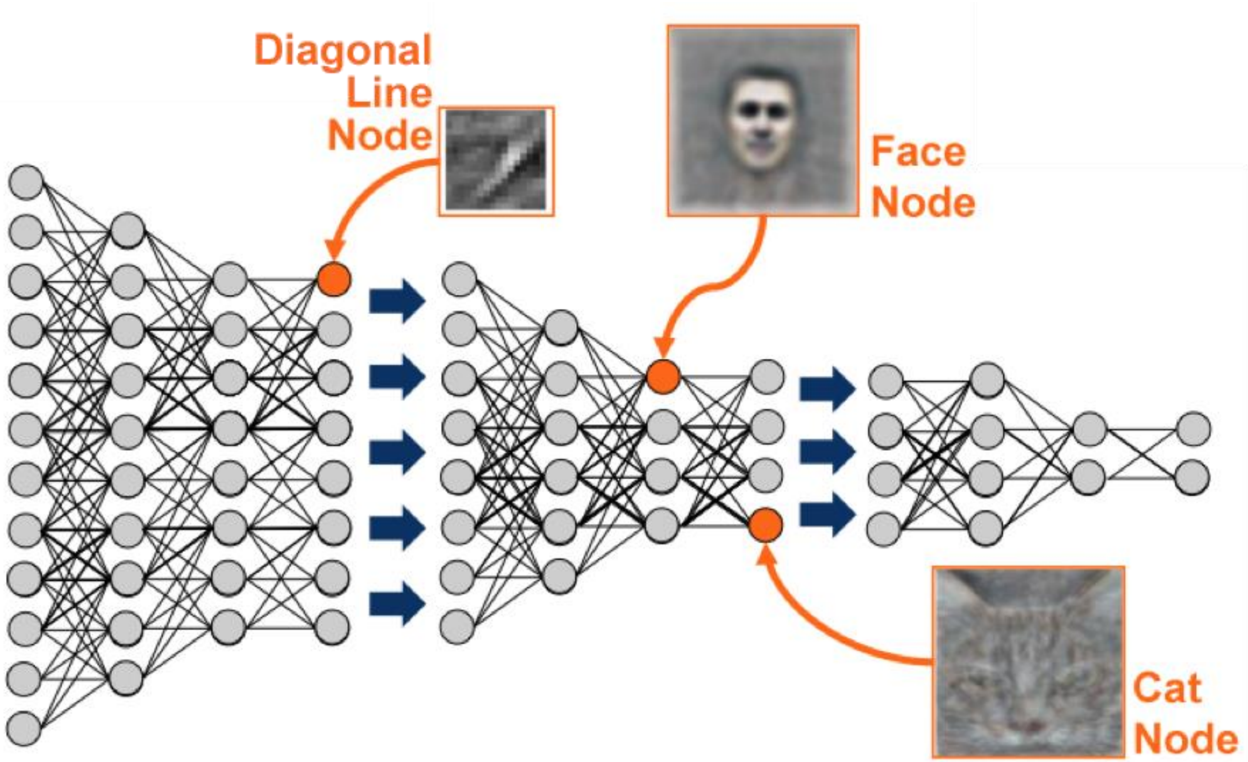
UBC MLRG (Fall 2016): Deep Learning

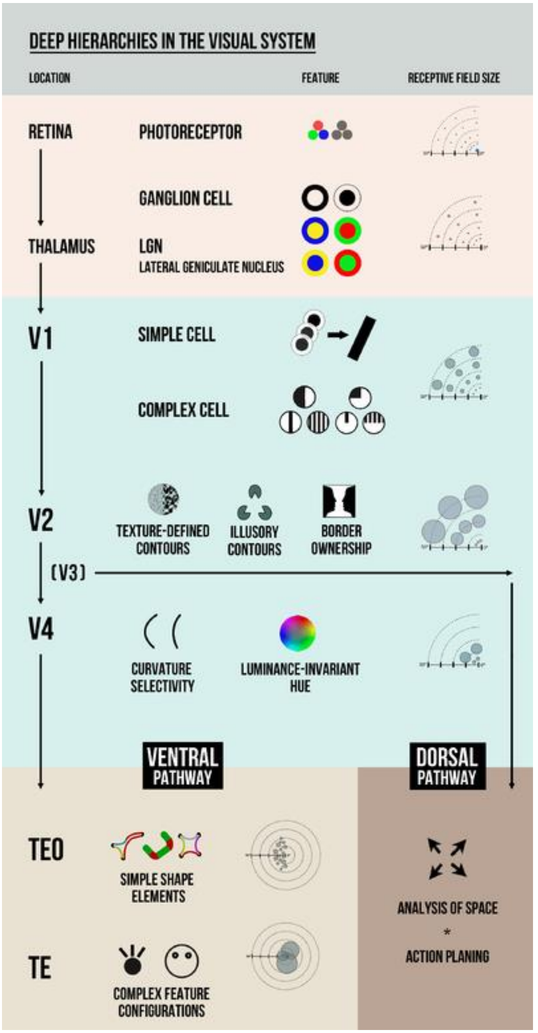
Recent MLRG History

- Topics covered in recent MLRG terms:
  - Fall 2014: Deep learning and Bayesian optimization.
  - Winter 2015: Causality, bandits, reinforcement learning.
  - Summer 2015: Probabilistic graphical models.
  - Fall 2015: Convex optimization.
  - Winter 2016: Bayesian statistics.
  - Summer 2016: Miscellaneous.
  - Fall 2016: Deep learning .
  - Winter 2016: Reinforcement learning?

