Deep Learning

CPSC 470 – Artificial Intelligence

Brian Scassellati

2019 Turing Award

Yann LeCun, Geoffrey Hinton, Yoshua Bengio

A Brief History

- 1950-60s: modeling biological neurons (Rosenblatt, Hebb, etc.)
- 1969: research stagnates after Minsky & Papert show limits of perceptrons
- 1990s: Convolutional neural networks (LeCun) and Recurrent networks (Schmidhuber) return
- 1990s: interests fade as few real-world results hold up
- 2006: revival of deep networks (Hinton, et al.)
- 2013-: massive industrial interest

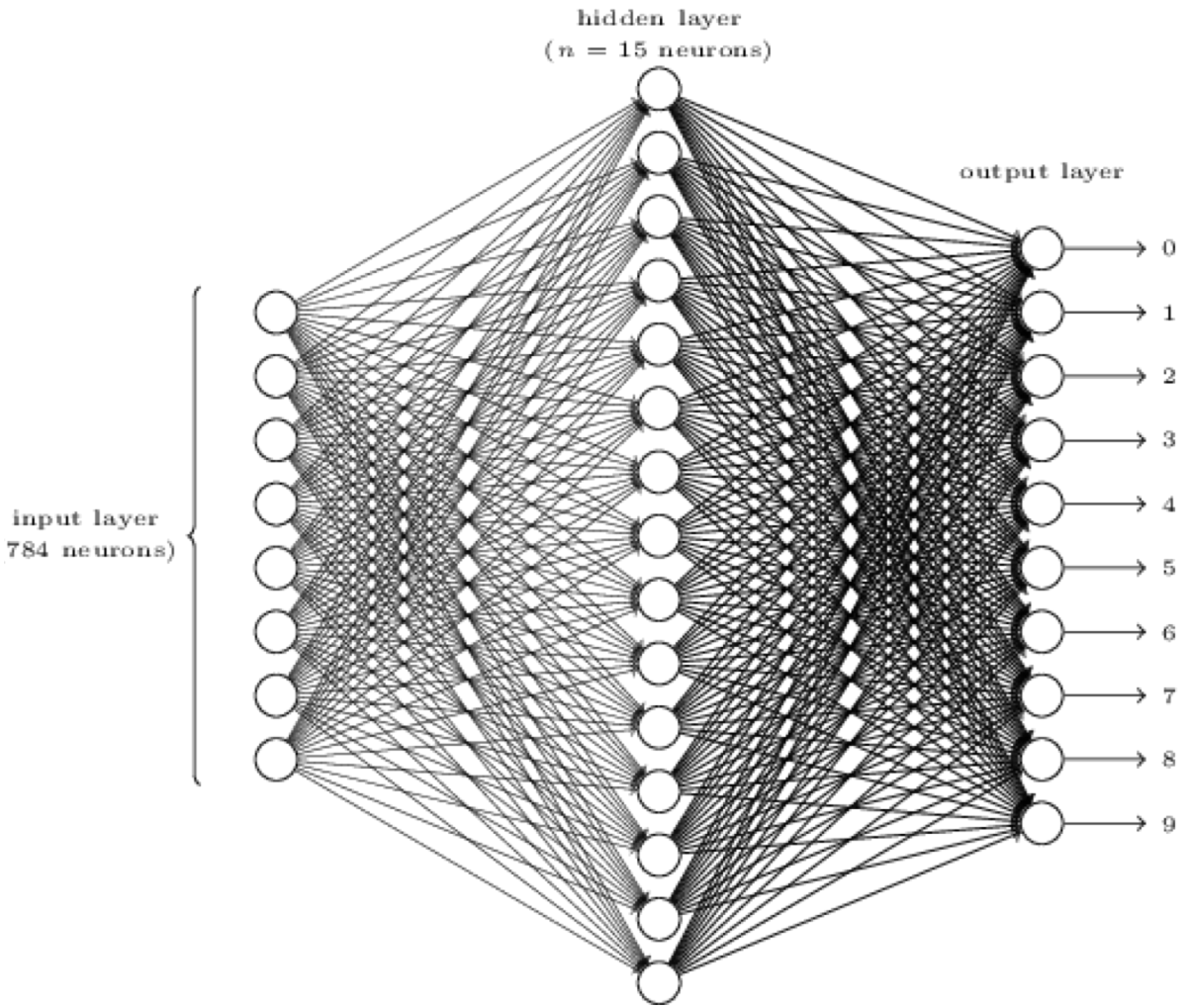
Why now?

- Multilayer networks have been around for 25+ years... What is so different now?
  - More data
  - More compute power
  - Better algorithms

Handwriting Recognition

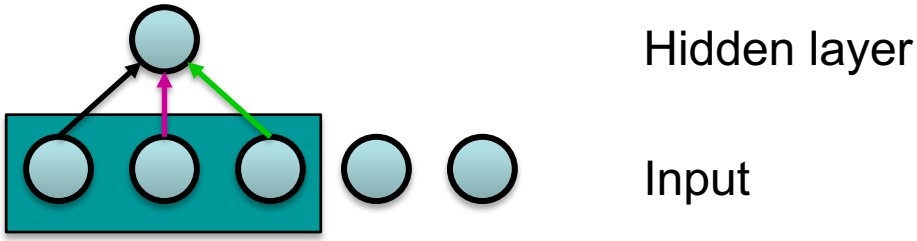
- Classification problem:
  - output should be a digit 0-9
  - Input is a 28x28 pixel image
- MNIST data set (60,000 examples)

