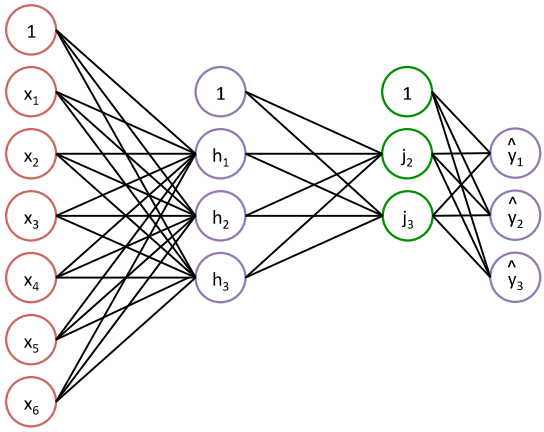
Neural Nets (cont)

Administrivia

- In class presentations start tomorrow
  - there will be Qs on the final based on the in class presentations
- Dec 2: Final reports due on connex
- Dec 2: class cancelled, I am traveling
  - Also, no office hours Dec 2.
- Caleb will set up a review session for you
  - Dec 2 or 5 (to be determined)

Neural Networks



Regularization

- We are learning a much larger number of parameters
- We need to regularize

Regularization at the Weight Level

$$\frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \frac{\lambda}{2}\sum_{\forall w} w_j^2$$

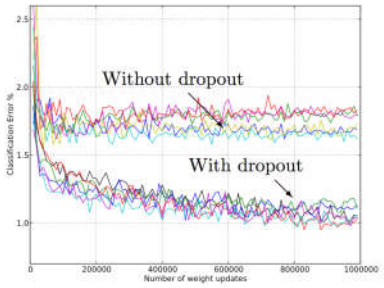
Regularization at the Node Level

$$\sum_{i=1}^{N}\sum_{j=1}^{p}(x_i - \hat{x}_{i,j})^2 + \sum_{i=1}^{N}\sum_{j=1}^{k} g(W_j x_i)$$

g is a regularizer that wants only a few neurons to fire for each picture, i.e. only a few of the hidden nodes (h) to be non-zero

Regularization at the Network Level

- Dropout
  - randomly remove nodes from the network
    - hidden and visible
  - encourages redundant connections
  - reduces overfitting



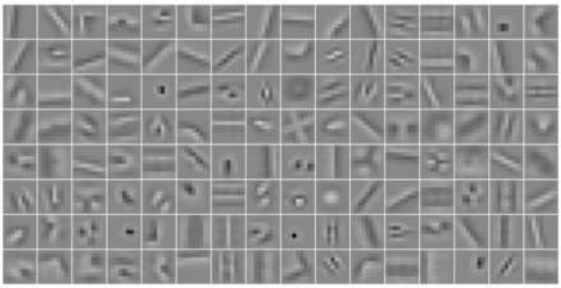
https://www.cs.toronto.edu/~hinton/absps/JMLRdropout.pdf

Convolutional neural nets

- Built to capture the invariances we see in images
  - objects can appear at any place in the image
- Learn filters that respond to particular visual features
  - usually edges
- Look for those visual features anywhere in the image
  - i.e. convolve the filter with many patches of the image
- Hidden layers combine these filters to create more complex shapes
  - e.g. straight edges combined to form curves, combined to form the handle on a coffee mug

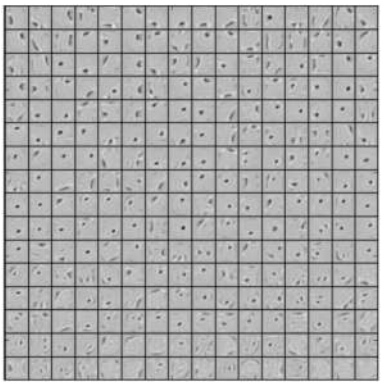
What do CNNs learn?

