DISTRIBUTED MACHINE LEARNING – IMAGE CLASSIFICATION FOR SUPERCOMPUTERS

Michael Steyer
Technical Consulting Engineer
Intel Corporation

Legal Disclaimer & Optimization Notice

INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS”. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO THIS INFORMATION INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

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.

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.

Notice revision #20110804

Agenda

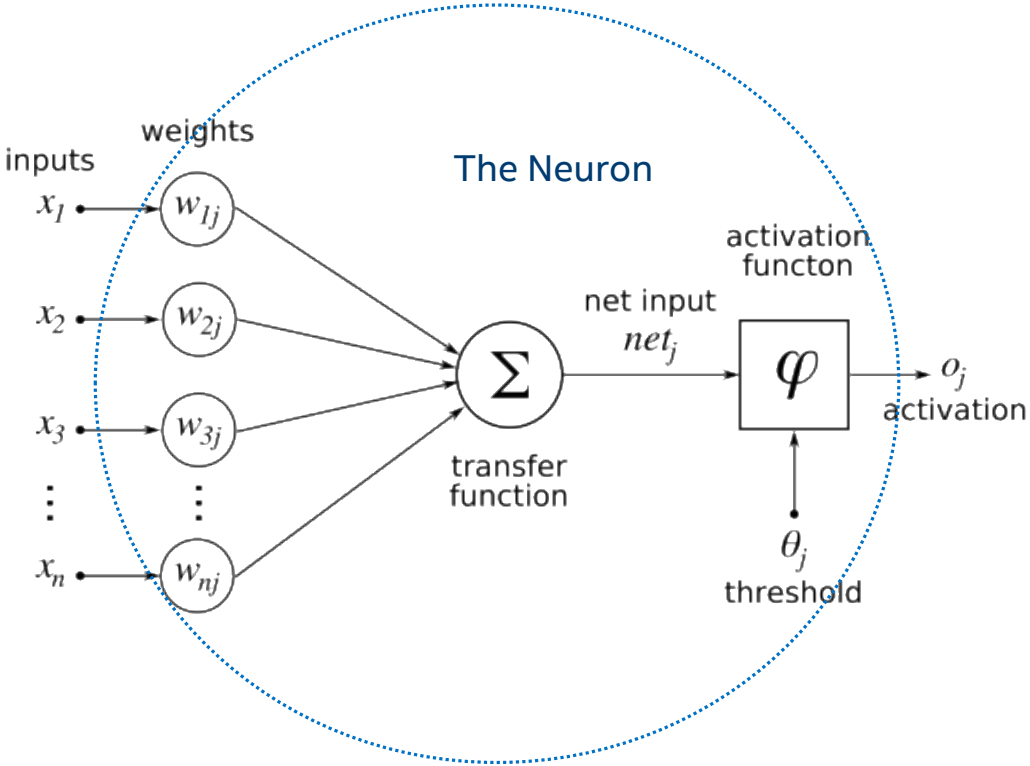
1) Introduction
2) Steps towards distributed image classification
    1. Getting Intel Caffe
    2. Installing Caffe prerequisites
    3. Compiling Caffe
    4. The test dataset images
    5. Transforming the dataset
    6. Defining the network topology
    7. Training the network with Caffe
    8. Inference / Scoring of an individual image

Introduction
Demo

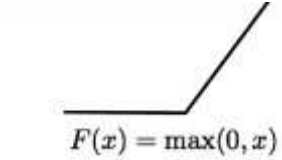
What this IS	What this IS NOT
• An Introduction how to use Intel Caffe on Clusters • A Proof Of Concept • Just one sample approach how to use Caffe, while there are many …	• An introduction to Deep Learning • A complete tutorial • A parameter tuning session for Deep Learning topologies • A Benchmarking session / performance comparison

DEMO – KICK OFF TRAINING

Recap - Neural Network

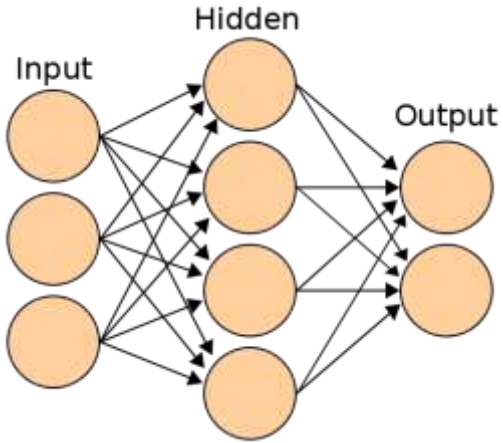


Activation Function



← ReLU (rectified linear unit)