Simple Feed-forward Network for Handwriting Recognition

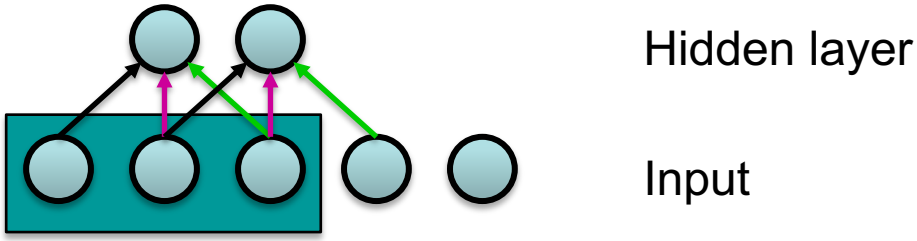


- 2-layer feed forward NN, trained with backprop: error rate of 3%
- Ignores any structure in the image
- Next step: bigger, deeper networks
  - Best NN variant: error rate of 0.25%

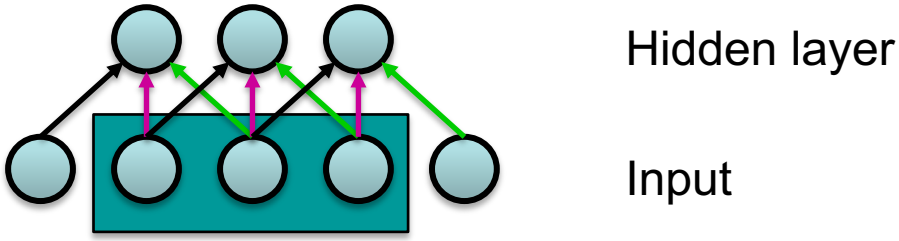
Basic Idea of CNNs (Convolutional Neural Networks)



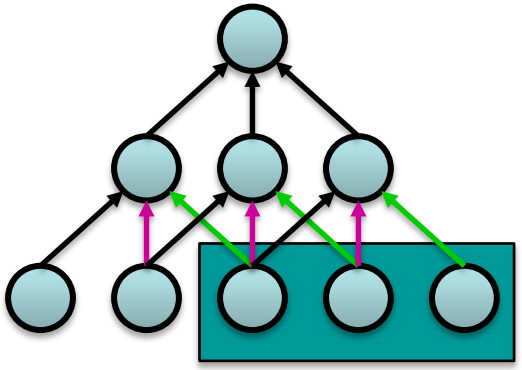
Basic Idea of CNNs (Convolutional Neural Networks)



Basic Idea of CNNs (Convolutional Neural Networks)



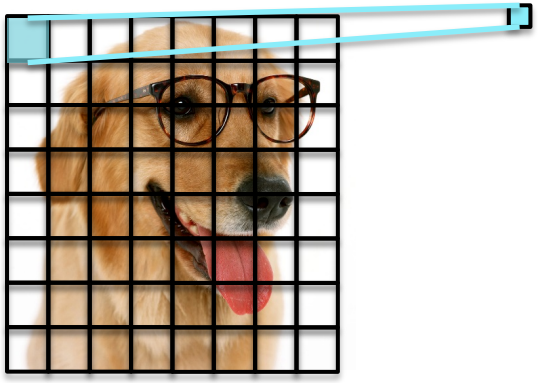
Basic Idea of CNNs (Convolutional Neural Networks)



