Deep Learning on SHARCNET: Tools you can use

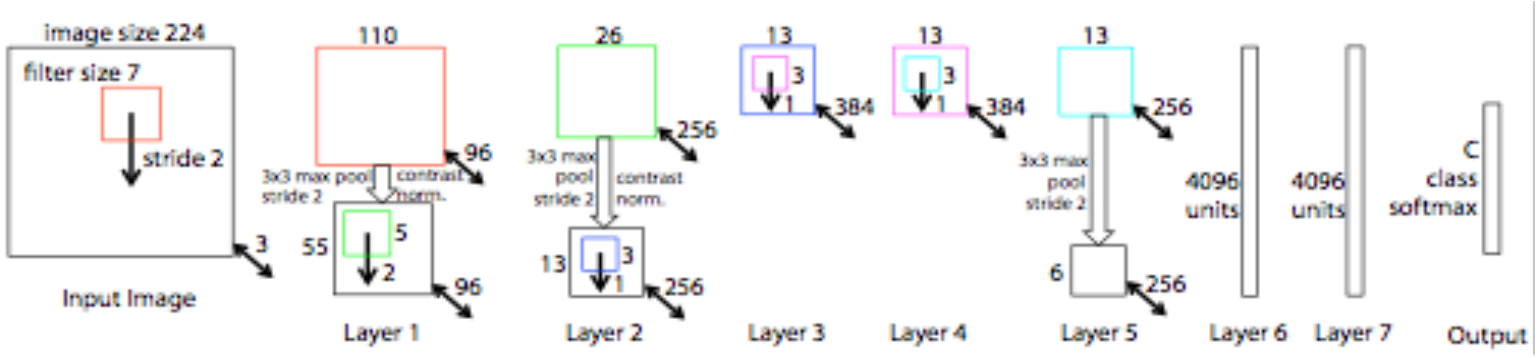
Presenter: Fei Mao, HPC Consultant, SHARCNET

Outlines

- Deep Learning
- Hardware in SHARCNET
- Software tools
- Common issues

Deep Learning

Example: A Convolutional Net
"An Astounding Baseline for Recognition"

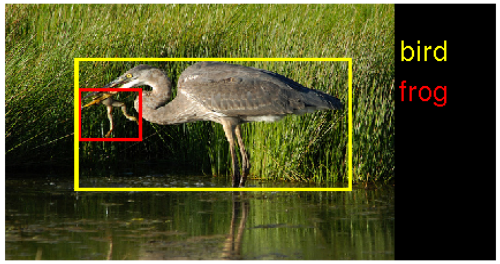


- Learn multiple layers of representation, corresponding to different levels of abstraction
  - theory on the advantage of depth (Hasted et al 1986 & 1991), (Bengio et al. 2007), (Bengio and Delalleau 2011), (Braverman 2011)
  - exploiting composition gives an exponential gain in representational power (humans organize ideas and concepts hierarchically!)
  - biologically inspired learning - the brain is deep! (Bengio 2015)

Image: Convolutional net architecture (Zeiler and Fergus, 2013)

