

UBC MLRG (Fall 2017):
Deep Learning meets Graphical Models

Machine Learning Reading Group (MLRG)

- Machine learning reading group (MLRG) format:
 - Each semester we pick a general topic.
 - Each week someone leads us through a tutorial-style lecture/discussion.
 - So it's organized a bit more like a "topics course" than reading group.
- We use this format because ML has become a huge field.

Machine Learning Reading Group (MLRG)

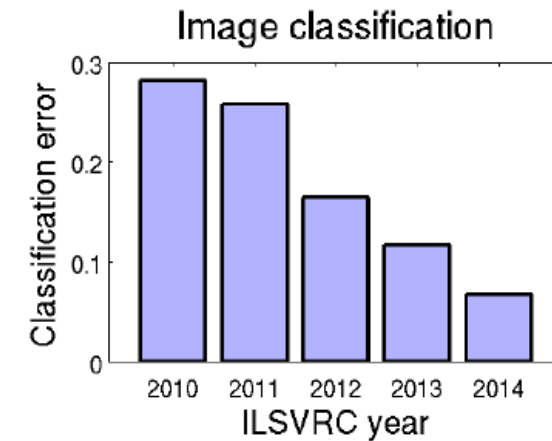
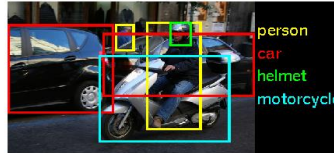
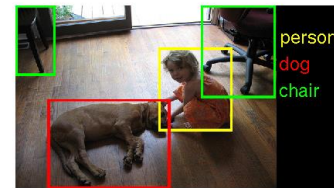
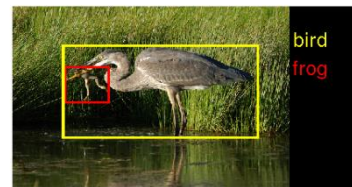
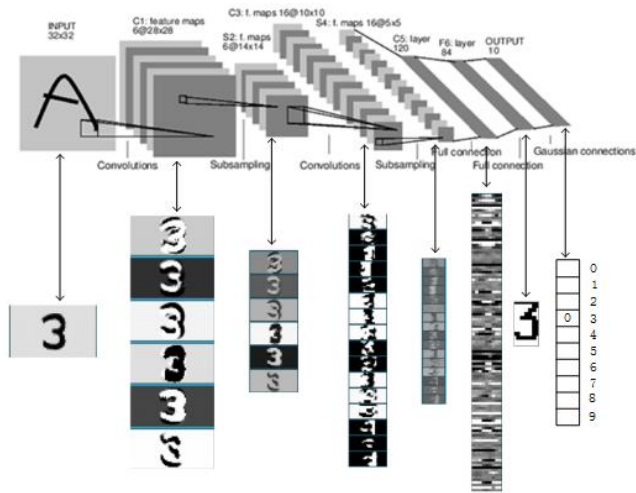
- I've tried to pack as much as possible into the two ML courses:
 - CPSC 340 covers most of the most-useful methods.
 - CPSC 540 covers most of the background needed to read research papers.
- This reading group covers **topics that aren't yet in these course**.
 - Aimed at people who have taken CPSC 340, and are comfortable with 540-level material.
- This may change as we start offering more ML courses next term...

Recent MLRG History

- Topics covered in recent tutorial-style MLRG sessions:
 - Summer 2015: Probabilistic graphical models.
 - Fall 2015: Convex optimization.
 - Winter 2016: Bayesian statistics.
 - Summer 2016: Miscellaneous.
 - Fall 2016: Deep learning.
 - Winter 2016: Reinforcement learning.
 - Summer 2017: Online, active, and causal learning.
 - Fall 2017: [Deep learning meets graphical models.](#)

Why “Deep Learning meets Graphical Models”

- **Deep learning** is one of the hottest topics in machine learning.
 - Recent non-trivial improvements in speech, vision, and language.
 - Variety of other interesting applications (artistic style, image colouring).




- **Learning features** rather than hand-tuning them.
- A huge/complicated **non-linear transform** of data.