Hidden layer

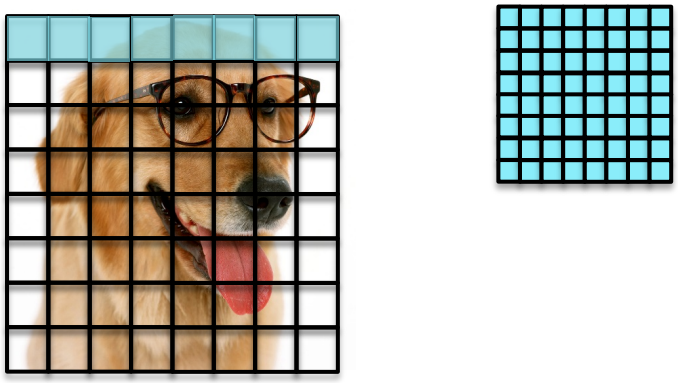
Input

CNN for Image Classification



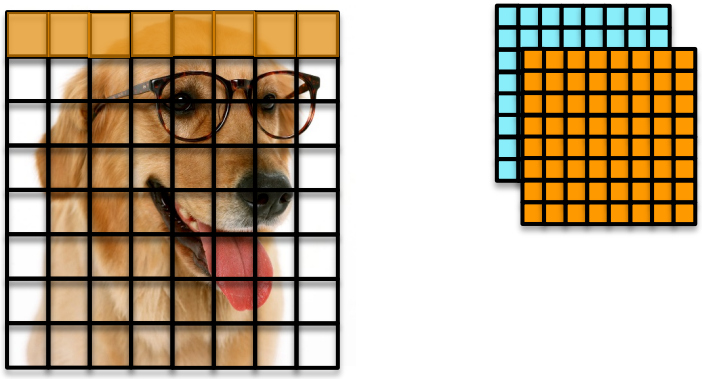
Convolutional Layer

CNN for Image Classification



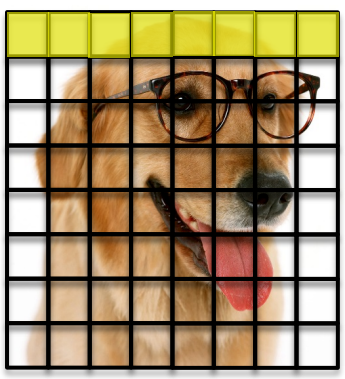
Convolutional Layer

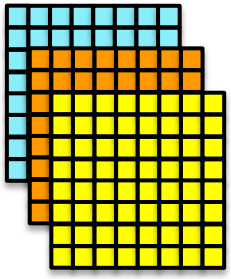
CNN for Image Classification



Convolutional Layer

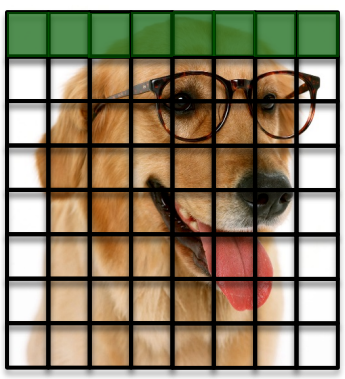
CNN for Image Classification

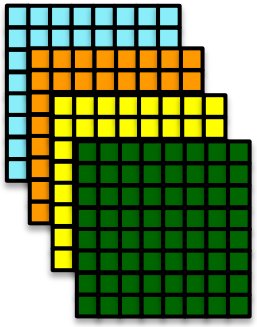




Convolutional Layer

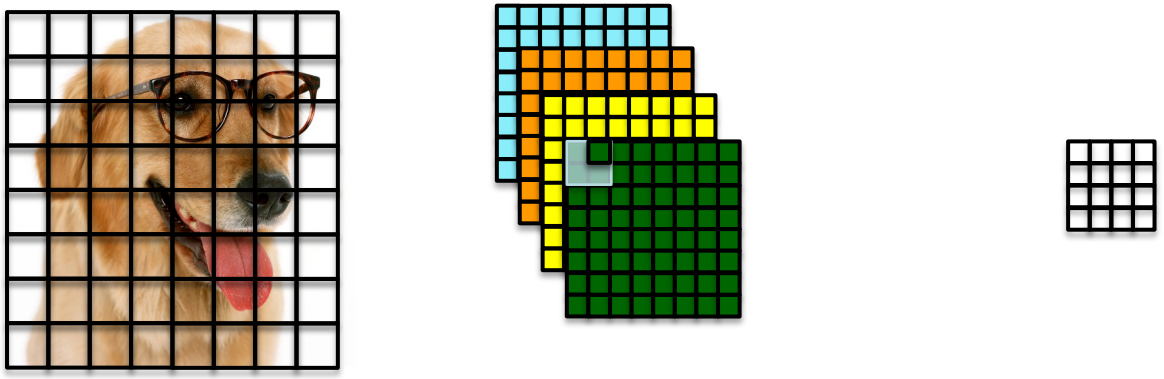