DL Successes

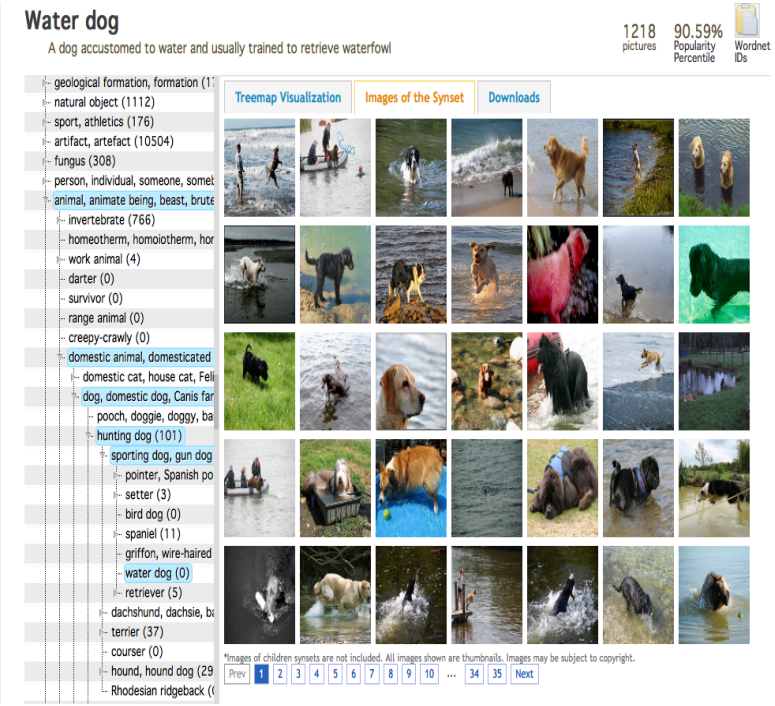
- Vision
  - Object recognition and detection
  - Pose estimation and activity recognition
- Speech
  - Speech recognition
  - Speaker authentication
- Natural Language Processing
  - Question answering
  - Machine translation
- Gaming
  - Mastering the game of Go



IMAGENET

Large Scale Visual Recognition Challenge (ILSVRC)

- A benchmark in object category classification and detection on hundreds of object categories and millions of images
  - ILSVRC classification task: 1000 object classes and approximately 1.2 million training images
- The first "real" benchmark where deep learning beat sophisticated computer vision systems
  - ILSVRC 2010: 28.2% top5 err rate, NEC
  - ILSVRC 2012: 16.4%, AlexNet (deep convnet)
  - ILSVRC 2014: 6.7%, GoogLeNet (very deep convnet)
  - ILSVRC 2015: 3.57%, Deep Residual Net(152 layers)



DL with HPC

- DL particularly well-suited to parallelization
  - Data parallelism inherent in pixel-based inputs (e.g. images and videos). Each core (e.g. GPU) is responsible for a chunk of input data
  - Task parallelism inherent in redundant processing units (neurons). Each core (e.g. GPU) is responsible for part of the architecture
- Hardware accelerators (e.g. GPUs, FPGAs)

  - Theano, Caffe all provide GPU support by way of CUDA
  - NV DIGITS provides multi-GPU support
- Distributed frameworks (e.g. Google Tensorflow)

GPU Clusters

- Mosaic
  - 20 GPU nodes
  - One K20m per node
  - 20 CPU cores per node
  - 256GB mem per node
  - IB QDR
  - mos1 as dev-node
- Copper
  - 8 GPU nodes
  - 4 K80 cards per node (8 GPU units)
  - 16 CPU cores per node
  - 96GB mem per node
  - IB FDR

Tools for DL

- Popular Deep Learning Tools
  - Google Tensorflow(Python)
    - https://www.sharcnet.ca/help/index.php/Tensorflow
  - Caffe/ NV DIGITS (C++ w/ Python & Matlab wrappers)
    - https://www.sharcnet.ca/help/index.php/Caffe
    - https://www.sharcnet.ca/help/index.php/DIGITS
  - Theano/Pylearn2/Lasagne (Python)
    - https://www.sharcnet.ca/help/index.php/Theano

Common issues

- Number of CPU cores
  - Tensorflow, >5 cores
  - DIGITS, whole node
  - Caffe, >2 cores
  - Theano, 1 or 2 cores
- GPUs (support cuDNN)
  - K20 (Mosaic, ~4.7GB mem)
  - K80 (Copper, ~11.5GB mem)
  - 750Ti (Angel, 2GB mem)
- File I/O
  - /tmp for small files (e.g. images)
  - /scratch or /work for database files
- Internet access
  - only on login node
- Multi-GPU
  - Shared Memory (e.g. Copper)
  - Distributed (Copper and Mosaic)
- Security
  - NV DIGITS, open port to public