Intro to **Deep Learning**

Prof. Joydeep Ghosh
ECE/UT
www.ideal.ece.utexas.edu/~ghosh
ghosh@ece.utexas.edu

Google I/O 2017 3:20

AI, Deep Learning, and Machine Learning: A Primer

by Frank Chen

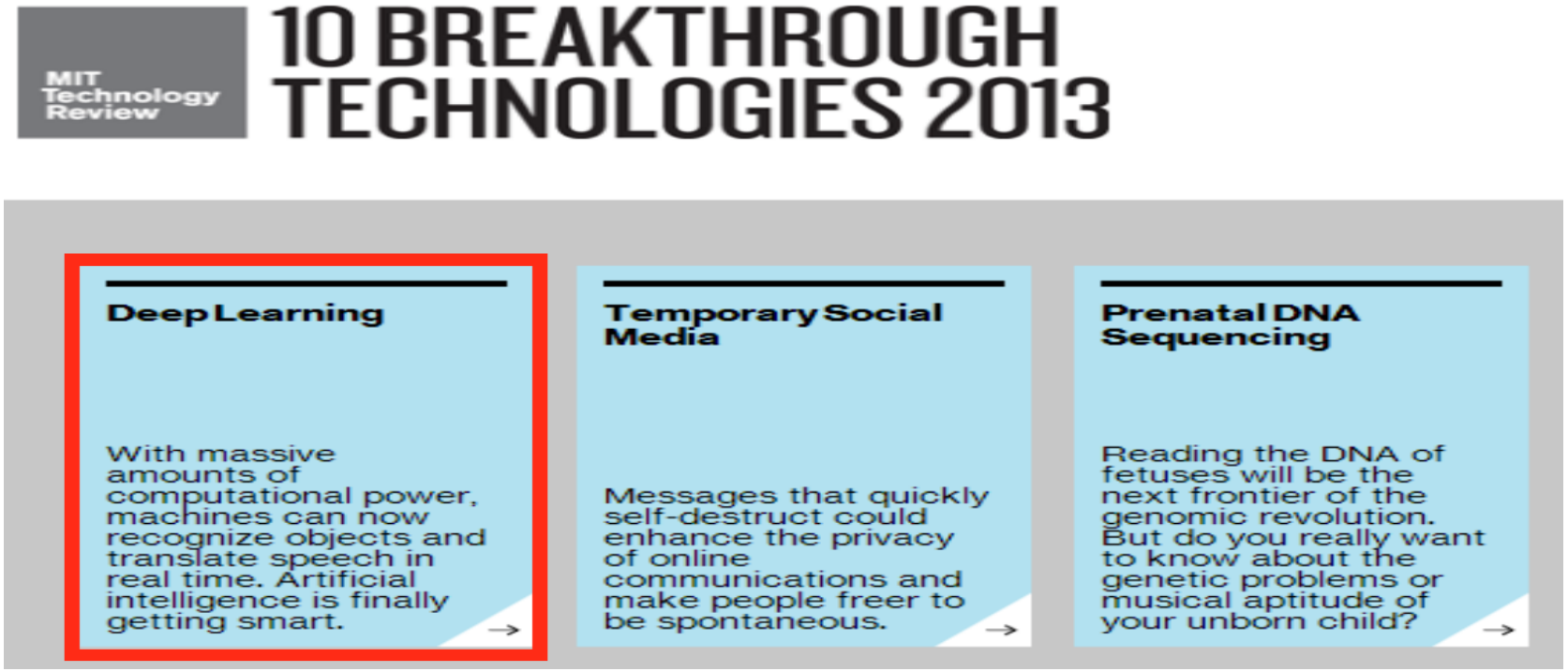
start at 32:00

Also see The Promise of AI" by Frank Chen
https://vimeo.com/215926017
a ~45-minute narrated walkthrough of what companies are doing with AI today and what's bubbling up from the research community that's just a few years out.

http://aiplaybook.a16z.com/docs/guides/dl

Deep Learning

- Amazing improvements in speech recognition, NLP, recognizing objects in images,…
- See http://deeplearning.net/



Going Deep

See tutorial at: http://www.iro.umontreal.ca/~bengioy/talks/mlss-austin.pdf

From Facebook's Deepface paper
https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf

Our method reaches an accuracy of 97.35% on the Labeled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 27%, closely approaching human-level performance.

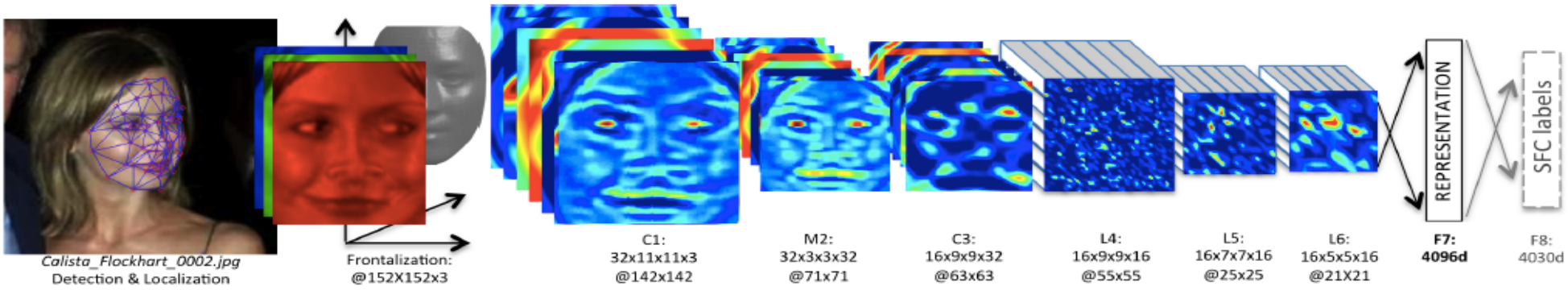


Figure 2. Outline of the *DeepFace* architecture. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate outputs for each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

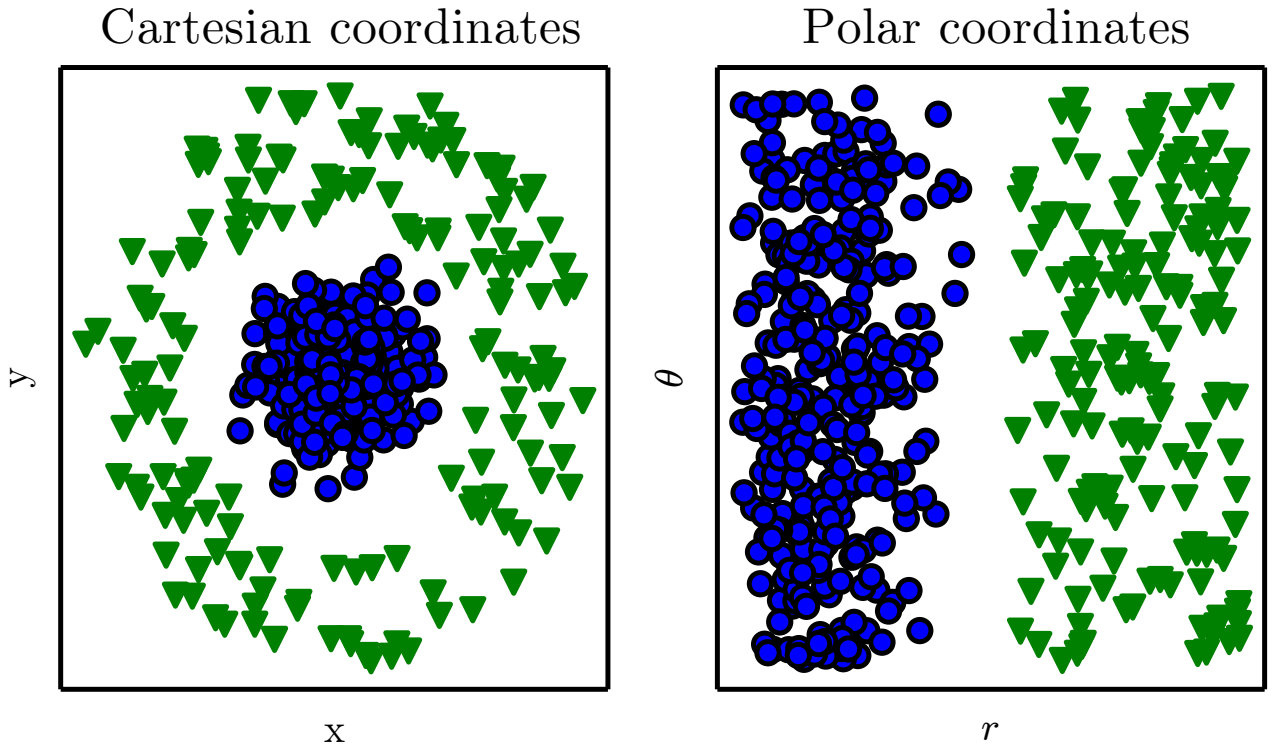
- Deep Learning: machine learning algorithms based on learning multiple levels of representation/abstraction

Key Ingredients:

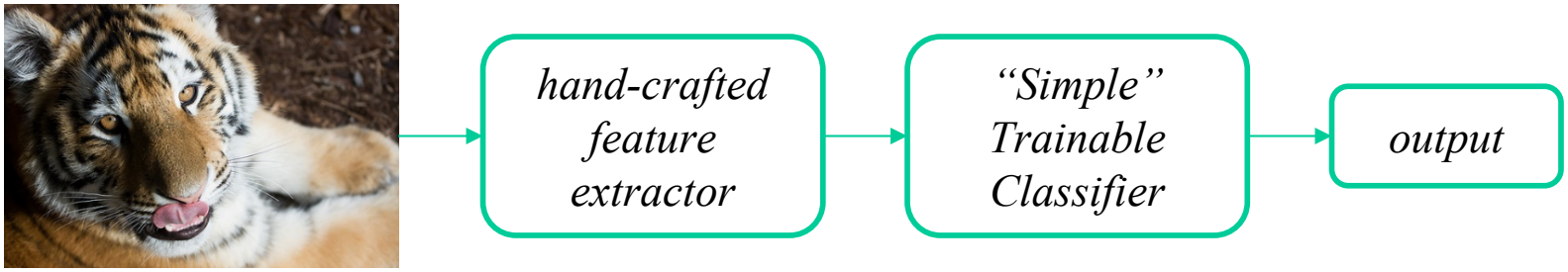
1. Lots & lots of data
2. Very flexible models (with lots of knobs that are adjusted during training)
3. Enough computing power
4. Powerful priors that can defeat the curse of dimensionality

From MLPs to Deep Nets

- MLPs can mimic any continuous function
  - Why go deeper?
  - Representations Matter



Traditional pattern recognition models use hand-crafted features and relatively simple trainable classifier.



