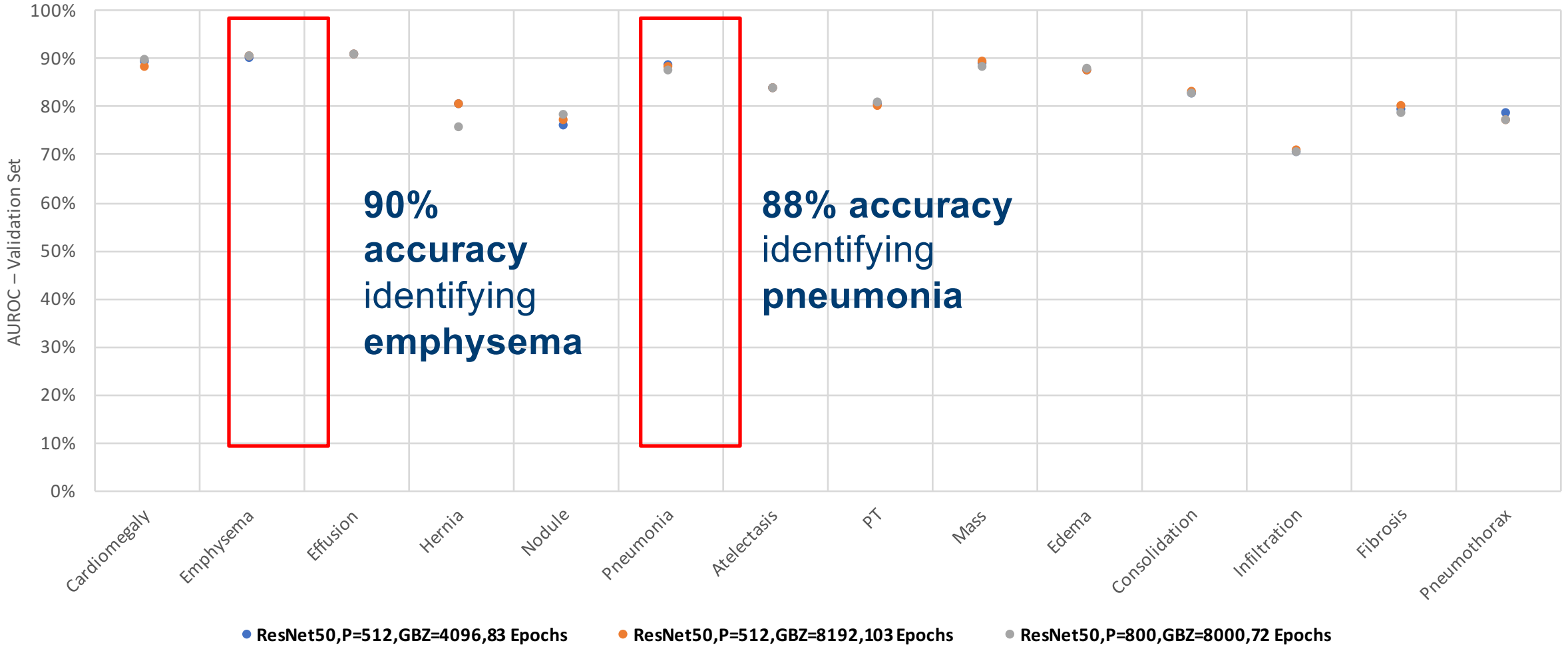
VGG has no batch normalization, and ResNet less batch normalization than DesnseNet

Why not try another topology?

