Deep Learning for AI

Yoshua Bengio

August 28th, 2017 @ DS3

Data Science Summer School

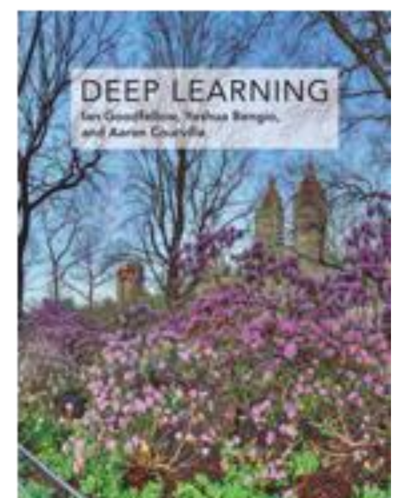
MILA

CIFAR CANADIAN INSTITUTE FOR ADVANCED RESEARCH

Université de Montréal

IVADO

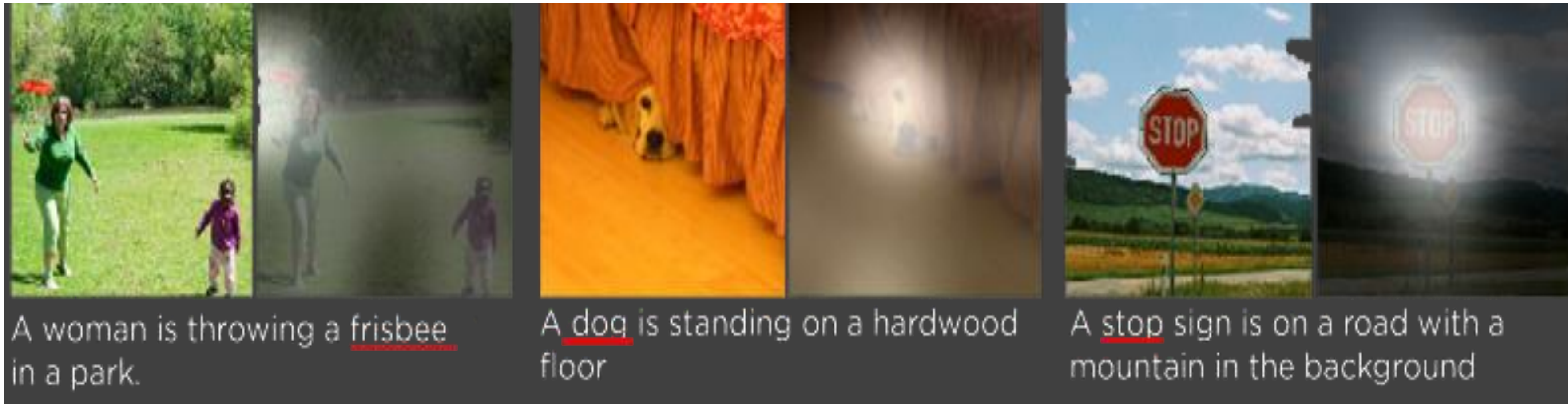
PLUG: **Deep Learning**, MIT Press book is out, chapters will remain online



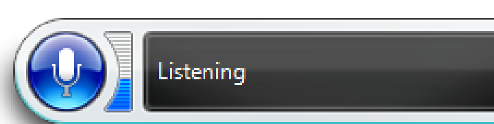
A new revolution seems to be in the work after the industrial revolution.

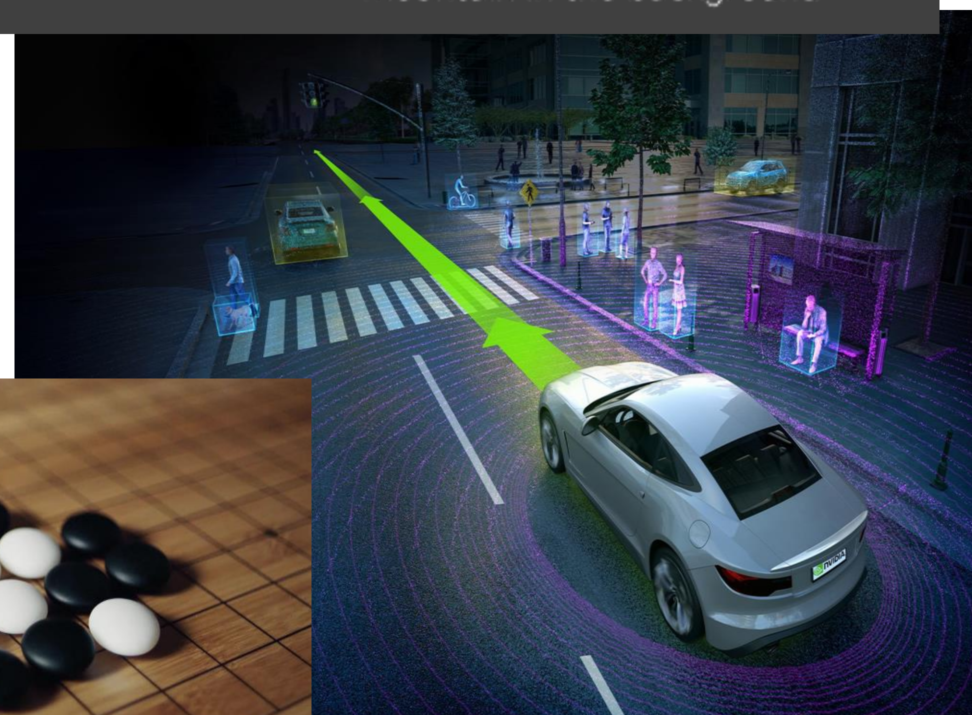
And Machine Learning, especially Deep Learning, is at the epicenter of this revolution.

Deep Learning Breakthroughs



Computers have made huge strides in perception, manipulating language, playing games, reasoning, ...





Intelligence Needs Knowledge

- Learning:

  powerful way to transfer knowledge to intelligent agents

  - Failure of classical AI: a lot of knowledge is **intuitive**

  - Solution: get knowledge from data & experience

Machine Learning, AI & No Free Lunch

- Five key ingredients for ML towards AI
  1. Lots & lots of data
  2. Very flexible models
  3. Enough computing power
  4. Computationally efficient inference
  5. **Powerful priors that can defeat the curse of dimensionality**

Bypassing the curse of dimensionality

We need to build compositionality into our ML models

Just as human languages exploit compositionality to give representations and meanings to complex ideas

Exploiting compositionality can give an **exponential** gain in representational power

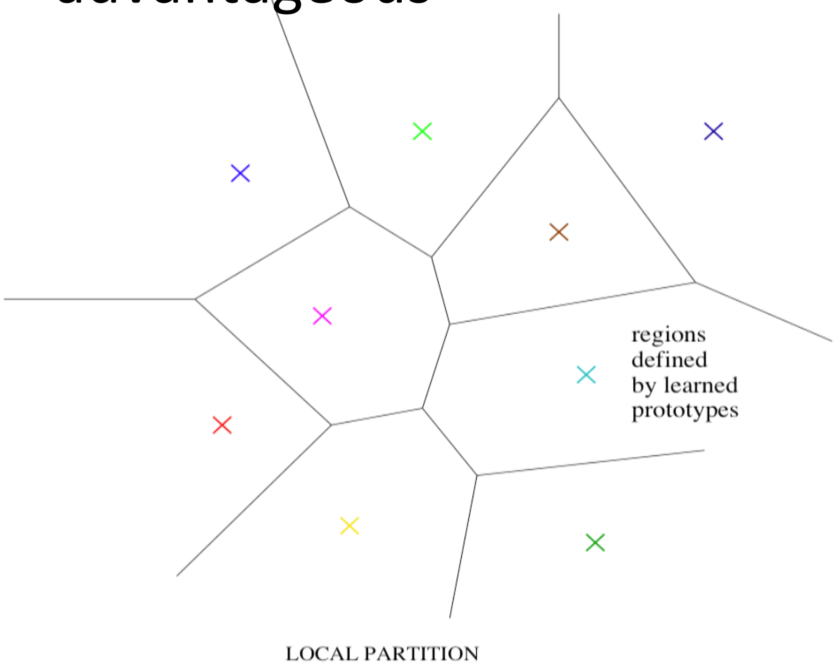
Distributed representations / embeddings: feature learning

Deep architecture: multiple levels of feature learning

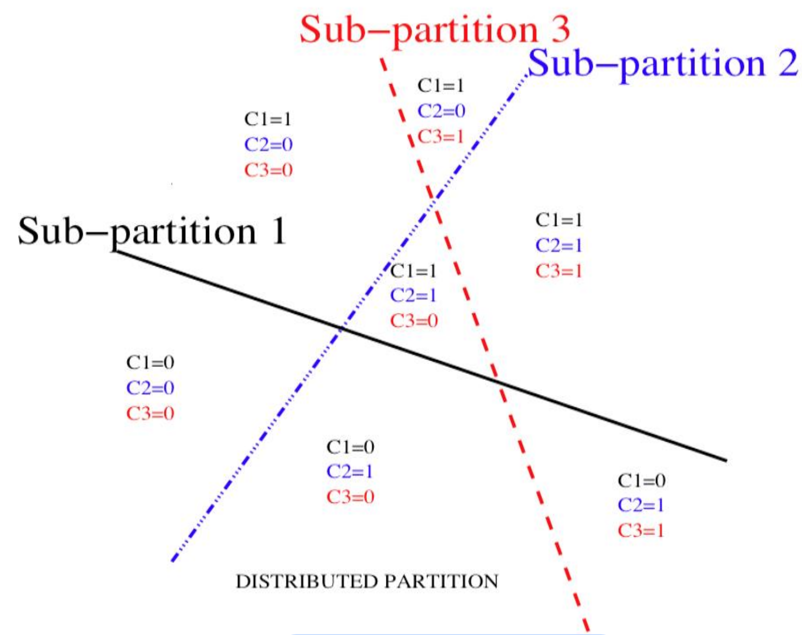
Prior assumption: compositionality is useful to describe the world around us efficiently

Distributed Representations: The Power of Compositionality – Part 1

- Distributed (possibly sparse) representations, learned from data, can capture the **meaning** of the data and state
- Parallel composition of features: can be exponentially advantageous