However:

Hand-crafted features are usually highly dependent on the application

Hand-crafted features need human input and perhaps several trials, so may take time

Depth: Repeated Composition

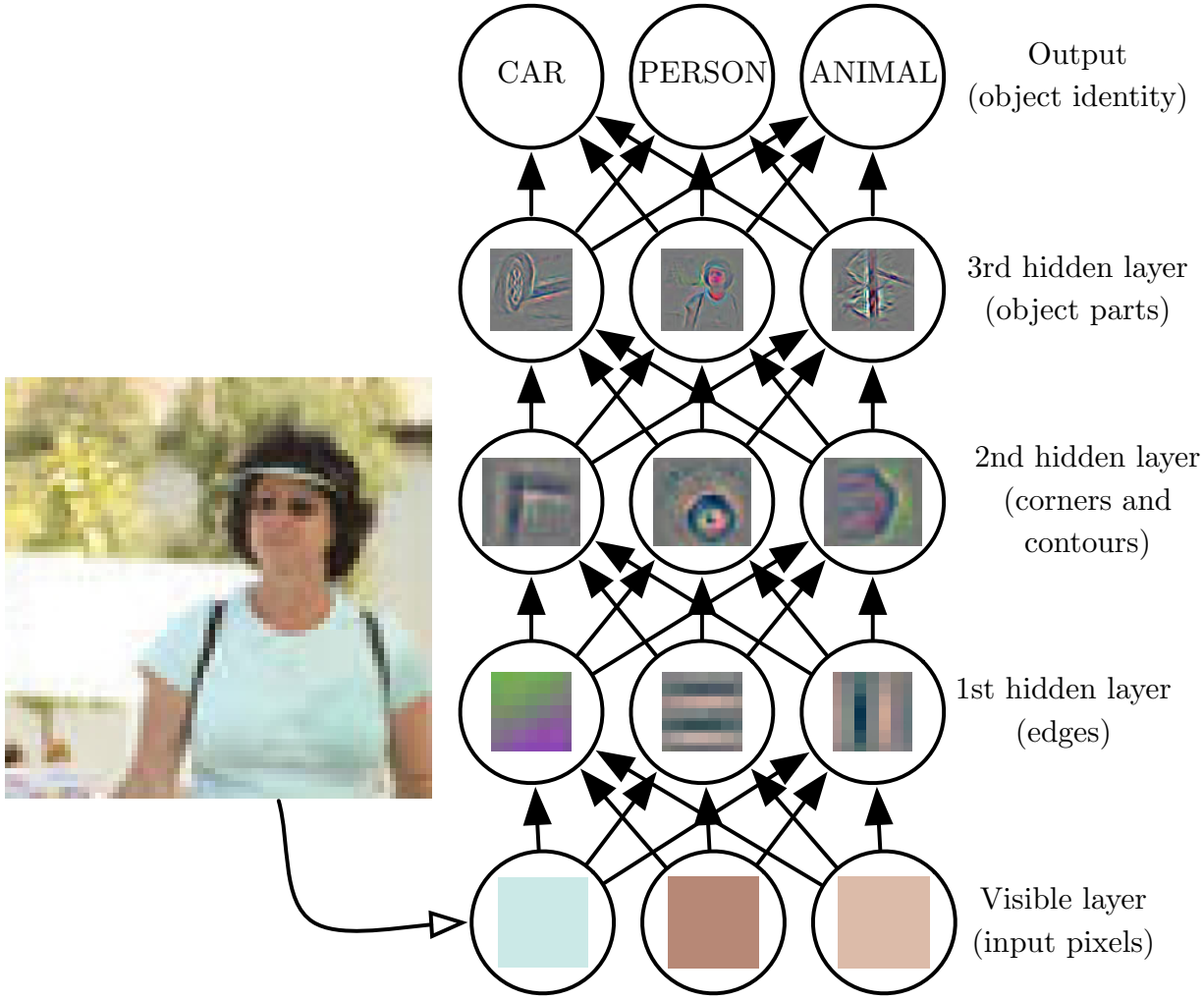
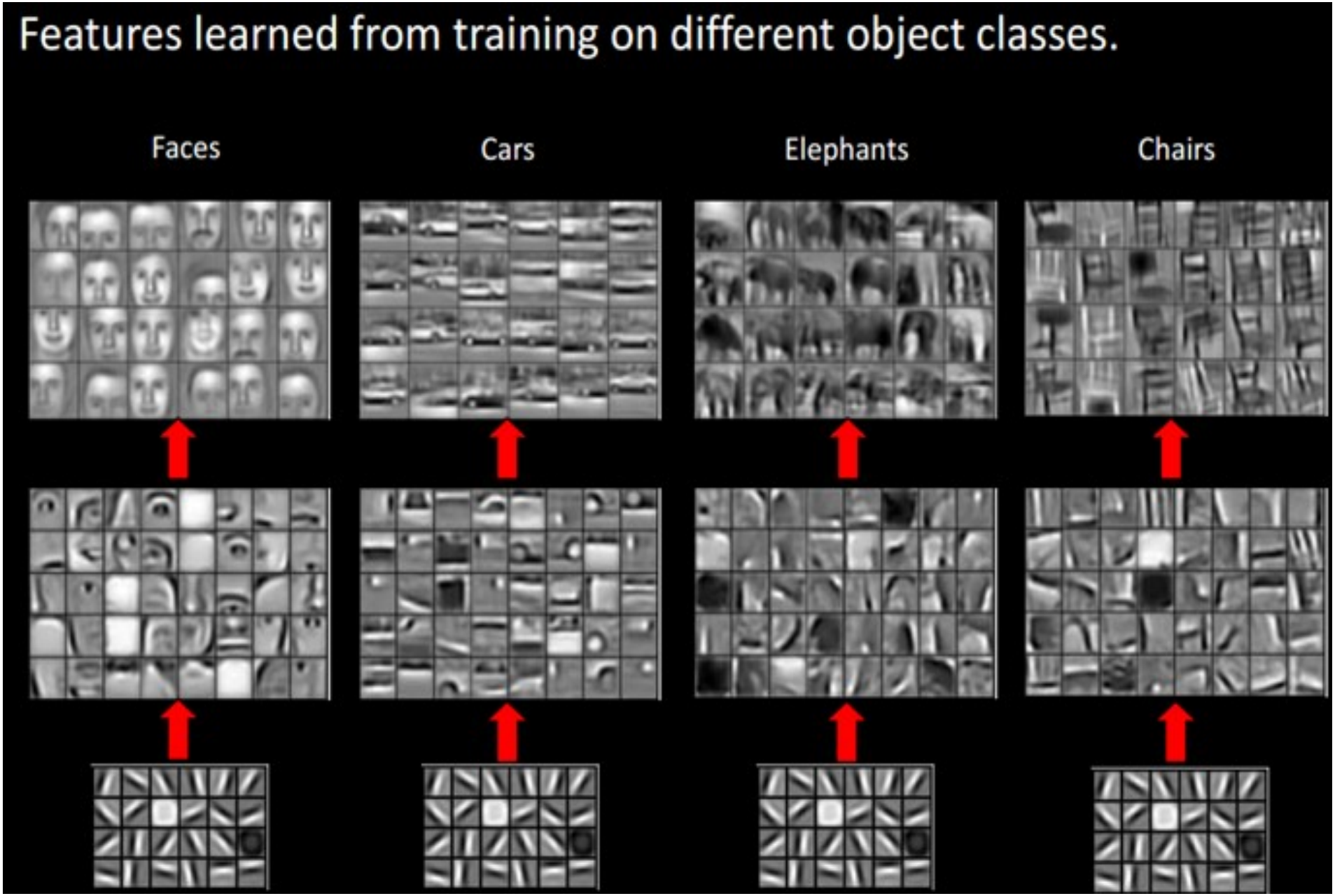


Figure 1.2


Features learned from training on different object classes.
Faces
Cars
Elephants
Chairs

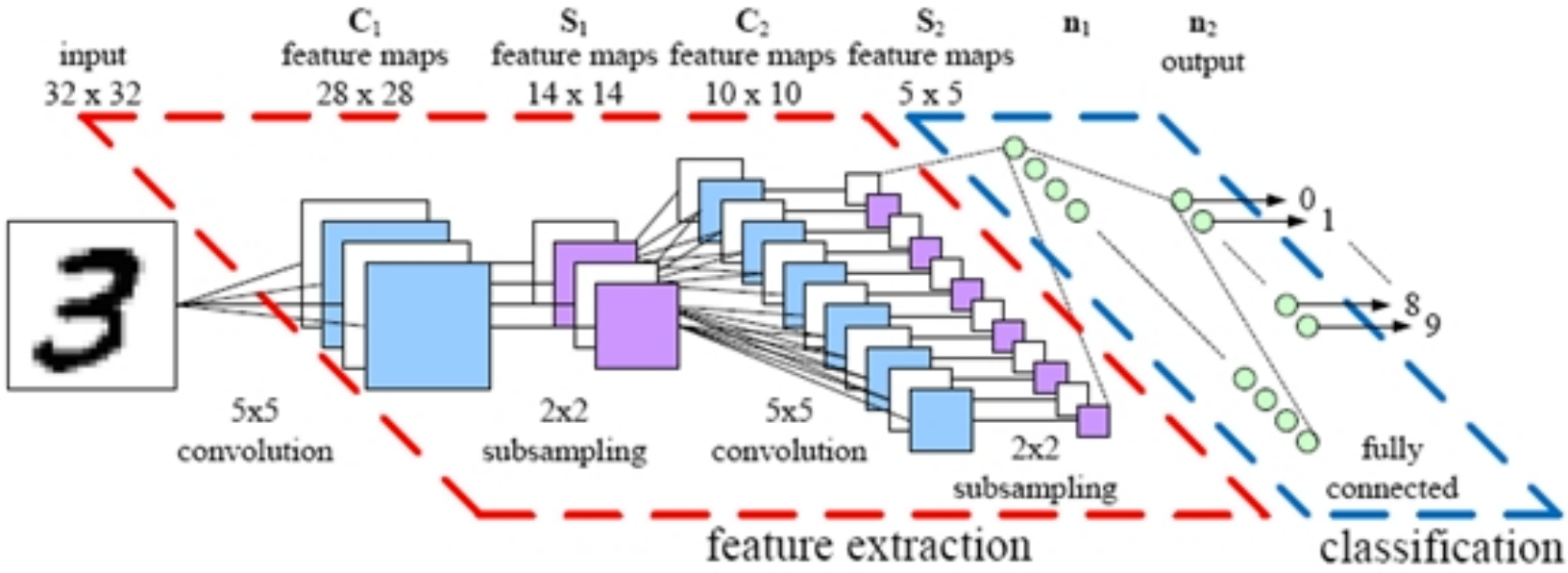
Deep Learning for Image Analysis

- How we teach computers to understand pictures | Fei Fei Li
- https://www.youtube.com/embed/40riCqvRoMs?start=243&end=900

- UT undergrads' smart-car project

Convolutional Neural Nets (CNNs)

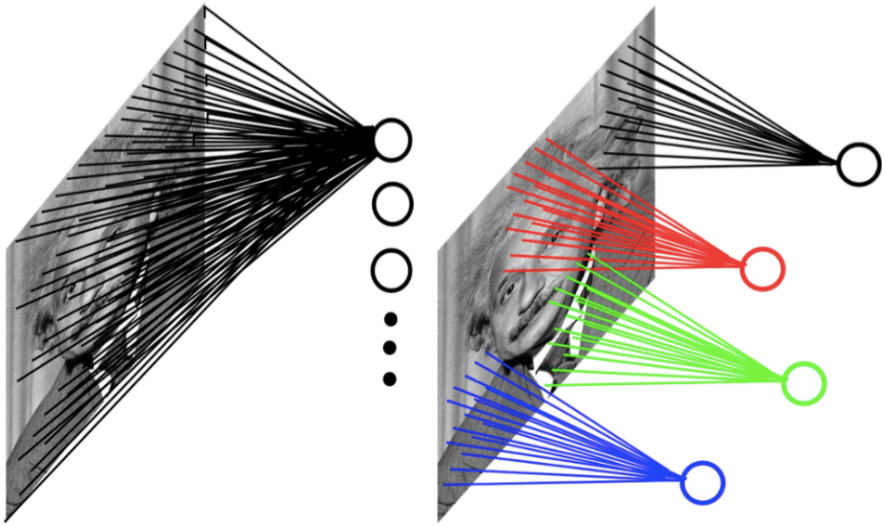
- How do they work? (excellent blog by Brandon Rohrer)



From: http://parse.ele.tue.nl/cluster/2/CNNArchitecture.jpg

CNNs

- The hidden units in a convolution layer are only connected to a local receptive field.
- Their weights may be shared.
- The number of parameters needed by CNNs is much smaller.

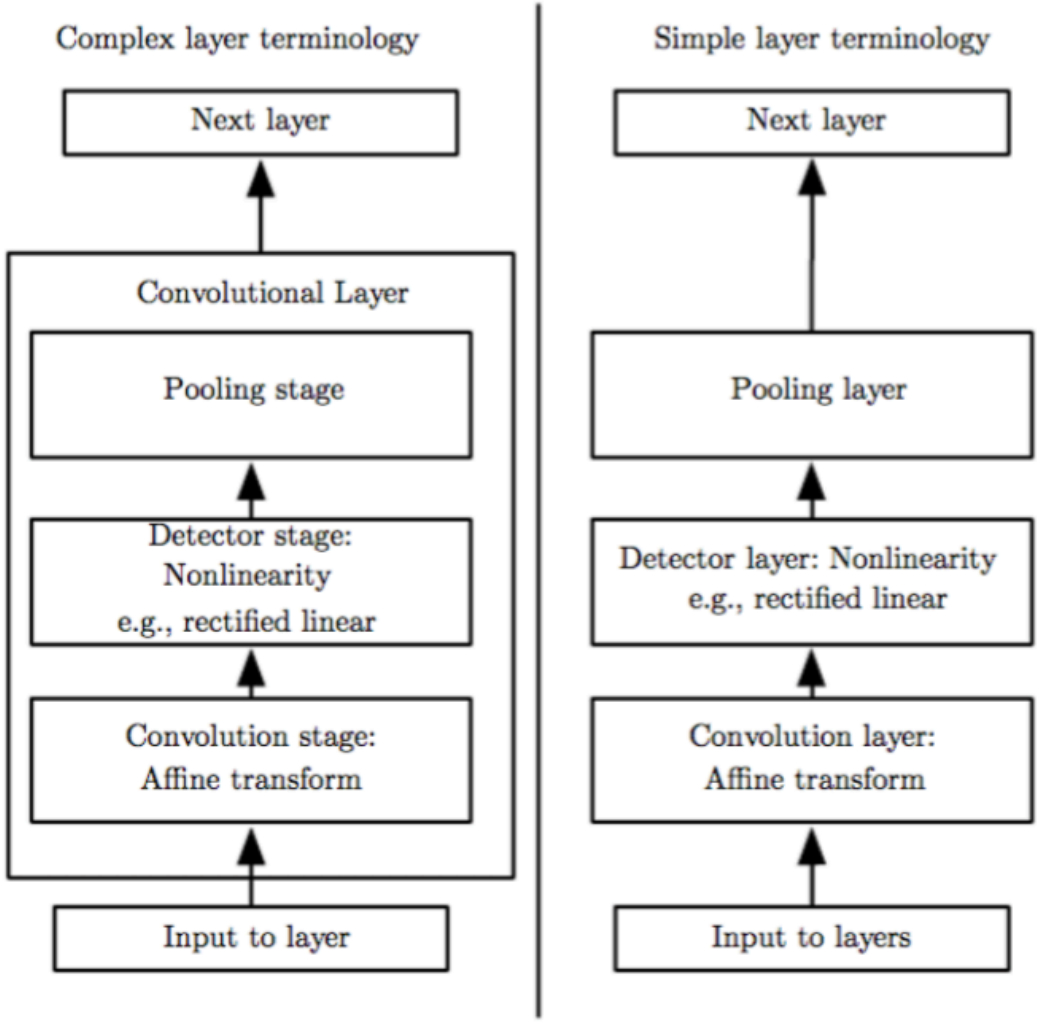


Example: 200x200 image

a) fully connected: 40,000 hidden units => 1.6 billion parameters

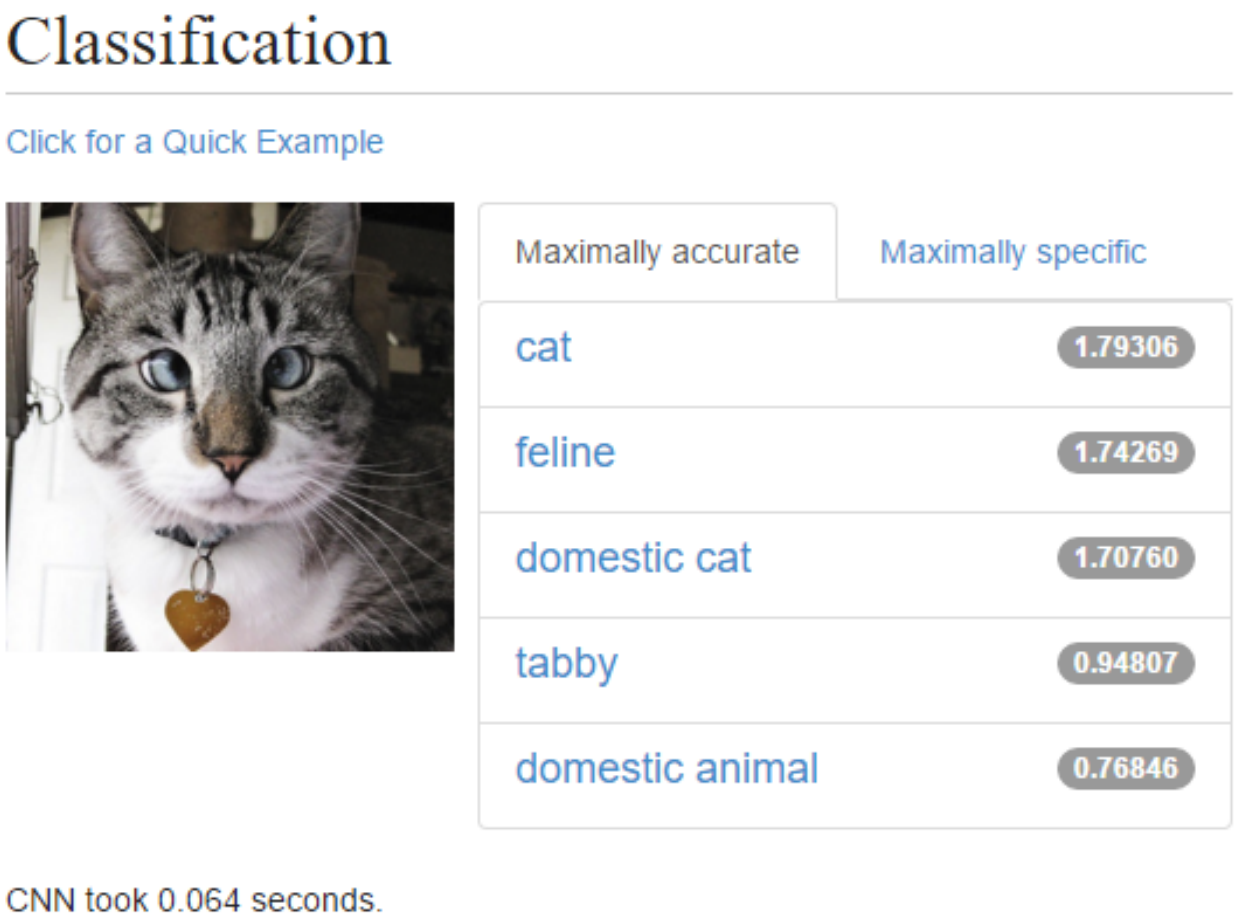
b) CNN: 5x5 kernel, 100 feature maps => 2,500 parameters

THREE STAGES OF A CONVOLUTIONAL LAYER



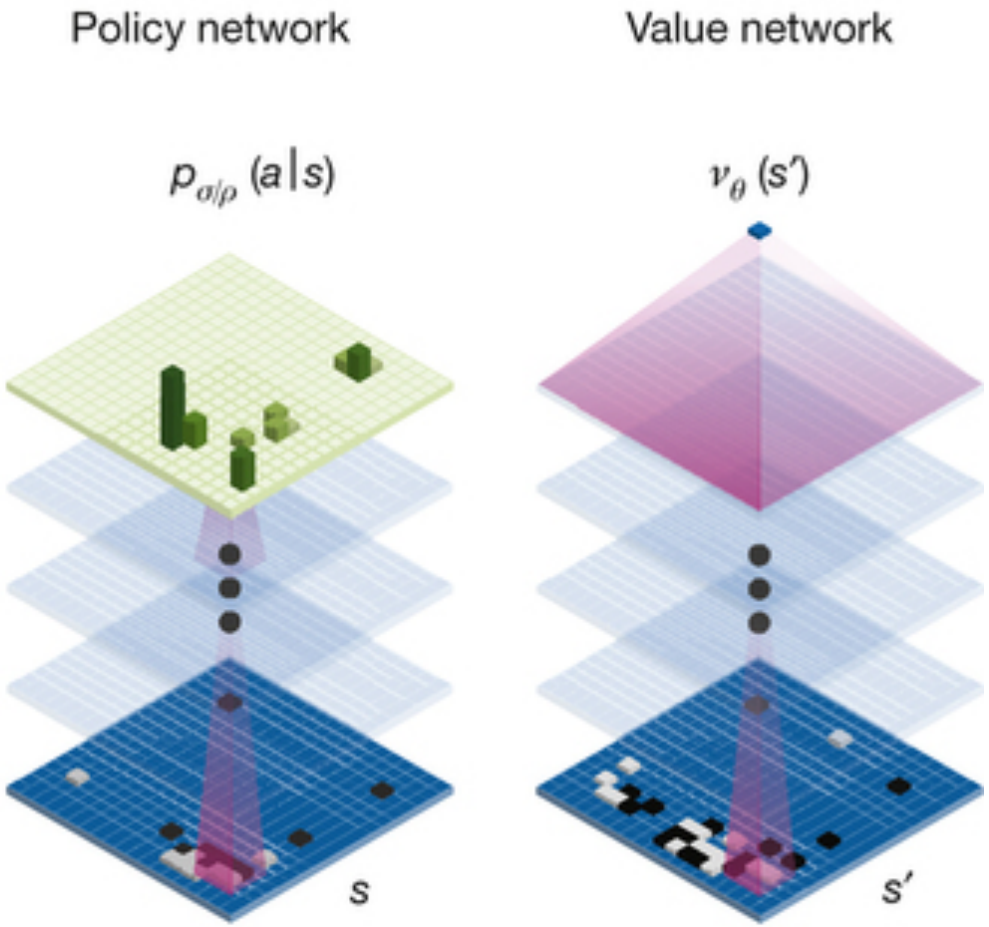
1. Convolution stage
2. Nonlinearity: a nonlinear transform such as rectified linear or tanh
3. Pooling: output a summary statistics of local input, such as max pooling and average pooling

DEEP CNN FOR IMAGE CLASSIFICATION



Try out a live demo at
http://demo.caffe.berkeleyvision.org/

DEEP CNN IN ALPHAGO



(Silver et al, 2016)

Policy network:

Input: 19x19, 48 input channels

Layer 1: 5x5 kernel, 192 filters

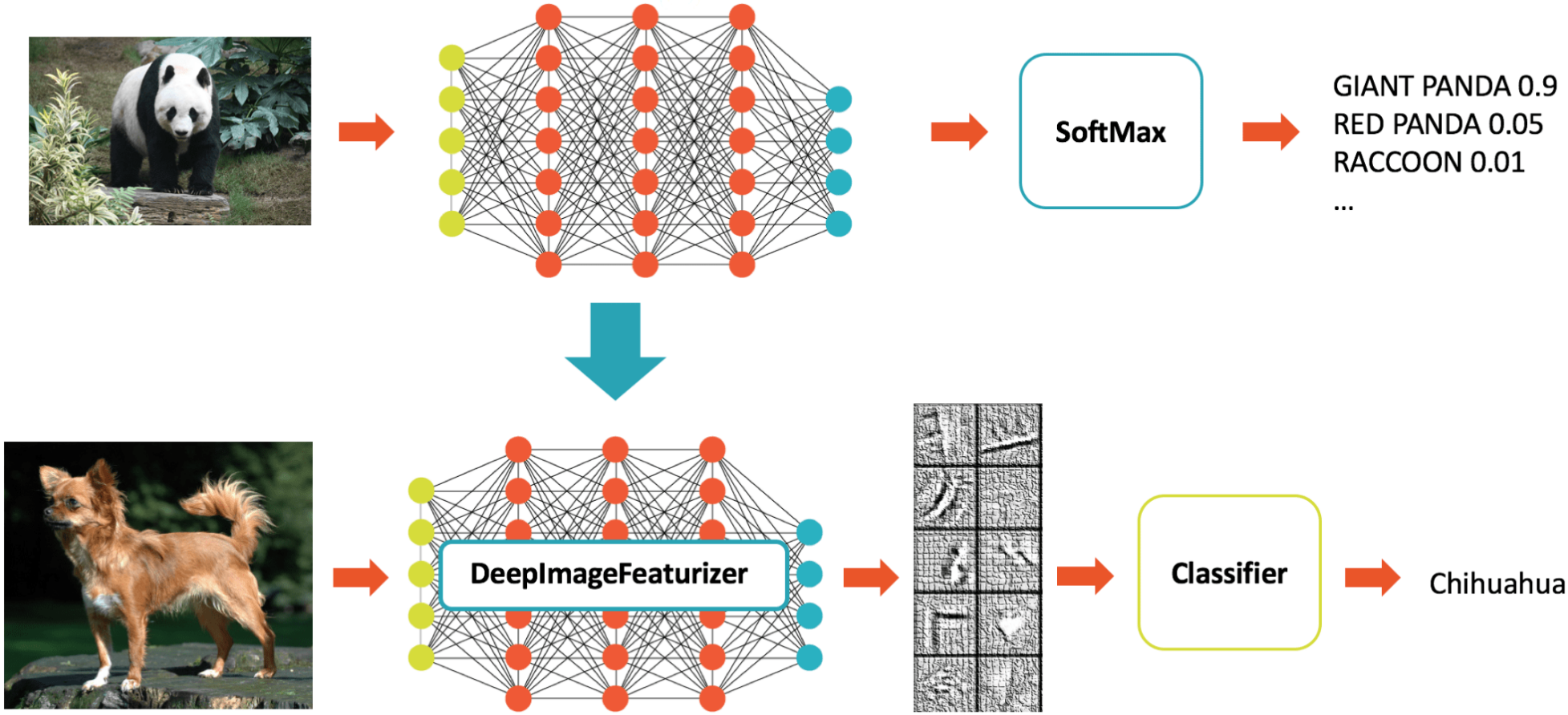
Layer 2 to 12: 3x3 kernel, 192 filters

Layer 13: 1x1 kernel, 1 filter

Value network has similar architecture to policy network

Transfer Learning with CNNs

- From https://databricks.com/blog/2017/06/06/databricks-vision-simplify-large-scale-deep-learning.html
- Examples: Cancer detection: https://www.nature.com/nature/journal/v542/n7639/full/nature21056.html
- Metamind (now powering Salesforce CRM)

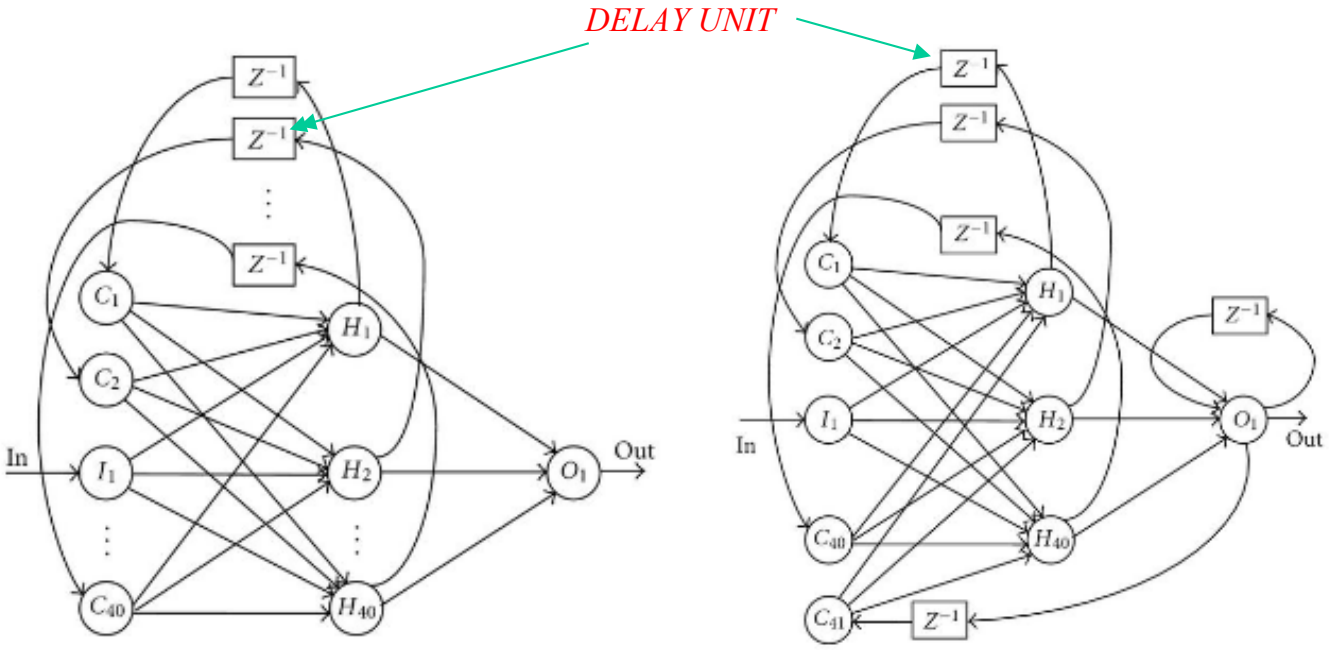


Recurrent Neural Networks (RNNs) and Language

(adding memory)

Recurrent Neural Networks (RNNs)

- Feedback connections lead to a (distributed) internal state that represents past history
  - Allows for temporal models



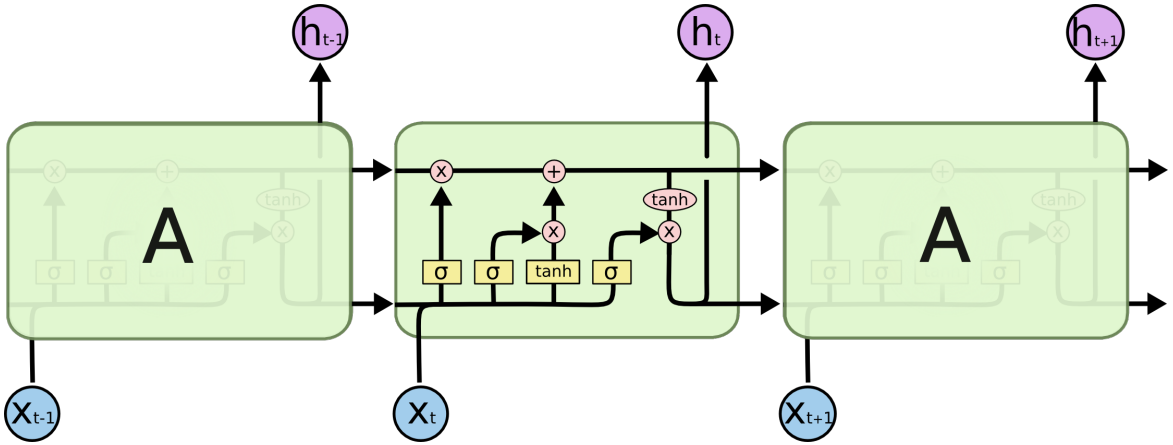
- https://tryolabs.com/blog/2016/12/06/major-advancements-deep-learning-2016/