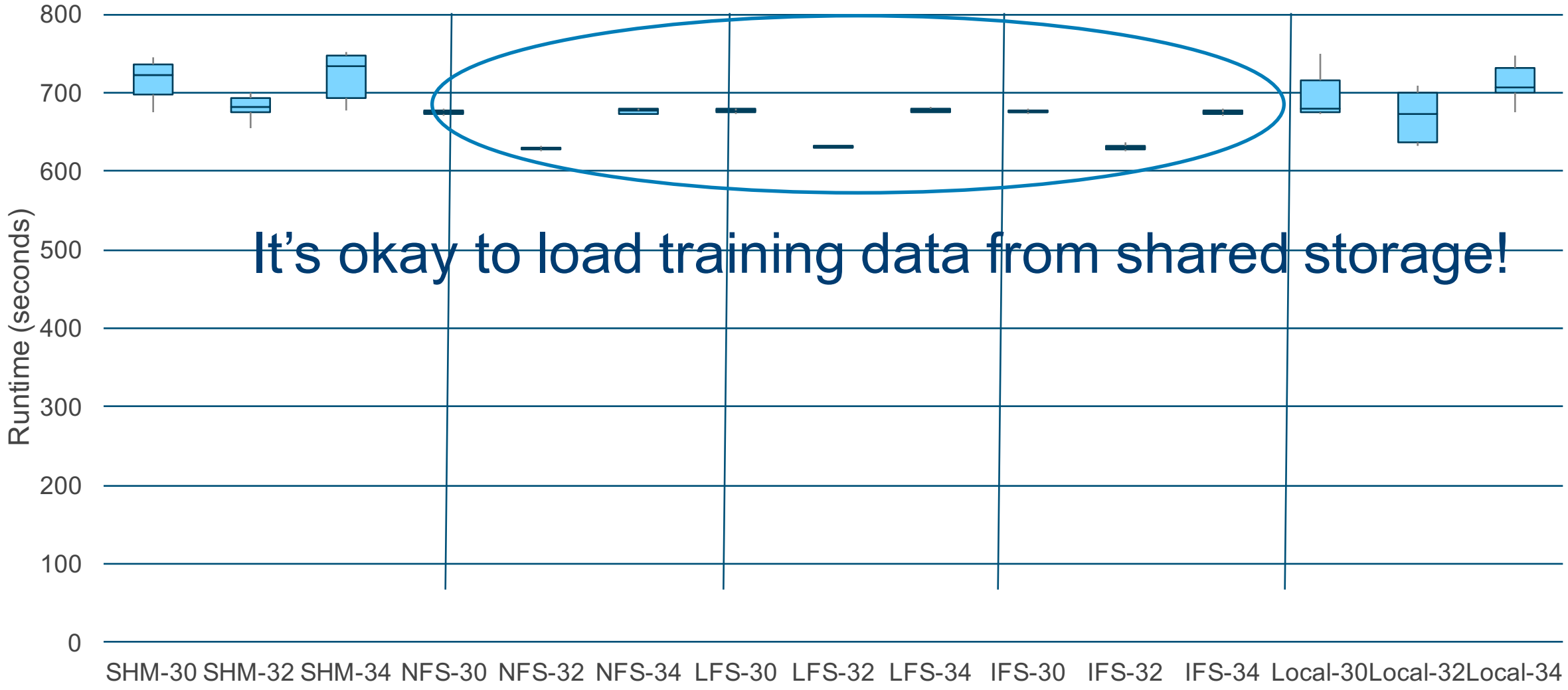
Accuracy of VGG16 and ResNet50 relative to DenseNet121



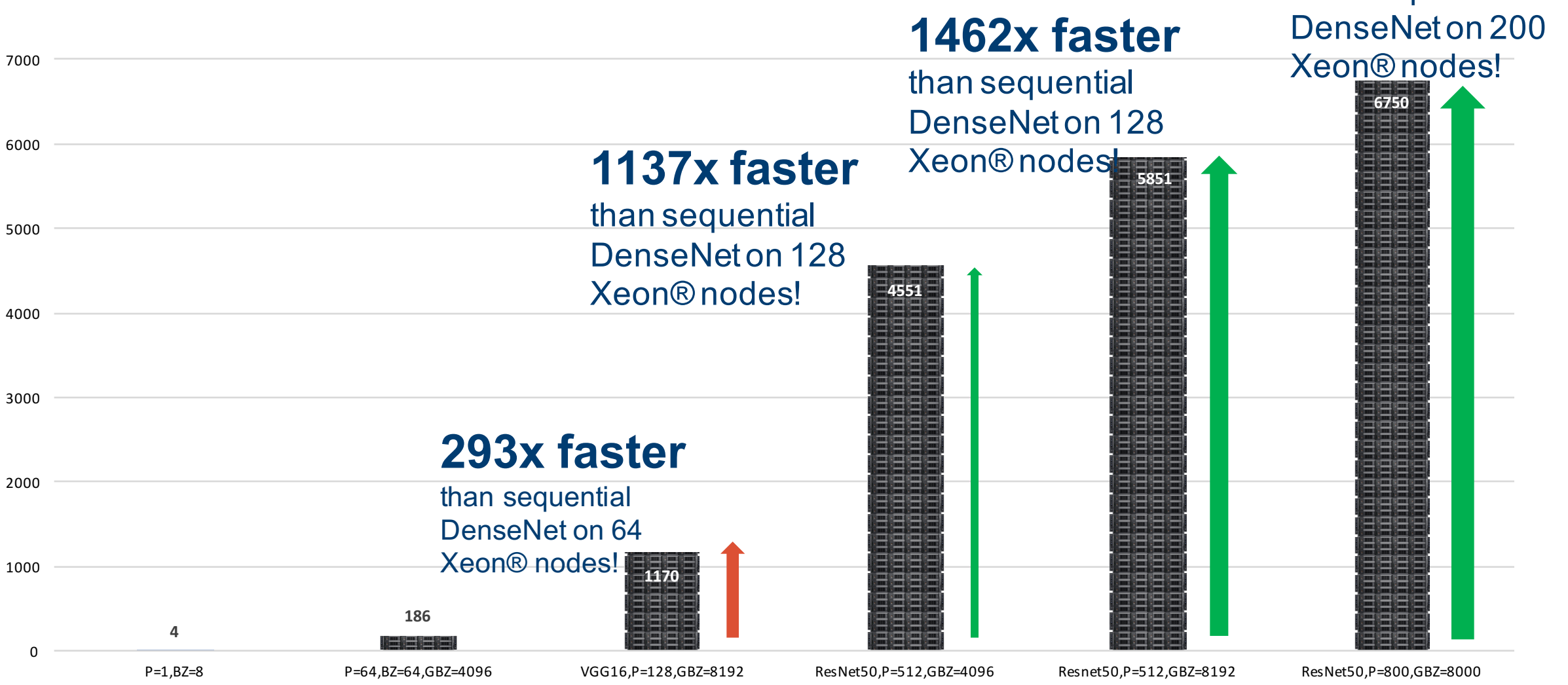
Categorical Accuracy of ResNet50-based AI Radiologist



Training CheXNet – Storing Training Data



Training Throughput with VGG and ResNet



Dell EMC Poweredge C6420 with dual Intel® Xeon® Scalable Gold 6148 on Intel ® Omni-Path network. ResNet50 tests performed with TensorFlow+Horovod

The Power to Do More

Benefits

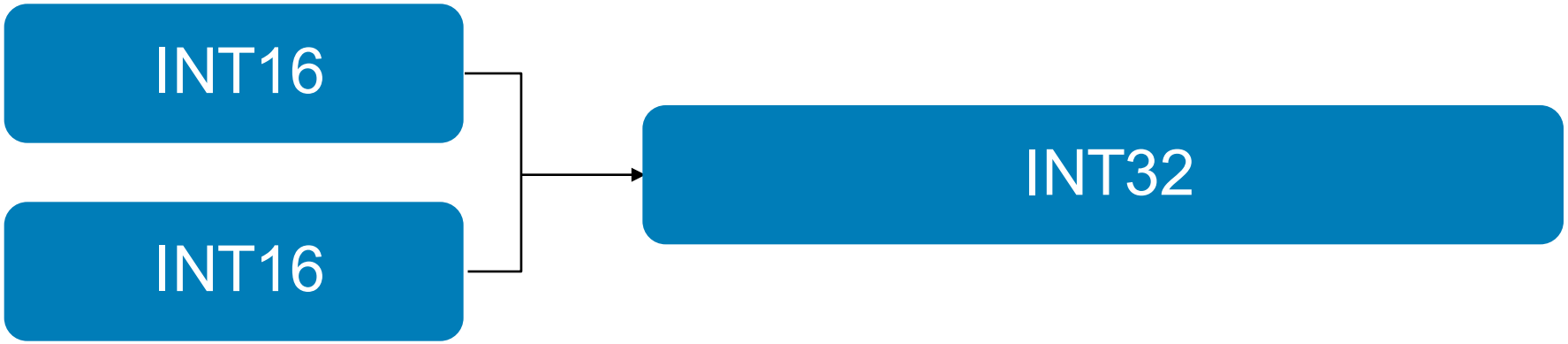
1. Delivers **faster time to deployment** with and end-to-end platform

2. Design a modular building block for high content screening, using Intel Xeon Scalable Processor and Intel Omni-Path Architecture

3. Achieve 1688x greater deep learning training performance with the Intel optimized version of TensorFlow + Horovod

Downstream - VNNI

Intel's next-generation "Cascade Lake" Xeon CPUs will feature Vector Neural Network Instructions (VNNI), which will provide reduced-precision vector operations in silicon.



The AI Engineering team will be investigating performance/accuracy tradeoffs on our existing models, including the AI Radiologist, when VNNI-enabled silicon is available for testing.

THE DELL EMC HPC VISION

"We put tools in the hands of people who are going to solve the world's biggest problems."

Michael Dell

We will achieve this global impact by building expertise to design, develop, deploy, and support the most capable, cost-effective portfolio of HPC solutions that integrate Dell EMC innovations with community standards.

Michael Dell at CERN


DELL EMC