CNN for Image Classification





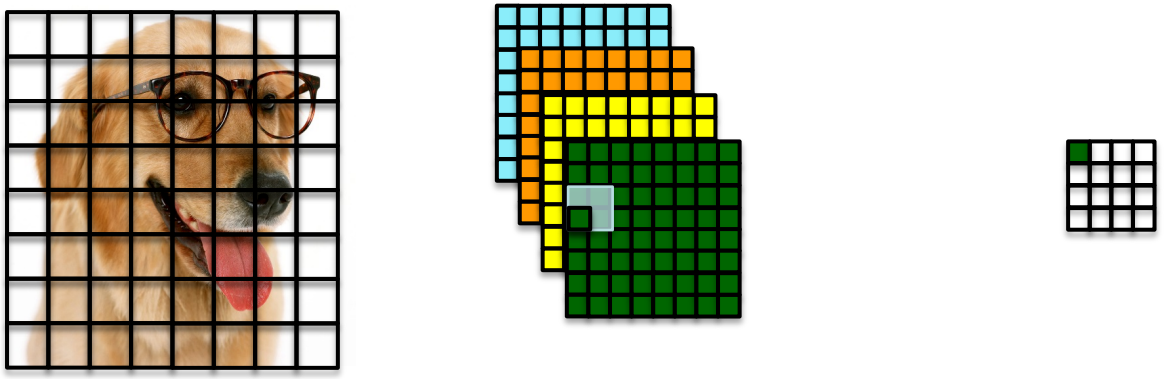
Convolutional Layer

CNN for Image Classification



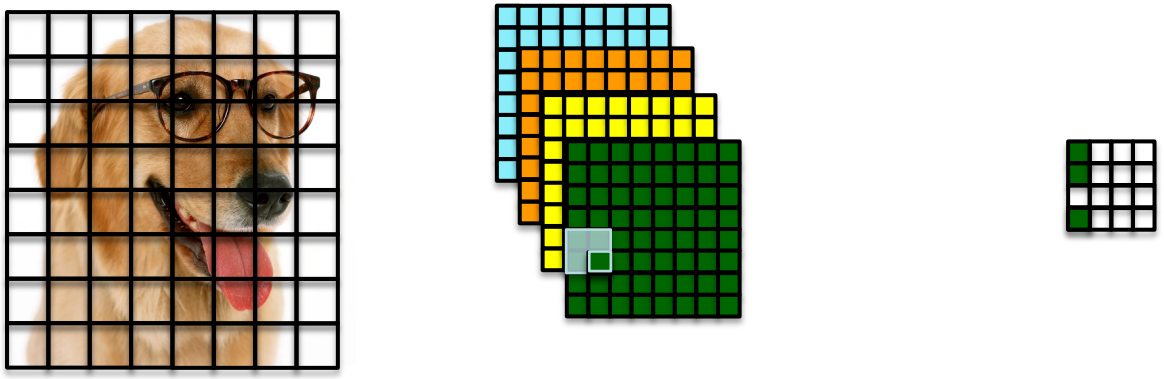
Max Pooling

CNN for Image Classification



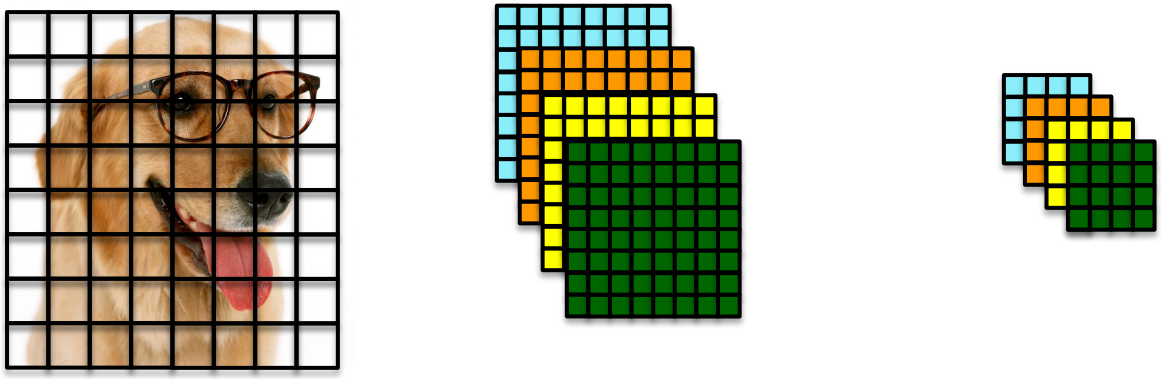
Max Pooling

CNN for Image Classification



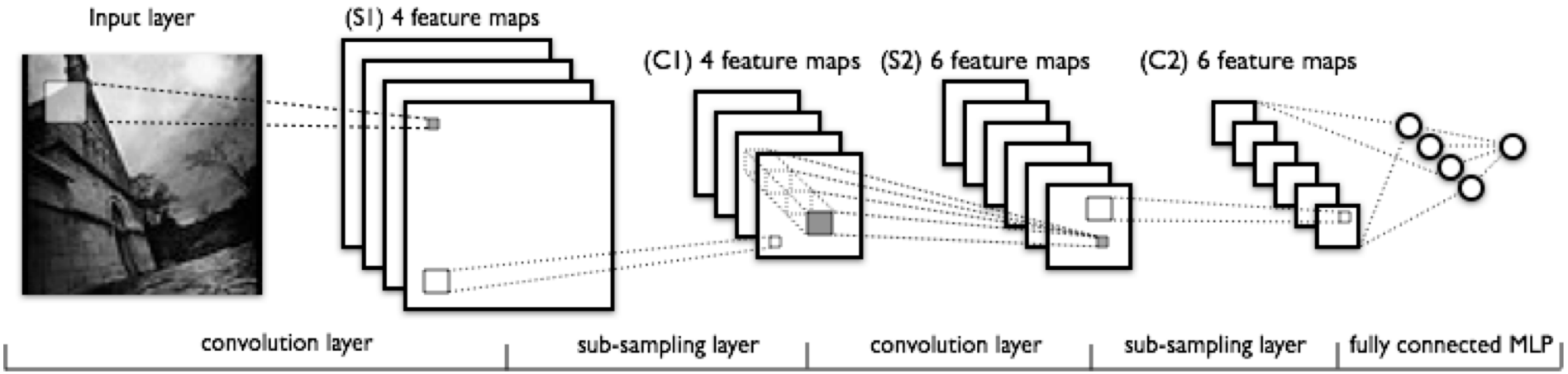
Max Pooling

CNN for Image Classification



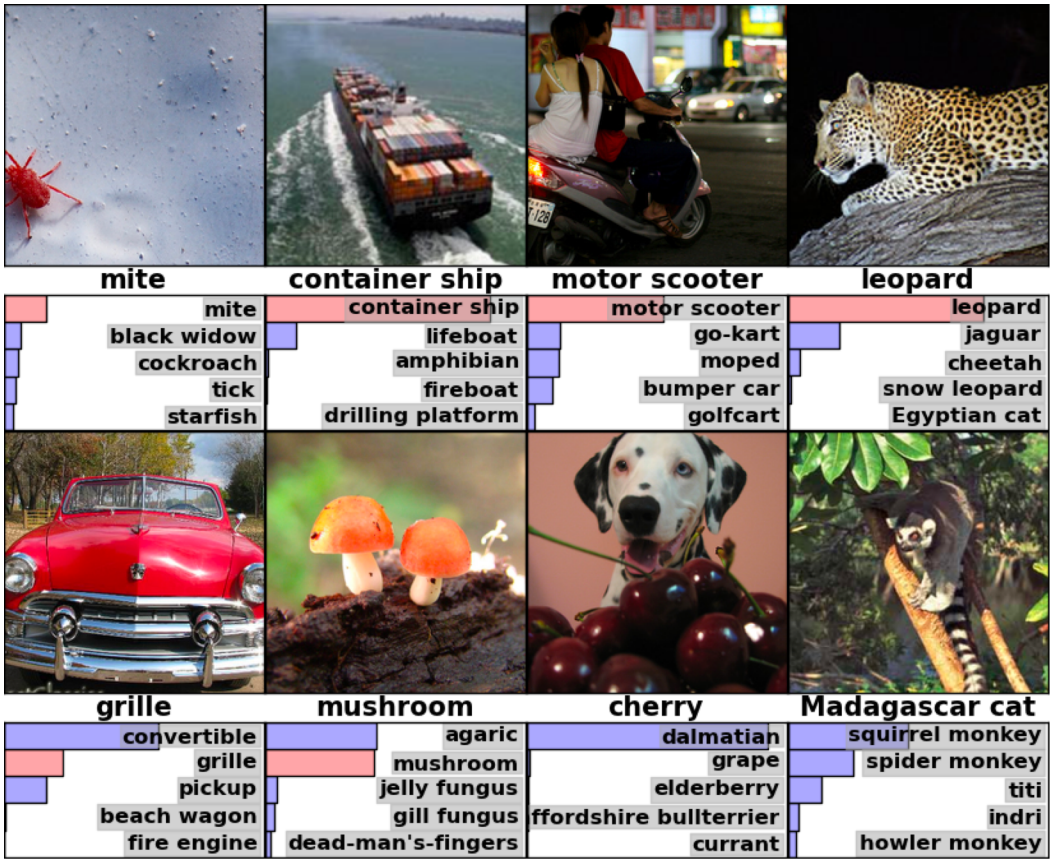
Max Pooling

CNN for Image Classification



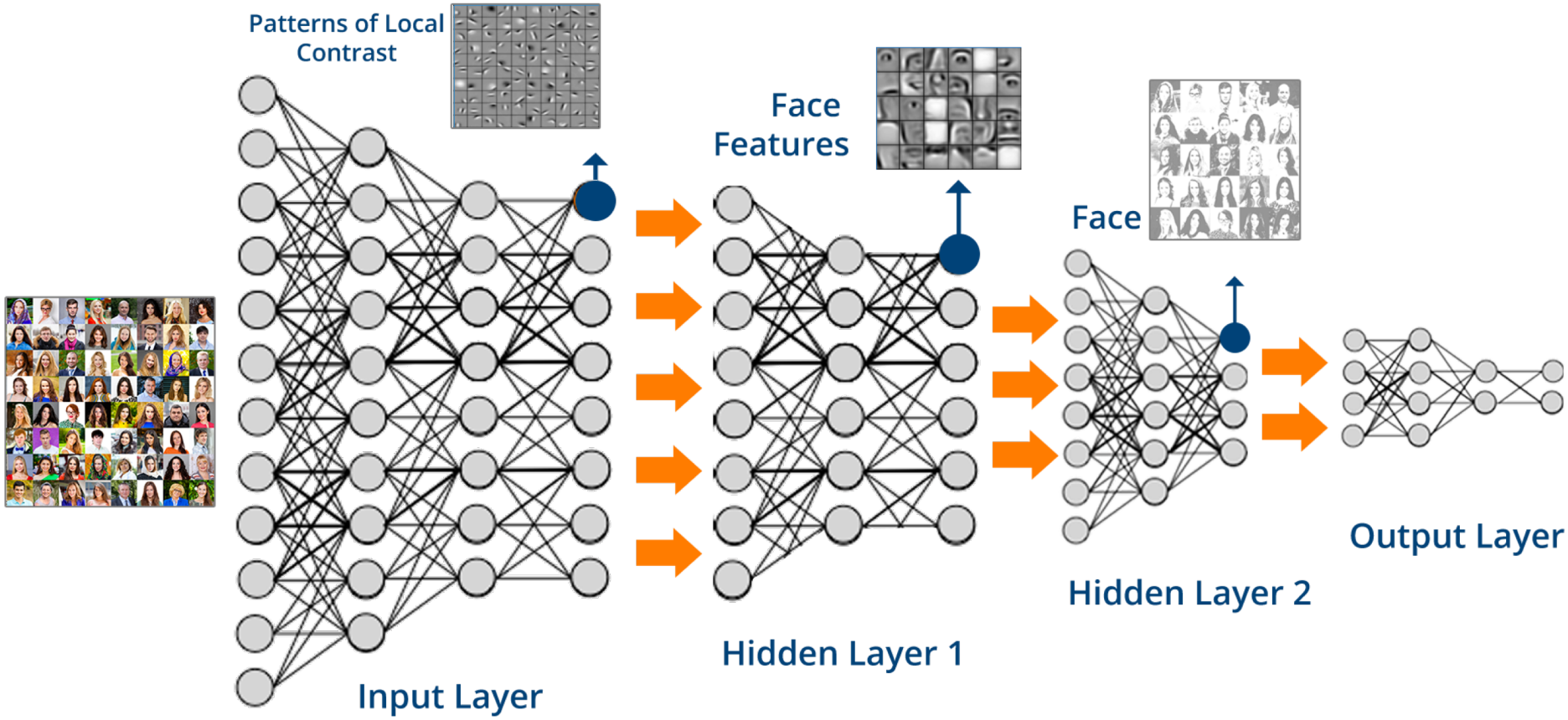
ImageNet Contest