- https://www.forbes.com/sites/quora/2016/12/30/these-were-the-best-machine-learning-breakthroughs-of-2016/#5591ed845f84

- Recurrent Neural Networks (RNN) https://www.youtube.com/embed/WCUNPb-5EYI?start=32
- Long Short-Term Memory (LSTM) https://www.youtube.com/embed/WCUNPb-5EYI?start =405

LSTMs

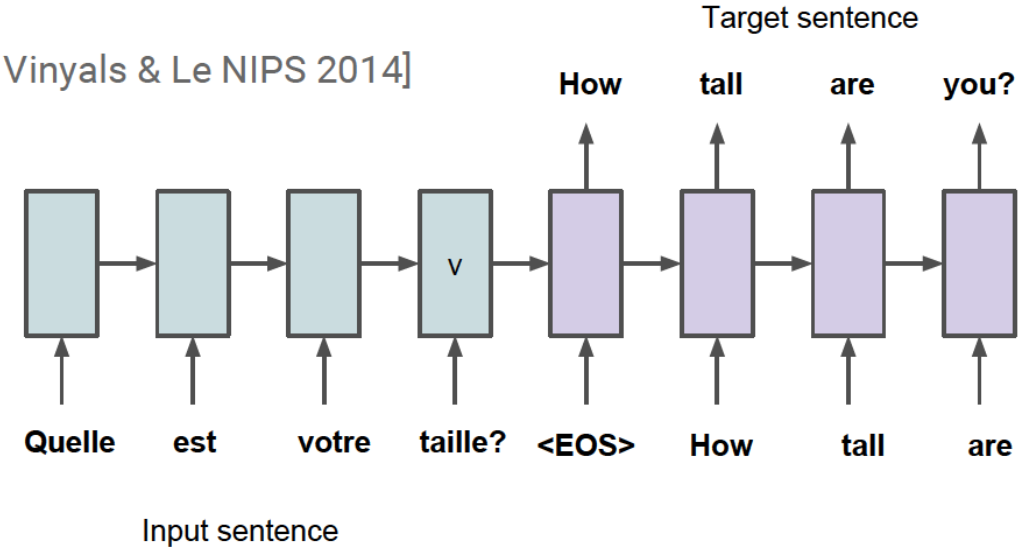
- Particularly useful for time series modeling when there are very long time lags of unknown size between important events
- Key Applications
  - Speech recognition
  - Handwriting recognition
  - Human motion prediction



http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Sequence-to-Sequence Model: Machine Translation

[Sutskever & Vinyals & Le NIPS 2014]



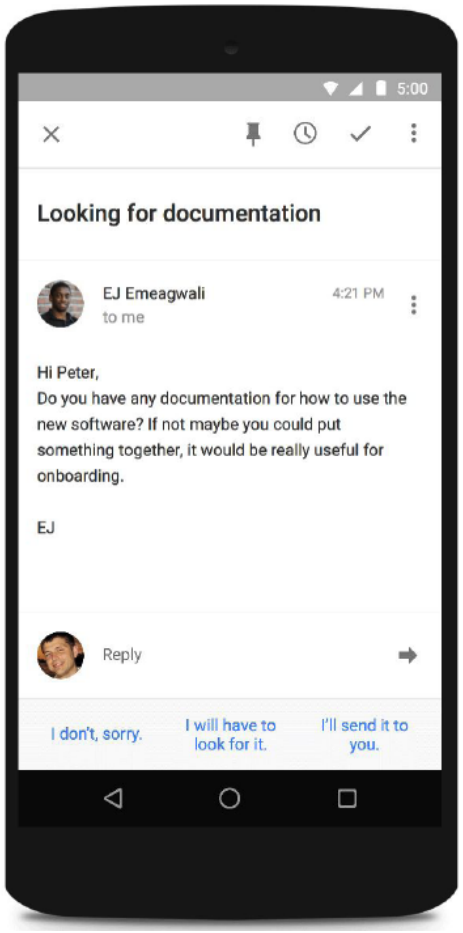
Inbox by Gmail

Smart Reply

April 1, 2009: **April Fool's Day joke**

Nov 5, 2015: **Launched Real Product**

Feb 1, 2016: **>10% of mobile Inbox replies**

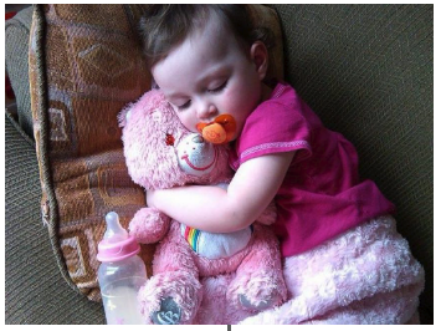


Research at Google

Combining Vision and Text

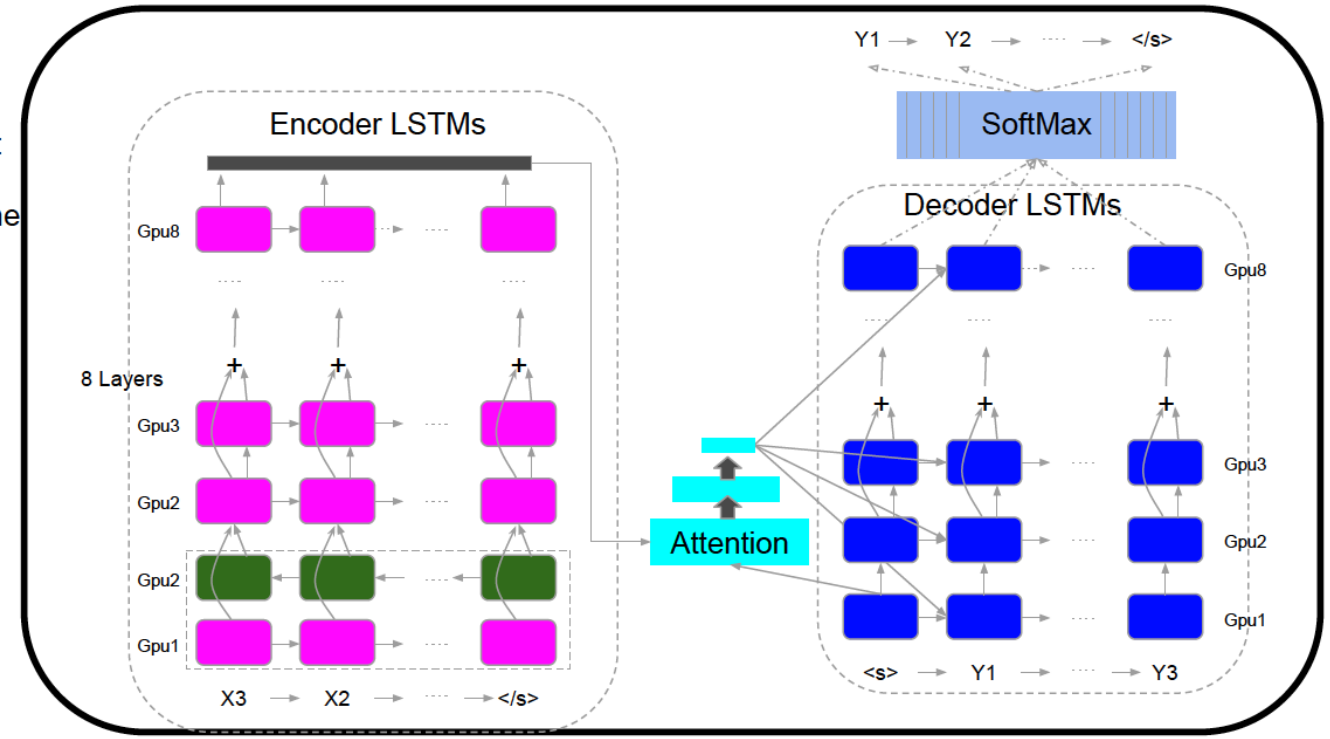
Image Captioning

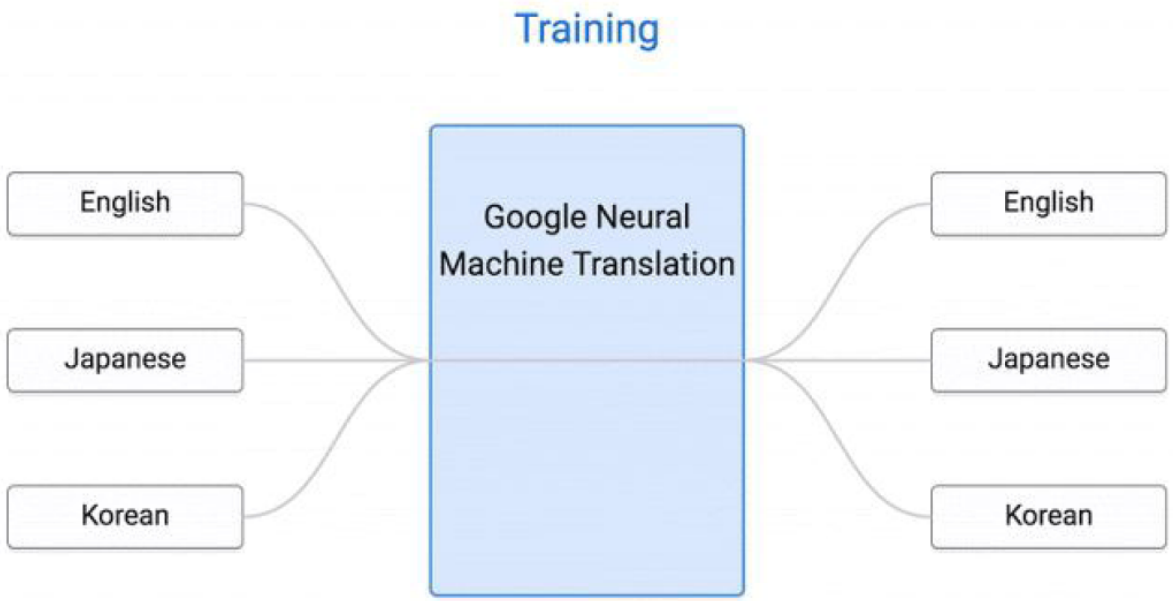
[Vinyals *et al.*, CVPR 2015]



Google Neural Machine Translation Model

One model replica: one machine w/ 8 GPUs





Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation,
Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean
https://arxiv.org/abs/1611.04558

https://research.googleblog.com/2016/11/zero-shot-translation-with-googles.html

Understanding NLP Capabilities via APIs (courtesy a16z)

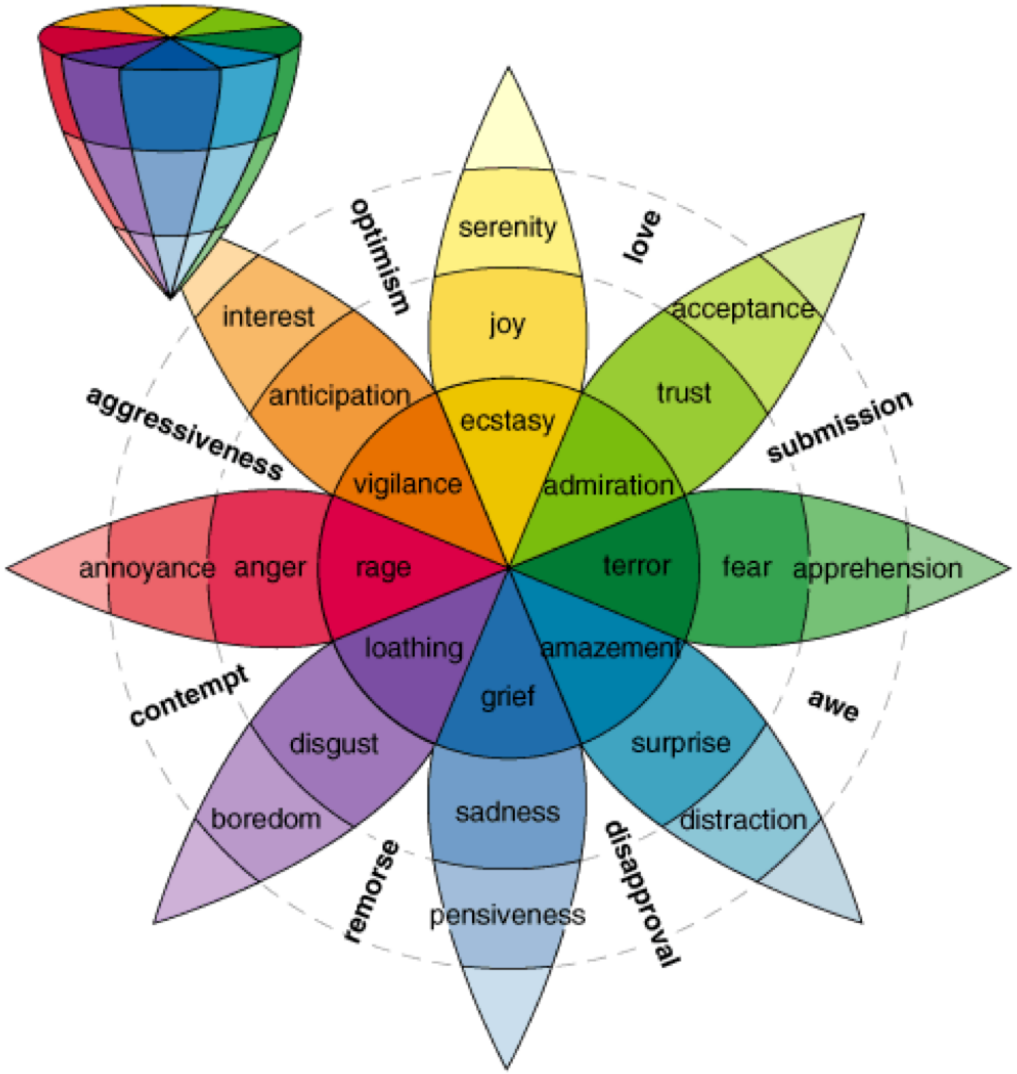
1. Sentiment analysis

- E.g. "Seeing the F-117 Nighthawk was cool."
- "Airplanes with the shape of hawks are cool."
- "Seeing a hawk flying at night is cool."

- All use similar nouns, have significantly different meanings, but communicate a similar feeling, or *sentiment*.

- **Application**: a system can monitor conversations of customers with technical support and derive a general degree of "happiness", "anger" or other emotions expressed in the communication.

  - Try out http://aiplaybook.a16z.com/test/phrase/sentiment-analysis



From tutorial at http://sentiment.christopherpotts.net/index.html

2. Entity Analysis

- Another important area in NLP deals with language *entities*. Types of analysis range from understanding syntax to extracting actual entities, identifying and labeling them by type (e.g. person, organization, location, events, product, etc). As in other cases the response data, structure, and accuracy varies wildly from one service to the next.

- http://aiplaybook.a16z.com/test/phrase/entity-analysis

3. Language Analysis and Detection

- One can simply speak or write and a service can use that information to automatically route to the correct language.

- http://aiplaybook.a16z.com/test/phrase/language-analysis

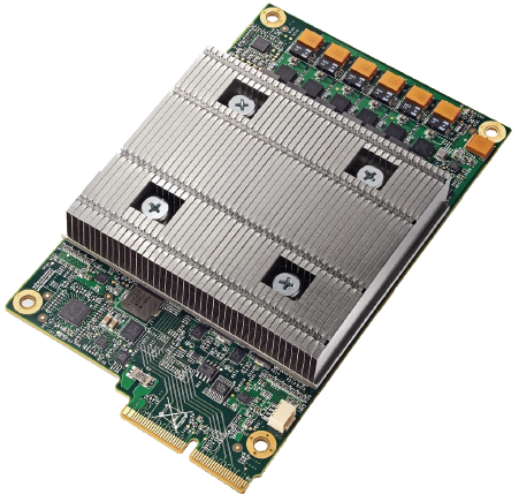
Limitations of Deep Learning

- Brute Force:
  - "Current supervised perception and reinforcement learning algorithms require lots of data, are terrible at planning, and are only doing straightforward pattern recognition." By contrast, humans "learn from very few examples, can do very long-term planning, and are capable of forming abstract models of a situation and manipulate these models to achieve extreme generalization."- Francois Chollet
- Reflect Biases in Training Data
  - If the term "doctor" is more associated with men than women, then an algorithm might prioritize male job applicants over female job applicants for open physician positions.
- Difficult to Comprehend
  - New emphasis on Explainable AI, e.g. see https://youtu.be/hUnRCxnydCc https://www.youtube.com/watch?v=axt7sOJP_e4

Extras

Tensor Processing Unit

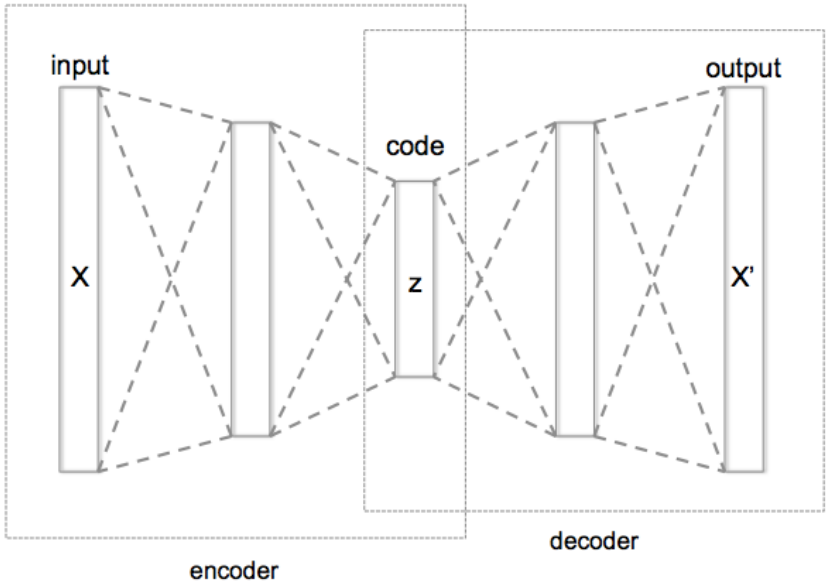
Custom Google-designed chip for neural net computations



In production use for >20 months: used on every search query, for neural machine translation, for AlphaGo match, ...

Auto-Encoders

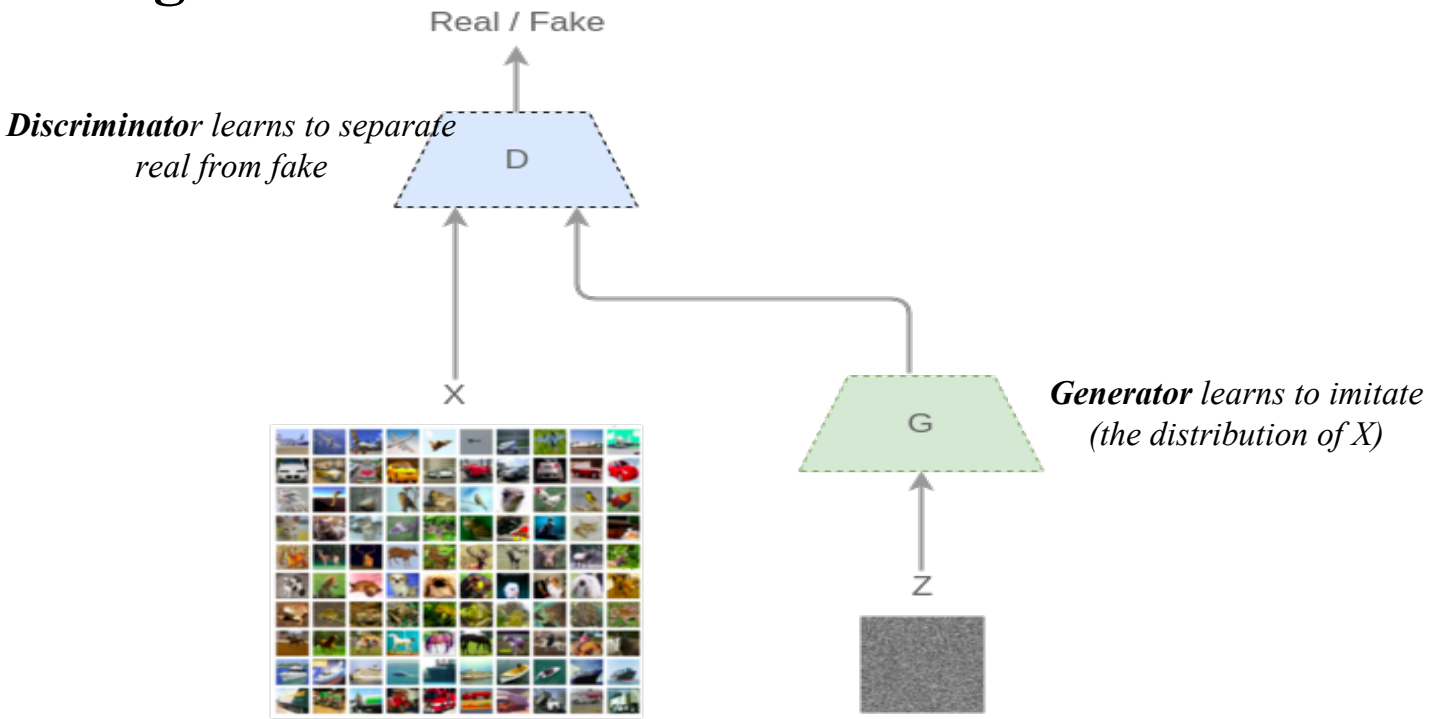
(unsupervised) Non-linear compression



From https://en.wikipedia.org/wiki/Autoencoder

Generative Adversarial Networks (GANs)

- **Generating Fake but Realistic Data**



From https://tryolabs.com/blog/2016/12/06/major-advancements-deep-learning-2016/

How to train GANs?