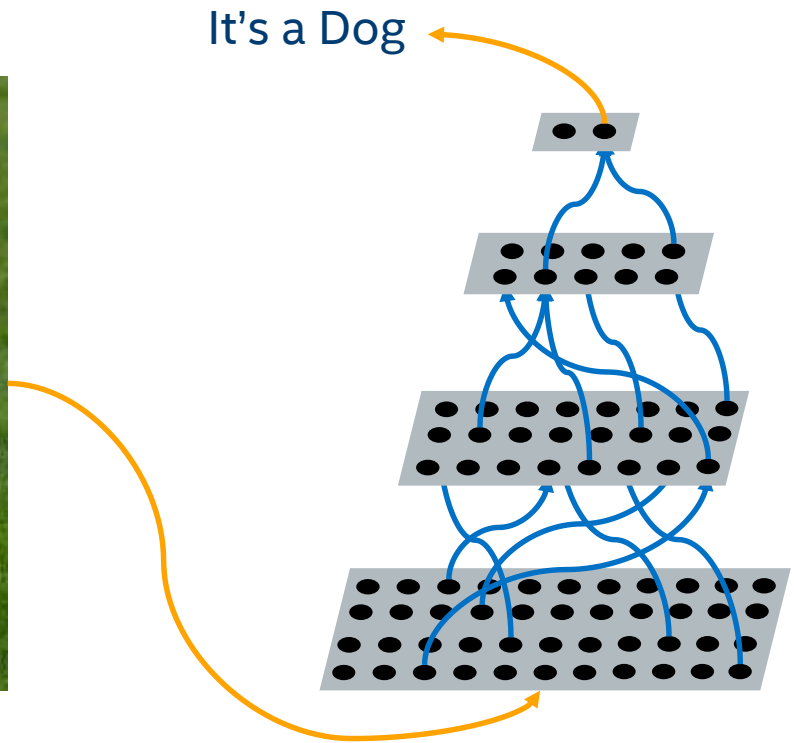
F: a non-linear differentiable function



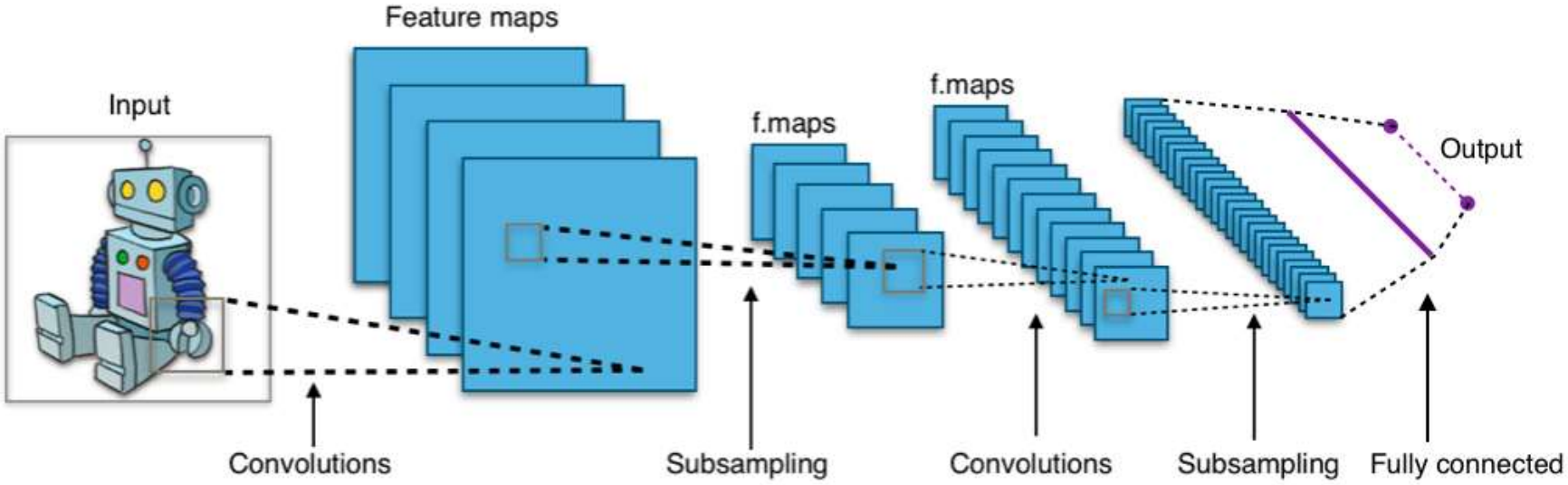
https://en.wikibooks.org/wiki/Artificial_Neural_Networks/Print_Version

