

# TRAINING MULTISCALE CNN FOR LARGE MICROSCOPY IMAGE CLASSIFICATION IN ONE HOUR

**Kushal Datta, Intel® Artificial Intelligence Products Group**

Co-Authors: Imtiaz Hossain<sup>2</sup>, Sun Choi<sup>1</sup>, Vikram Salelore<sup>1</sup>, Kyle Ambert<sup>1</sup>,  
William Godinez<sup>2</sup>, Xian Zhang<sup>2</sup>

1. Intel® Artificial Intelligence Products Group
2. Novartis Institutes for Biomedical Research



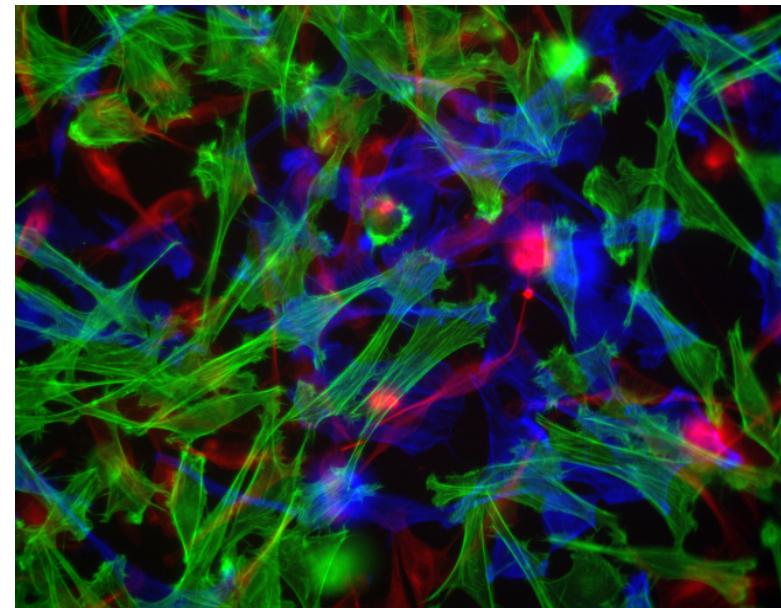
# MOTIVATION

**26x**

## ImageNet

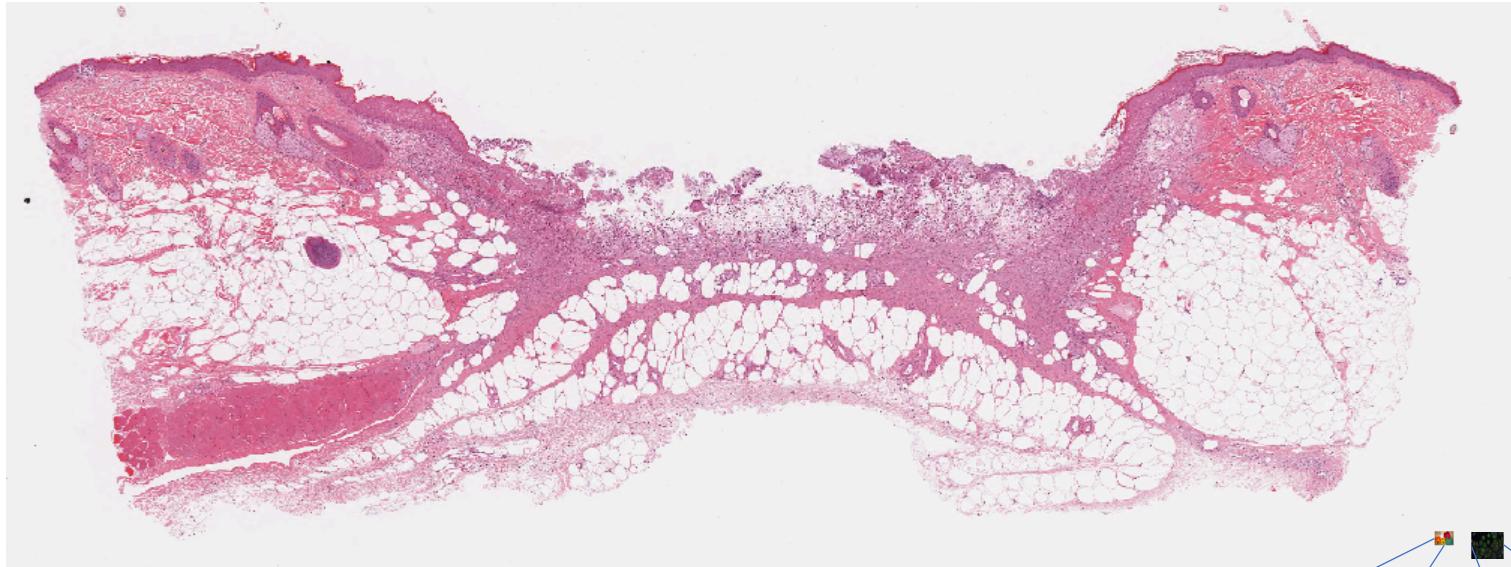


**224x224x3**

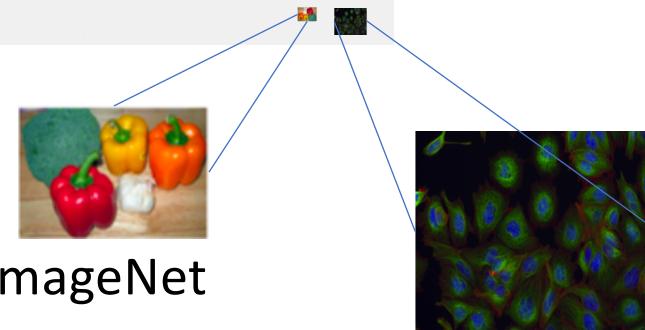


**1024x1280x3**

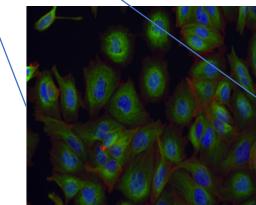
# IMAGES IN BIOMEDICINE



Typical histology slide image at 40X

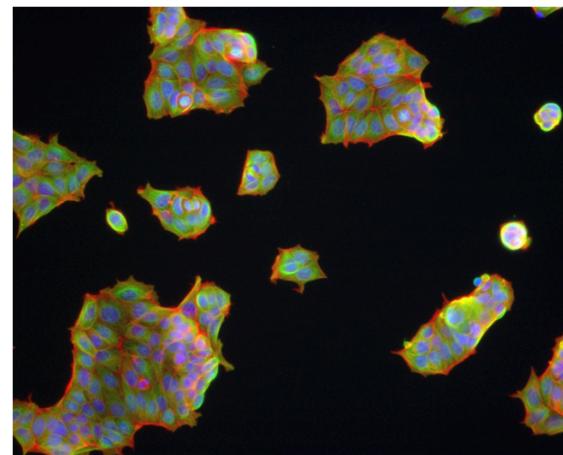
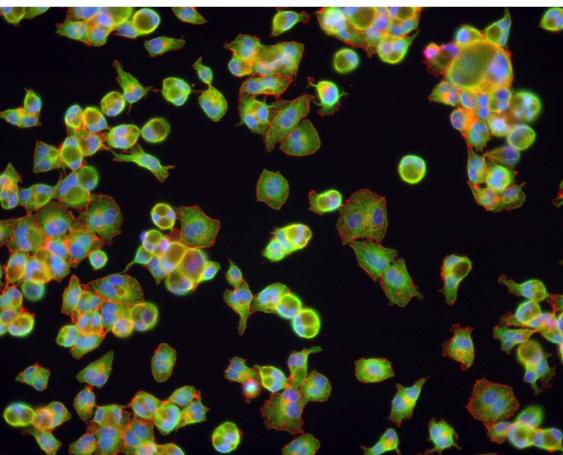
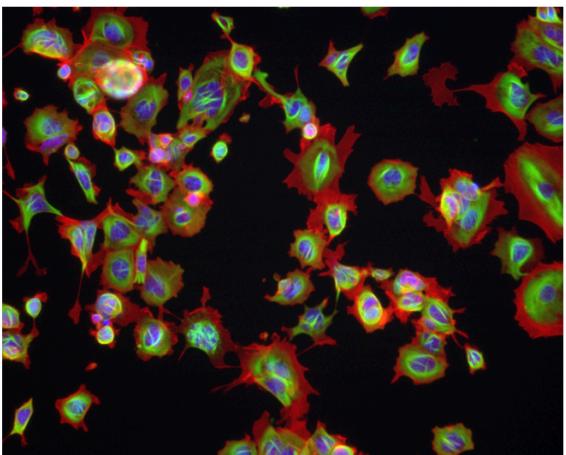
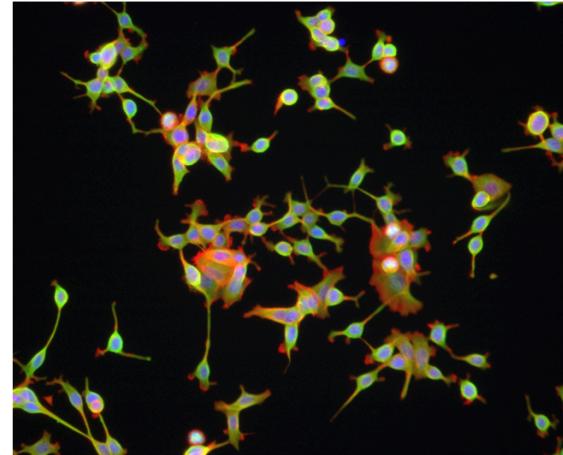
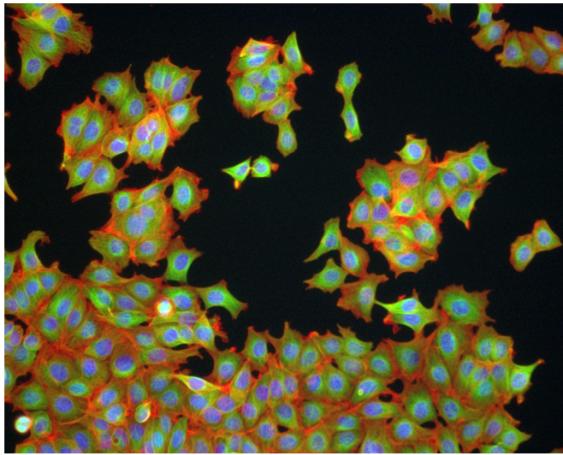
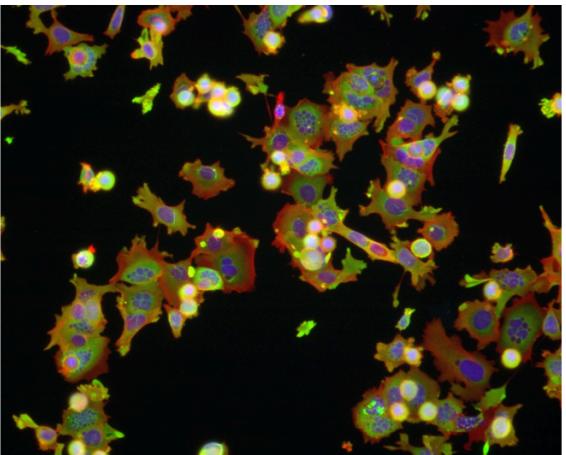