- Total number of images: 14,197,122
- Number of images with bounding box annotations: 1,034,908



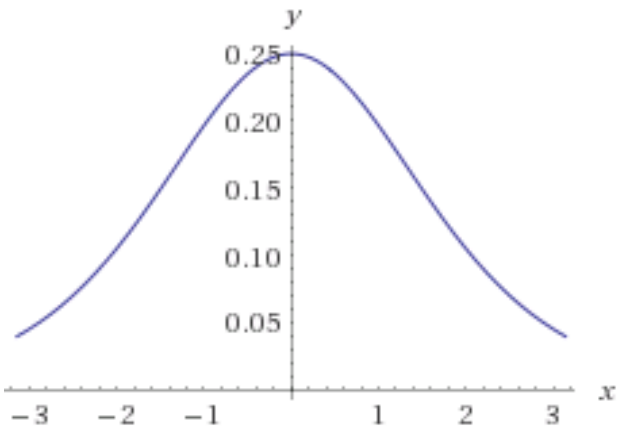
In 2012, CNNs scored a 15.3% error rate, compared to a 26.2% second-place finish

Central Idea of Deep Networks



The Vanishing Gradient Problem

- Deep neural networks use backpropagation.
- Back propagation uses the chain rule.
- The chain rule multiplies derivatives.
- Often these derivatives between 0 and 1.
- As the chain gets longer, products get smaller
- until they disappear.