- When trained on images



http://yann.lecun.com/exdb/publis/pdf/lecun-eccv-12.pdf

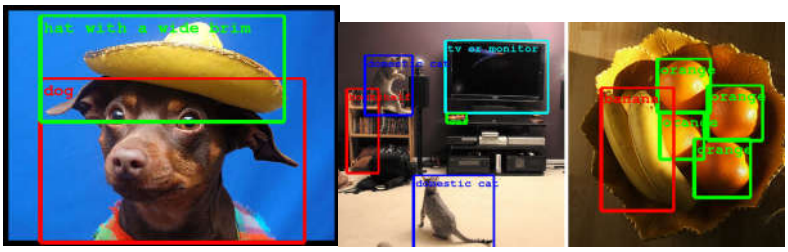
What do CNNs Learn?



https://www.cs.toronto.edu/~hinton/absps/JMLRdropout.pdf

ImageNet

- Manually annotated 15 million images with 20,000 categorical labels



https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures?language=en

But what does it mean to UNDERSTAND and image?

Scene Description Generation

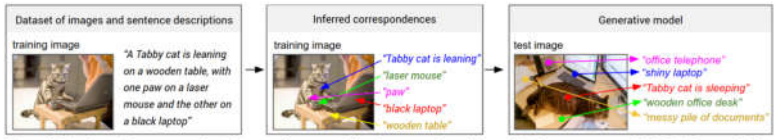


Figure 2. Overview of our approach. A dataset of images and their sentence descriptions is the input to our model (left). Our model first infers the correspondences (middle, Section 3.1) and then learns to generate novel descriptions (right, Section 3.2).

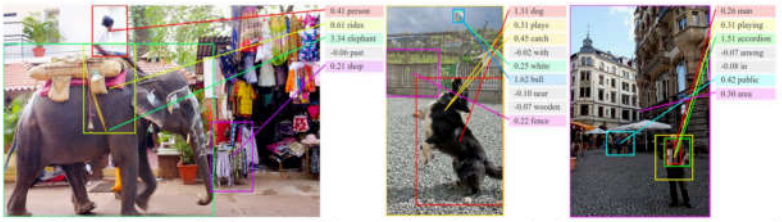


Figure 5. Example alignments predicted by our model. For every test image above, we retrieve the most compatible test sentence and visualize the highest-scoring region for each word (before MRF smoothing described in Section 3.1.4) and the associated scores ($v_i^T s_t$). We hide the alignments of low-scoring words to reduce clutter. We assign each region an arbitrary color.

Two pieces:
- Generate image representation
- Generate sentence representation

Learn representations such that related image-sentence pairs have similar representations

http://cs.stanford.edu/people/karpathy/cvpr2015.pdf

A Recurrent Neural Net (RNN) knows how to generate output (words) conditioned on the context (image)

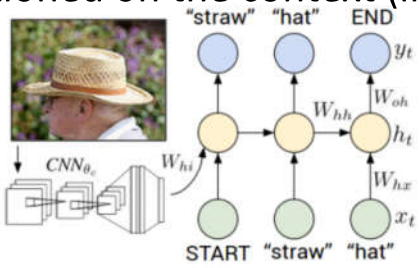


Figure 4. Diagram of our multimodal Recurrent Neural Network generative model. The RNN takes a word, the context from previous time steps and defines a distribution over the next word in the sentence. The RNN is conditioned on the image information at the first time step. START and END are special tokens.

More examples (with a few funny mistakes)
http://cs.stanford.edu/people/karpathy/deepimagesent/

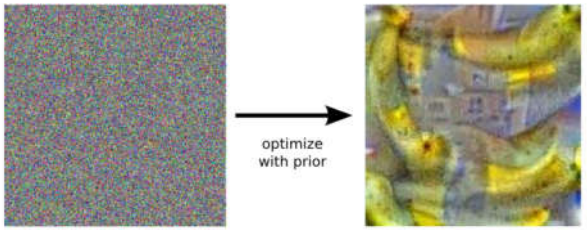
Digging Deeper

- One of the largest criticisms of deep learning is that the models are very complex and are essentially black boxes
- Some recent work has tried to open the box