ImageNet



Cellular Microscopy

# Context is valuable

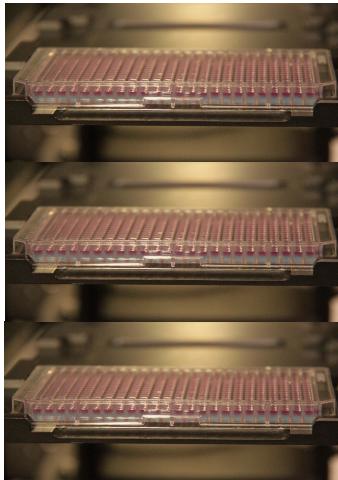


# HIGH-CONTENT SCREENING (HCS) IN DRUG DISCOVERY



Cellular Experiment

Multiple experiments and lead discovery projects

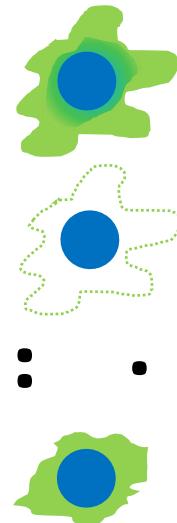


Miniaturized High Throughput Screening setup

1000s to >1million compound treatments per experiment

Treatment A  
Treatment B  
⋮  
Treatment Z

Systematic perturbations



Phenotypes



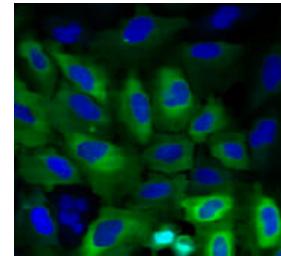
Microscopy Images

Multiple channels per image (different subcellular components)  
Typically 2560x2160 images, 2-4 channels  
Additional multipliers: Dose variations, Time-dependent and 3D measurements.

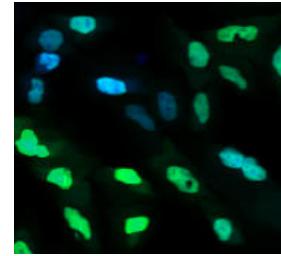
# HIGH-CONTENT SCREENING (HCS) IN DRUG DISCOVERY



Cellular Experiments



Positive control



Negative control



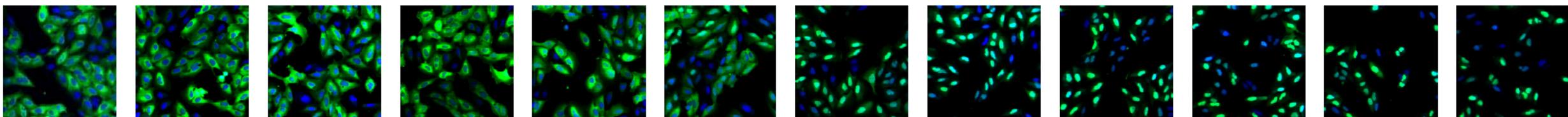
Negative control: Wild type



Positive control: Fat

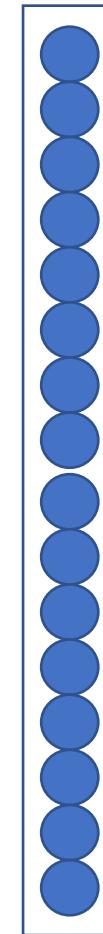
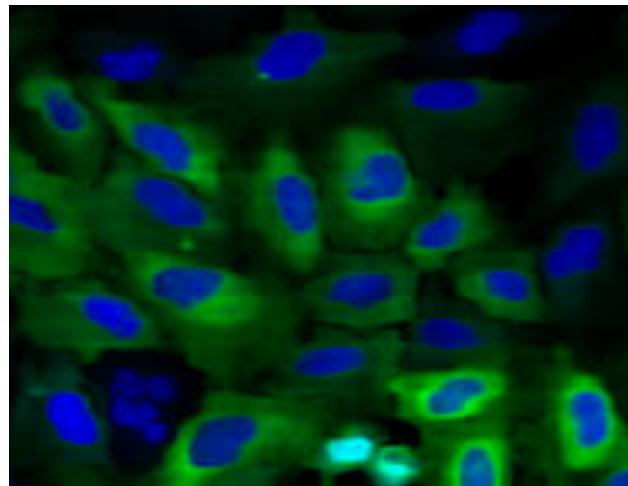
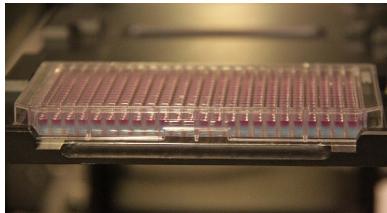
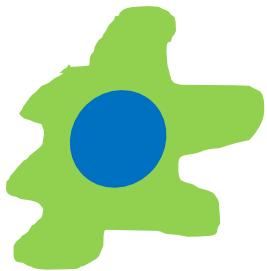
BBBC013: Cytoplasm to nucleus translocation of the Forkhead (FKHR-EGFP) fusion protein in stably transfected human osteosarcoma cells, U2OS

BBBC011: RNA interference screen for genes affecting fat mass in the *C.elegans*. The animals are stained with oil red O, a non-fluorescent dye



Example from BBBC013 dataset: compound concentration dependent phenotype change

# HIGH-CONTENT SCREENING (HCS) IN DRUG DISCOVERY



## High Content Feature Set

### Morphometric features

- Nucleus/Cell shape and size
- Position and orientation
- Geometric properties
- Contour – Curvature and Fourier descriptors
- Invariant moment

### Densitometric features expressing total intensity

- Intensity features from different regions
- Invariant moments from the extinction image

### Textural features

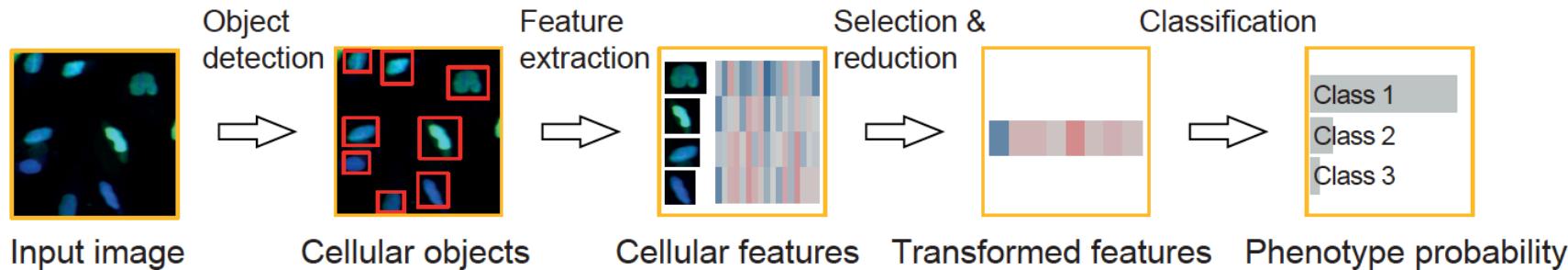
- Gradient/Laplace/Flat texture image features
- Topological gradients and rice fields

### Others

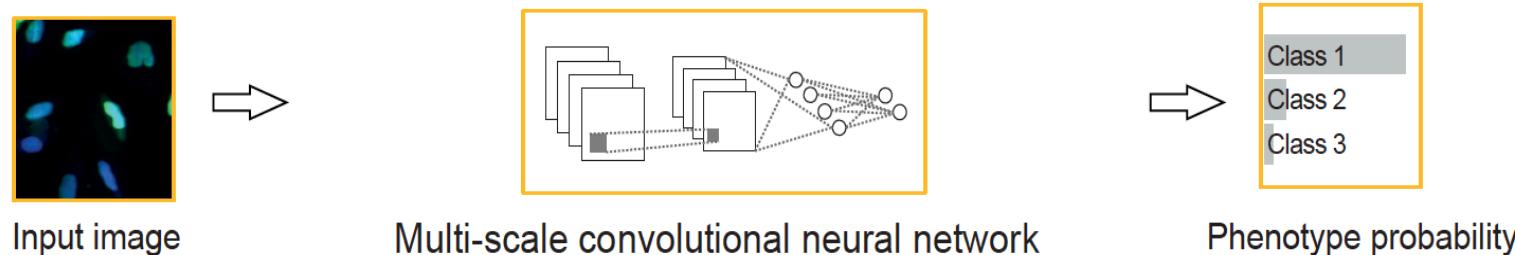
Orlov, Nikita et al. "WND-CHARM: Multi-Purpose Image Classification Using Compound Image Transforms.", *Pattern recognition letters*, 2008  
Rodenacker K, Bengtsson E. A Feature Set for Cytometry on Digitized Microscopic Images. *Analytical Cellular Pathology : the Journal of the European Society for Analytical Cellular Pathology*, 2003  
Carpenter, Anne E et al. "CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes." *Genome Biology*, 2006

# CONVENTIONAL VS NEURAL NETWORK IN HCS

## Conventional HCS analysis pipeline

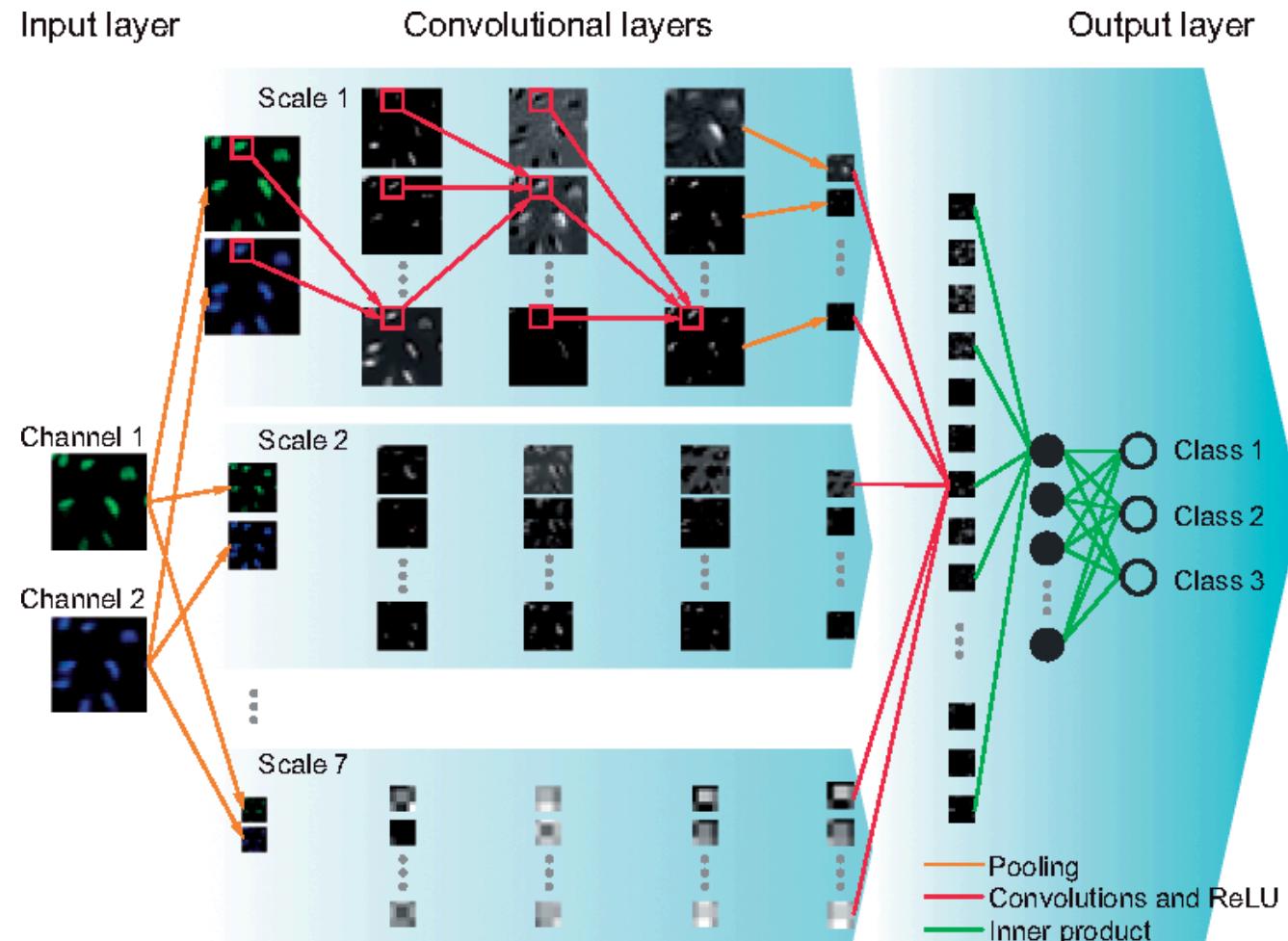


## HCS analysis using Multi-Scale Convolutional Network<sup>1</sup>



<sup>1</sup> Godinez et al, A multi-scale convolutional neural network for phenotyping high-content cellular images. Bioinformatics, 2017

# MULTI-SCALE CNNS

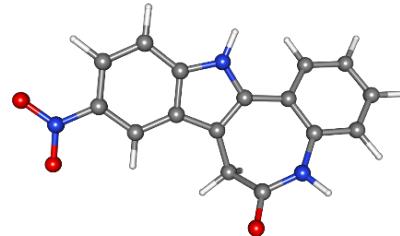


Godinez et al, A multi-scale convolutional neural network for phenotyping high-content cellular images. Bioinformatics, 2017

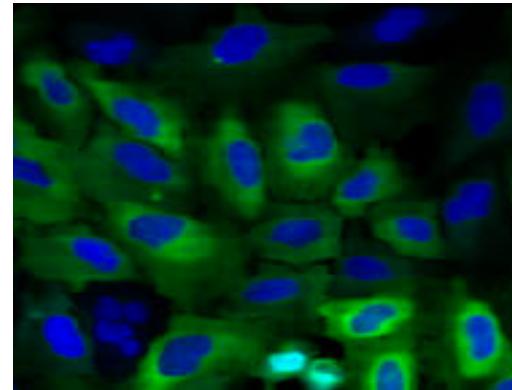
Scalable Data Analytics in Scientific Computing, ISC 2019

# DATASET A

## BROAD BIOIMAGE BENCHMARK COLLECTION - 21



+ concentration



(Phenotype)

103 such compound concentration pairs

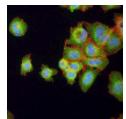
38 compounds in all

2-3 replicates per treatment

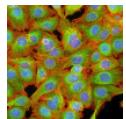
10,104 total images

**Label: Mechanism of Action**  
(visual inspection or from literature)

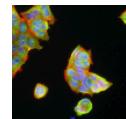
### Mechanisms of action (12 + neutral)



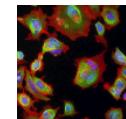
Actin  
Disruptors



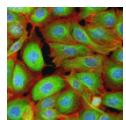
DNA  
Damage



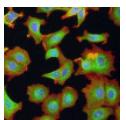
Epithelial



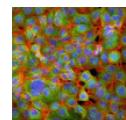
Microtubule-  
stabilizers



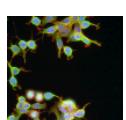
Aurora  
Kinase  
Inhibitor



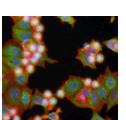
DNA  
Replication



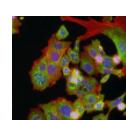
Kinase-  
inhibitors



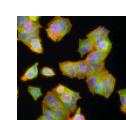
Cholesterol  
Lowering



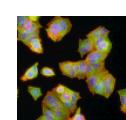
Eg5-inhibitors



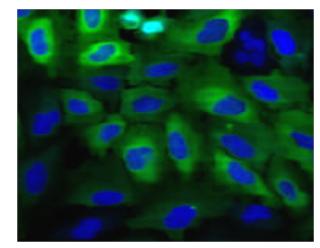
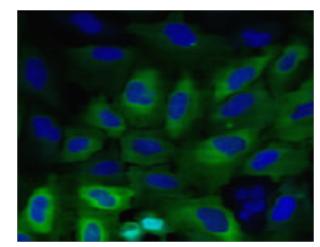
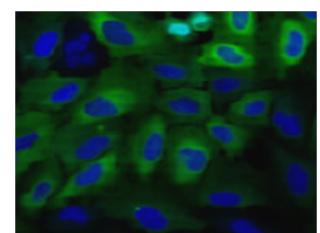
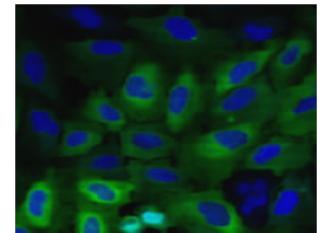
Microtubule-  
destabilizers



Protein-  
degradation



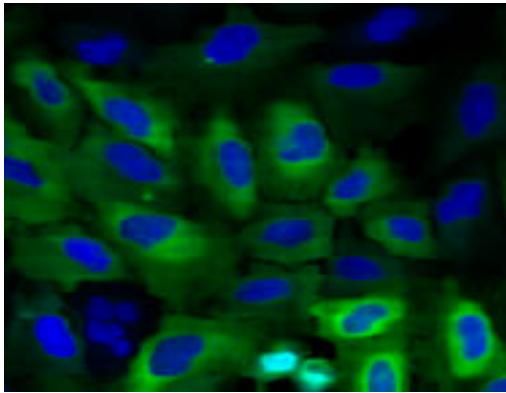
Protein-  
synthesis



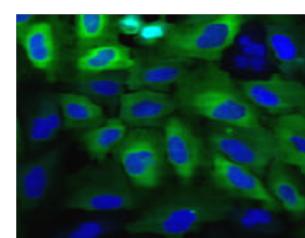
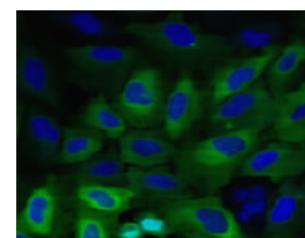
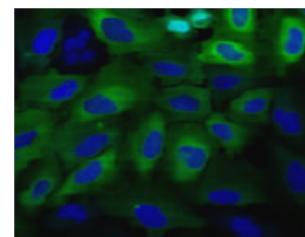
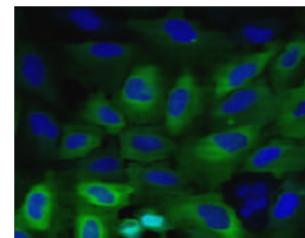
5 augmentations

# DATASET B

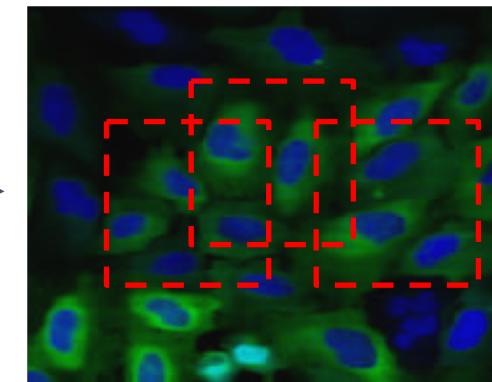
## BROAD BIOIMAGE BENCHMARK COLLECTION - 21



1024x1280x3



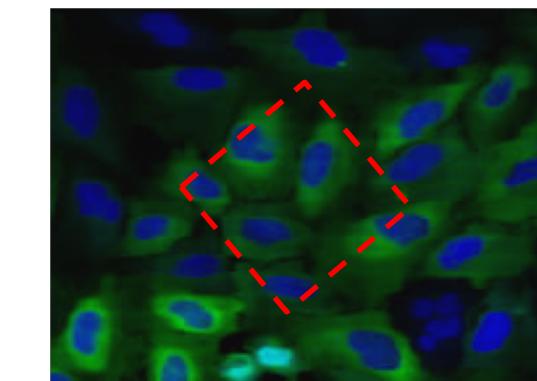
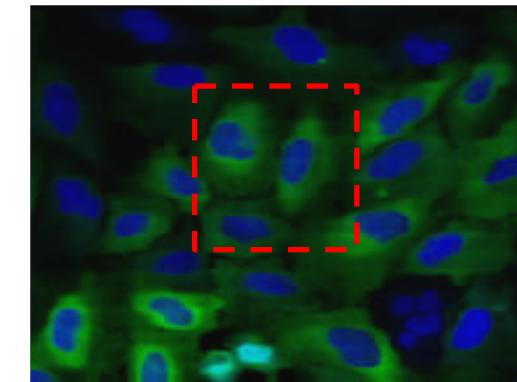
5 flips and mirroring