Why “Deep Learning meets Graphical Models”

- 10 years ago, some of the hottest topics were:
 - Structured prediction:
 - Used to model **dependencies** between random variables.
 - Bayesian learning:
 - Used to model **uncertainty** in statistical inferences.
- Common theme:
 - **Modeling whole distributions** rather than just input:output mappings.
- Significant recent attention on **deep learning for distributions**.

Structured Prediction Examples

Classical **supervised learning** focuses on predicting single discrete/continuous label:

Input: 

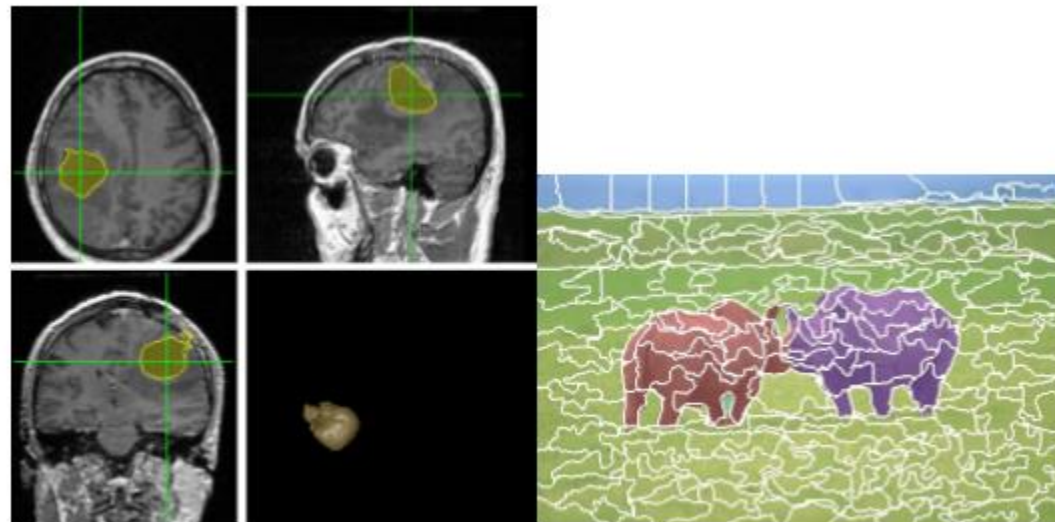
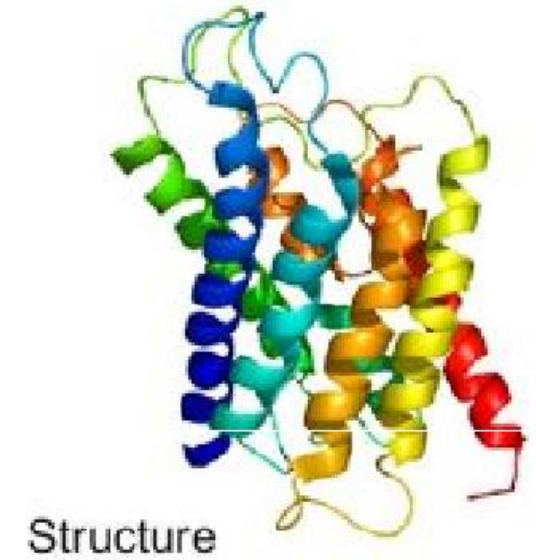
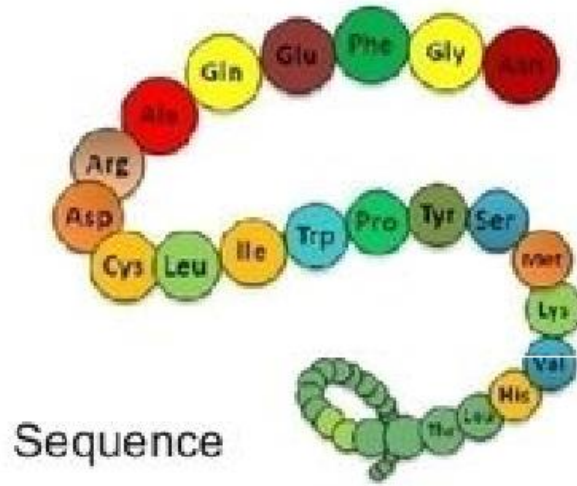
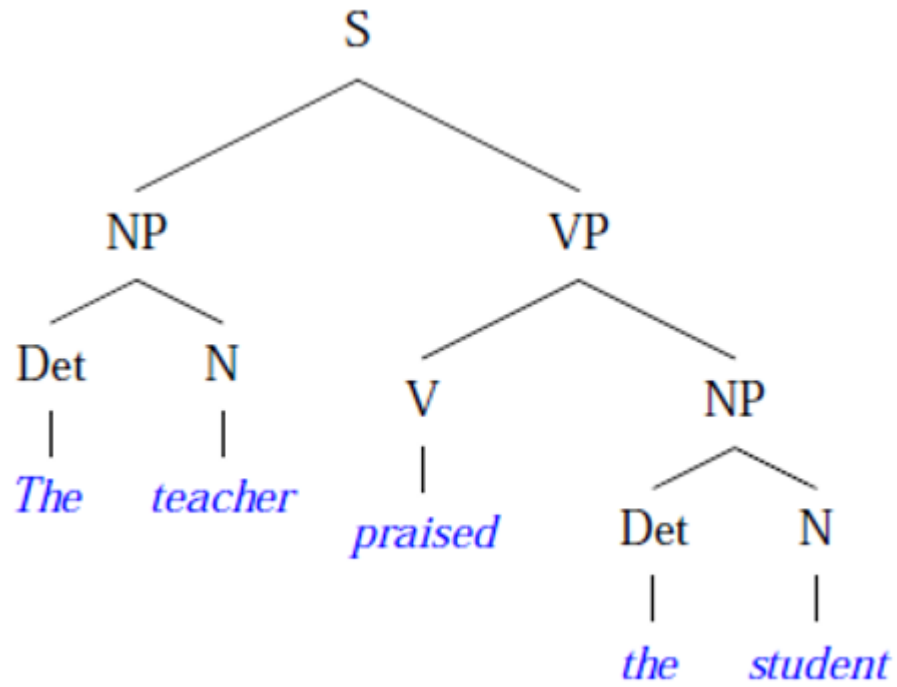
Output: "P"

