Last Time

- The components of convolutional neural networks
- Alexnet, VGG

Today

- Understanding what's going on inside the networks (David: ~10 minutes)
- An overview of the caffe toolbox (Rohit)

Logistics

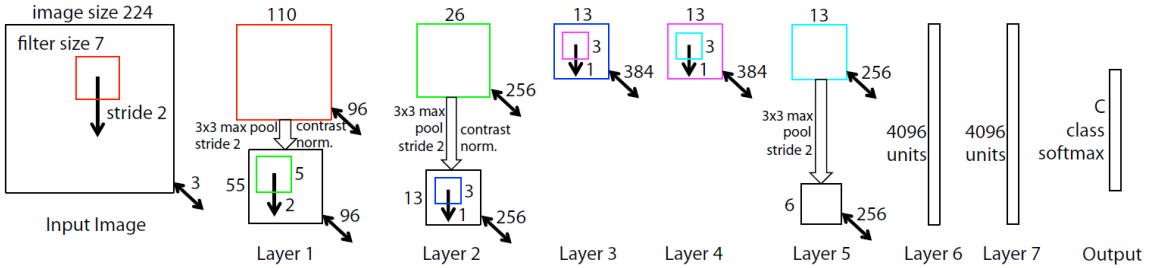
- Blog post: Tuesday 10PM at latest
  - No strict format
  - Approximately two paragraphs
- Project Proposal Deadline: Feb 15, Noon
  - 2 pages (max)
  - Team, problem definition, plan

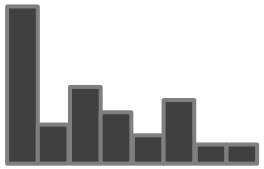
Visualizing and Understanding Convolutional Networks

David Fouhey

Many figures from Matt Zeiler

Review





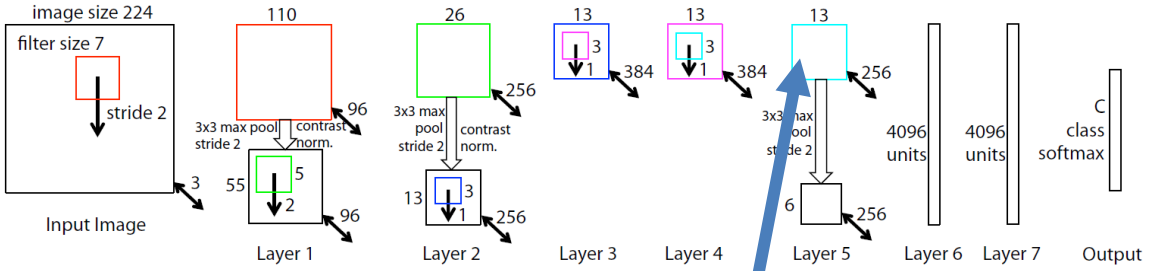
Image

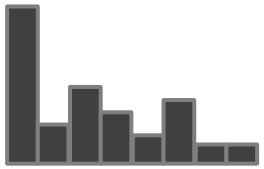
P(Class|Image)

When I started

- “It’s a black box!”
- “Nobody understands what’s going on!”
- “Conv1 is gabor filters, but what’s actually going on?!”
- “Sure, LeCun and Hinton know how to make them work, but it’s magic.”

Goal





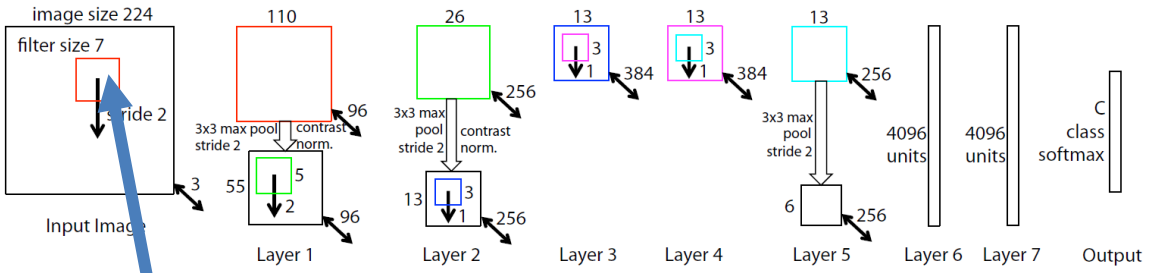
Image

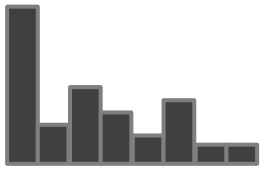
P(Class|Image)

What does this neuron mean?