Deep Learning

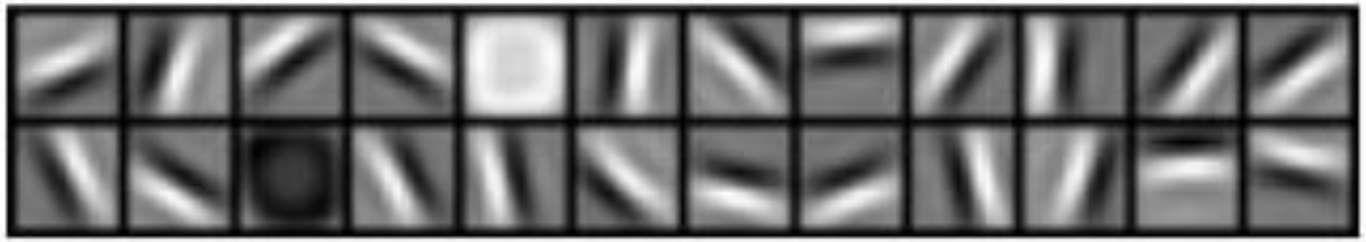
- The hottest topic in machine learning right now is deep learning.
  - Models inspired by hierarchies in the brain.
  - Model complex concepts using “layers” of simple units.





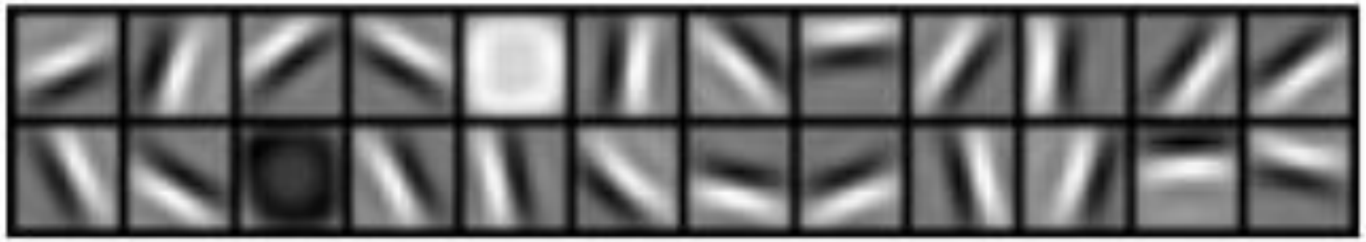
Deep Learning

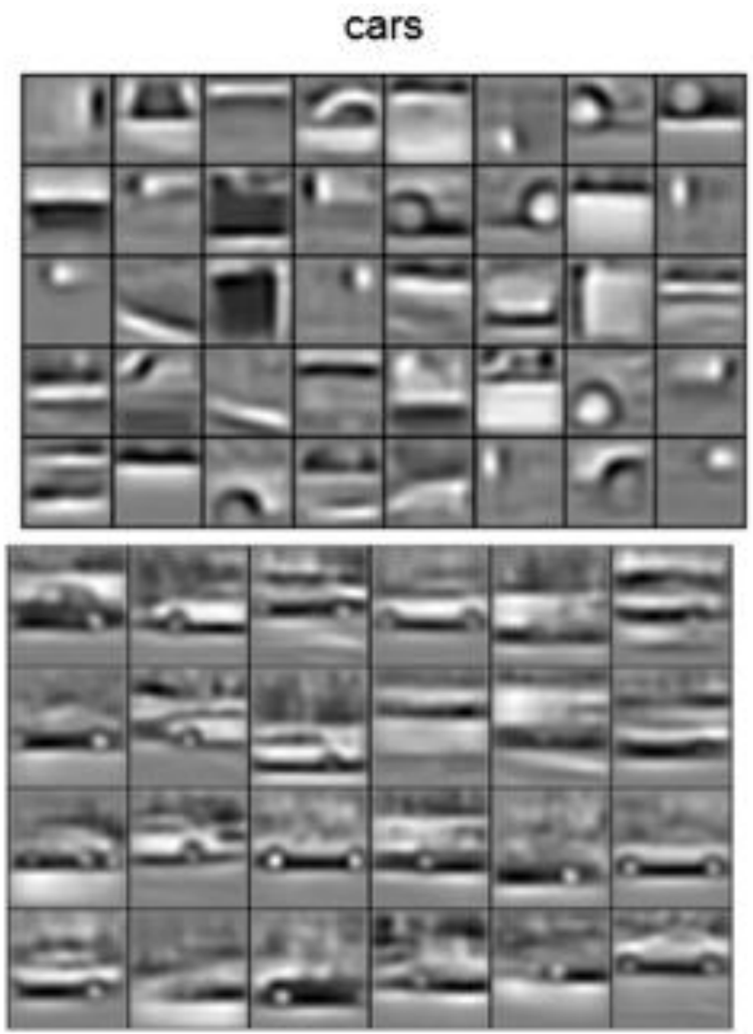
- The hottest topic in machine learning right now is deep learning.



Deep Learning

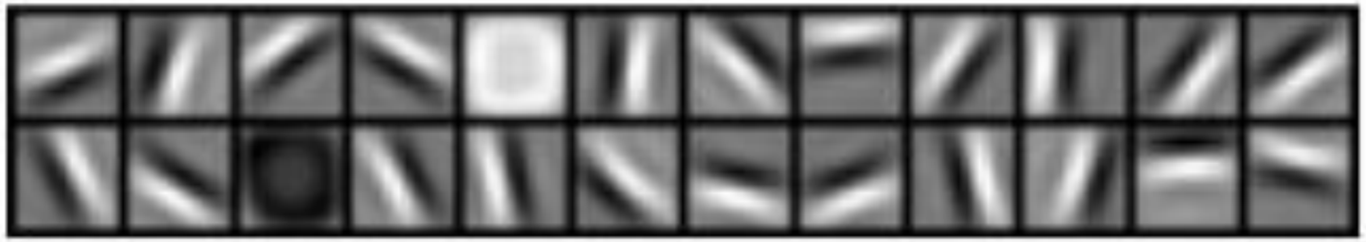
- The hottest topic in machine learning right now is deep learning.





Deep Learning

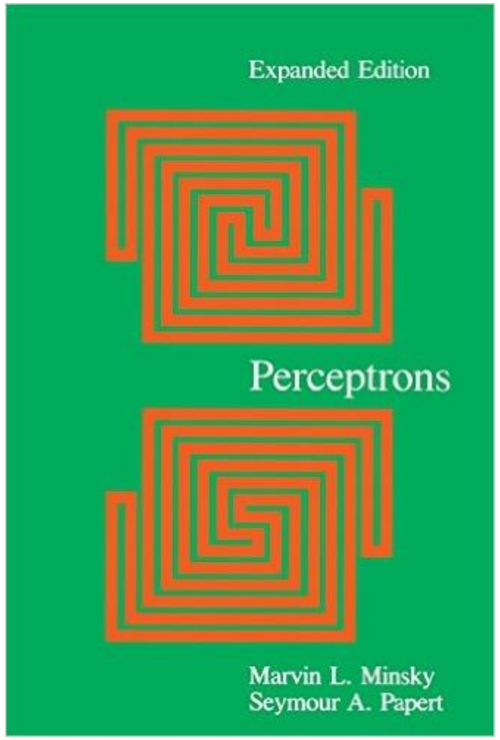
- The hottest topic in machine learning right now is deep learning.