3 sub selections

724x724x3

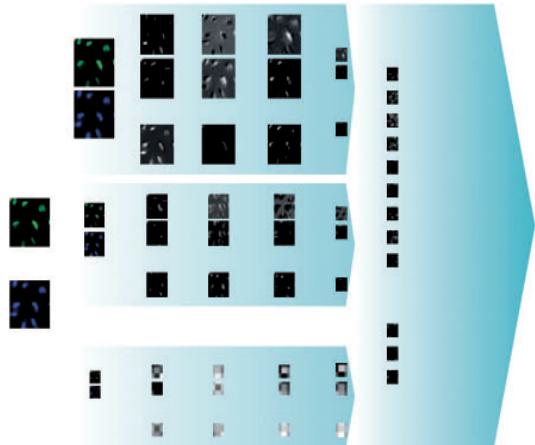
- 313,282 training images
- 35,306 validation images



15 degree rotated crops

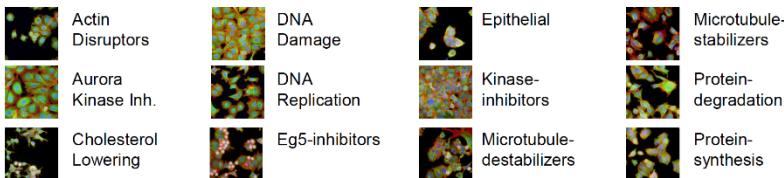
# INITIAL TRAINING TIMES

Intel® Xeon Phi™ 7290F  
processor, Caffe\* 2017

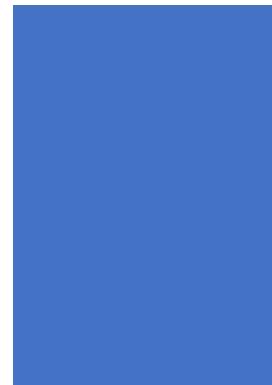


M-CNN

<b>Cell line</b>	Human breast cancer cells (MCF7)
<b>Modality</b>	Fluorescence
<b>Channels</b>	Actin (red) Tubulin (green), DNA (blue)
<b>Treatment</b>	113 compounds at 8 concentrations in triplicate
<b>Task</b>	<u>Predict mechanism-of-action (MoA) of compounds</u>
<b>Number of images</b>	13200
<b>Validation images</b>	1292 - 103 compound-concentration pairs with annotated MoA



Training Time: 11hrs  
(Batch Size: 8)



Broad Bioimage Benchmark  
Collection 021

1 Node

# M-CNN TRAINING TIME <1HR HELPS

**Better hyper-parameter exploration**

**Typical compound screens consist of millions of images and unknown mechanisms of action**

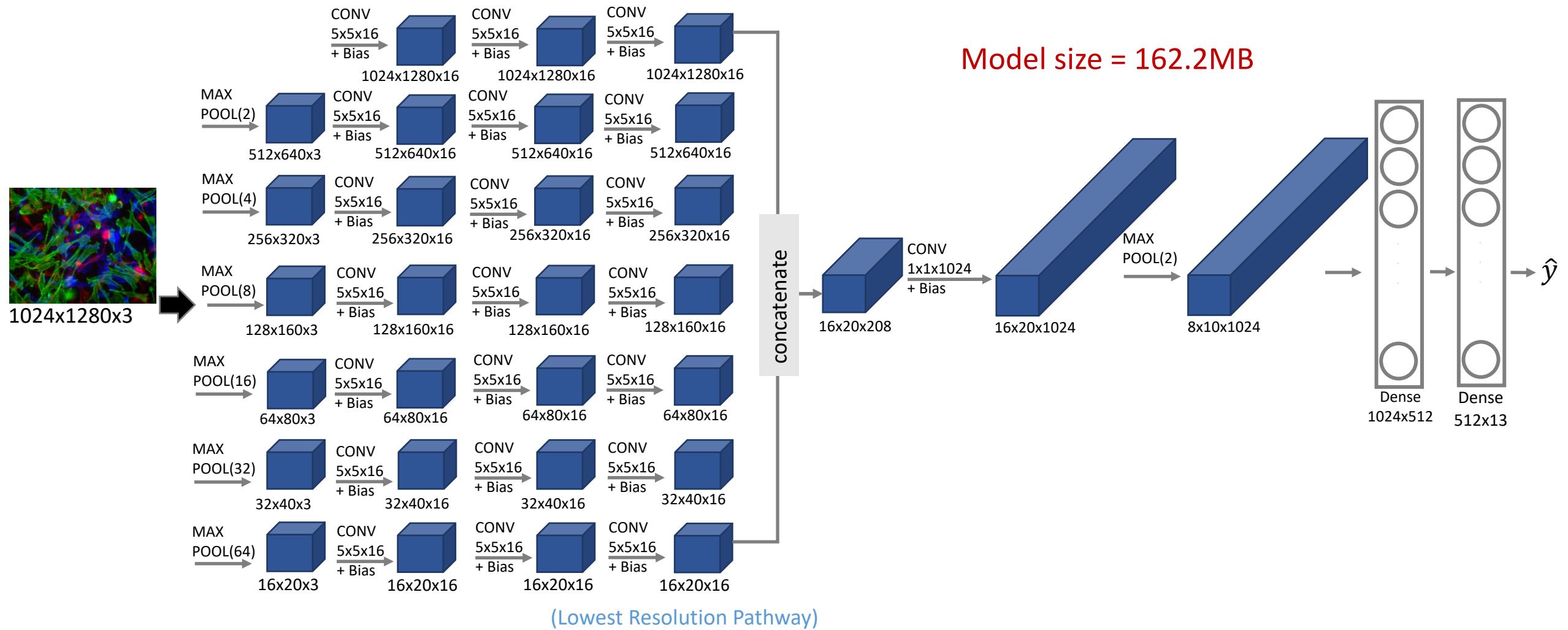
**Unsupervised learning with performant architectures enables exploration of discovering interesting compounds**

# STRATEGY

- IMPROVE MAXIMIZE PERFORMANCE ON A SINGLE NODE
- SCALEOUT PERFORMANCE ON MULTIPLE NODES
  - USE LARGE BATCHES
  - ADJUST LEARNING RATE

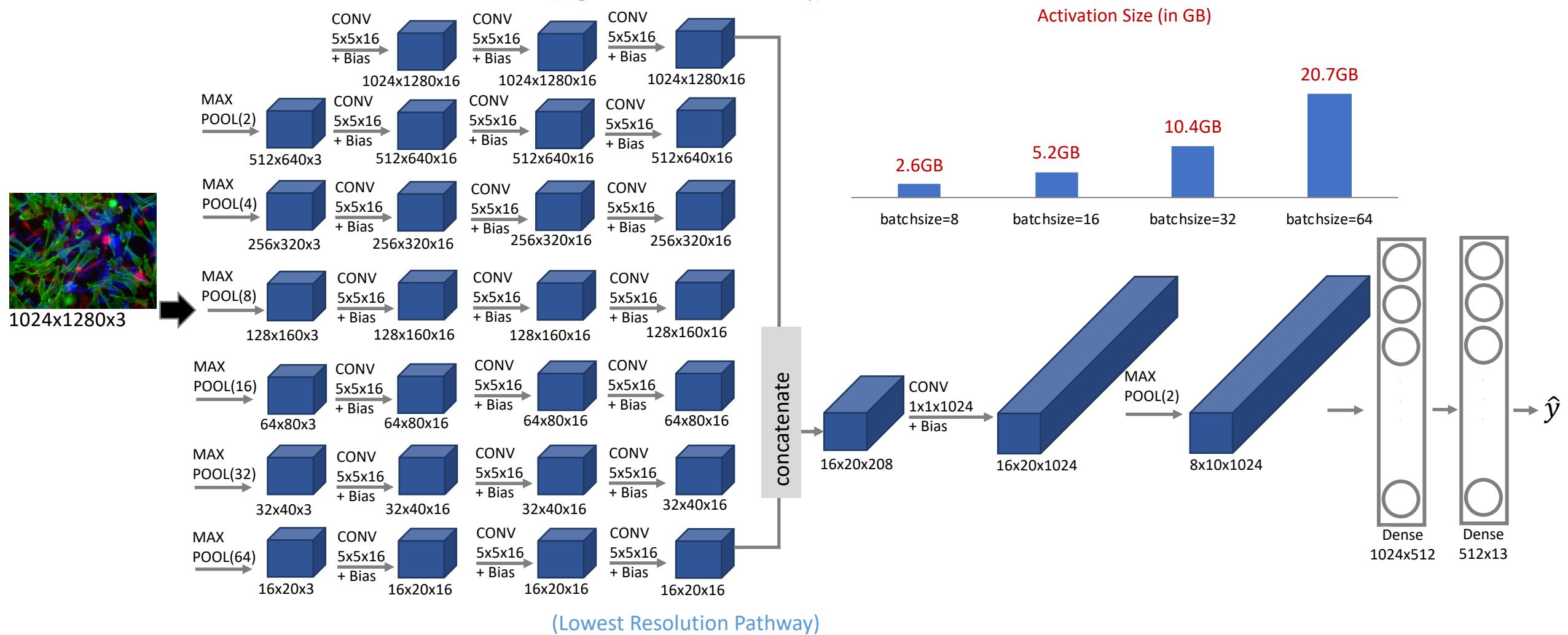
# OPERATIONS AND KERNELS IN M-CNN

(Highest Resolution Pathway)