Cat or Dog ???

https://commons.wikimedia.org/wiki/American_Eskimo_Dog



Recap - Convolutional Neural Networks (CNN)



https://en.wikipedia.org/wiki/Convolutional_neural_network

Getting Intel Caffe

https://github.com/intel/caffe

This fork of BVLC/Caffe is dedicated to improving performance of this deep learning framework when running on CPU, in particular Intel® Xeon processors (HSW+) and Intel® Xeon Phi processors

Installing Caffe Prerequisites

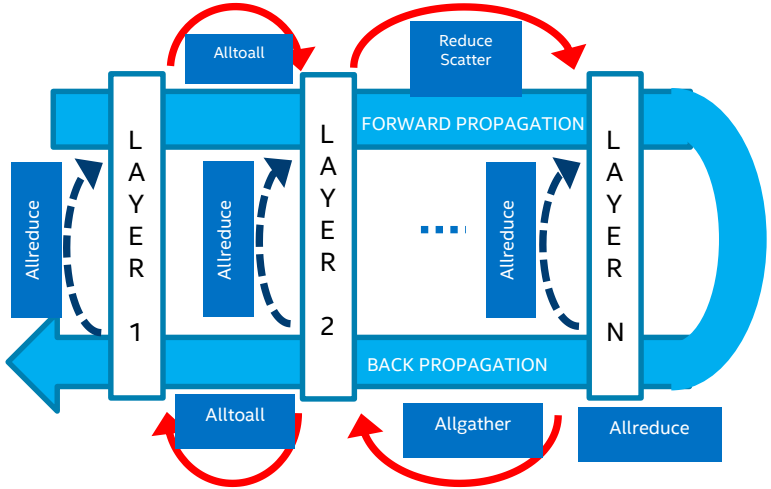
- Official:
  http://caffe.berkeleyvision.org/installation.html
- Works for me (RHEL* environment):
  - protobuf-3.0.2
  - opencv-3.1.0
  - lmdb-LMDB_0.9.18
  - leveldb-1.19
  - hdf5-1.8.17
  - glog-0.3.3
  - gflags-2.1.2
  - cmake-3.0.0
  - libunwind-1.1
  - boost_1_62_0
  - jpeg-8d
- Python …

Intel® Machine Learning Scaling Library (MLSL)

https://github.com/01org/MLSL/releases

Some of the Intel MLSL features include:

- Built on top of MPI, allows for use of other communication libraries
- Optimized to drive scalability of communication patterns
- Works across various interconnects: Intel® Omni-Path Architecture, InfiniBand*, and Ethernet
- Common API to support deep learning frameworks (Caffe*, Theano*, Torch*, etc.)



Compiling Caffe

https://github.com/intel/caffe/wiki/Multinode-guide

Some parameters:

```
$ vim Makefile.config
…
CPU_ONLY := 1
…
CUSTOM_CXX := icpc
…
USE_MLSL :=1
…
BLAS := mkl
…
ANACONDA_HOME := /opt/intel/python/3.5_2017u1.0.035/intelpython35
…
```

DEMO – ENVIRONMENT & DEPENDENCIES

Test Dataset Images

Flowers dataset - 5 different categories, 3.6k images

What is possibly more interesting than … ?

Flowers!?

daisy/3957488431_52a447c0e8_m.jpg CC-BY by Tristan Martin - https://www.flickr.com/photos/mukumbura/3957488431/

tulips/3454461550_64d6e726bf_m.jpg CC-BY by minniemouseaunt - https://www.flickr.com/photos/minniemouseaunt/34544461550/

daisy/5673728_71b8cb57eb.jpg CC-BY by Teo - https://www.flickr.com/photos/teo/5673728/

dandelion/3730618647_5725c692c3_m.jpg CC-BY by Michael Ruiz - https://www.flickr.com/photos/simax/3730618647/

dandelion/14053173516_a00150a919_m.jpg CC-BY by UnknownNet Photography - https://www.flickr.com/photos/threar/14053173516/

daisy/144603918_b9de002f60_m.jpg CC-BY by Andreas. - https://www.flickr.com/photos/124330160/144603918/

roses/475936554_a2b38aaa8e.jpg CC-BY by peasap - https://www.flickr.com/photos/peasap/475936554/

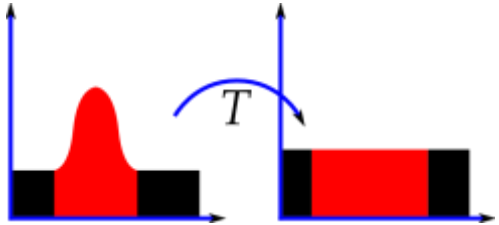
roses/3667366832_7a8017c528_n.jpg CC-BY by Parvin - https://www.flickr.com/photos/55948751@N00/3667366832/

sunflowers/4895720722_8247f2015b_n.jpg CC-BY by F Delventhal - https://www.flickr.com/photos/krossbow/4895720722/

Dataset Transformation

1) Histogram Equalization

2) Resizing -> 227x227



https://en.wikipedia.org/wiki/Histogram_equalization

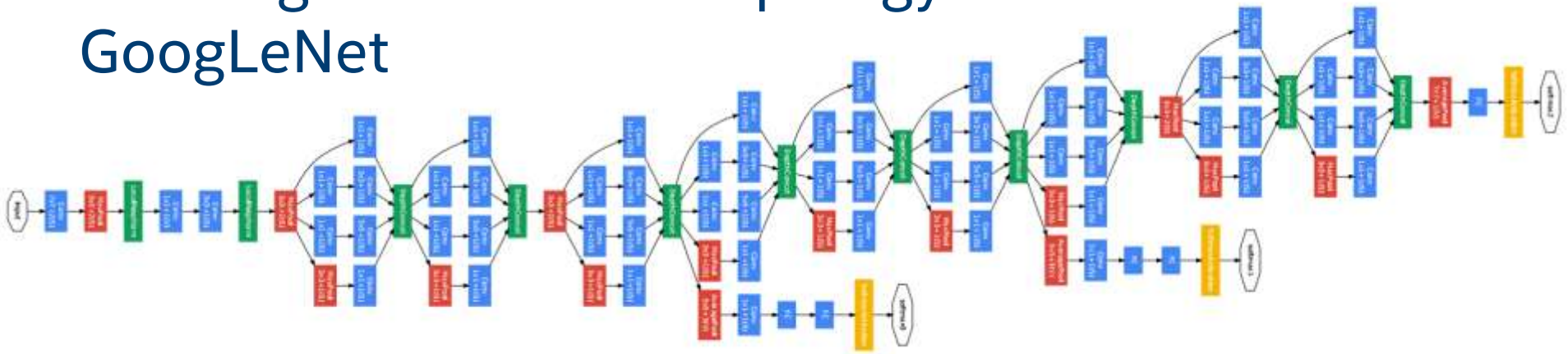
1 & 2

daisy/5673728_71b8cb57eb.jpg CC-BY by Teo - https://www.flickr.com/photos/teo/5673728/

3) Storing images into LMDB using the caffe toolchain

DEMO – DATASET & TRANSFORMATION

Defining the Network Topology – GoogLeNet



Going deeper with convolutions - http://ieeexplore.ieee.org/document/7298594/

Network Definition	Others
1) Define input (Data) layers 2) Define number of classes in output layers 3) Define batch size	1) Define Solver 2) Generate Deployment Network version

The Solver
Stochastic Gradient Descent (SGD)