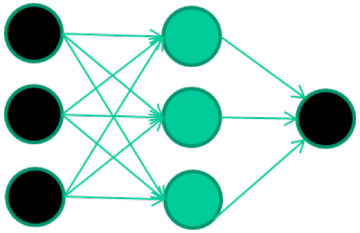
Derivative of sigmoid function

Or do they explode?

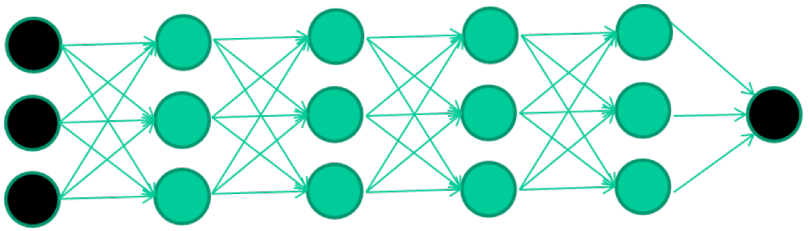
- With gradients larger than 1,
- you encounter the opposite problem
- with products becoming larger and larger
- as the chain becomes longer and longer,
- causing overlarge updates to parameters.
- This is the exploding gradient problem.

Algorithmic Improvements

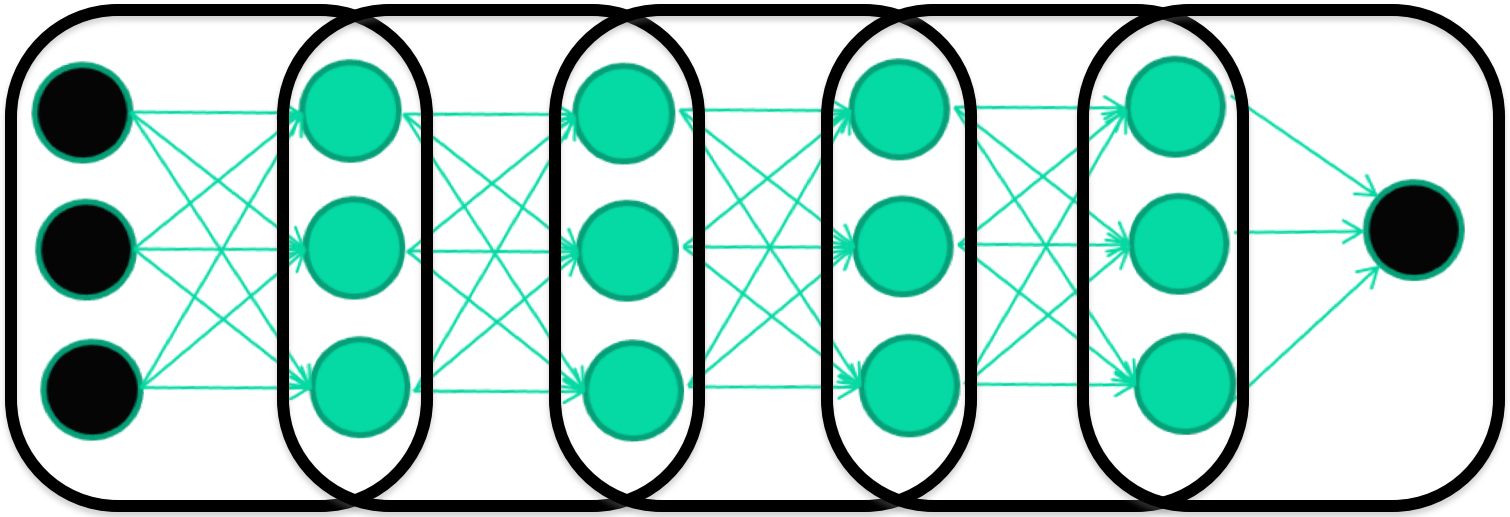
- Back propagation (and other similar algorithms) are good for learning weights in a single hidden layer, or a few layers deep



- But these algorithms are not good at learning the weights for networks with more hidden layers



New Training Technique



Train this layer...