Structured prediction allows **general objects** as labels:

Input: 

Output: "Paris"

Structured Prediction Examples

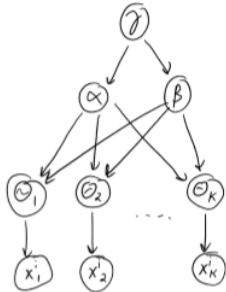


Bayesian Modeling Examples

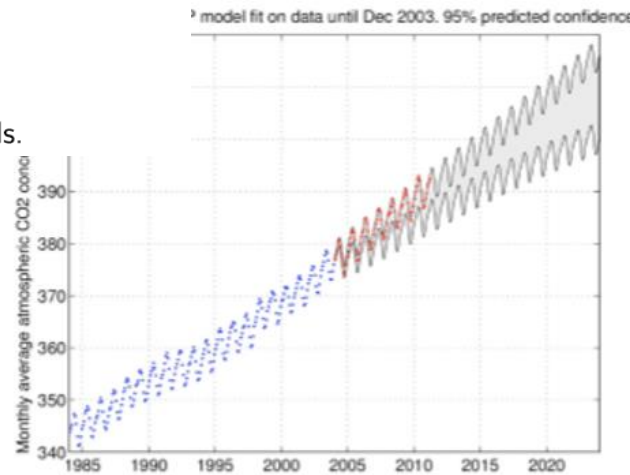
- Bayesian framework allows us to specify probabilities in hierarchical settings, and allow confidence in predictions.

Hierarchical Bayes for Non-IID Data

- Consider treating α and β as random variables and using a hyperprior:



- Now there is a dependency between the different θ_j .
- You combine the non-IID data across different hospitals.
- Data-rich hospitals inform posterior for data-poor hospitals.