Caffe ImageNet Rules of Thumb:

http://caffe.berkeleyvision.org/tutorial/solver.html

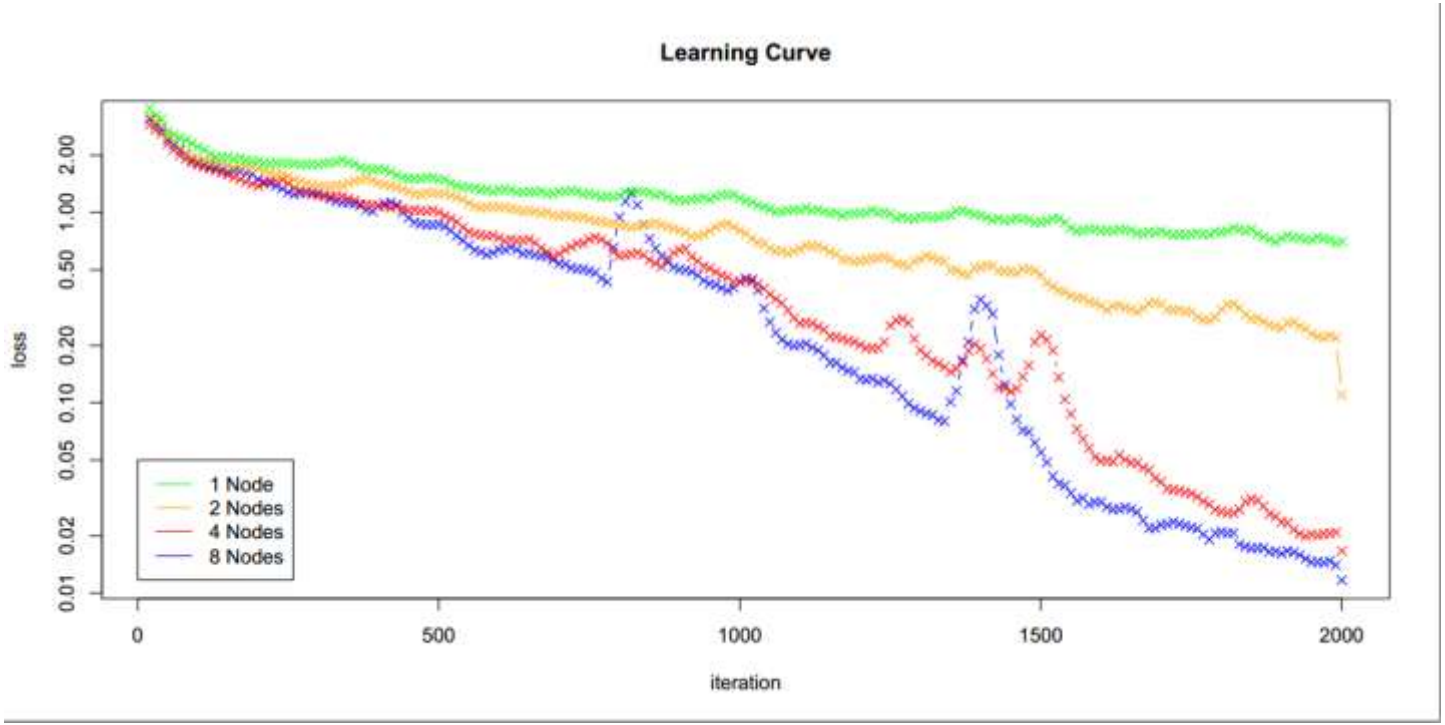
```
base_lr: 0.01              # begin training at a learning rate of 0.01 = 1e-2
lr_policy: "step"          # learning rate policy: drop the learning rate in "steps"
                           # by a factor of gamma every stepsize iterations
gamma: 0.1                 # drop the learning rate by a factor of 10
                           # (i.e., multiply it by a factor of gamma = 0.1)
stepsize: 100000           # drop the learning rate every 100K iterations max_iter: 350000
                           # train for 350K iterations total
momentum: 0.9
```

DEMO – OVERVIEW NETWORK & SOLVER

Training the network with Caffe

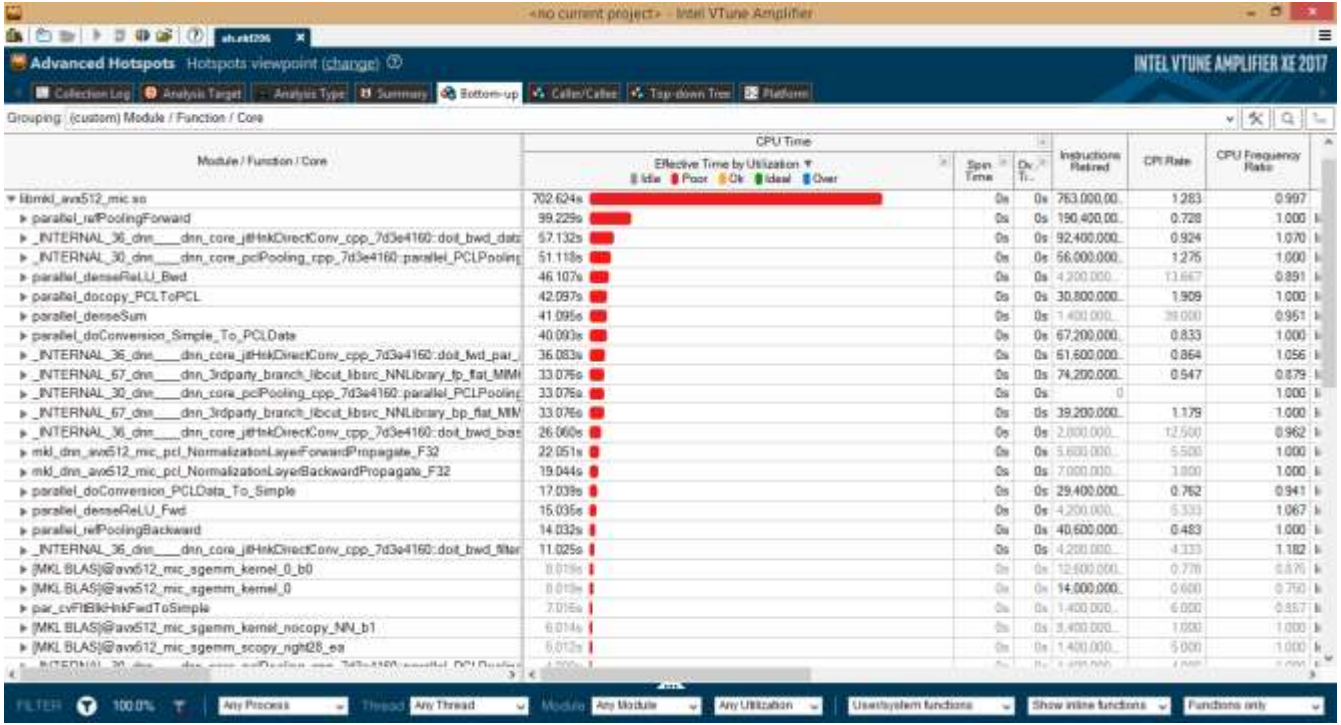
```
$ mpirun -n 8 -ppn 1 -f $HOSTFILE caffe train -engine "MKL2017" \
    --solver=./solver.prototxt --param_server=mlsl
…
I0129 09:20:45.530730 25291 net.cpp:499] Network initialization done.
I0129 09:20:45.533807 25291 solver.cpp:112] Solver scaffolding done.
I0129 09:20:45.534752 25291 caffe.cpp:310] Configuring multinode setup
I0129 09:20:45.535677 25291 caffe.cpp:325] Starting Multi-node
Optimization in MLSL environment
W0129 09:20:45.535863 25291 MlslSync.hpp:137] RUN: DITRIBUTED WEIGHT
UPDATE IS DISABLED, PER LAYER TIMINGS ARE DISABLED, SINGLE DB SPLITTING
IS DISABLED
W0129 09:20:45.535989 25291 MlslSync.hpp:77] synchronize_params: bcast
I0129 09:20:45.576264 25236 solver.cpp:489] Solving GoogleNet
I0129 09:20:45.576364 25236 solver.cpp:490] Learning Rate Policy: step
…
```

Training 1/2/4/8 Nodes with Batch-size 64 each



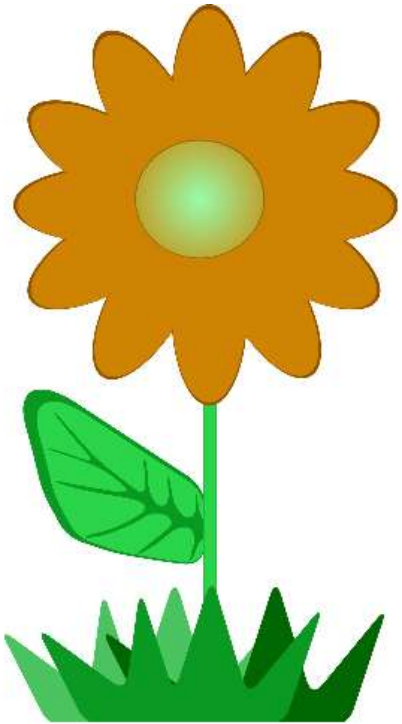

```
Iteration / Epoch = (3600 * 5/6) / (64 * (1|2|4|8)) = 46.88 | 23.44 | 11.72 | 5.86
```

Is Caffe Leveraging the Intel® MKL DNN Library? Intel® VTune™ Amplifier XE



DEMO – DEPLOYMENT & INFERENCE

What about a flower from … The Internet!



OR

?

Sunflower

sunflowers/19595718862_c68896370c_m.jpg CC-BY by jd_09 - https://www.flickr.com/photos/jer_09/19595718862/

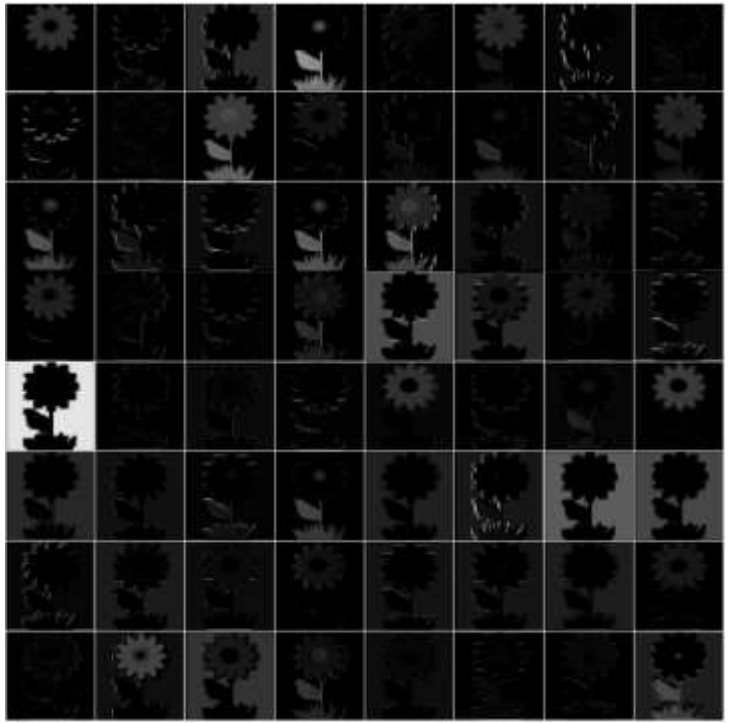
Daisy

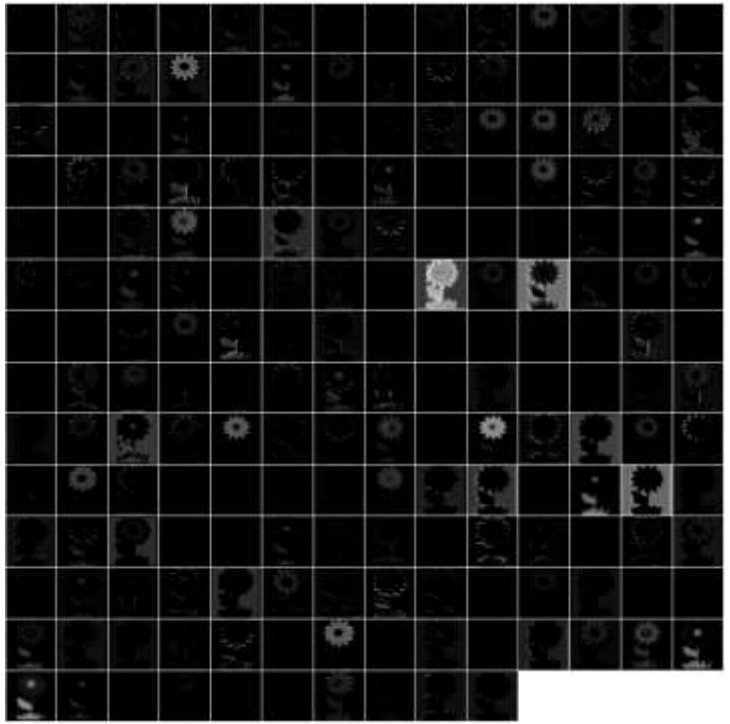
daisy/14073784469_ffb12f3387_n.jpg CC-BY by Bridget Leyendecker - https://www.flickr.com/photos/b_leyende/14073784469/