Not Distributed



Distributed

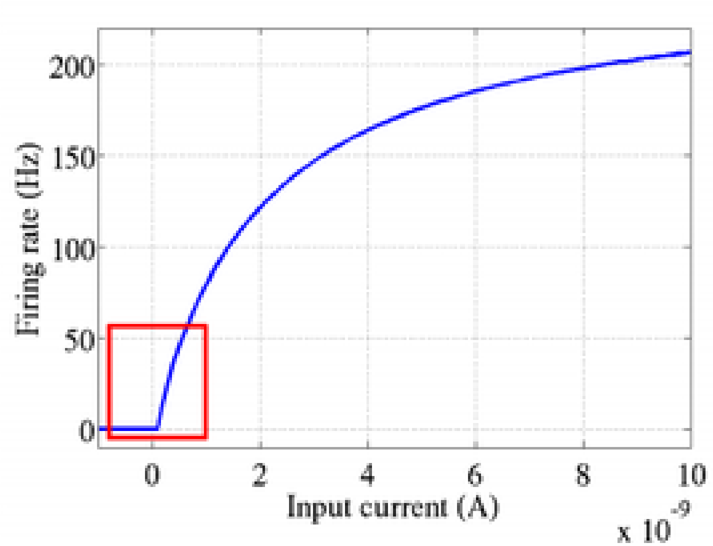
Deep Representations: The Power of Compositionality – Part 2

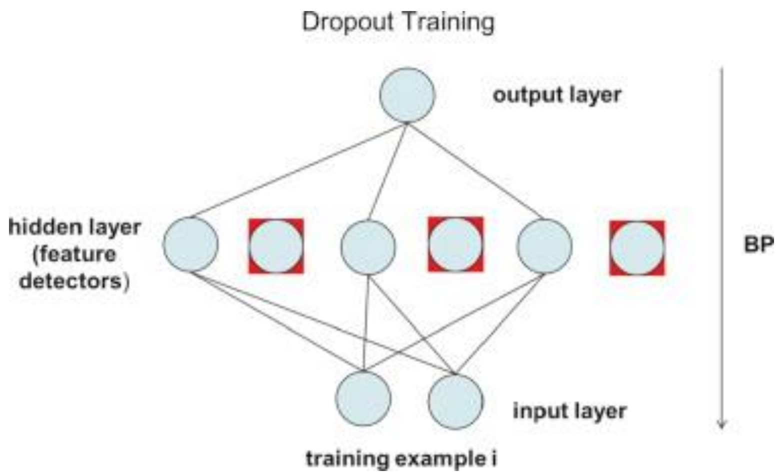
- Learned function seen as a composition of simpler operations, e.g. inspired by neural computation
- Hierarchy of features, concepts, leading to more abstract factors enabling better generalization
- Again, theory shows this can be exponentially advantageous

Why multiple layers? The world is compositional

Anything New with Deep Learning since the Neural Nets of the 90s?

- Rectified linear units instead of sigmoids, enable training much deeper networks by backprop (Glorot & Bengio AISTATS 2011)