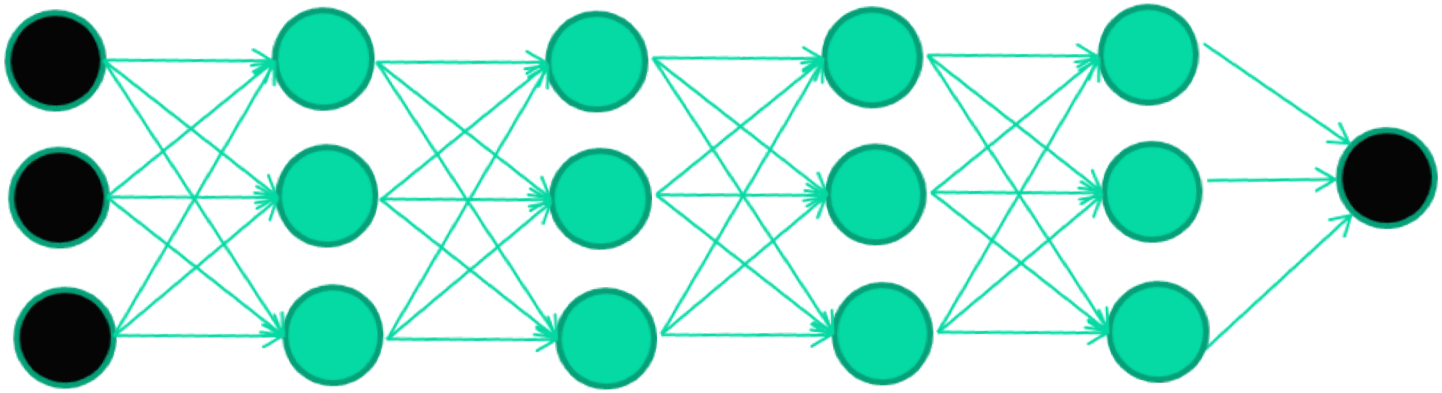
Train this layer...

Train this layer...

Train this layer...

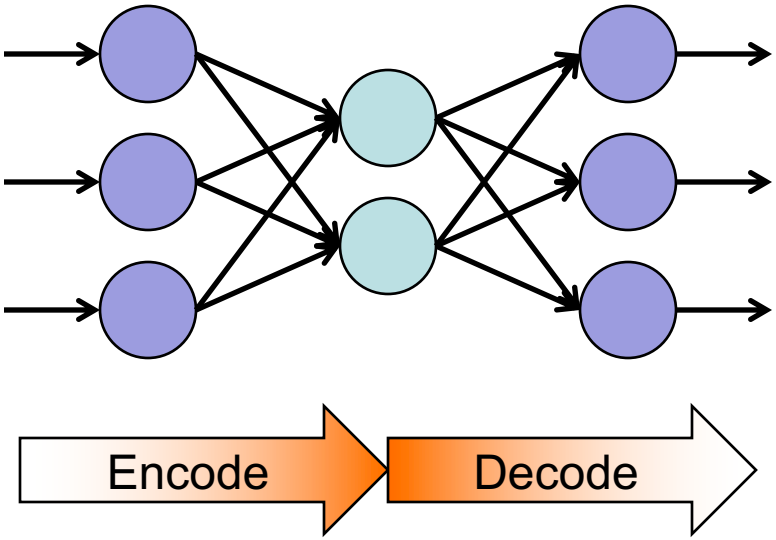
Train this layer...

New Training Technique



- How do we figure out what is a good internal representation?
- What do we use to train? We have no internal values...

Deep Learning Insight



- Train each (non-output) layer to be an auto-encoder
  - reproduce the input and train using a standard weight training algorithm (like back propagation)
- Learn what makes a good internal representation – anything that can reconstruct the input

Deep Face

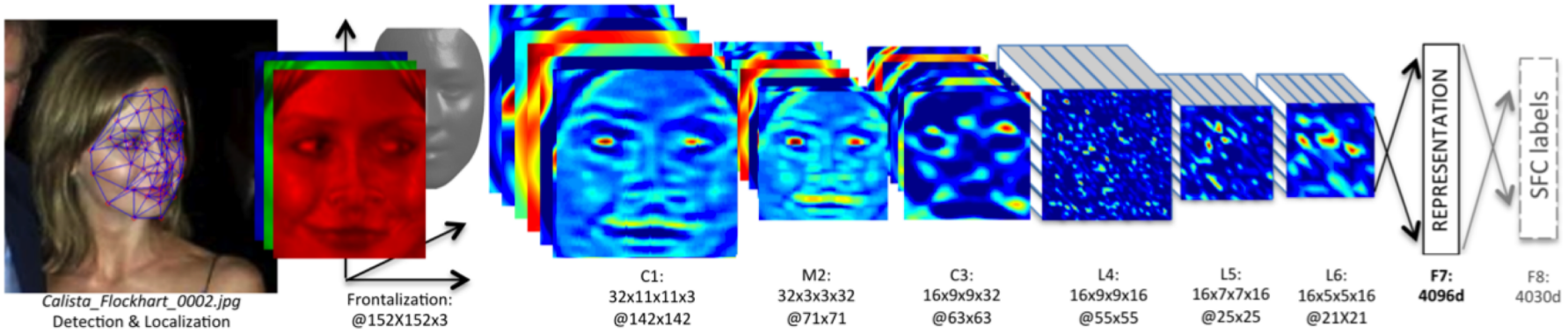


Figure 2. **Outline of the *DeepFace* architecture**. A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate feature maps produced at each layer. The net includes more than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

Face Recognition: Detect → Align → Represent → Classify

- Labeled Faces in the Wild dataset with 4M images
- Deep Face: 97.35% accuracy with 120M parameters
- Human performance: 97.5% accuracy

Taigman,Yang, Ranzato, & Wolf (Facebook), CVPR 2014

ImageNet Architecture

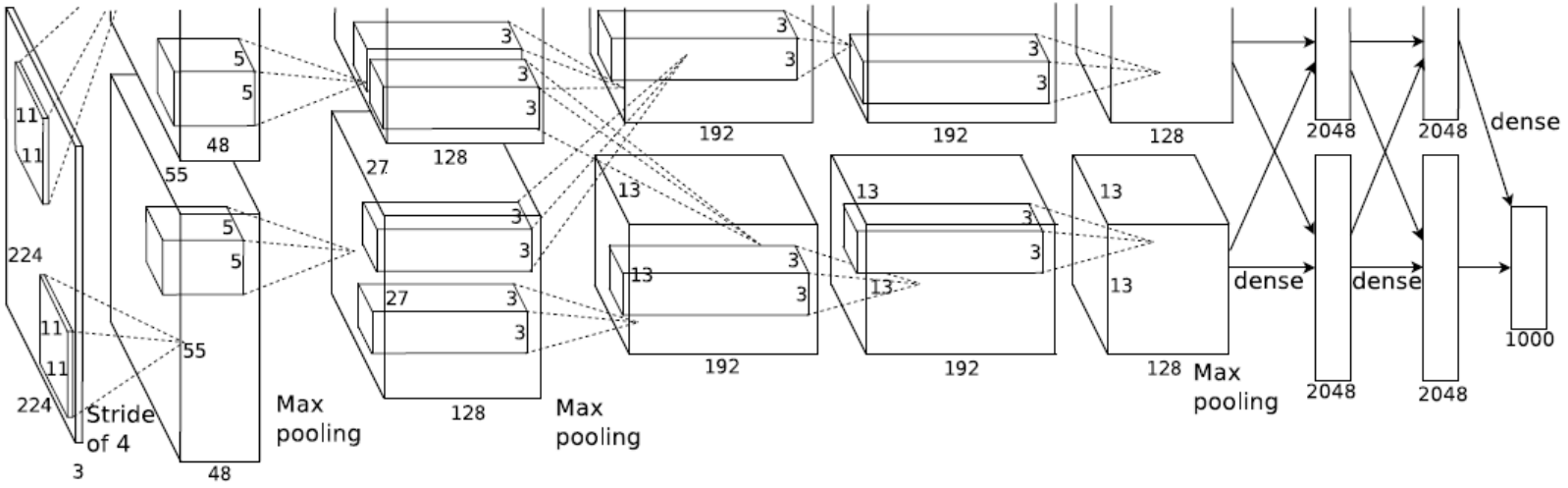


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

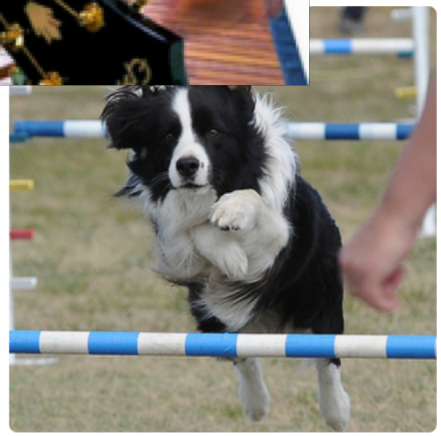
Current (2018) best scores on ImageNet yield a 2% error rate.

rker in
est is
ad."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

"girl in pink dress is jumping in air."



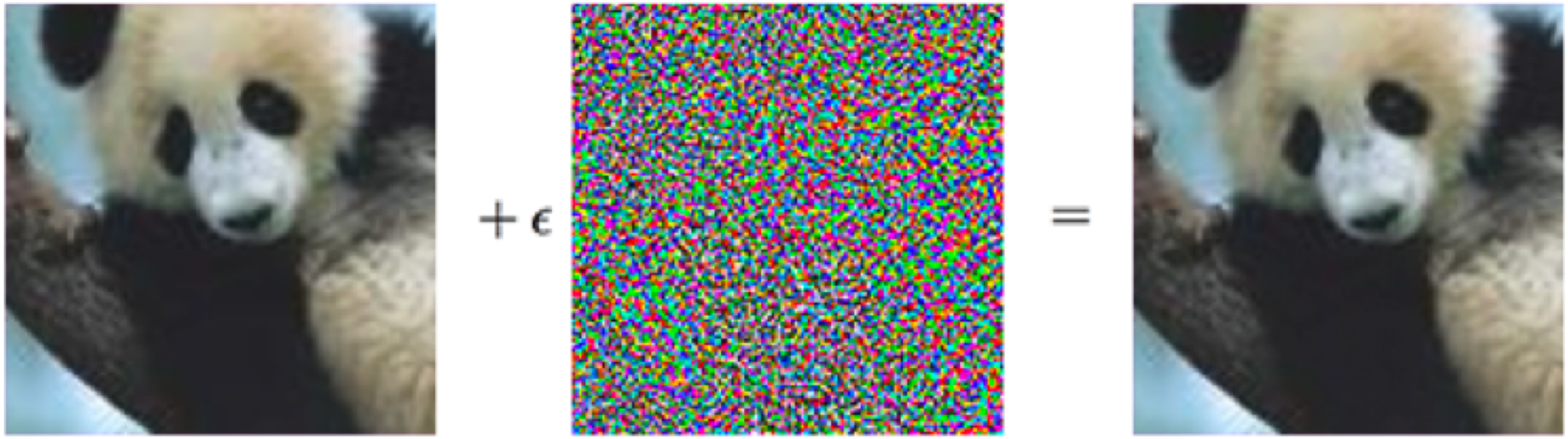
"black and white dog jumps over bar."

"young girl in pink shirt is swinging on swing."

"man in blue wetsuit is surfing on wave."

Deep Visual-Semantic Alignments for Generating Image Descriptions (CVPSR 2015). Andrej Karpathy, Li Fei-Fei.

Deep Network Problems



https://xkcd.com/1838/