faces, cars, airplanes, motorbikes

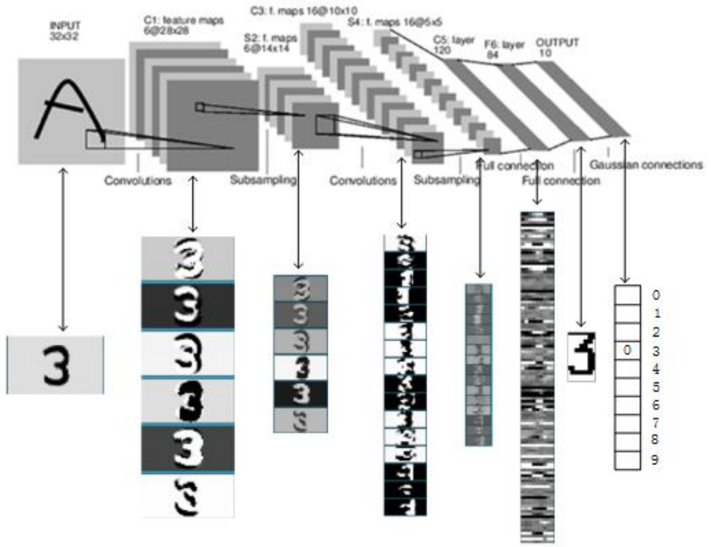
ML and Deep Learning History

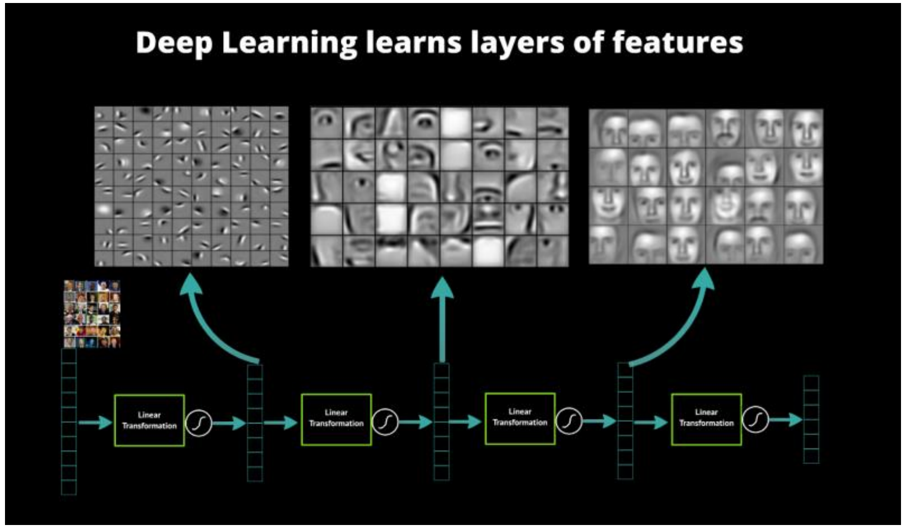
- 1950 and 1960s: Initial excitement (then drop in popularity):
  - Roughly: a linear classifier trained with stochastic gradient.
  - "the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence." New York Times (1958).
  - Quickly realized limitations of linear models.
  - Widrow/adaline video:
    - https://www.youtube.com/watch?v=IEFRtz68m-8

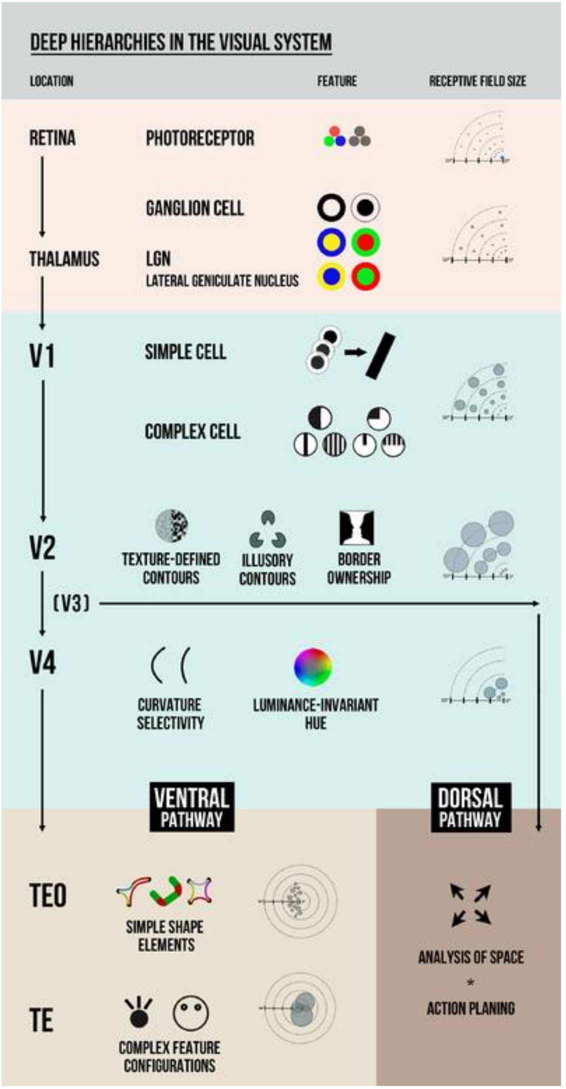


ML and Deep Learning History

- 1970 and 1980s: Connectionism (brain-inspired ML)
  - Connected networks of simple units.
    - Use parallel computation and distributed representations.
  - Adding hidden layers (z_i) increases expressive power.
    - With 1 layer and enough sigmoid units, a universal approximator.
  - Success in optical character recognition Multiple layers.







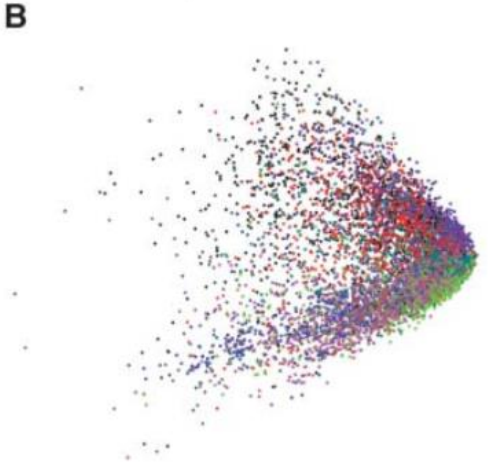
ML and Deep Learning History

- 1990s and early-2000s: drop in popularity.
  - It proved really difficult to get multi-layer models working robustly.
  - We obtained similar performance with simpler models:
    - Rise in popularity of logistic regression and SVMs with regularization and kernels.
  - ML moved closer to other fields:
    - Reproducing-kernel Hilbert spaces.
    - Convex optimization.
    - Probabilistic graphical models.
    - Bayesian methods.

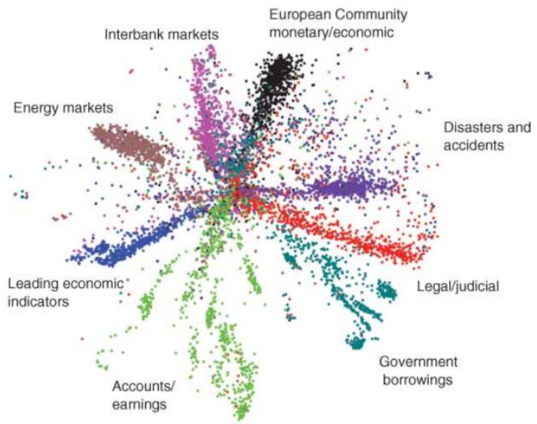
ML and Deep Learning History

- Late 2000s: push to revive connectionism as "deep learning".
  - Canadian Institute For Advanced Research (CIFAR) NCAP program:
    - "Neural Computation and Adaptive Perception".
    - Led by Geoff Hinton, Yann LeCun, and Yoshua Bengio ("Canadian mafia").
  - Unsupervised successes: deep belief networks and autoencoders.
    - Could be used to initialize deep neural networks.

PCA



Autoencoder



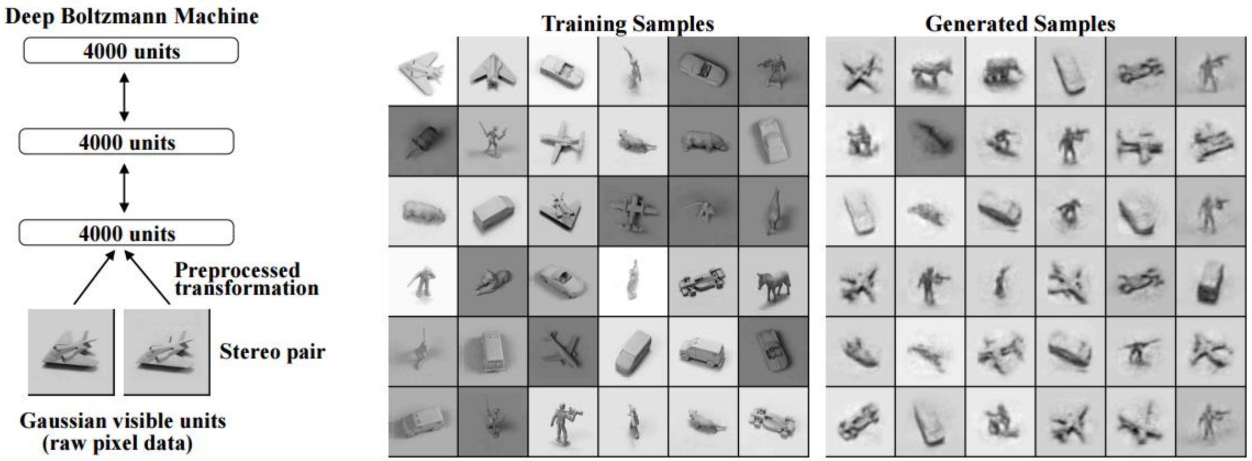
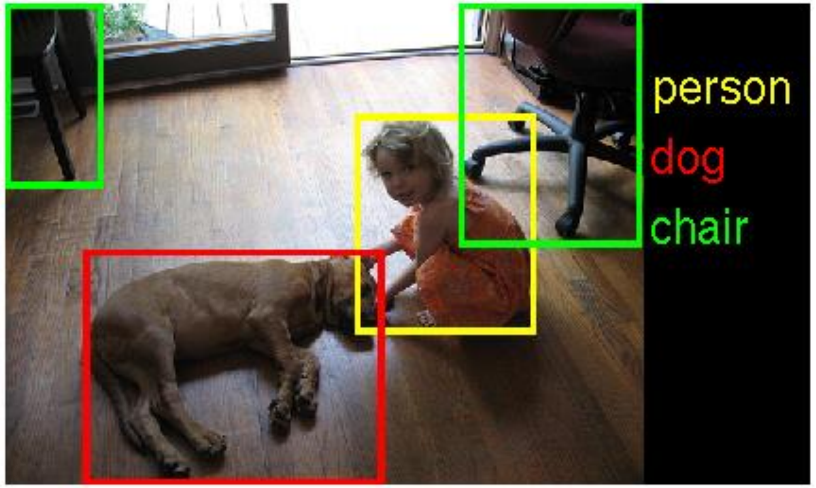


Figure 5: **Left:** The architecture of deep Boltzmann machine used for NORB. **Right:** Random samples from the training set, and samples generated from the deep Boltzmann machines by running the Gibbs sampler for 10,000 steps.

https://www.youtube.com/watch?v=KuPai0ogiHk

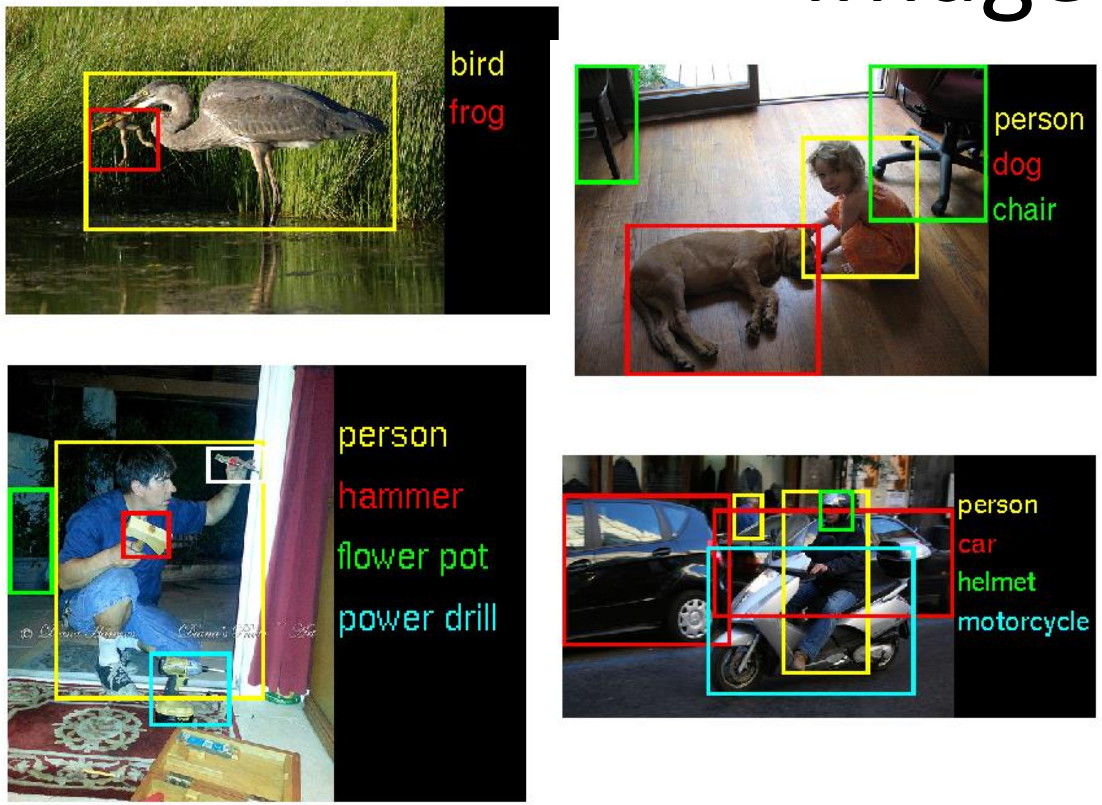
2010s: DEEP LEARNING!!!

- Bigger datasets, bigger models, parallel computing (GPUs/clusters).
  - And some tweaks to the models from the 1980s.
- 2009: huge improvement in speech recognition.
  - All phones use deep learning.
- 2012: huge improvement in computer vision.
  - This is now finding its way into products.



- Media hype:
  - "How many computers to identify a cat? 16,000", New York Times (2012).
  - "Why Facebook is teaching its machines to think like humans", Wired (2013).
  - "What Is 'deep learning' and why should businesses care?", Forbes (2013).
  - "Computer eyesight gets a lot more accurate" New York Times (2014).
- 2015: huge improvement in language understanding.

ImageNet Challenge



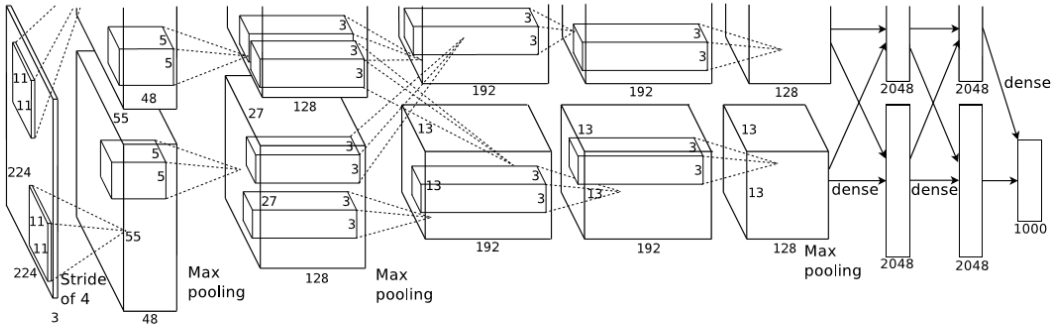
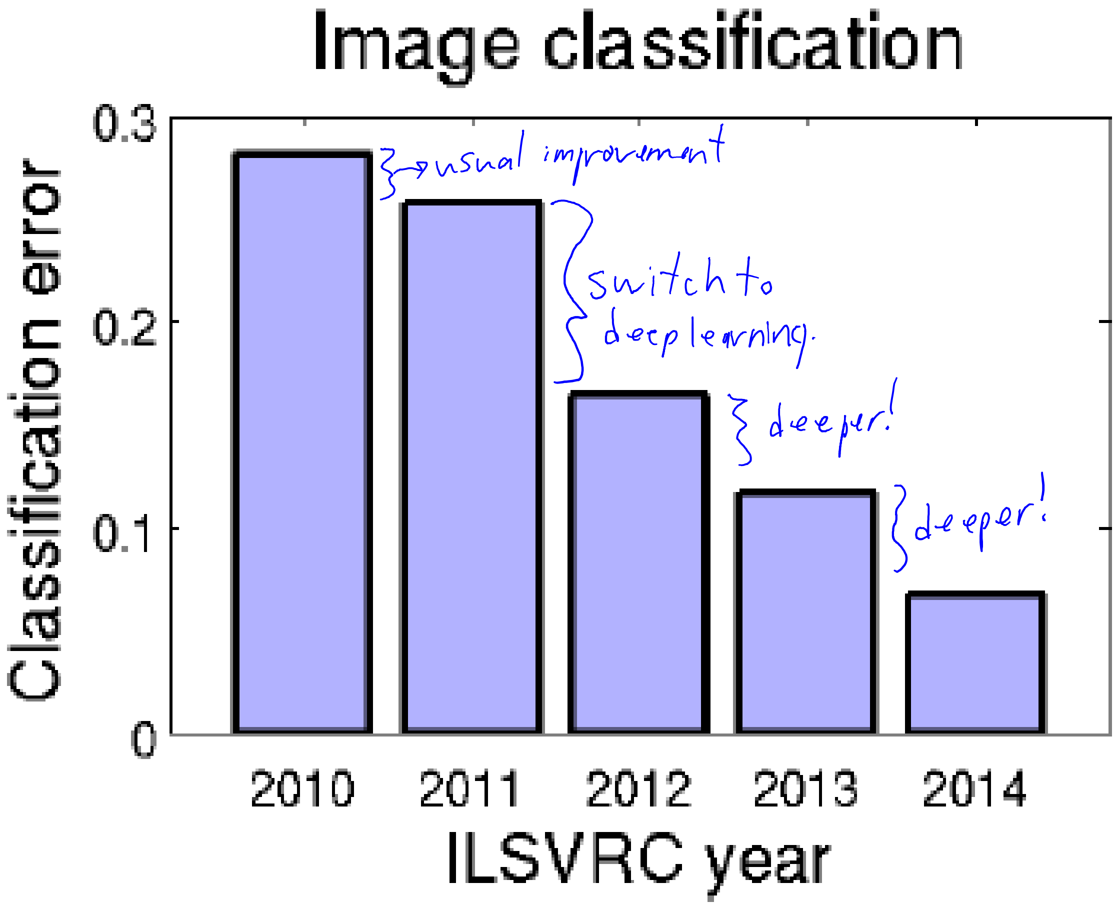
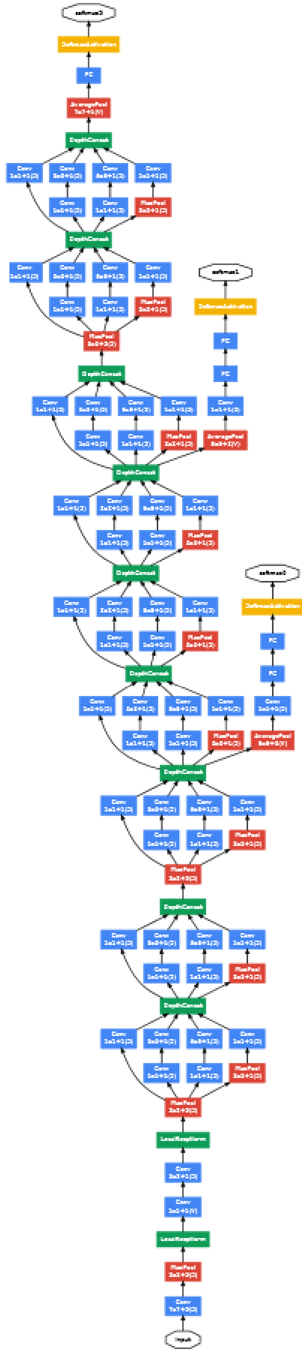


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts





Deep Learning Tricks

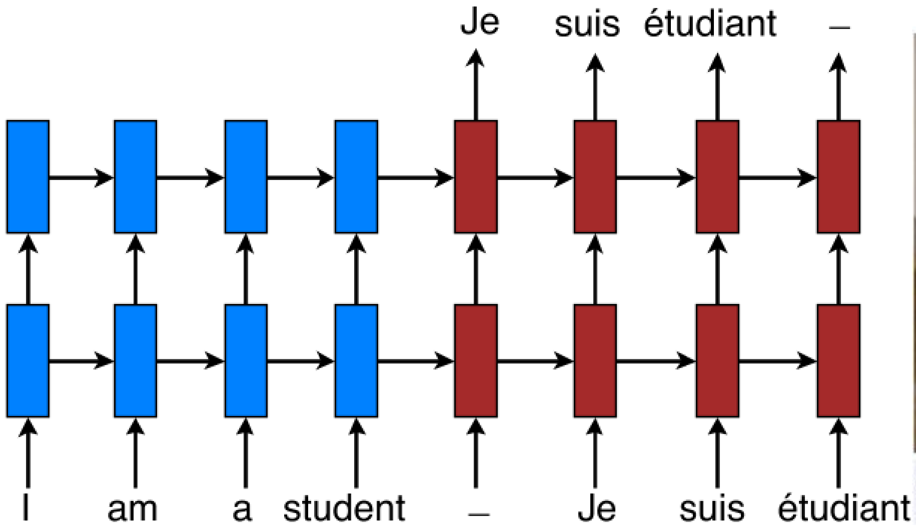
- Standard training method is stochastic gradient (SG):
  - Getting SG to work for convex problems is tricky.
  - For deep neural networks, naïve methods do not work well.
- Are local mimima the problem?
  - There is some empirical/theoretical evidence that local minima are good.
  - But naïve stochastic gradient often does not even find local mimima.
    - Most time is spent near saddle points.
- We've discovered 'tricks' to train deep models:
  1. Different non-linear transformations (ReLU).
  2. Step-size strategies.
  3. Regularization, early stopping, dropout.
  4. Initialization.
  5. Special network structures (CNNs, RNNs).
  6. Batch normalization.

Applications

- Mimicking artistic styles and inceptionism:

Recursive Neural Networks and Beyond

- Recursive neural networks: doing amazing things with sequences.

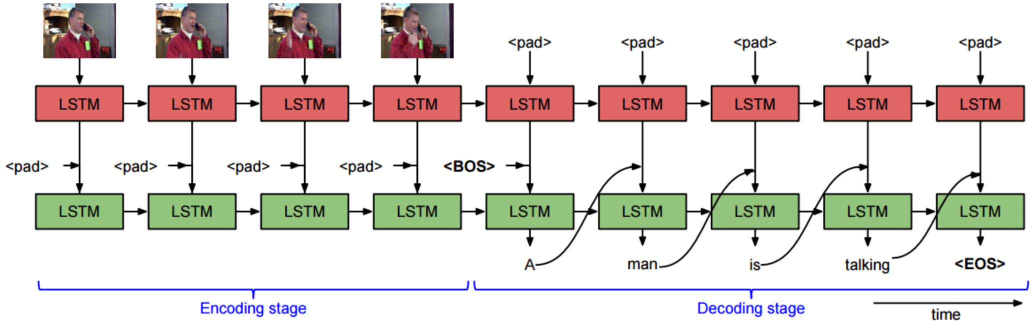


a cat is sitting on a toilet seat
logprob: -7.79

a display case filled with lots of different types of donuts
logprob: -7.78

a group of people sitting at a table with wine glasses
logprob: -6.71

- Some interesting recent advances:
  - Methods with explicit read/write memory.
  - Generative adversarial models.



https://www.youtube.com/watch?v=mLxsbWAYIpw

Schedule

Date	Topic	Presenter
Sept 21	Motiation/Overview	Mark
Sept 28	Feedforward neural nets, backpropagation	Julie
Oct 5	Network-independent tricks (initialization, regularization, early stopping, dropout, batch normalization,…)	Mohamed
Oct 12	ImageNet tricks (CNNs, inception modules, highway networks, residual networks,…)	Issam
Oct 19	Graphical models (Boltzmann machines, deep belief networks)	Jason
Oct 26	Artistic style transfer	Saif
Nov 2	Recursive neural nets (LSTMs, GRUs)	Nasim
Nov 9	Recursive neural nets 2 (pixel-recurrent, decoupled)	Stephen/Kevin
Nov 16	Variational autoencoders and Bayesian dark knowledge	Ricky
Nov 23	Generative adversarial networks	Reza
Nov 30	Memory nets, neural Turing, stack-augmented RNNs	Alireza