99.99% Daisy

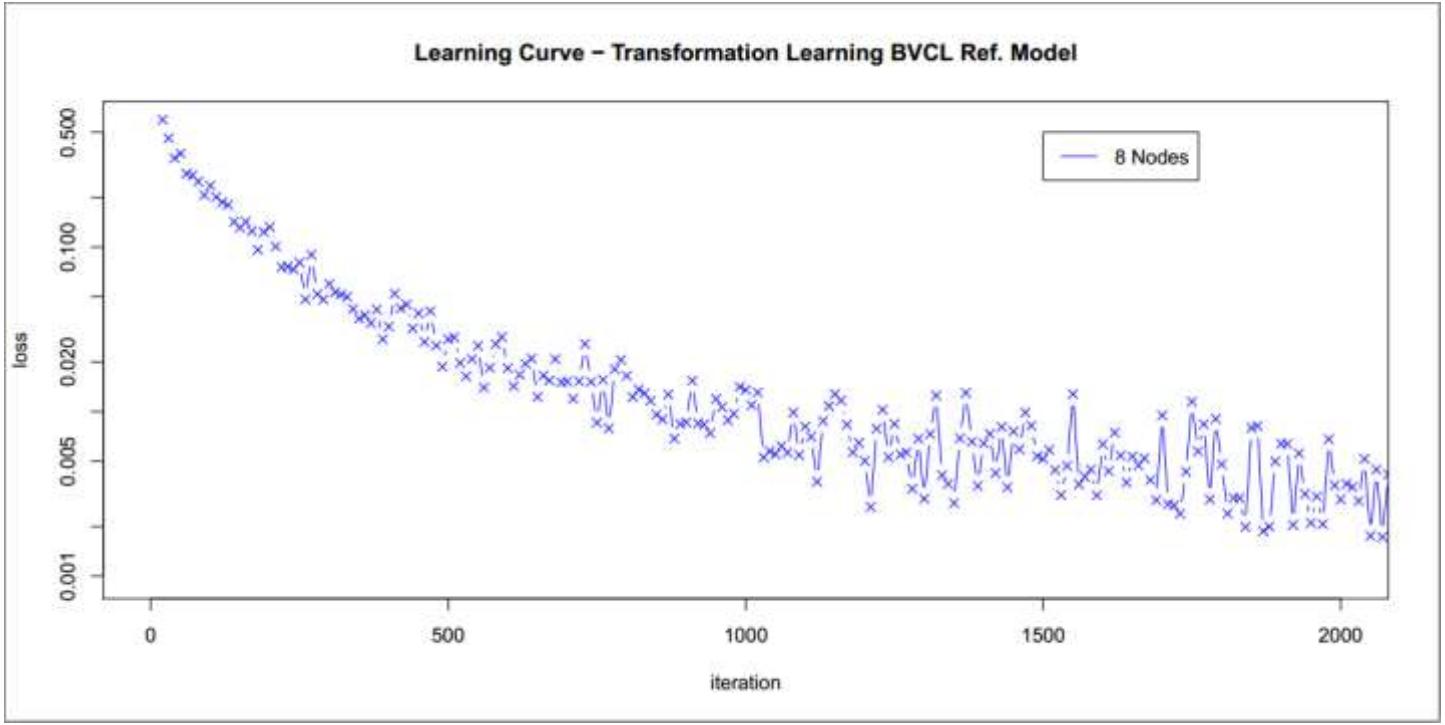
Source: http://www.publicdomainpictures.net/view-image.php?image=77758&picture=flower-clip-art-2

Why 99.99% Daisy?
The answer sits somewhere in the features …





Faster Training?
Transform Learning!



QUESTIONS?


intel®