- Objective of generative network - increase the error rate of the discriminative network.
- Objective of discriminative network – decrease binary classification loss.
- Discriminator training - backprop from a binary classification loss.
- Generator training - backprop the negation of the binary classification loss of the discriminator.

Generated bedrooms. Source: "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks" https://arxiv.org/abs/1511.06434v2

InfoGan

- Also provides a latent vector space that separates label-informative factors of variation from other

(a) Azimuth (pose)

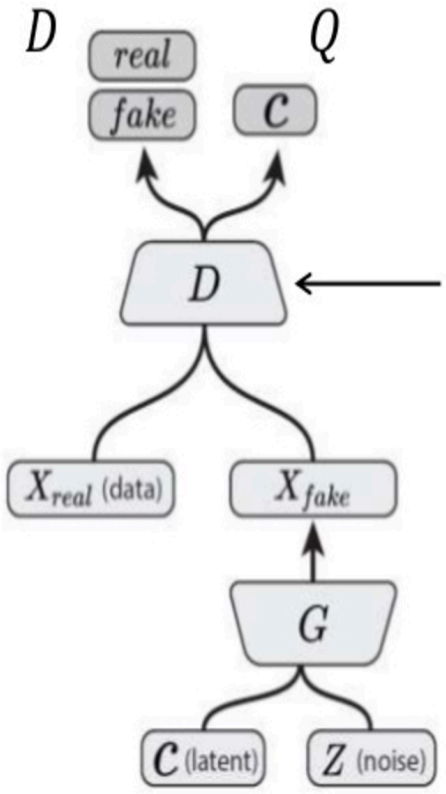
(b) Presence or absence of glasses

(c) Hair style

(d) Emotion

Figure 6: **Manipulating latent codes on CelebA:** (a) shows that a categorical code can capture the azimuth of face by discretizing this variation of continuous nature; in (b) a subset of the categorical code is devoted to signal the presence of glasses; (c) shows variation in hair style, roughly ordered from less hair to more hair; (d) shows change in emotion, roughly ordered from stern to happy.

InfoGAN



Convolution layers of the discriminator

Given DCGANs,

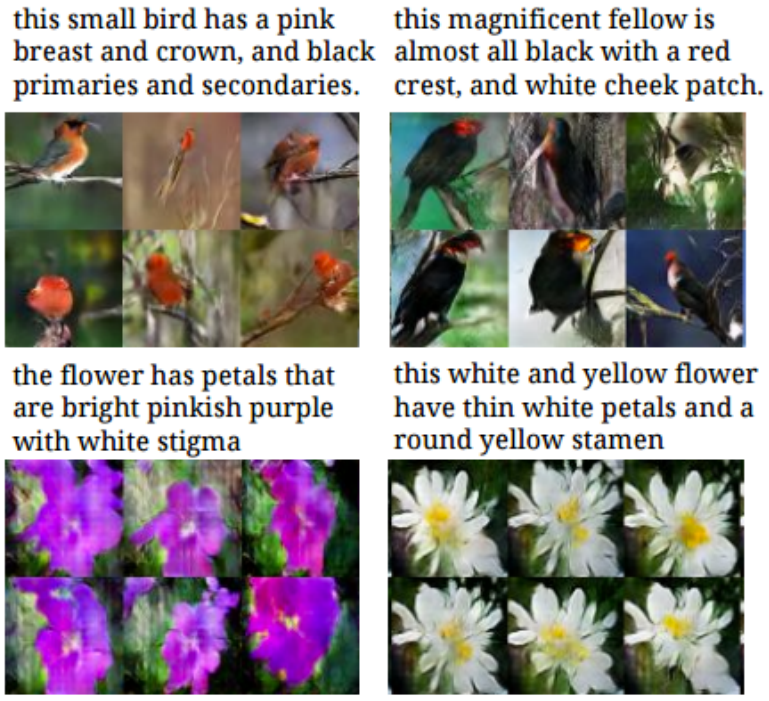
InfoGAN comes for negligible additional costs!

Image from Odena, *et al.*, arXiv:1610.09585.

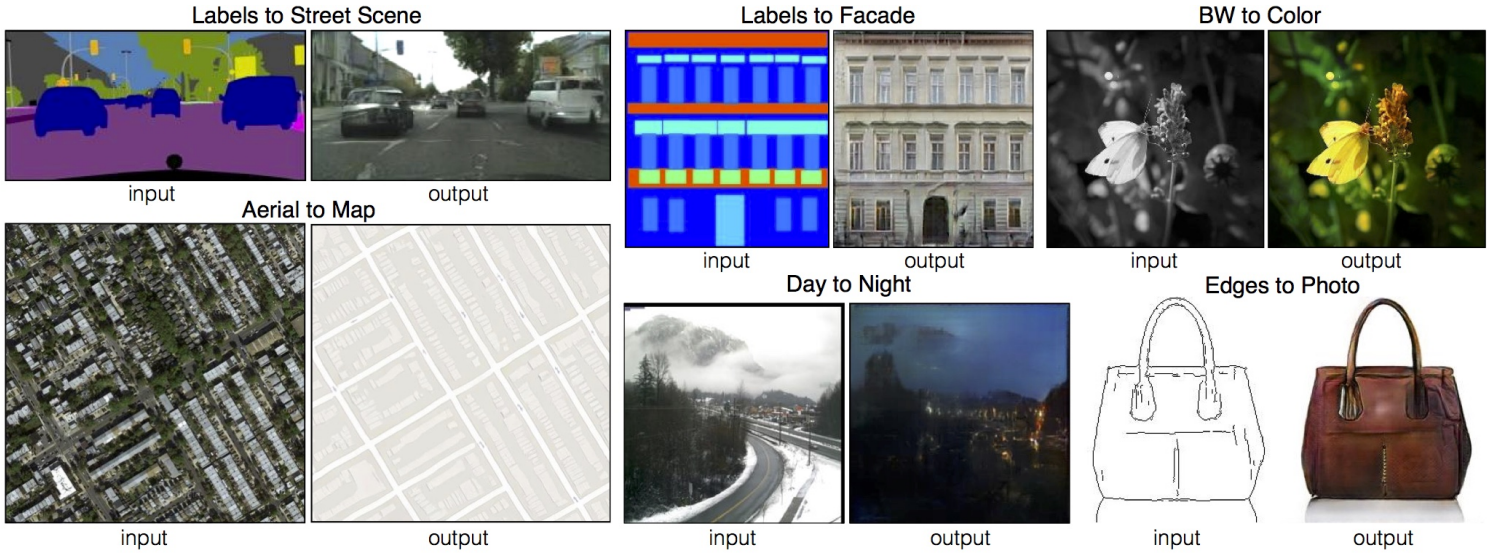
Conditional GANs

- These models are able to generate samples taking into account external information (class label, text, another image), using it to force *G* to generate a particular type of output.
- Applications:

1. **Text to Image**

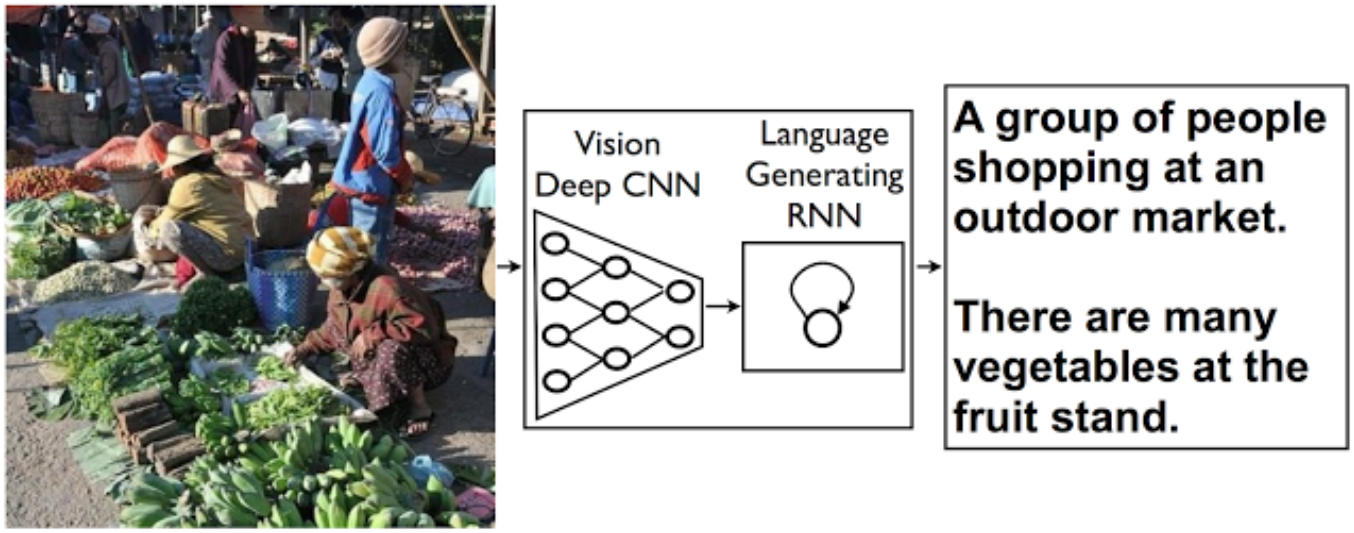


CGANs for Image 2 Image



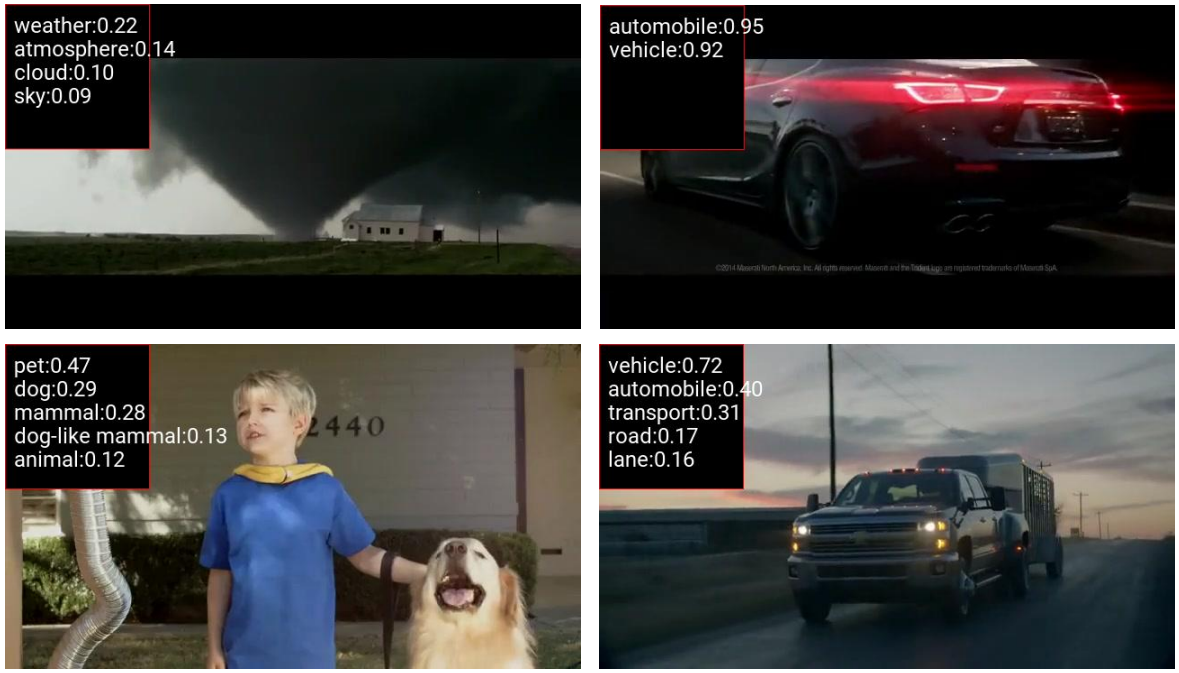
Automatic Textual Description of Images

- https://research.googleblog.com/2014/11/a-picture-is-worth-thousand-coherent.html



The model combines a vision CNN with a language-generating RNN so it can take in an image and generate a fitting natural-language caption.

Video Attribution



Food Photo → Recipe!

- Story with link to paper and video
- Try it out at http://tuesday.csail.mit.edu:4242/

Query Image	True ingrs.	Retrieved ingrs.	Retrieved Image
	whole milk half - and - half cr white sugar lemon extract ground cinnamon frozen blueberries vanilla wafers ice cubes	berries strawberry yogurt banana milk white sugar	
	butter garlic cloves all - purpose flour kosher salt milk chicken broth mozzarella cheese parmesan cheese onion	1 box any pasta you ground beef 1 envelope taco seas water 1/2 packages cream c cheese	
	cooked white rice salt shrimp Broccolini mayonnaise nori	sushi rice salmon avocado cream cheese nori	
	mayonnaise onion cider vinegar sugar celery seeds green cabbage carrot salt & freshly groun ground chuck	yellow onion coarse salt ground pepper ground chuck buns eggs ketchup canned beets lettuce leaves	

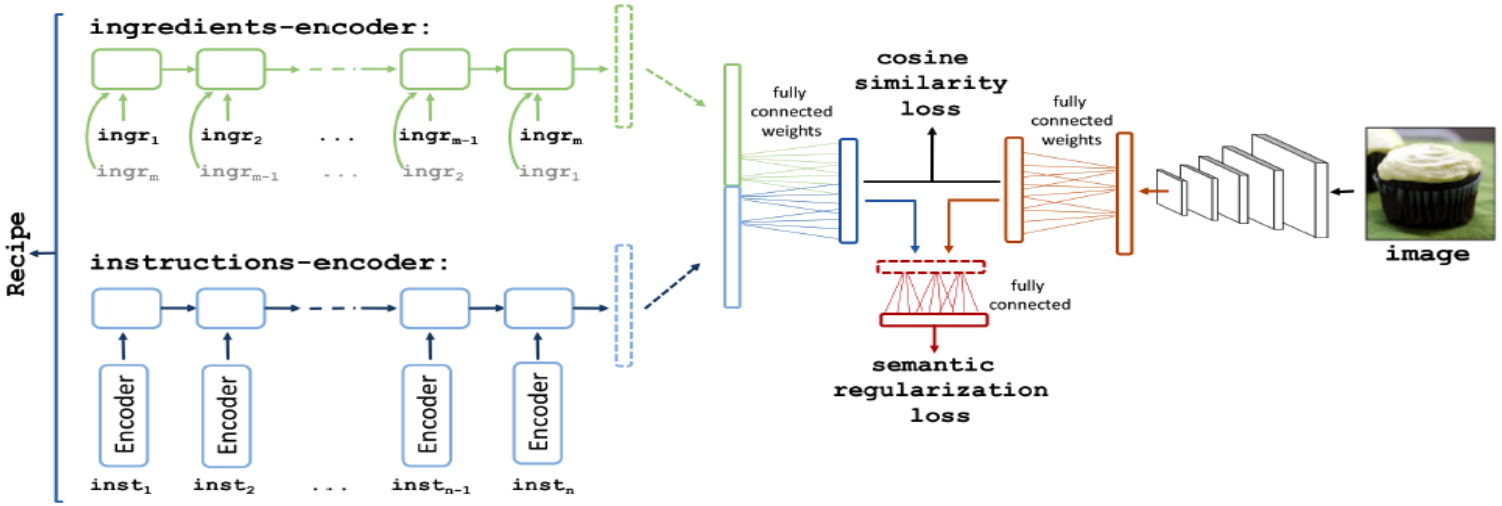
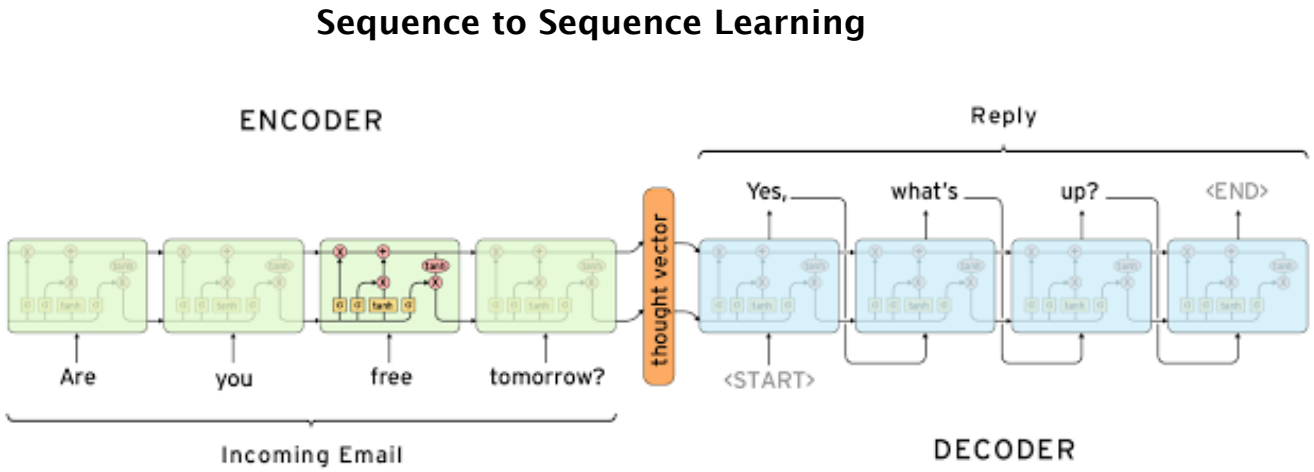


Figure 3: **Joint neural embedding model with semantic regularization.** Our model learns a joint embedding space for food images and cooking recipes.

Partition	# Recipes	# Images
Training	720,639	619,508
Validation	155,036	133,860
Test	154,045	134,338
Total	1,029,720	887,706

Table 1: **Recipe1M dataset.** Number of samples in training, validation and test sets.

Deep Learning for Chatbots?



- Most of the value of deep learning today is in narrow domains where you can get a lot of data. Here's one example of something it cannot do: have a meaningful conversation. There are demos, and if you cherry-pick the conversation, it looks like it's having a meaningful conversation, but if you actually try it yourself, it quickly goes off the rails.- Andrew Ng, 2016
  - http://www.seattletimes.com/business/baidu-research-chief-andrew-ng-fixed-on-self-taught-computers-self-driving-cars/

Deep Learning Software

- Long list at http://deeplearning.net/software_links/

Major efforts include

- Caffe (Berkeley)

- **CNTK (Msft)**
- **Deeplearning4j (Java Based)**
- **TensorFlow (Google); Keras as high-level interface.**
- **Theano (Python)**
- **Torch (Facebook and others)**

Comparison Table at https://en.wikipedia.org/wiki/Comparison_of_deep_learning_software

Issues to think about

- Architecture design
  - number and type of layers (conv, pool)
  - kernel sizes for (conv, pool)
  - activation function (relu, sigmoid, softmax, softplus)
  - number of feature maps (conv) or neurons (fully connected)
  - loss function choice (cross entropy, regression)
  - Advanced architecture choices (BatchNorm, Local Response Normalization (AlexNet paper), Residual (ResNet), Skip)
- Proper initialization
  - random, xavier (scaling)
  - unsupervised pre-training
- Learning
  - SGD, SGD+Momentum, AdaGrad, Adam, RMSProp
  - weight decay, learning rate, momentum
  - Dropout, gradient clipping
- Hardware
  - mapping onto GPUs
  - heterogeneous distributed (synchronous, asynchronous)

References

- Courses/videos
  - https://www.udacity.com/course/deep-learning--ud730
  - https://cs231n.github.io/
  - http://cs224d.stanford.edu/
  - http://videolectures.net/deeplearning2016_montreal/
- Blogs
  - http://colah.github.io/
  - http://karpathy.github.io/
  - http://blog.otoro.net/
  - http://distill.pub/2016/augmented-rnns/
  - http://www.wildml.com/
- Podcasts
  - http://www.thetalkingmachines.com/

References

- Papers
  - http://www.arxiv-sanity.com/top
  - https://github.com/terryum/awesome-deep-learning-papers
  - https://deepmind.com/research/publications/
  - http://www.shortscience.org/about
  - https://docs.google.com/document/d/1IXF3h0RU5zz4ukmTrVKVotPQypChscNGf5k6E25HGvA/edit

- Code and datasets
  - http://www.gitxiv.com/page/about
  - https://github.com/caesar0301/awesome-public-datasets

- Books
  - http://www.deeplearningbook.org/