- http://googleresearch.blogspot.ca/2015/06/inceptionism-going-deeper-into-neural.html

Basic Idea

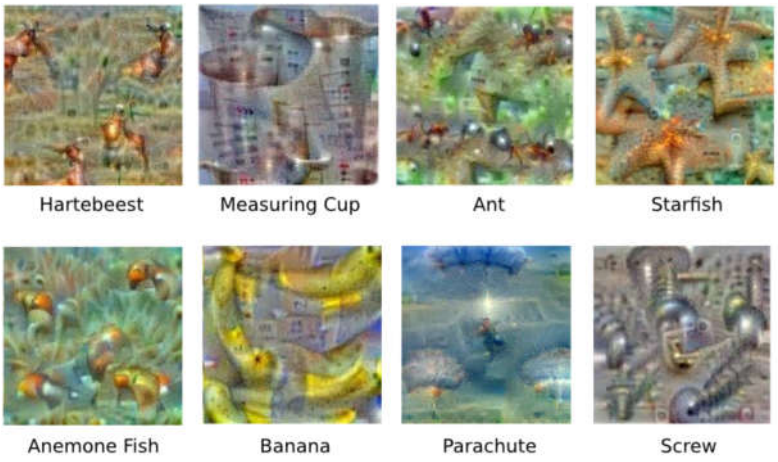
- Let's say we want to know what the network thinks a banana looks like
  - Start with noise
  - Tweak the image so that the network tells you it is seeing something more banana-like



Previous Attempts were not so good

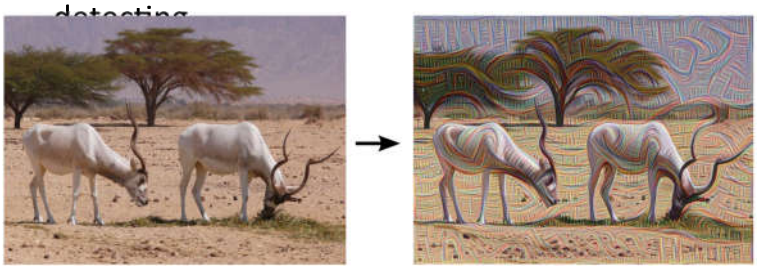
- Need to incorporate the idea of pixel correlation to get good images out of the network

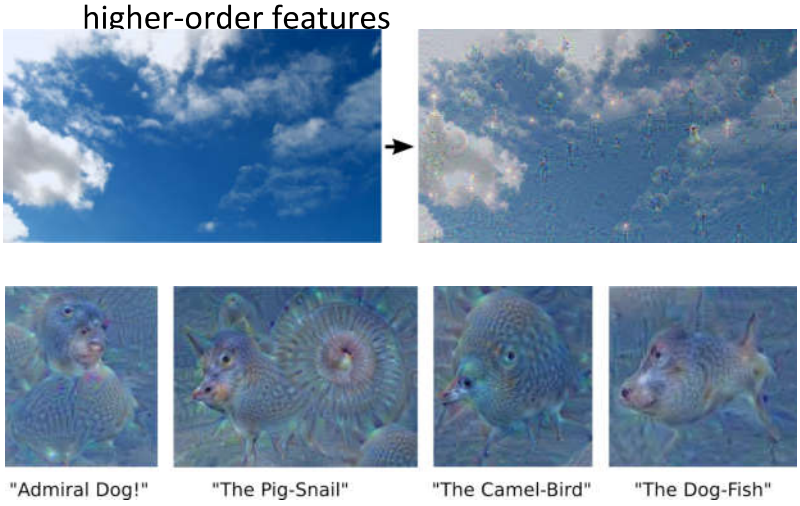
[Quoc Le, et al., *ICML*, 2012]



- Do this for higher level layers and enhance higher-order features

- Another technique, feed it any image, and ask the network to enhance the aspects it is detecting





The Grocery Trip

- https://www.youtube.com/watch?v=DgPaCWJL7XI

Vote for a talk

- https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are

- https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures