Deep Neural Networks

Nathan Sprague

CS444

The Deep Learning "Revolution"

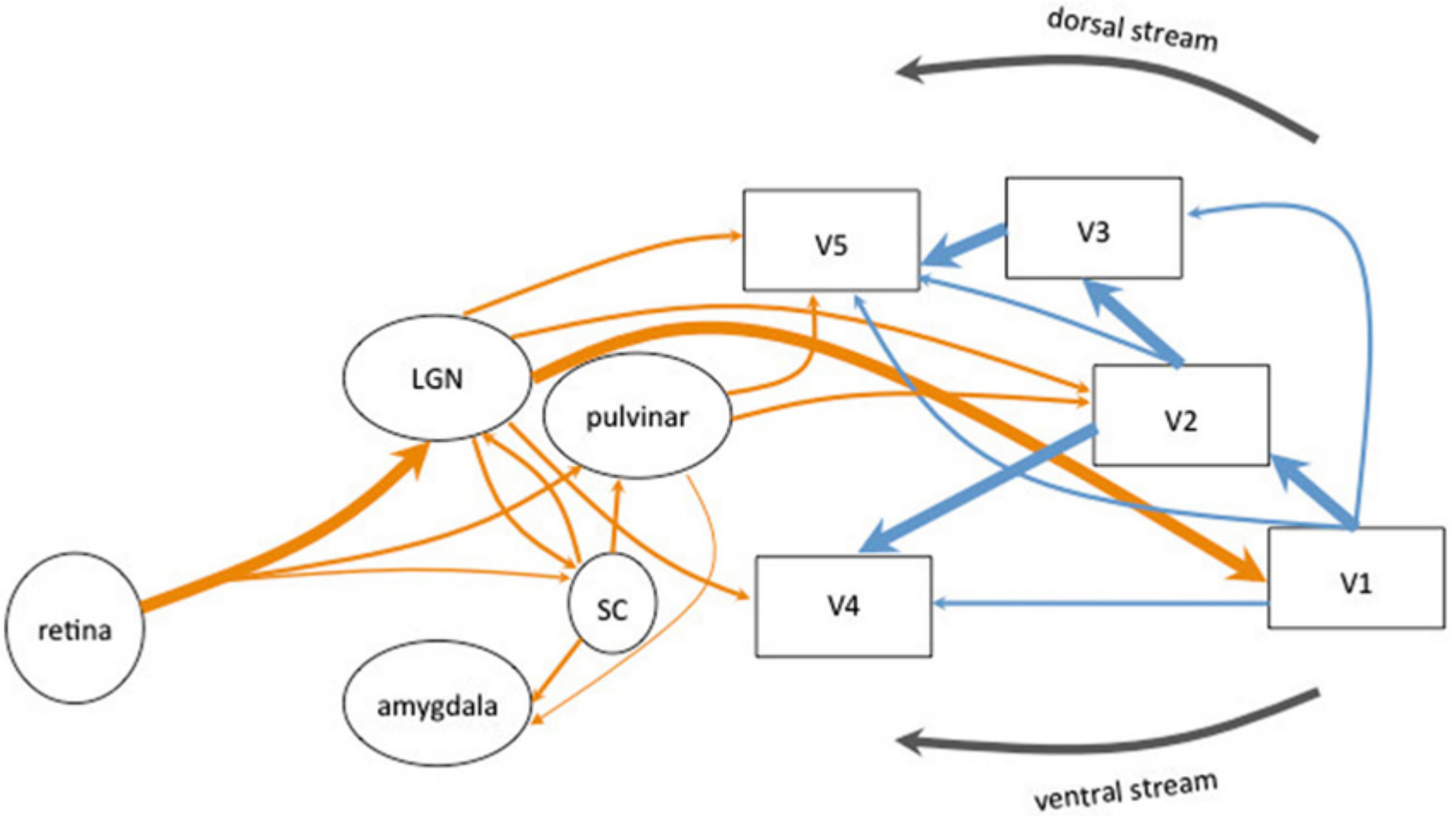
- Geoff Hinton introduced a simple idea in 2006
- Greedy, Layer-Wise, Unsupervised Pre-Training
  - Train the first hidden layer to re-represent the input.
  - Train the second hidden layer to re-represent the first hidden layer
  - ...
  - Fine-tune the entire network using backpropagation on labeled data

G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, pp. 1527–1554, 2006.

The Flood Gates Open

- Better Hardware
  - GPGPU
  - Cluster Computing
- Massive Data Sets
  - Street View House Numbers
  - Kaggle
  - ImageNet
  - KITTI
- Better Training Algorithms
  - Batch Normalization
  - RMSProp
  - Dropout
  - Adam
  - Adadelta
- New Architectures
  - Maxout
  - Rectified Linear Units
  - Resnets

Human Visual System



Urbanski, Marika, Olivier A. Coubard, and Clémence Bourlon. "Visualizing the blind brain: brain imaging of visual field defects from early recovery to rehabilitation techniques." Neurovision: Neural bases of binocular vision and coordination and their implications in visual training programs (2014).

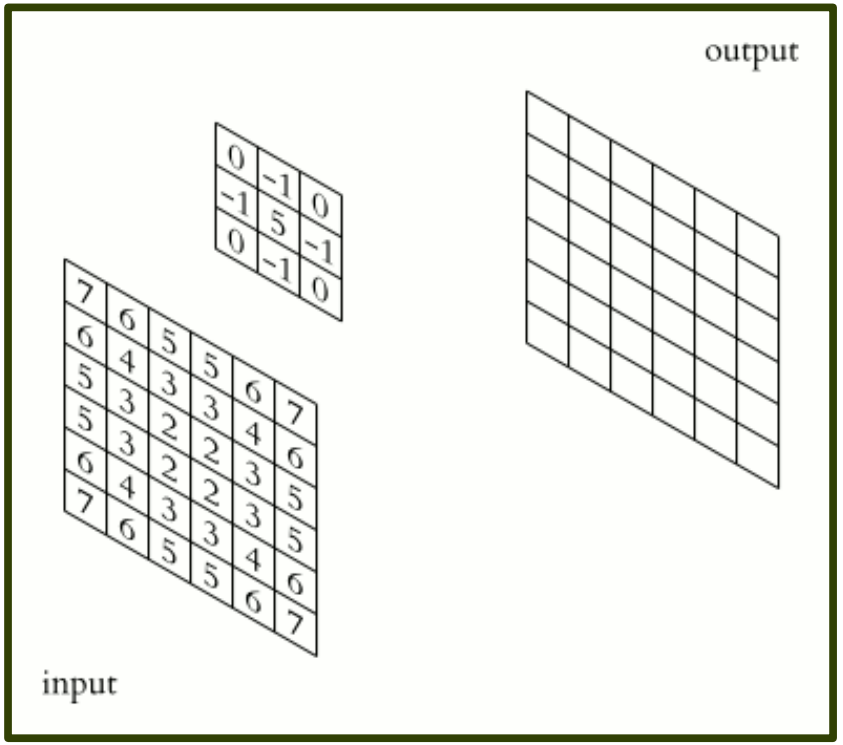
Convolutional Neural Networks

- Convolutional neural networks use the same trick of learning layers of localized features...
- CNN's were actually being used by Yann Lecun at Bell Labs around 1990
- (He would probably argue that "deep learning" is not so new)

Convolutions

Grayscale Image

1 convolutional filter

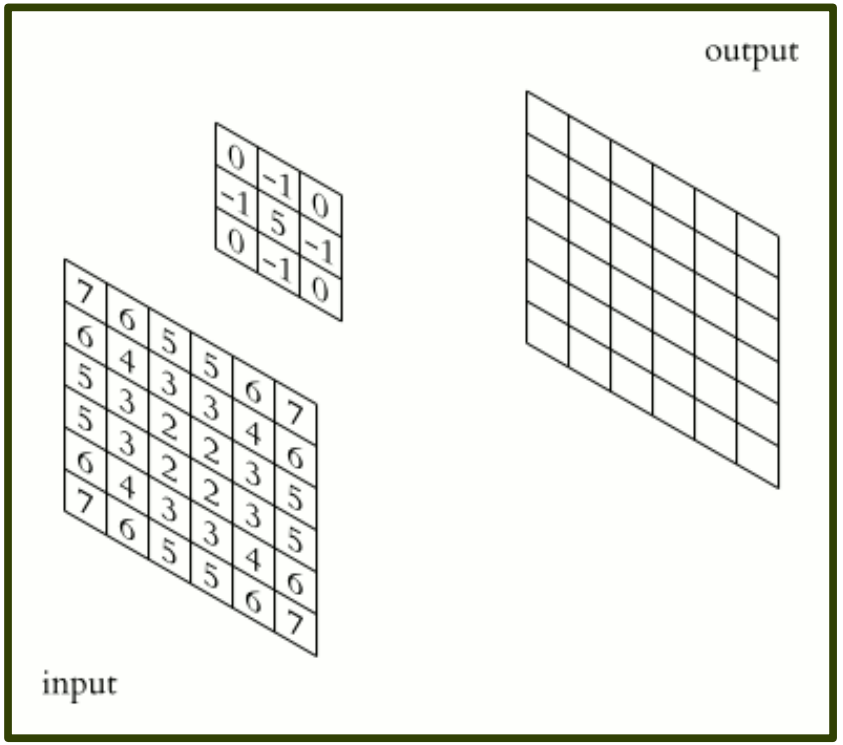


http://upload.wikimedia.org/wikipedia/commons/4/4f/3D_Convolution_Animation.gif
By Michael Plotke [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)

Convolutions

Grayscale Image

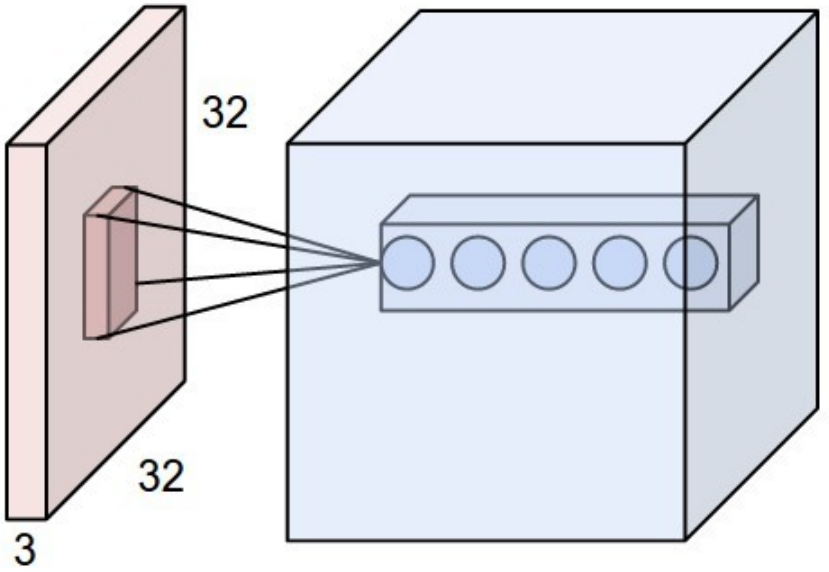
1 convolutional filter



http://upload.wikimedia.org/wikipedia/commons/4/4f/3D_Convolution_Animation.gif
By Michael Plotke [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)

Color Image

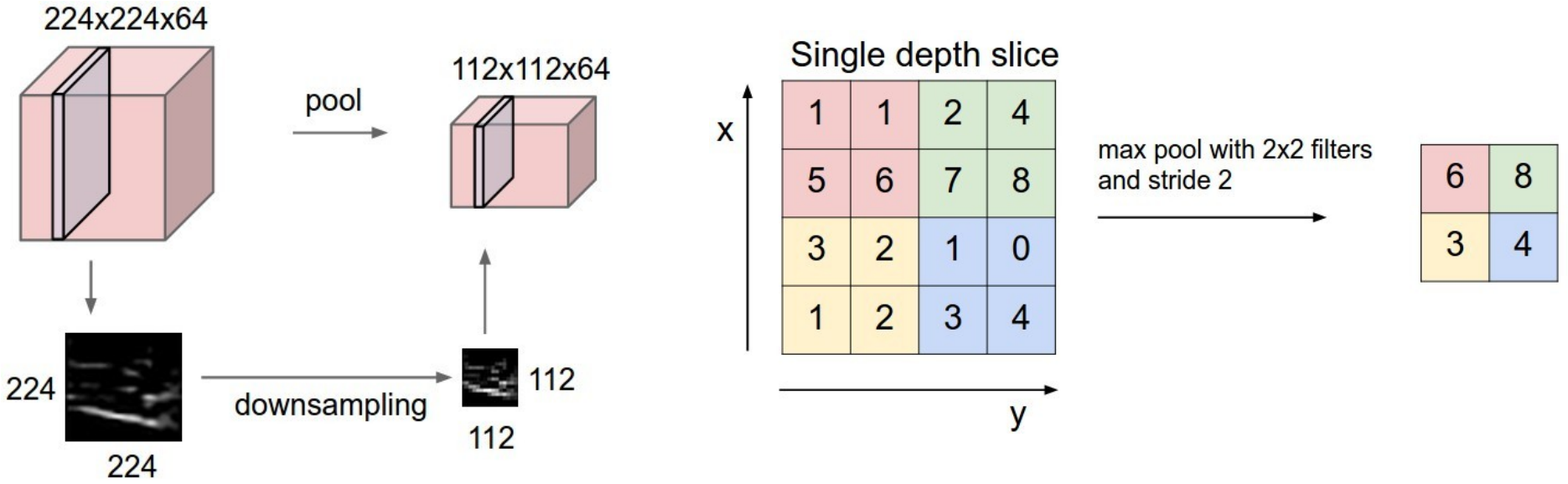
5 convolutional filters



http://cs231n.github.io/convolutional-networks/

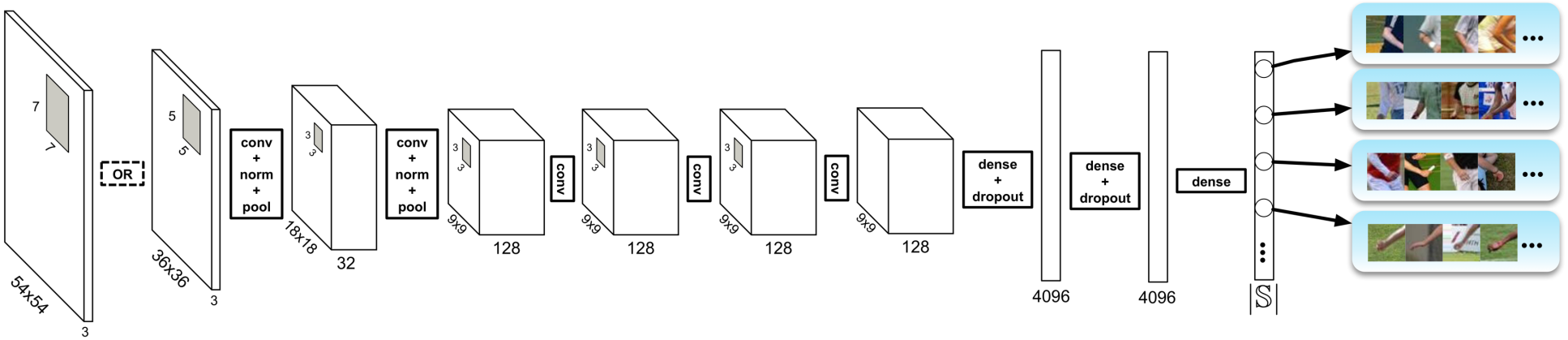
Pooling Layers

- Pooling layers down-sample the filter outputs to
  - Reduce dimensionality and computational requirements
  - Increase the spatial extent of subsequent filters



Complete Network

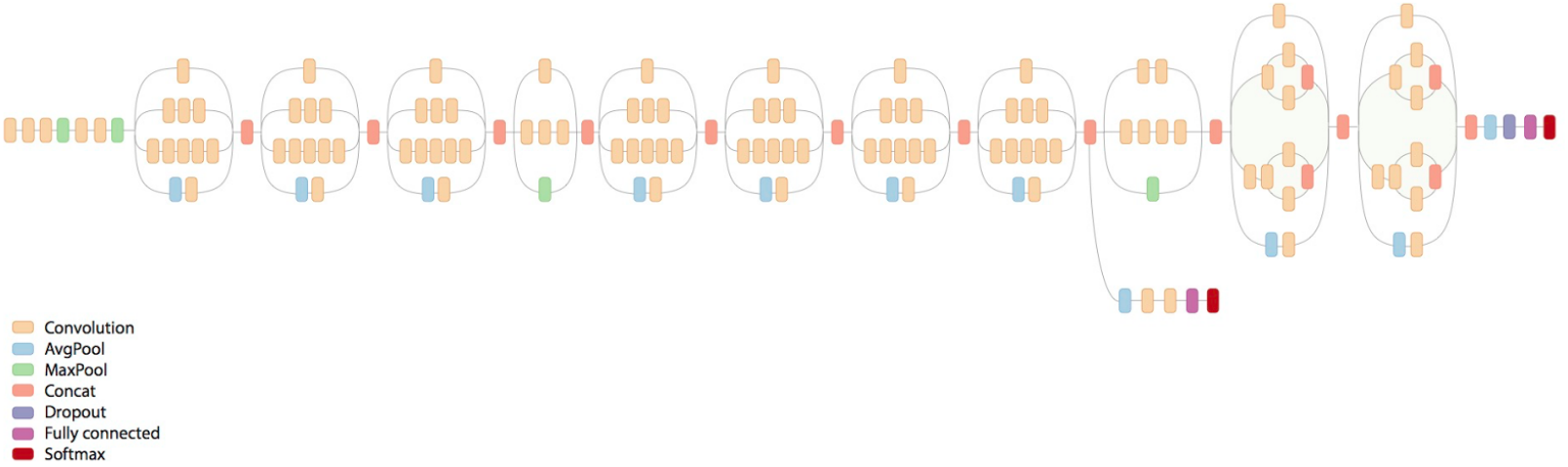
- A “traditional” CNN is composed of convolutional layers, each followed by non-linearities, followed by pooling layers, with a dense (non-convolutional) layer at the end:



Chen, Xianjie, and Alan L. Yuille. "Articulated pose estimation by a graphical model with image dependent pairwise relations." Advances in Neural Information Processing Systems. 2014.

Current State of The Art

- Current best-performing networks have somewhat more complicated architectures.
- GoogleNet for example:



Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.