Topics

| | |
|---------|------|
| gene | 0.04 |
| dna | 0.02 |
| genetic | 0.01 |
| ... | |

| | |
|----------|------|
| life | 0.02 |
| evolve | 0.01 |
| organism | 0.01 |
| ... | |

| | |
|--------|------|
| brain | 0.04 |
| neuron | 0.02 |
| nerve | 0.01 |
| ... | |

| | |
|----------|------|
| data | 0.02 |
| number | 0.02 |
| computer | 0.01 |
| ... | |

Documents

Topic proportions and assignments

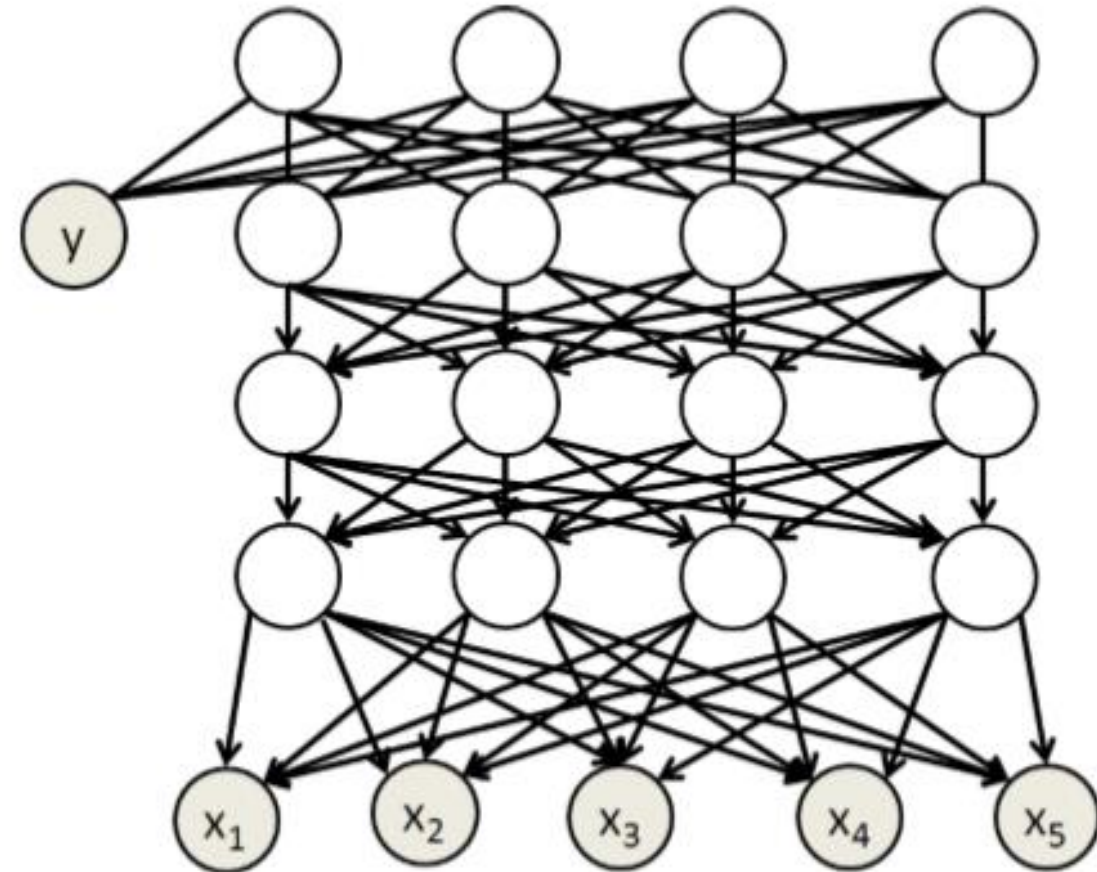
Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK— How many genes does an organism need to survive? Last week at the genome meeting here, two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough. Although the numbers don't match precisely, those predictions "are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson, a geneticist at the University in Sweden. "I'm amazed by the 800 number. It's particularly more and more genes are being discovered and sequenced. "It may be a way of organisms, any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an organism's genome to a reference genome can identify genes that are missing or duplicated. Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

Classic Deep Learning for Distributions

- 10 years ago there was work on deep learning for distributions:
 - Restricted Boltzmann machines.
 - Deep belief networks.
 - Deep Boltzmann machines.
- Graphical models with several layers of binary latent variables.
 - Trained using MCMC and SGD.
- Bayesian neural networks won NIPS “feature selection” challenge in 2003.