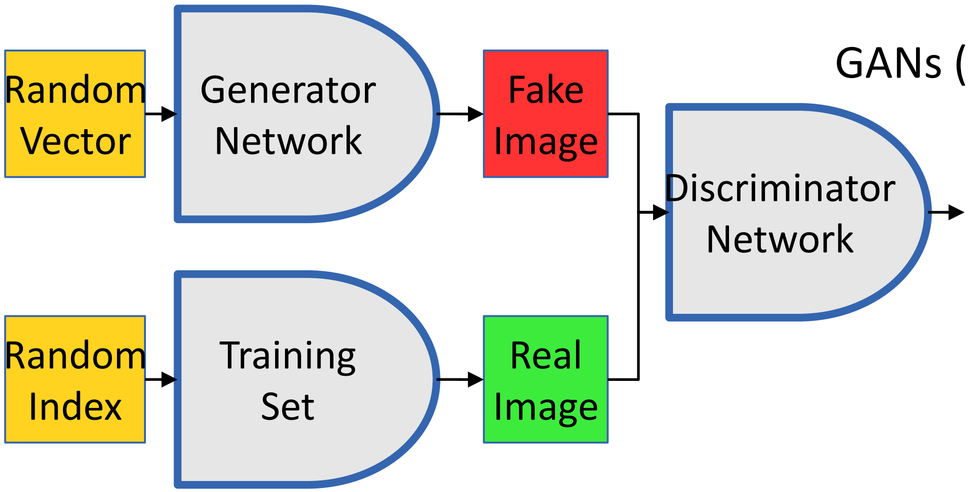




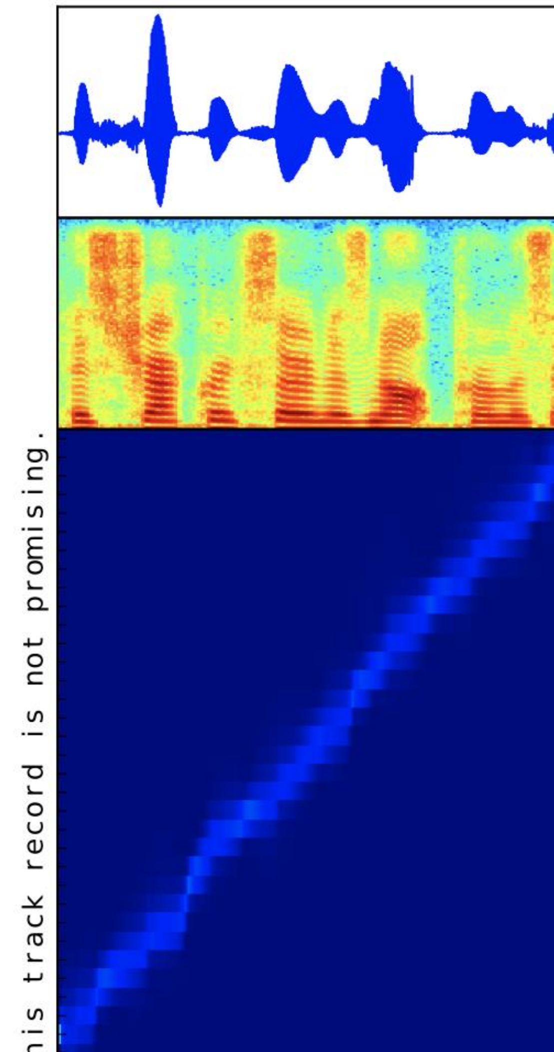
- Some forms of noise (like dropout) are powerful regularizers yielding superior generalization abilities
- Success of deep convnets trained on large labeled image datasets
- Success of recurrent nets with more memory, with gating units
- Attention mechanisms liberate neural nets from fixed-size inputs

What's New with Deep Learning?

- Progress in unsupervised generative neural nets allows them to synthesize a diversity images, sounds and text imitating unlabeled images, sounds or text



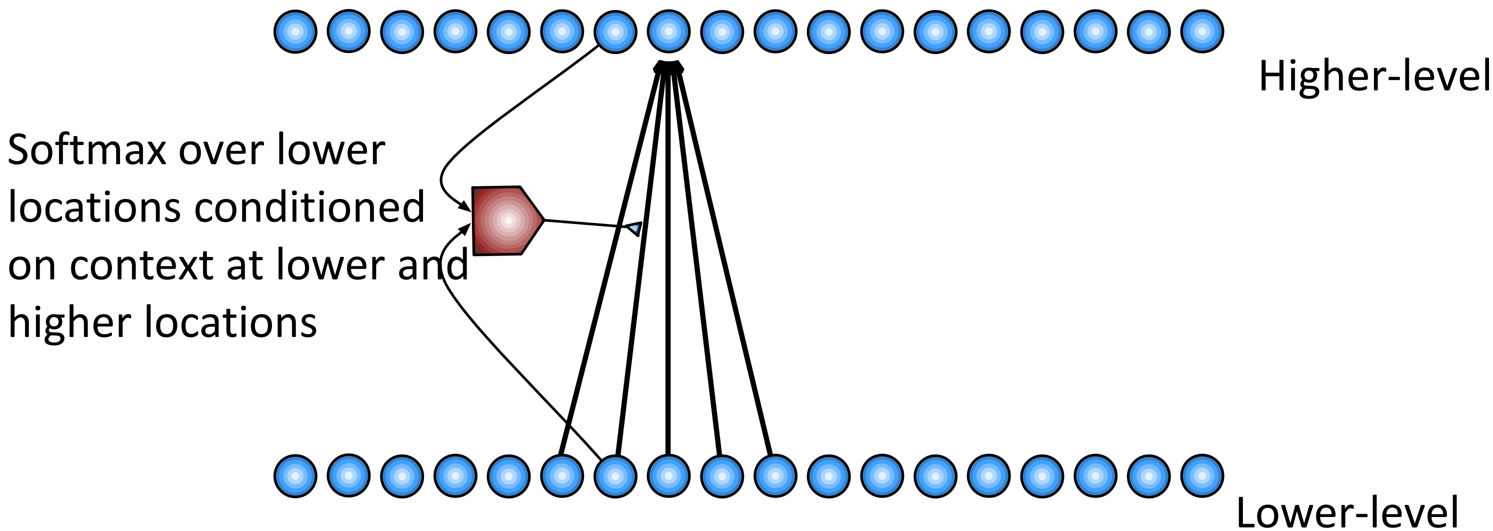
GANs (NIPS'2014)



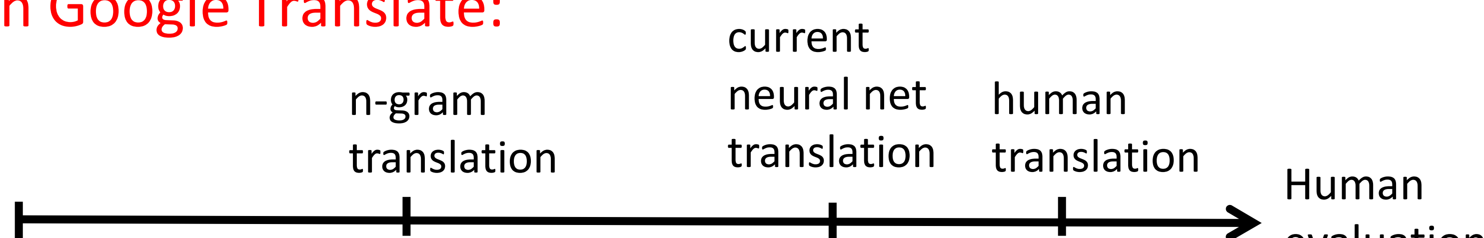
What's New with Deep Learning?