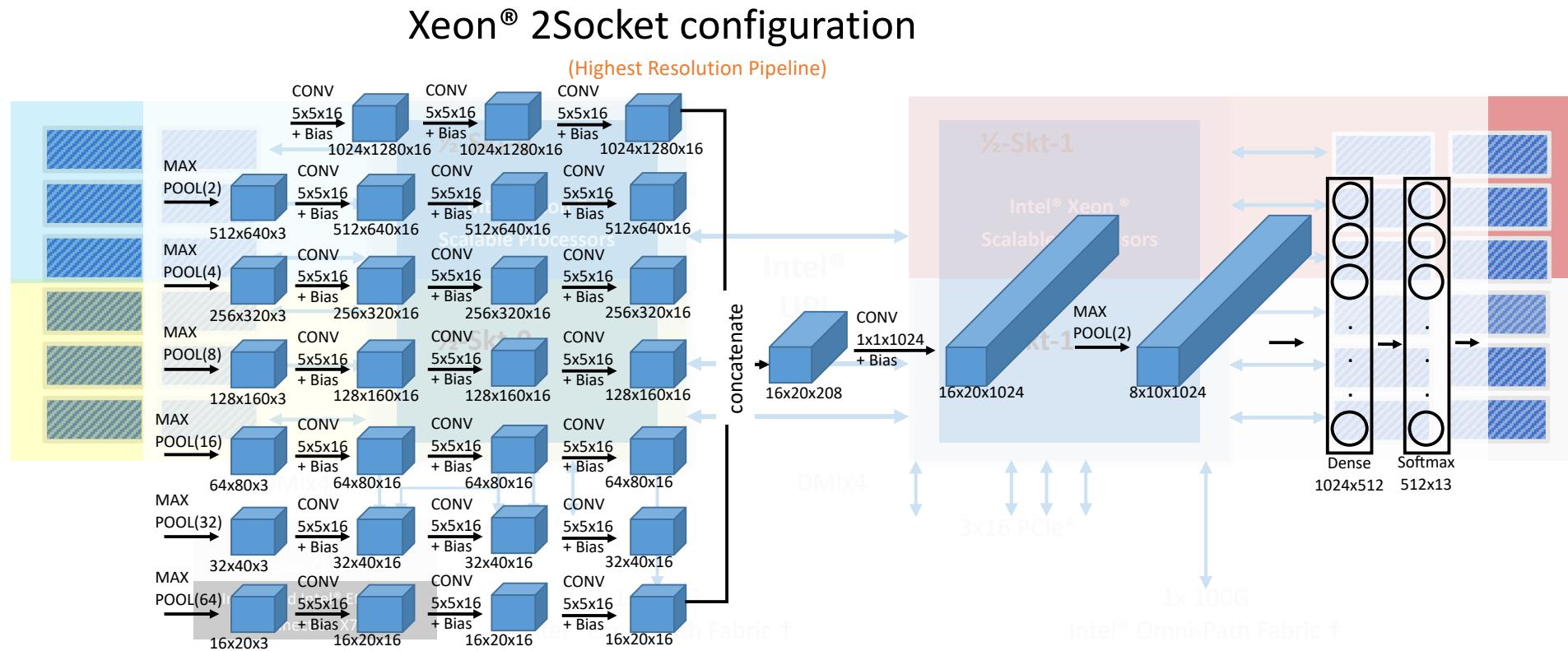


# SIZE OF ACTIVATION INCREASES LINEARLY WITH BATCH SIZE

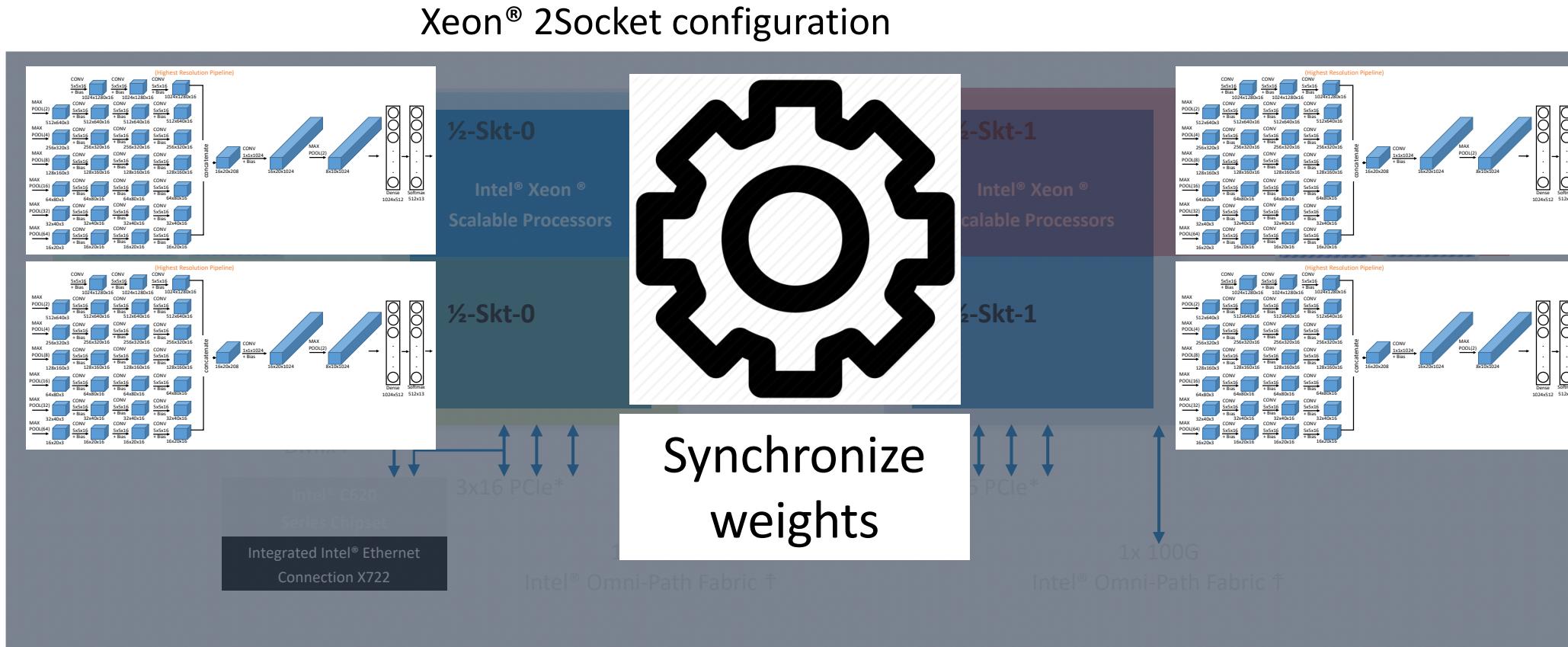
## (Highest Resolution Pathway)



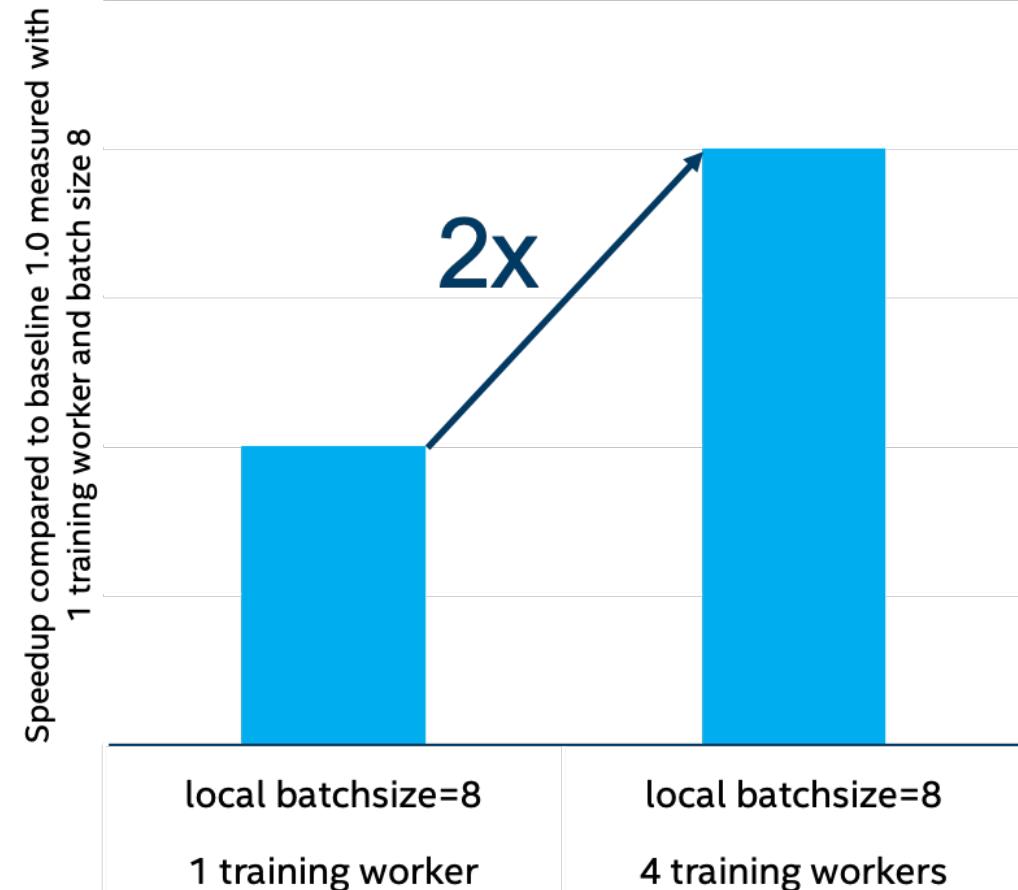
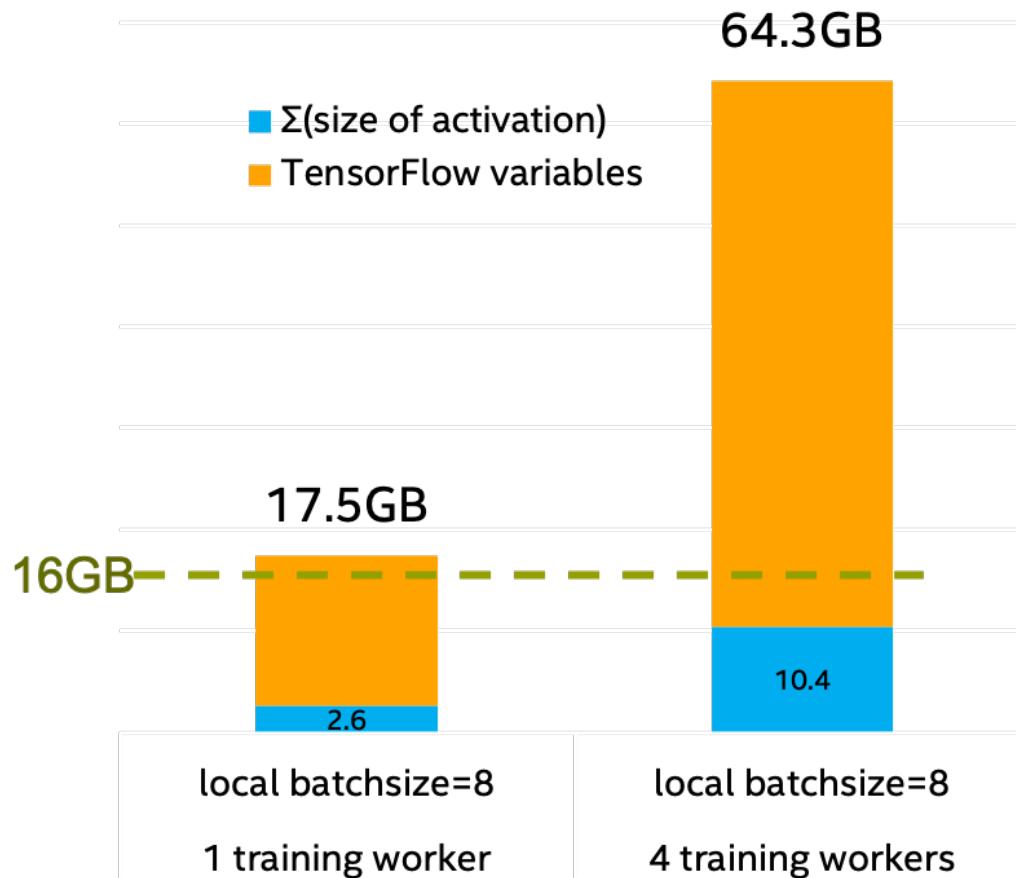
# SINGLE WORKER ON 2 SOCKET INTEL® XEON® GOLD 6148 PROCESSOR



# MULTI-WORKER ON NUMA CONFIGURATION IN TWO SOCKETS



# PERFORMANCE IMPROVED WITH 4 WORKERS 1-NODE 2S INTEL® XEON® SCALABLE GOLD PROCESSOR



TensorFlow: 1.7.0, Python: 2.7.5, Horovod: 0.12.1: OMP\_NUM\_THREADS=40/10

Performance results are based on testing as of dates shown in configuration and may not reflect all publicly available security updates. No product can be absolutely secure. See configuration disclosure for details. Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit: <http://www.intel.com/performance>

# SPEEDUP ON 1-NODE 2 SOCKET INTEL® XEON® SCALABLE GOLD PROCESSOR



# STRATEGY

- IMPROVE MAXIMIZE PERFORMANCE ON A SINGLE NODE
- SCALEOUT PERFORMANCE ON MULTIPLE NODES
  - USE LARGE BATCHES
  - ADJUST LEARNING RATE

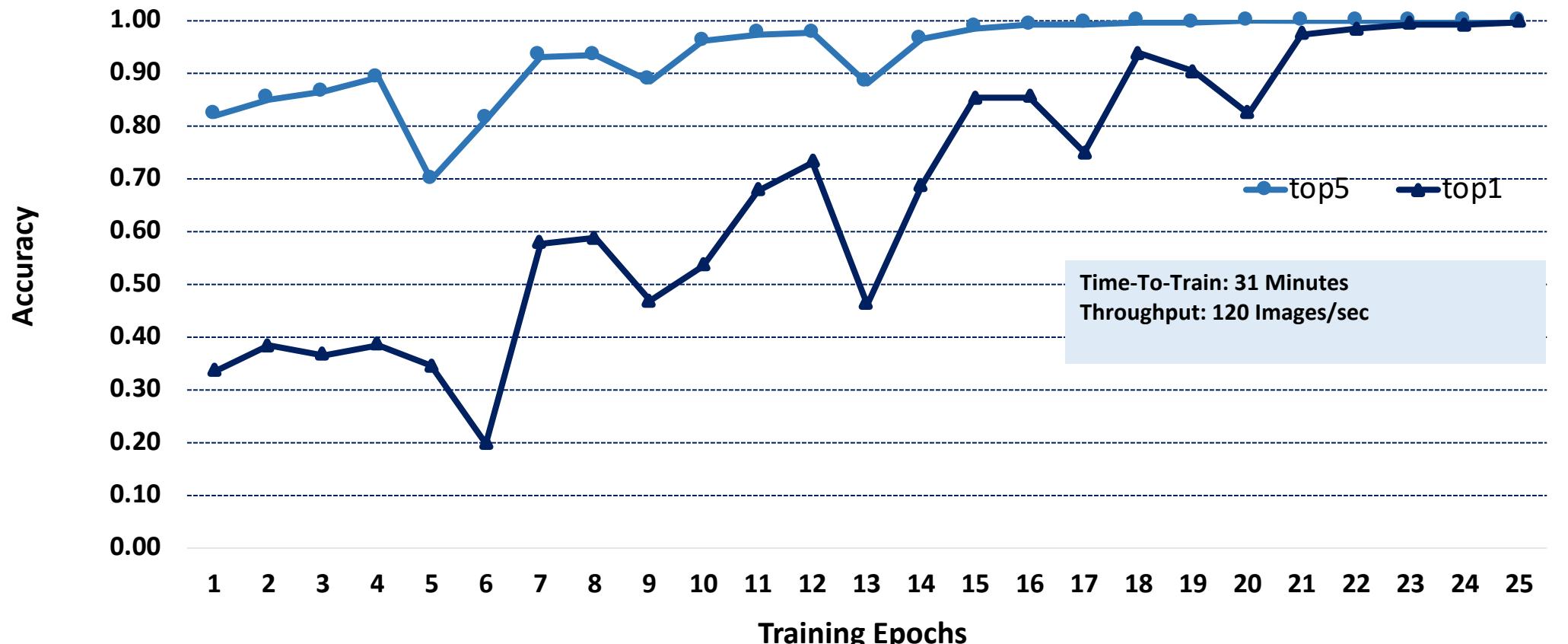


# METHOD TO SCALE TO MULTIPLE NODES

- **4 TRAINING WORKERS ON 8 NODE CLUSTER**
- **USE HOROVOD AND MPI TO SYNCHRONIZE GRADIENTS**
- **ADJUST LEARNING RATE TO ACHIEVE CONVERGENCE**

# LARGE BATCH TRAINING CONVERGES ON 8-NODES

High Content Screening/M-CNN Training with 10K images on 8-Node 2 Socket Intel® Xeon® Scalable Gold 6148 Processors; TensorFlow\* 1.7, Horovod\*, OpenMPI, BS=32/Node, GBS=256, Intel® Omni Path Fabric

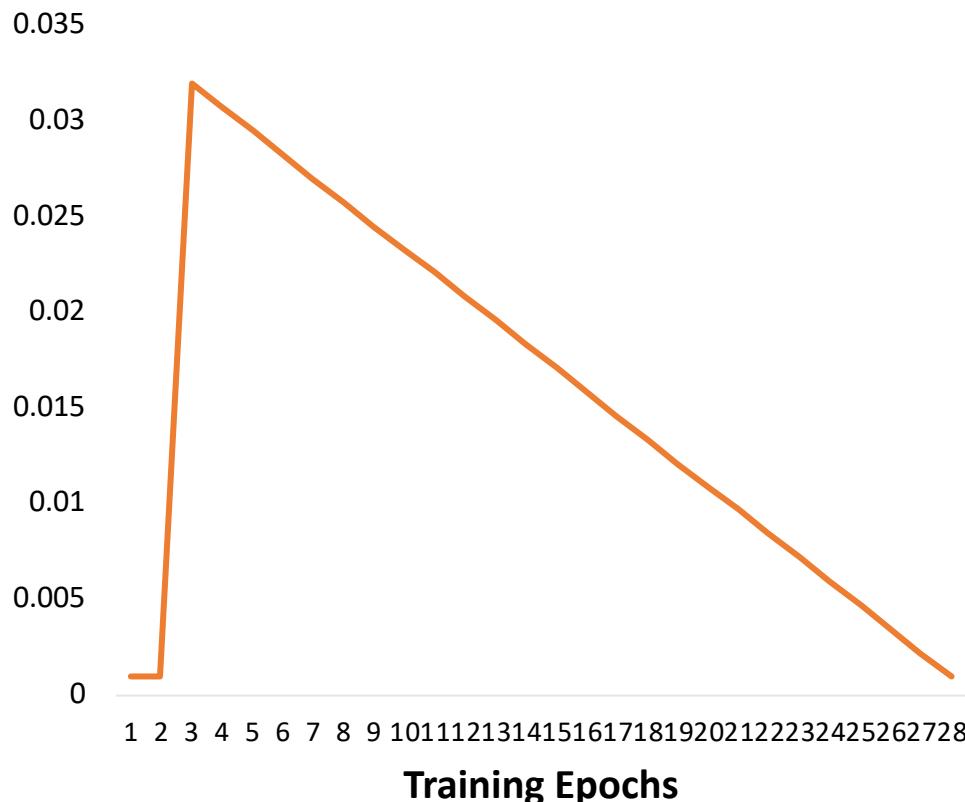


TensorFlow: 1.7.0, Python: 2.7.5, Horovod: 0.12.1: OMP\_NUM\_THREADS=10

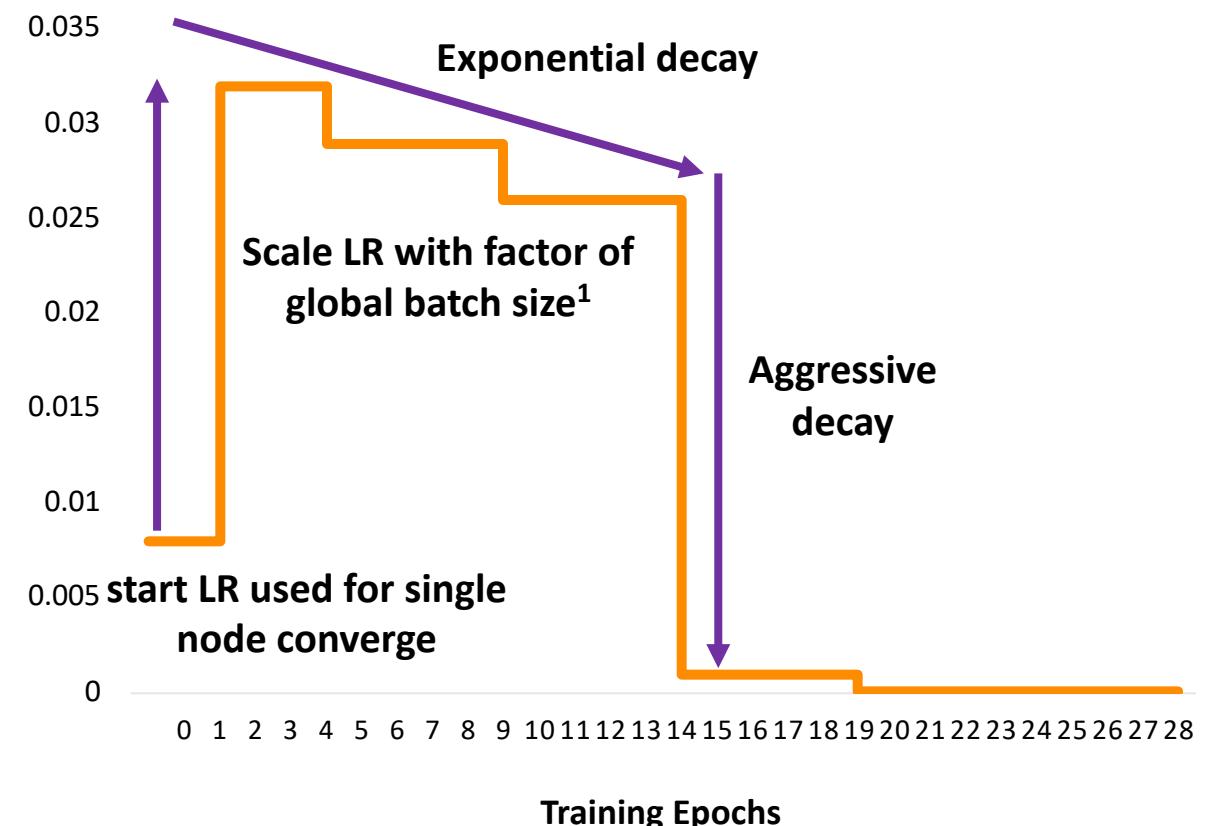
Performance results are based on testing as of dates shown in configuration and may not reflect all publicly available security updates. No product can be absolutely secure. See configuration disclosure for details. Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit: <http://www.intel.com/performance>

# LEARNING RATE SCHEDULE TO SCALE TO LARGE BATCHES

Established scaling rule<sup>1</sup>



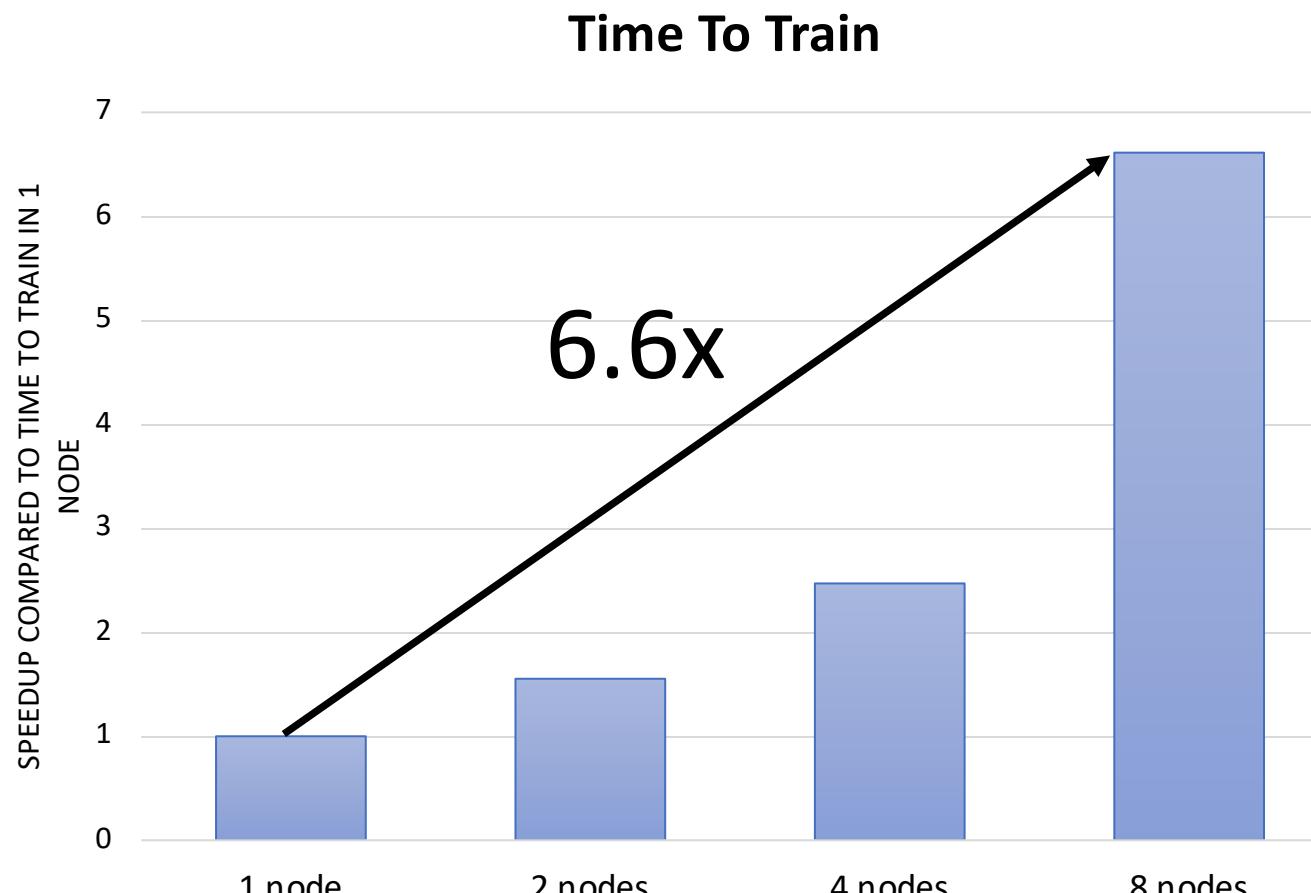
What worked for us



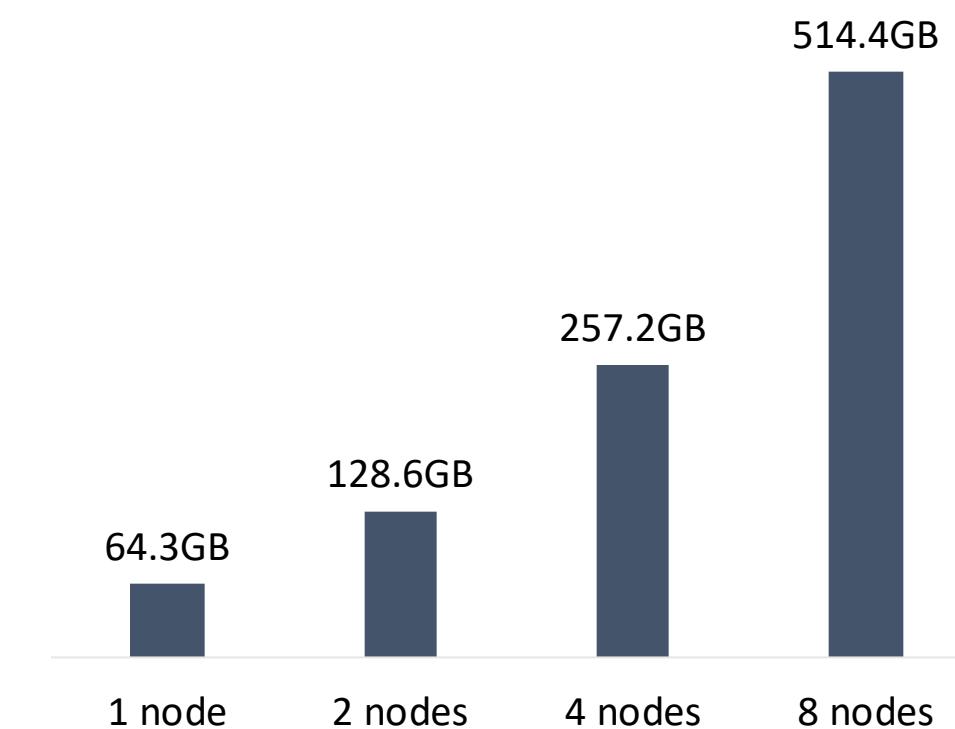
<sup>1</sup>Goyal et, 2017, "Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour"

# HIGH PERFORMANCE AT SCALE WITH INTEL® XEON® SCALABLE PROCESSOR

## M-CNN Training with 10K images



**Total Memory Used**  
192GB DDR4 per Intel® Xeon®

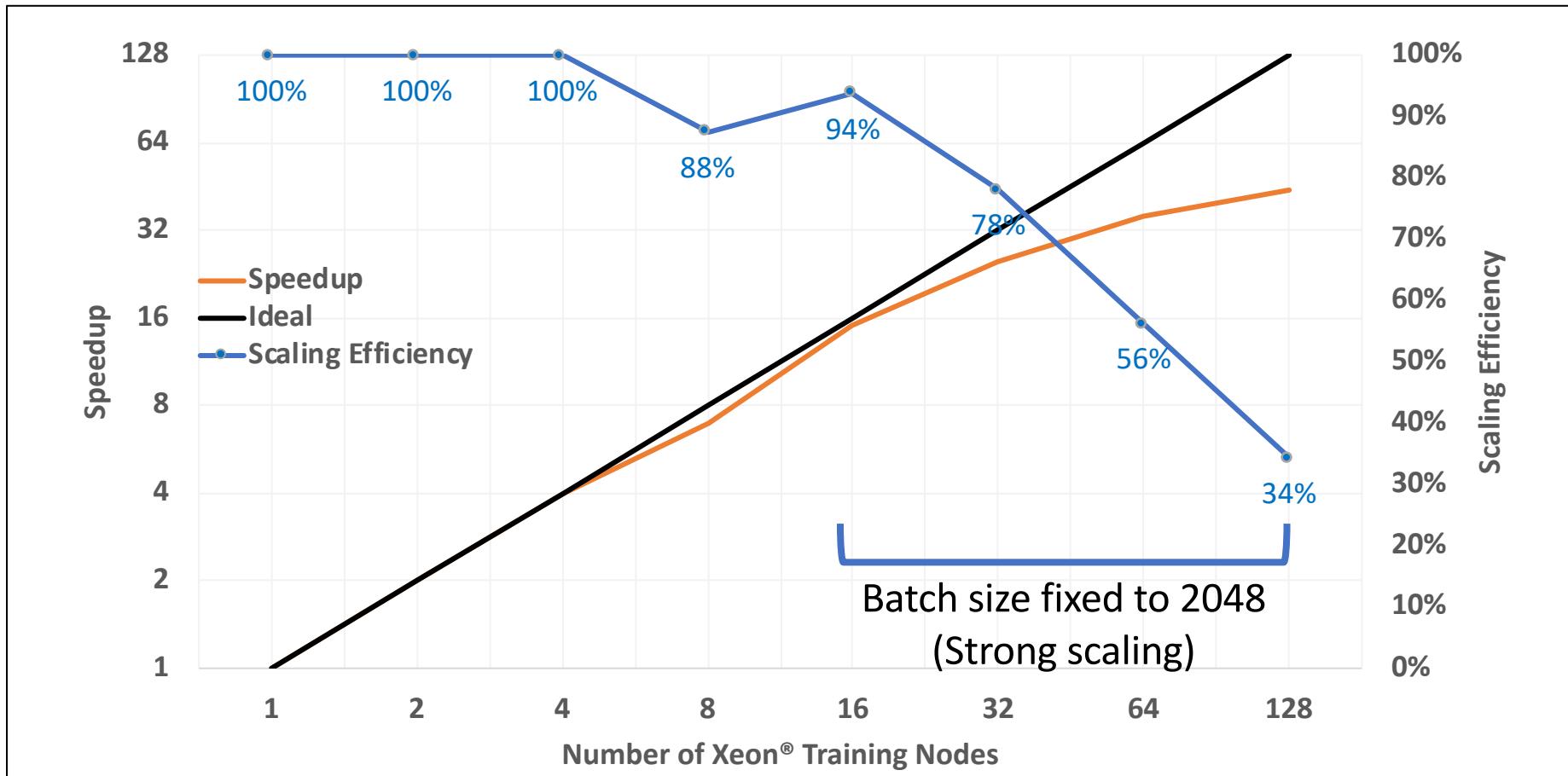


TensorFlow: 1.7.0, Python: 2.7.5, Horovod: 0.12.1, OpenMPI 3.0.0

Performance results are based on testing as of dates shown in configuration and may not reflect all publicly available security updates. No product can be absolutely secure. See configuration disclosure for details. Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit: <http://www.intel.com/performance>

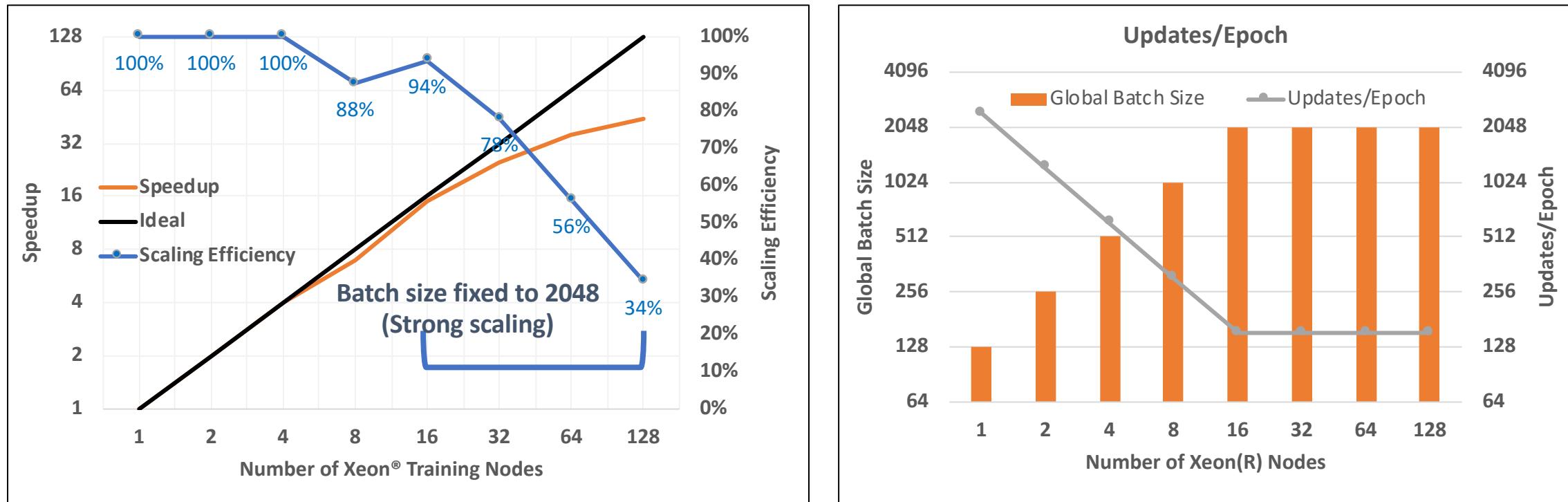
# LARGE BATCH TRAINING CONVERGES ON 128 INTEL® XEON® SCALABLE PROCESSORS

High Content Screening/M-CNN Training with 313K images on 128-Node 2 Socket Intel® Xeon® Scalable Gold 6148 Processors, TensorFlow\* 1.7, Horovod\*, OpenMPI, BS=32/Node, Intel® Omni Path Fabric



# WHY CAP LARGE BATCH SIZE TO 2048 ON 128 NODES

High Content Screening/M-CNN Training on 128-Node 2 Socket Intel® Xeon® Scalable Gold 6148 Processors, TensorFlow\* 1.7, Horovod\*, OpenMPI, BS=32/Node, Intel® Omni Path Fabric



Updates/Epoch > 150 for M-CNN training convergence

# 8/128 NODE CLUSTER CONFIGURATION DETAILS

**Compute Nodes:** 2 sockets Intel® Xeon® Gold 6148 CPU with 20 cores each @ 2.4GHz for a total of 40 cores per node. 2 Threads per core. L3 Cache: 27.3MB, 192GB of DDR4, Intel® Omni-Path Host Fabric Interface, dual-rail, Software: OpenMPI library 3.0.0, 100Gbps Intel® Intel® Omni-Path Host Fabric Interface, 480GB Intel® SSD OS drive, 1.6TB Intel® SSD data drive, CentOS\* Linux 7.3, Horovod 0.12.1, Python 2.7.5

**Top of the rack Switch:** 48-port Intel® Omni-Path Edge Switch 100 Series

**TensorFlow\***: Intel® optimized TensorFlow version 1.7.0 <https://github.com/tensorflow/tensorflow/tree/v1.7.0>

**Model:** As defined by Godinez et al, A multi-scale convolutional neural network for phenotyping high-content cellular images. Bioinformatics, 2017

## Performance Measured With:

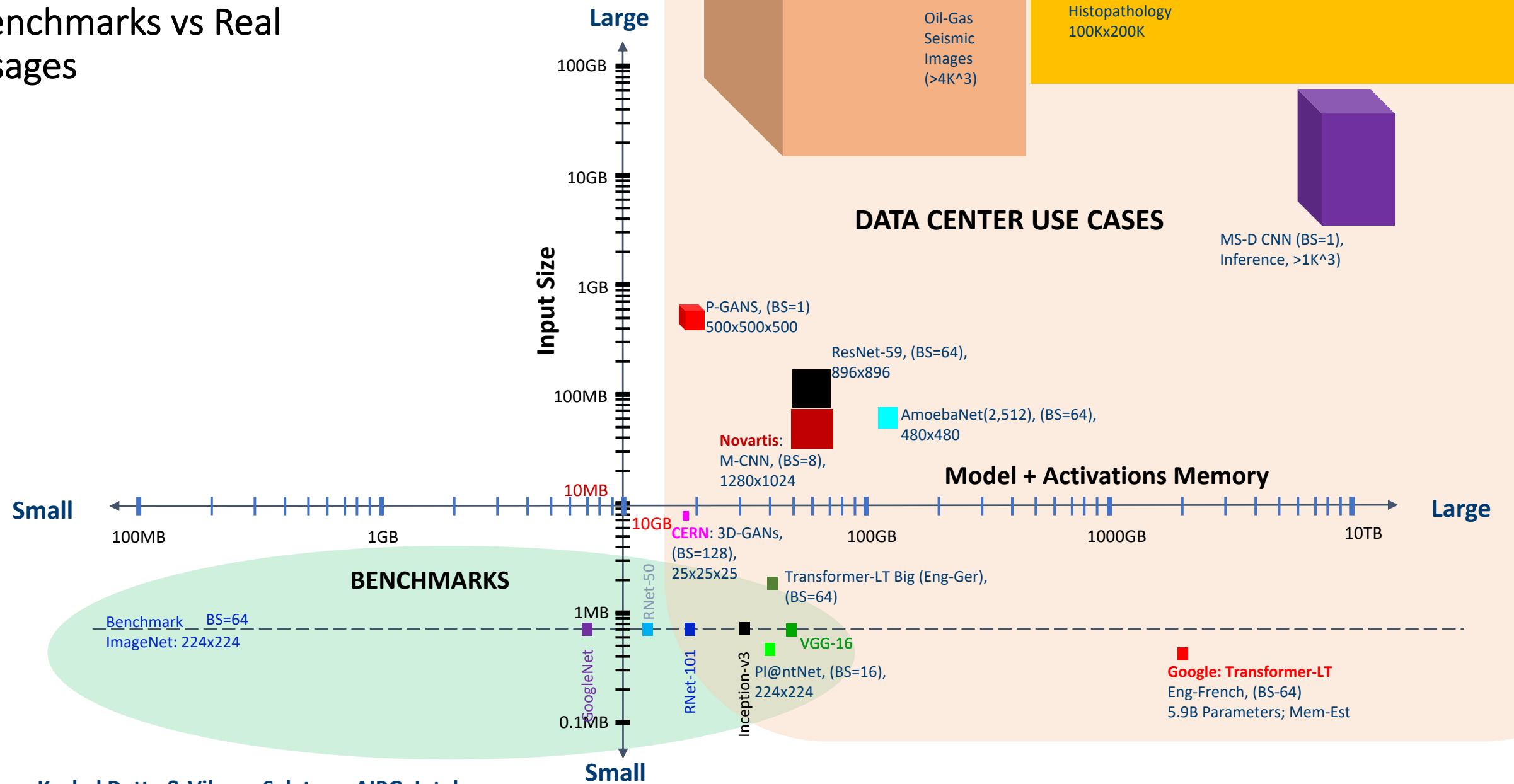
```
OMP_NUM_THREADS=10 mpirun -np 32 -cpus-per-proc 10 --map-by socket -hostfile HOSTFILE --report-bindings --oversubscribe -x LD_LIBRARY_PATH -x OMP_NUM_THREADS -x HOROVOD_FUSION_THRESHOLD numactl -l python tf_cnn_benchmarks.py --model=mcnn --batch_size=8 --data_format=NCHW --data_dir=INPUT_DATA_DIR --data_name=mcnn --num_intra_threads=10 --num_inter_threads=2 --num_batches=2000 --num_warmup_batches=70 --display_every=5 --momentum=0.9 --weight_decay=0.00005 --optimizer=momentum --resize_method=bilinear --distortions=False --sync_on_finish=True --device=cpu --mkl=True --kmp_affinity=="granularity=fine,compact,1,0" --variable_update=horovod --local_parameter_device=cpu --kmp_blocktime=1 --horovod_device=cpu --piecewise_learning_rate_schedule='0.008;2;0.032;5;0.029;10;0.026;15;0.001;20;0.0001' --train_dir=TRAIN_DATAWRITE_DIR --save_summaries_steps=1 --summary_verbosity=1
```

# STRATEGY

- IMPROVE MAXIMIZE PERFORMANCE ON A SINGLE NODE
- SCALEOUT PERFORMANCE ON MULTIPLE NODES
  - USE LARGE BATCHES
  - ADJUST LEARNING RATE



# Benchmarks vs Real Usages



Source: Kushal Datta & Vikram Saletore, AIPG, Intel

# Acknowledgments

## **Novartis Institutes for Biomedical Research (Basel & Cambridge):**

Stephen Litster

Michael Derby

Michael Steeves

Wolfgang Zipfel

## **Intel Corporation (AIPG, DCG, SMG):**

Michael Demshki

Deepthi Karkada

Adam Procter

Patrick Messmer

Krishna S. Muppalla

Bruno Riva

Hema Chamraj

# LEGAL NOTICES & DISCLAIMERS

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at [intel.com](http://intel.com), or from the OEM or retailer. No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit <http://www.intel.com/performance>.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

Statements in this document that refer to Intel's plans and expectations for the quarter, the year, and the future, are forward-looking statements that involve a number of risks and uncertainties. A detailed discussion of the factors that could affect Intel's results and plans is included in Intel's SEC filings, including the annual report on Form 10-K.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Performance estimates were obtained prior to implementation of recent software patches and firmware updates intended to address exploits referred to as "Spectre" and "Meltdown." Implementation of these updates may make these results inapplicable to your device or system.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, the Intel logo, Pentium, Celeron, Atom, Core, Xeon and others are trademarks of Intel Corporation in the U.S. and/or other countries.

\*Other names and brands may be claimed as the property of others.

© 2019 Intel Corporation.