Modern Deep Learning for Distributions

- More recent approaches:
 - Fully-convolutional networks (possibly with CRF at the top).

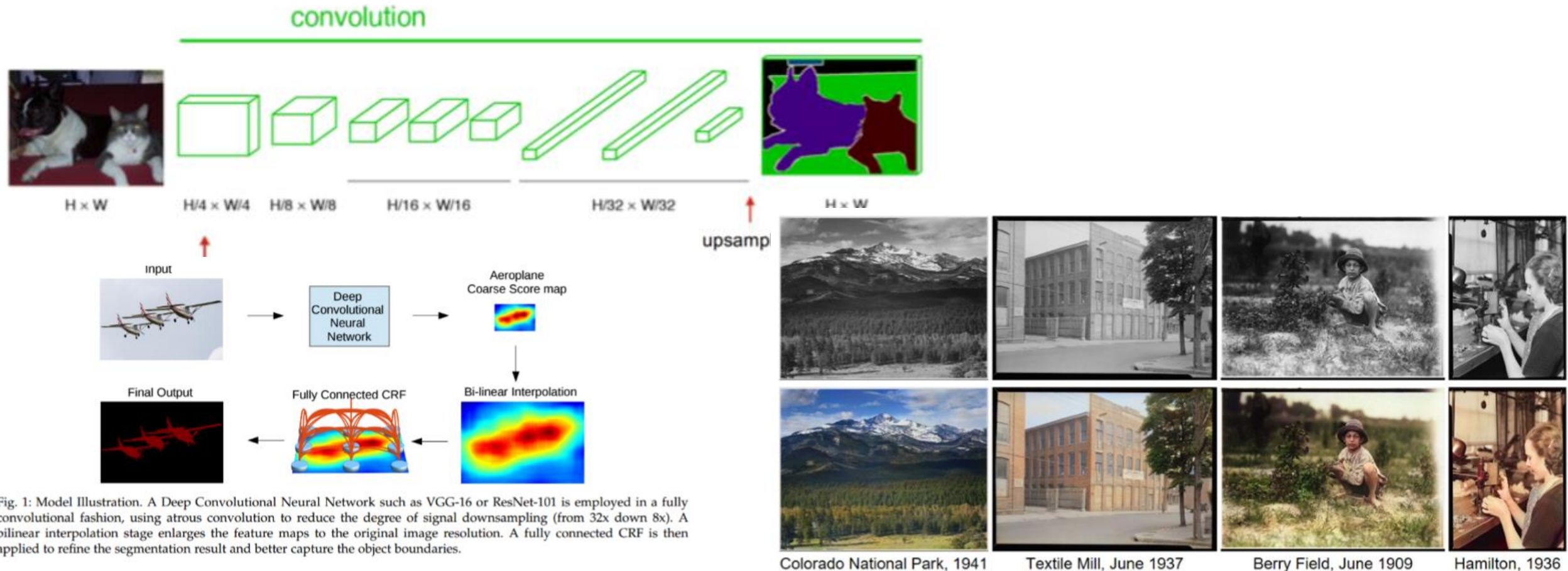


Fig. 1: Model Illustration. A Deep Convolutional Neural Network such as VGG-16 or ResNet-101 is employed in a fully convolutional fashion, using atrous convolution to reduce the degree of signal downsampling (from 32x down 8x). A bilinear interpolation stage enlarges the feature maps to the original image resolution. A fully connected CRF is then applied to refine the segmentation result and better capture the object boundaries.

Modern Deep Learning for Distributions

- More recent approaches:
 - Fully-convolutional networks (possibly with CRF at the top).
 - Recurrent and recursive neural networks that simulate chain/tree models.
 - <https://www.youtube.com/watch?v=LY7x2Ihqjmc>

And still I saw the Brooklyn stairs
With the shit, the ground, the golden haze
Of the frozen woods where the boat stood.
When I thought of shame and silence,
I was a broken skull;
I was the word which I called it,
And I saw the black sea still,
So long and dreary and true;
The way a square shook out my ground,
And the black things were worth a power,
To find the world in a world of reason,
And I saw how the mind saw me.