- Incorporating the idea of **attention, using GATING units,** has unlocked a breakthrough in machine translation:

Neural Machine Translation (ICLR'2015)

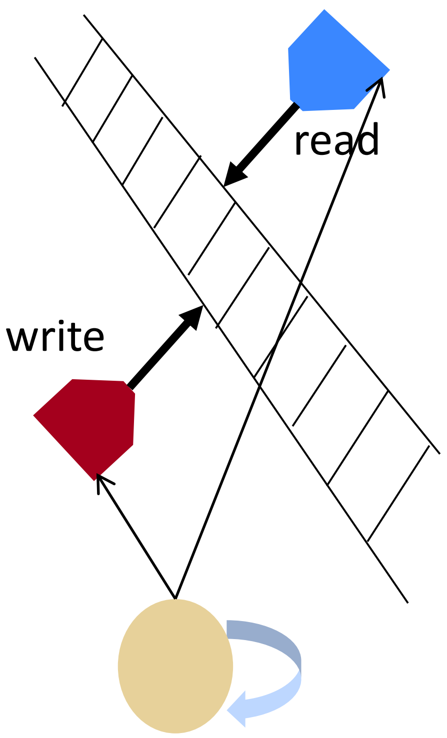


- Now in Google Translate:



What's New with Deep Learning?

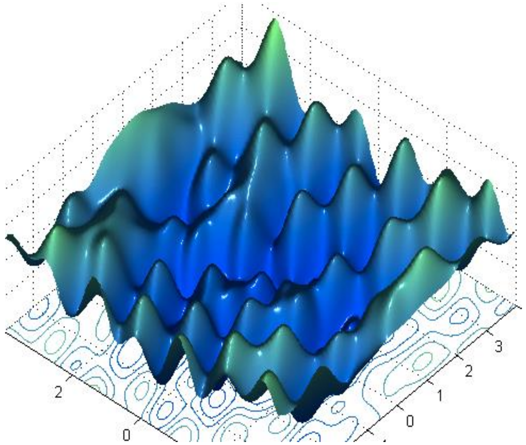
- Attention has also opened the door to neural nets which can write to and read from a memory
  - 2 systems:
    - Cortex-like (state controller and representations)
      - System 1, intuition, fast heuristic answer
    - Hippocampus-like (memory) + prefrontal cortex
      - System 2, slow, logical, sequential



- Memory-augmented networks gave rise to
  - Systems which reason
    - Sequentially combining several selected pieces of information (from the memory) in order to obtain a conclusion
  - Systems which answer questions
    - Accessing relevant facts and combining them

We are starting to better understand why deep learning is working

- **Generalization**:
  ICLR'2014
  - Distributed representations: (up to) exponential statistical advantage, **if the world is compositional**
  - Depth, multiple layers: similar story, on top
  NIPS'2014



- **Optimization**: *MYTHS BUSTED* NIPS'2014
  - Non-convexity & local min of the objective fn: not a curse
  - Stochastic gradient descent is very efficient
  - Additional human-inspired tricks: curriculum learning (ICML'2009)

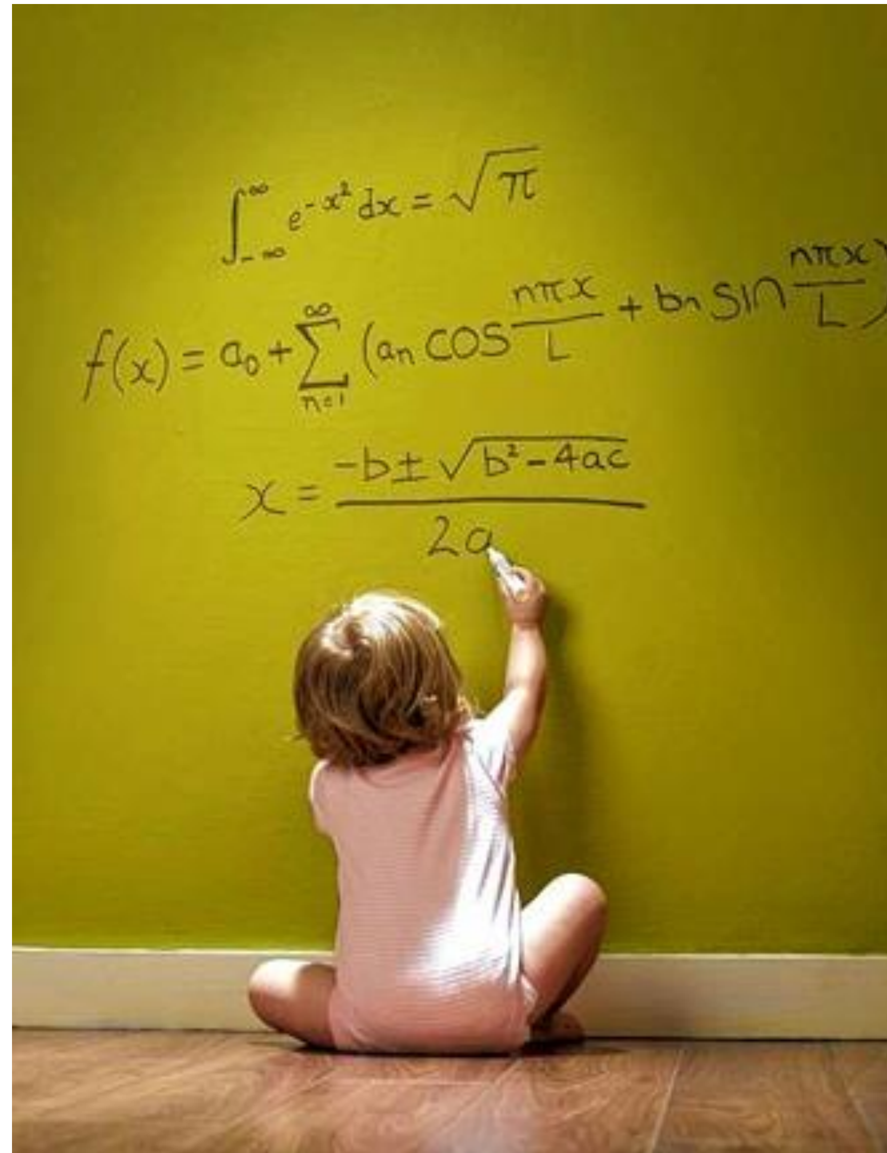
Still Far from Human-Level AI

- Industrial successes mostly based on **supervised** learning

- Learning superficial clues, not generalizing well outside of training contexts, easy to fool trained networks:
  - Current models cheat by picking on surface regularities
- Still unable to discover higher-level abstractions at multiple time scales, very long-term dependencies
- Still relying heavily on smooth differentiable predictors (using backprop, the workhose of deep learning)

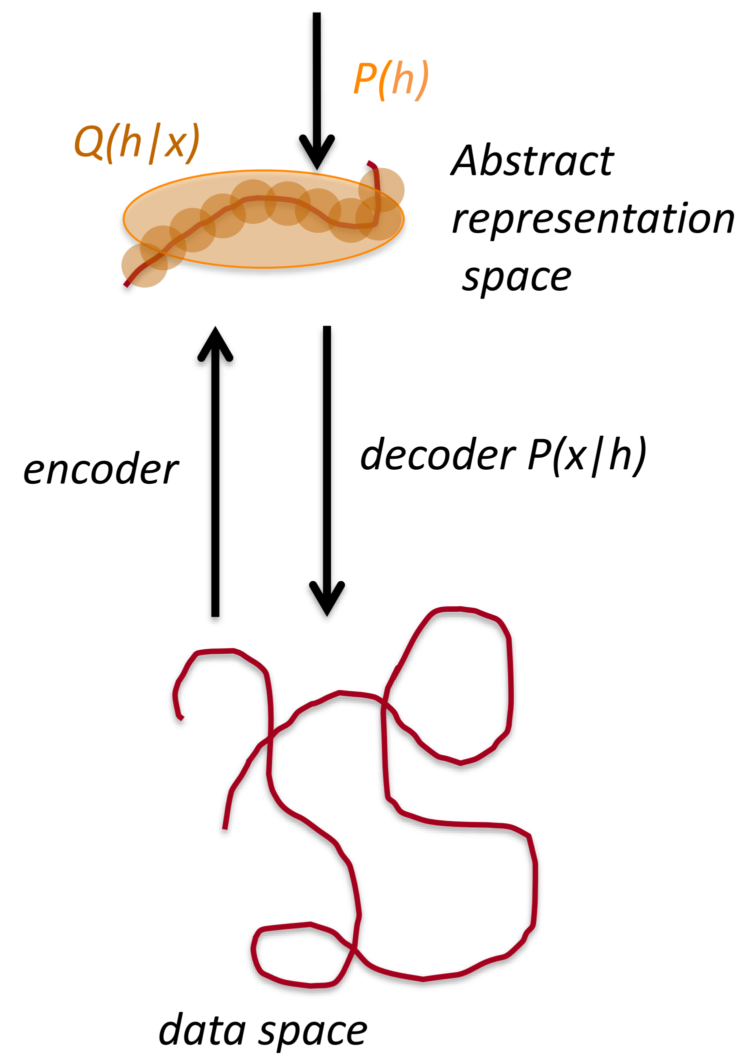
Humans outperform machines at unsupervised learning

- Humans are very good at unsupervised learning, e.g. a 2 year old knows intuitive physics
- Babies construct an approximate but sufficiently reliable model of physics, how do they manage that? Note that they interact with the world, not just observe it.



Latent Variables and Abstract Representations

- Encoder/decoder view: maps between low & high-levels
- Encoder does inference: interpret the data at the abstract level
- Decoder can generate new configurations
- Encoder flattens and disentangles the data manifold

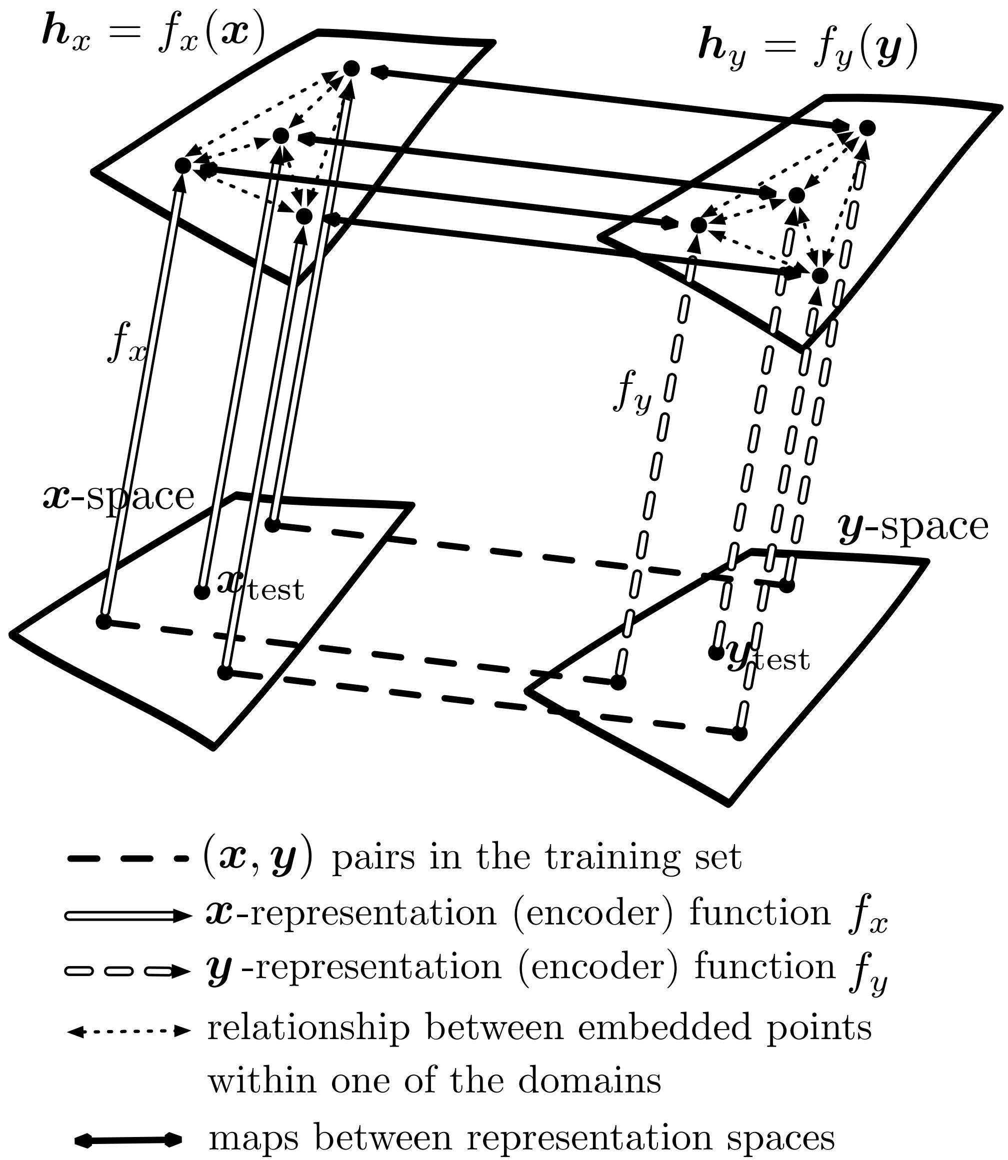


Maps Between Representations

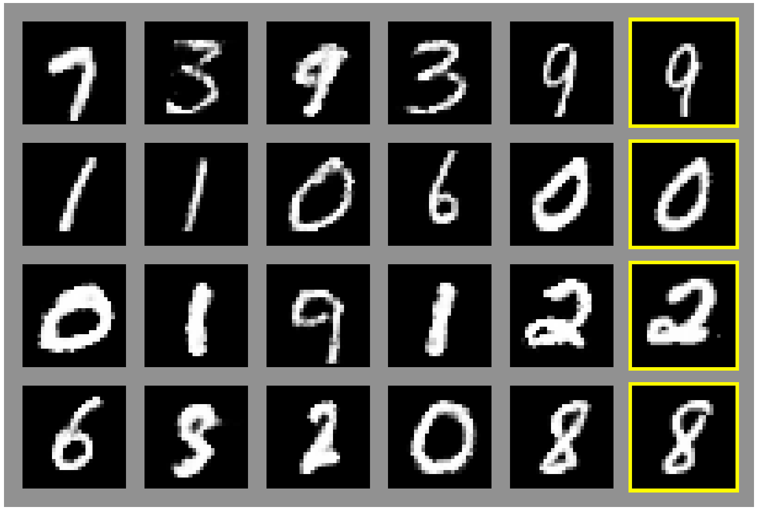
x and *y* represent different modalities, e.g., image, text, sound…

Can provide 0-shot generalization to new categories (values of *y*)

(Larochelle et al AAAI 2008)



Early Days of GAN Samples



MNIST

TFD

CIFAR-10 (fully connected)

CIFAR-10 (convolutional)

Convolutional GANs

(Radford et al, arXiv 1511.06343)

Strided convolutions, batch normalization, only convolutional layers, ReLU and leaky ReLU

GAN: Interpolating in Latent Space

If the model is good (unfolds the manifold), interpolating between latent values yields plausible images.

Combining Iterative Sampling from
Denoising Auto-Encoders with GAN

Plug & Play Generative Networks: Conditional Iterative Generation of Images in Latent Space

Anh Nguyen, Jason Yosinski, Yoshua Bengio, Alexey Dosovitskiy, Jeff Clune

(submitted to CVPR 2017) arXiv:1612.00005

227 x 227 ImageNet GENERATED IMAGES of category Volcano

(cheatinga bit by using lots of labeled data during training)

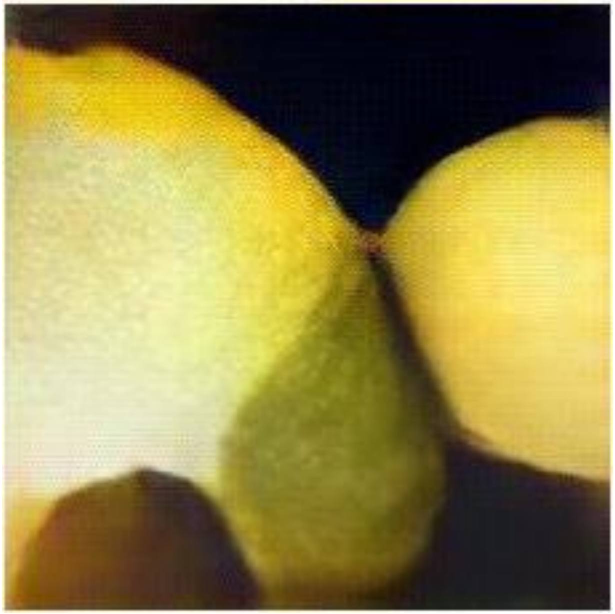
Plug & Play Generative Networks

High-Resolution
Samples
227 x 227

bird

ant

volcano

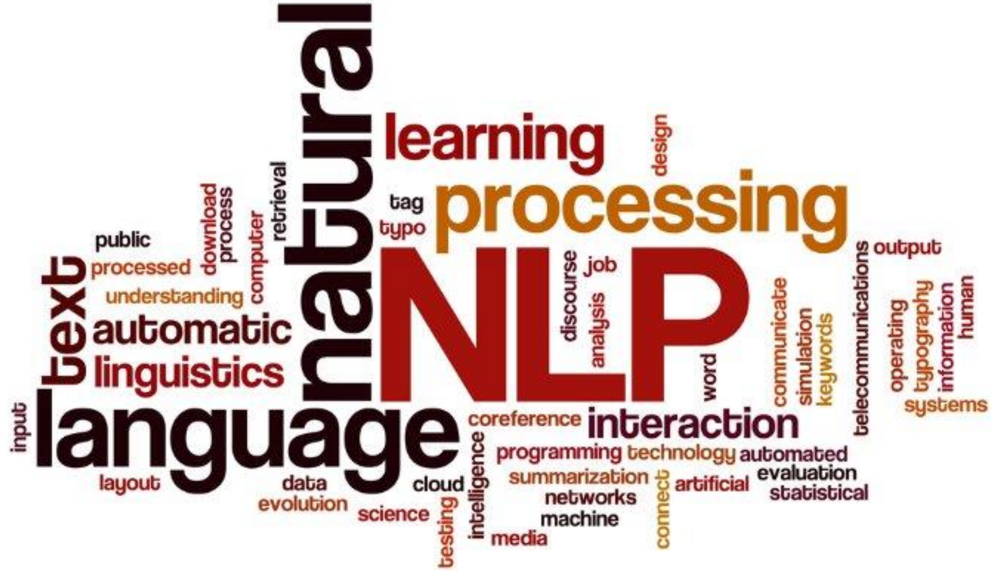


lemon

What's Missing

- More autonomous learning, better **unsupervised learning**
- Discovering the **underlying causal factors**
- Model-based RL which extends to completely new situations by **unrolling powerful predictive models which can help reason about rarely observed dangerous states**
- Sufficient **computational power** for models large enough to capture human-level knowledge

What's Missing

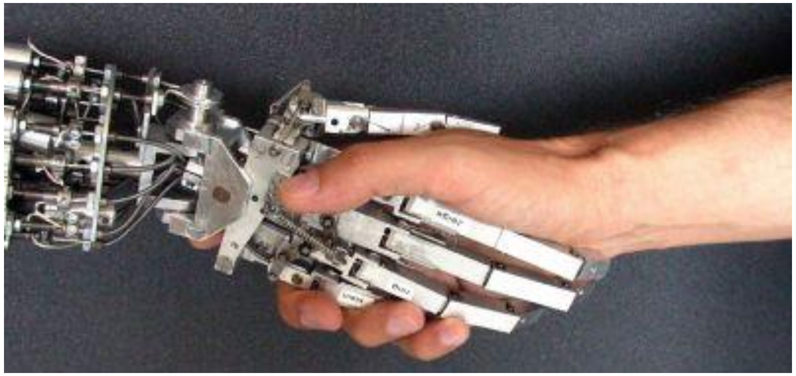


- Autonomously discovering **multiple time scales to handle very long-term dependencies**
- Actually **understanding language** (also solves generating), requiring enough world knowledge / commonsense
    - Neural nets which really understand the notions of object, agent, action, etc.
- Large-scale **knowledge representation** allowing one-shot learning as well as discovering new abstractions and explanations by 'compiling' previous observations

Acting to Guide Representation Learning

- *What is a good latent representation?*
- Disentangling the underlying factors of representation so that computers make sense of the world
- **Some factors (e.g. objects) correspond to 'independently controllable' aspects of the world**
- *Can only be discovered by acting in the world*

The Future of Deep AI



- Scientific progress is slow and continuous, but social and economic impact can be disruptive
- Many fundamental research questions are in front of us, with much uncertainty about when we will crack them, but we will
- Importance of continued investment in basic & exploratory AI research, for both practical (recruitment) short-term and long-term reasons
- Let us continue to keep the field open and fluid, be mindful of social impacts, and make sure AI will bloom for the benefit of all

Montreal Institute for Learning Algorithms

MILA

Université de Montréal