A man is sitting on the edge of the waters.
I should see him begin to stand at the throat of the graveyard
and my love is like a stairway in his left arm and a piece of the stairs,
and there is a girl in the doorway and she and I am a good time.
I want to see her the best thing with the footprints in the woods
and the candle shifts back to the shrine and the last late sun
the sky and the candle and the noise of the snow.

Modern Deep Learning for Distributions

- More recent approaches:
 - Fully-convolutional networks (possibly with CRF at the top).
 - Recurrent and recursive neural networks that simulate chain/tree models.
 - GANs and variational autodencers (density estimation).

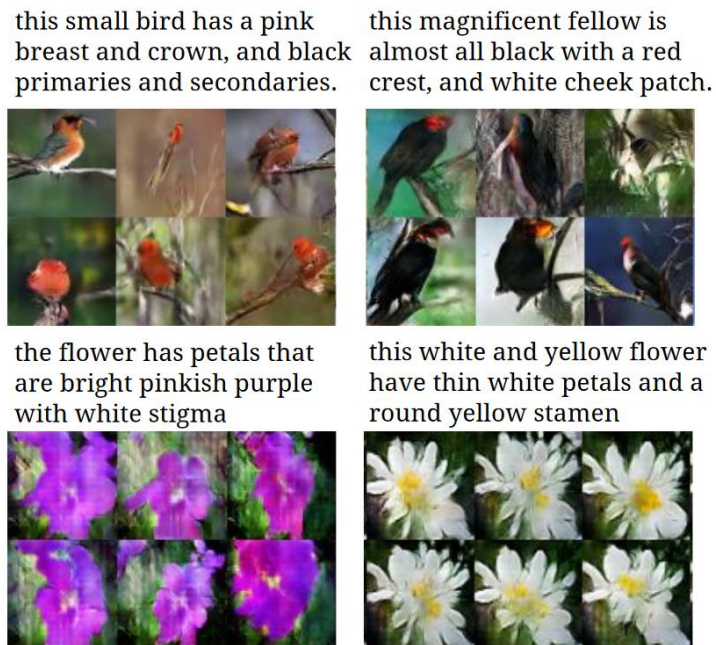


Figure 23: Text-to-image synthesis with GANs. Image reproduced from [Reed et al. \(2016b\)](#).



Figure 7: [Isola et al. \(2016\)](#) created a concept they called image to image translation, encompassing many kinds of transformations of an image: converting a satellite photo into a map, covering a sketch into a photorealistic image, etc. Because many of these conversion processes have multiple correct outputs for each input, it is necessary to use generative modeling to train the model correctly. In particular, [Isola et al. \(2016\)](#) use a GAN. Image to image translation provides many examples of how a creative algorithm designer can find several unanticipated uses for generative models. In the future, presumably many more such creative uses will be found.

Modern Deep Learning for Distributions

- More recent approaches:
 - Fully-convolutional networks (possibly with CRF at the top).
 - Recurrent and recursive neural networks that simulate chain/tree models.
 - GANs and variational autodencers (density estimation).
 - Various new flavour of Bayesian neural networks.
- Emphasis in most cases is a bit different from classic work:
 - **Feedforward network is used as tool** for generating non-linear transform.
 - Less effort on making network Bayesian, or making a deep graph model.
- Our goal: understand DL for distributions ideas (beyond 540).

Schedule

| Date | Topic | Presenter |
|--------|---|-------------|
| Sep 26 | Motivation/Overview | Mark |
| Oct 3 | FCNs and CRFs | Issam |
| Oct 10 | RNNs and language | Raunak |
| Oct 17 | Bayesian neural nets 1: sampling | Michael |
| Oct 24 | Bayesian neural nets 2: variational | Jason |
| Oct 31 | Variational autoencoders 1: basics | Devon |
| Nov 7 | Variational autoencoders 2: variations | Sharan |
| Nov 14 | Generative adversarial networks 1: basics | Mohamed |
| Nov 21 | Generative adversarial networks 2: variations | Alireza |
| Nov 28 | Beyond generative adversarial networks | Julie/Nasim |