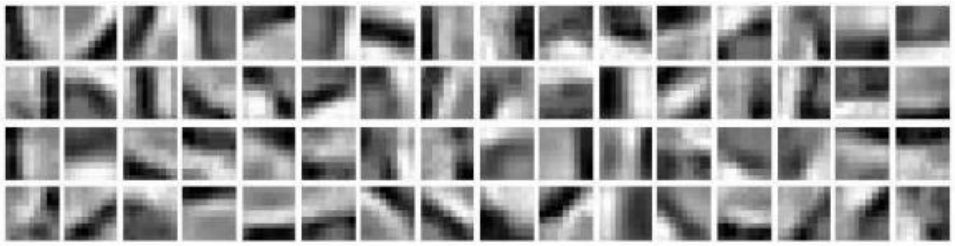
One Solution

Image



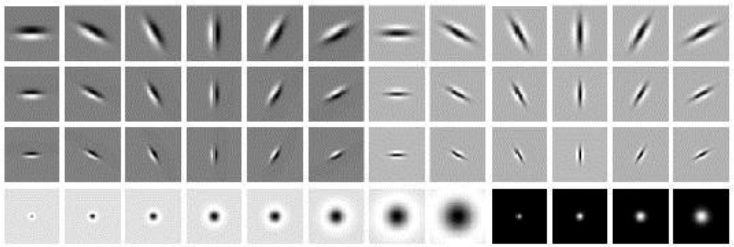


P(Class|Image)



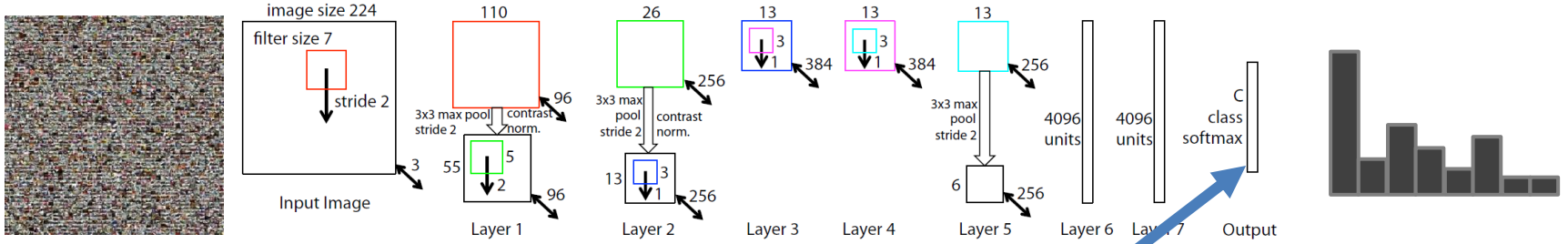
Ranzato et al. ‘07

Compare With:



Leung and Malik ‘01

One Simple Scheme



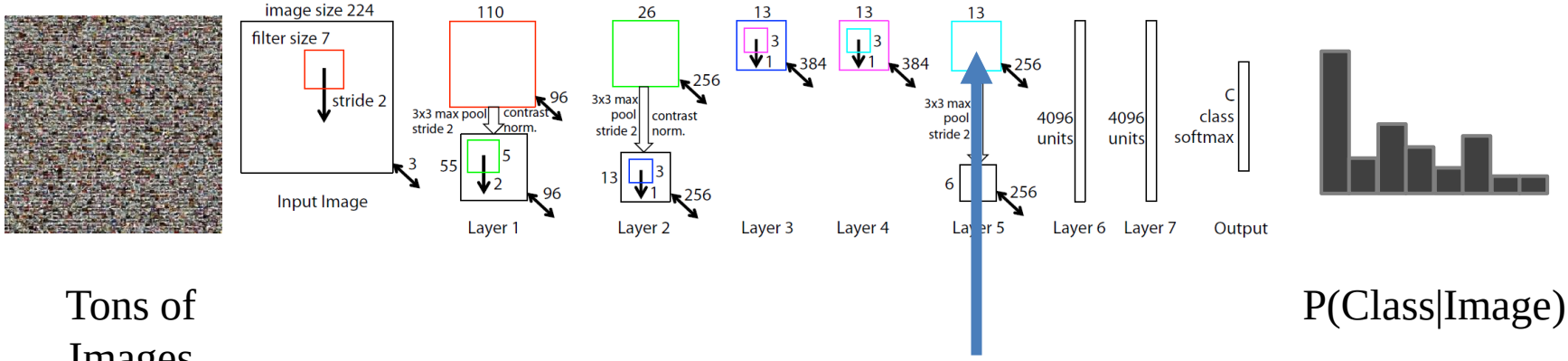
Tons of Images

P(Class|Image)

Most Wallaby-like

Least Wallaby-like

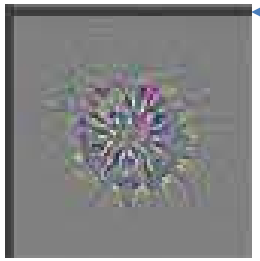
One Simple Scheme



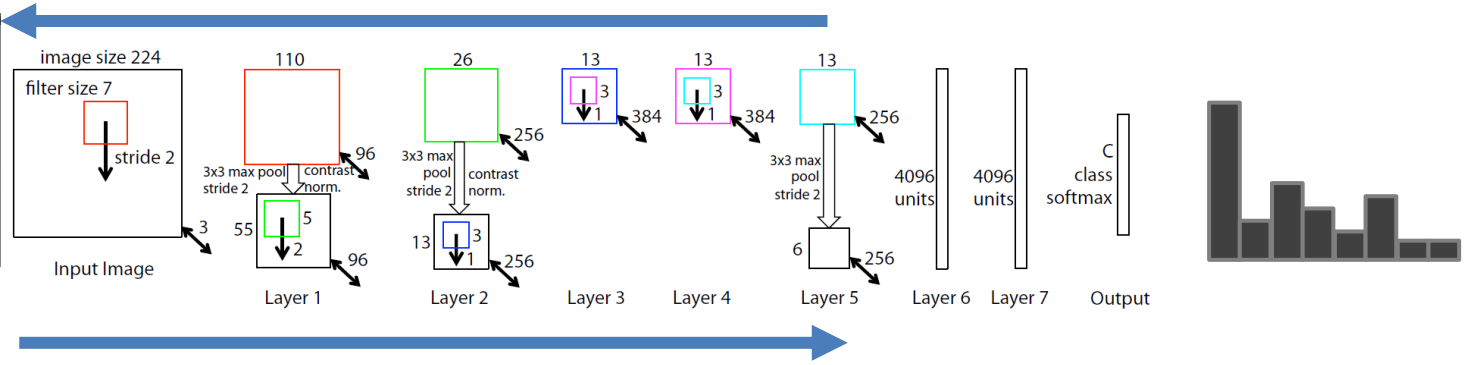
Tons of Images

P(Class|Image)

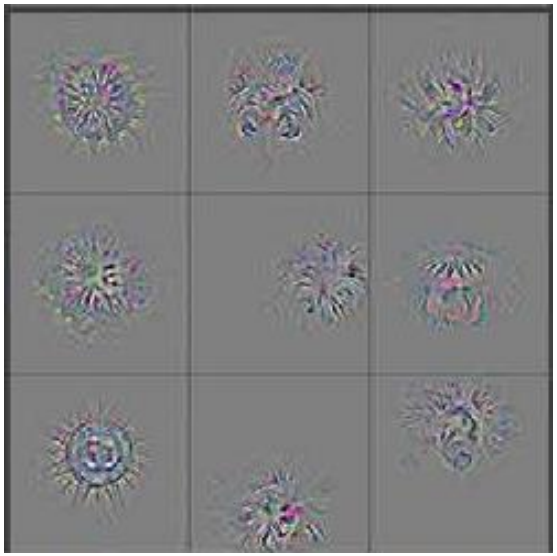
What's Really Going On?



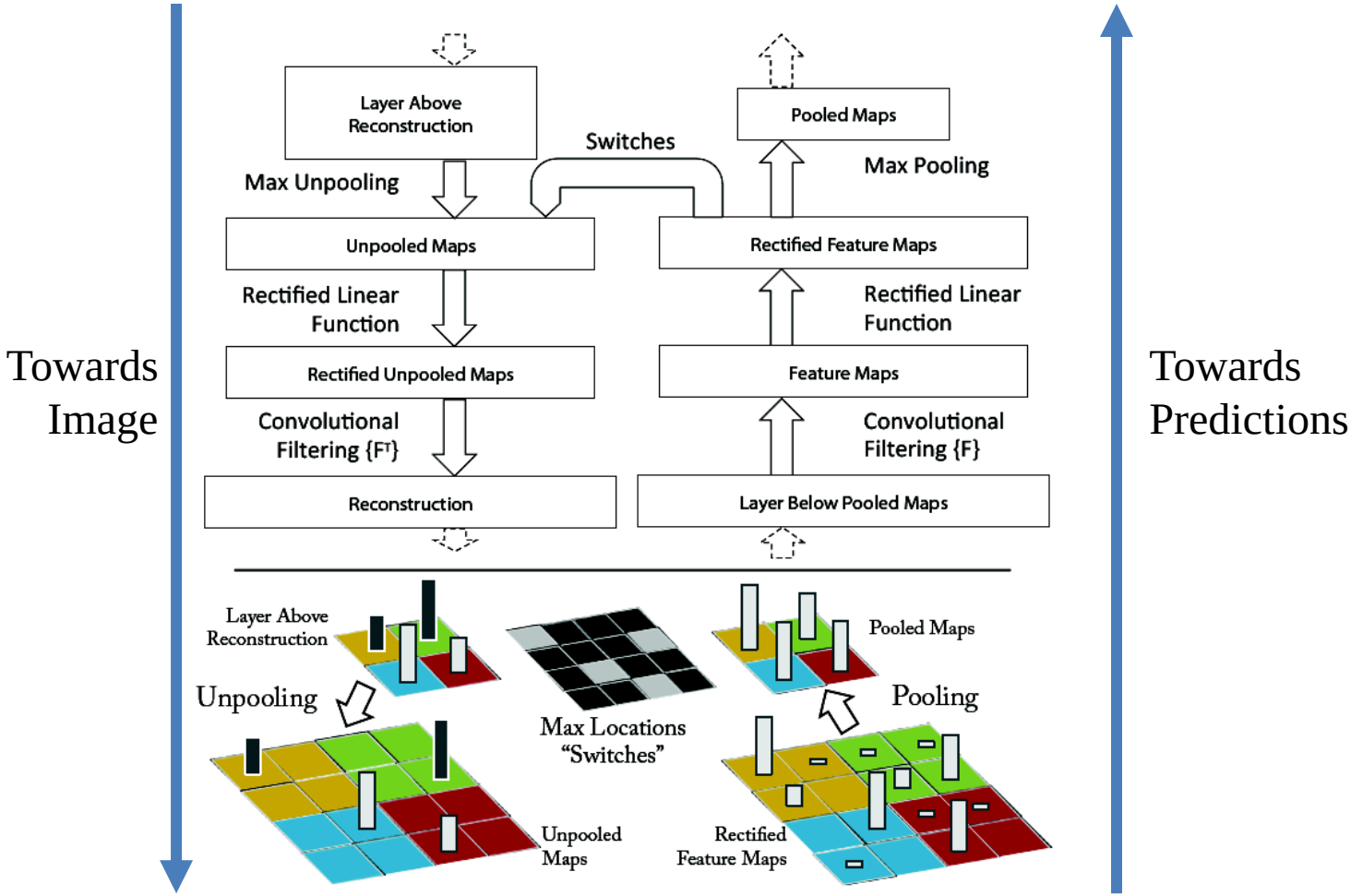
Max-Response Image



P(Class|Image)



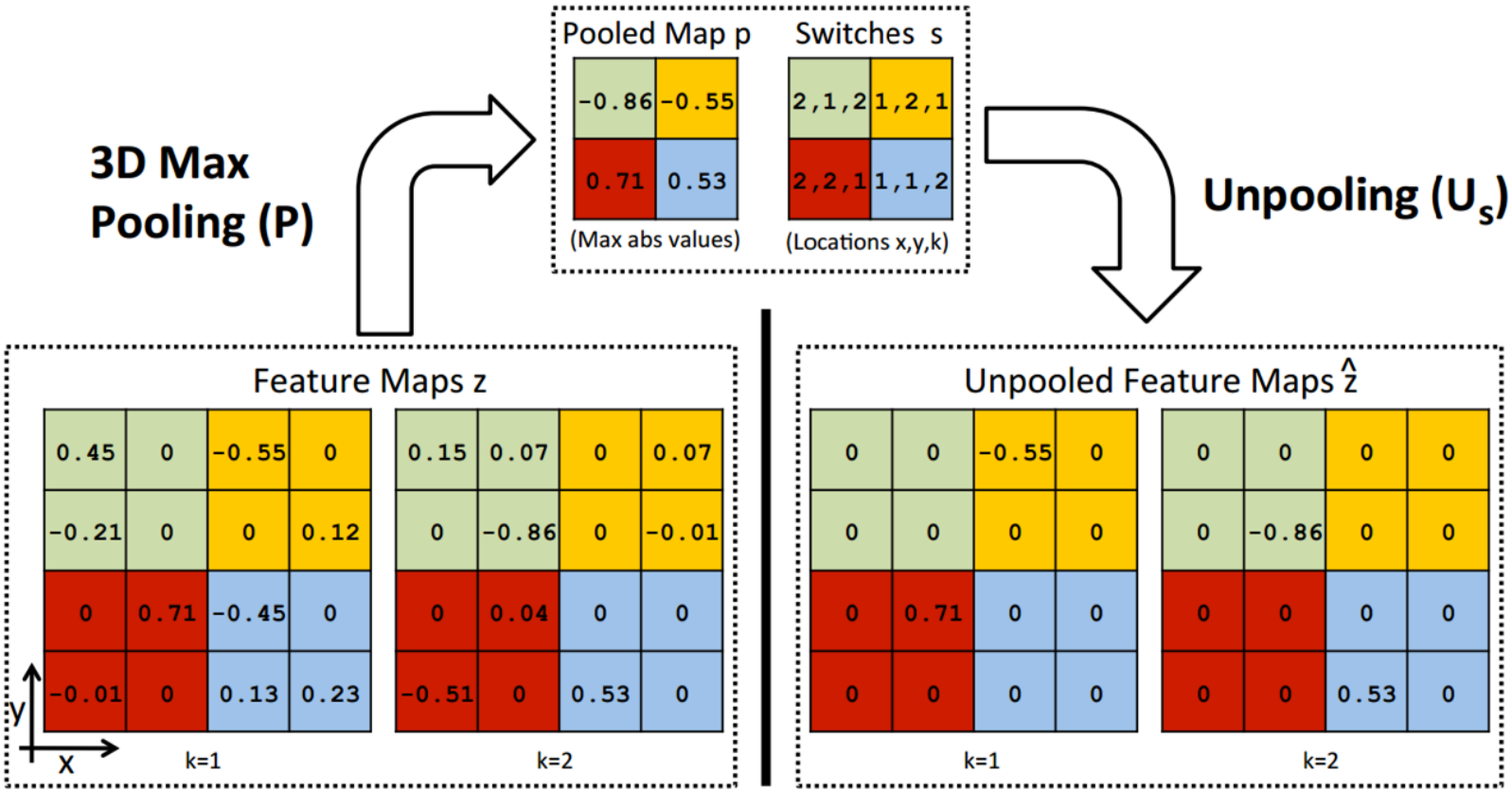
Going Back To The Image



Things to Invert

- Convolutions/Filtering
- Rectification/Non-linearity
- Pooling

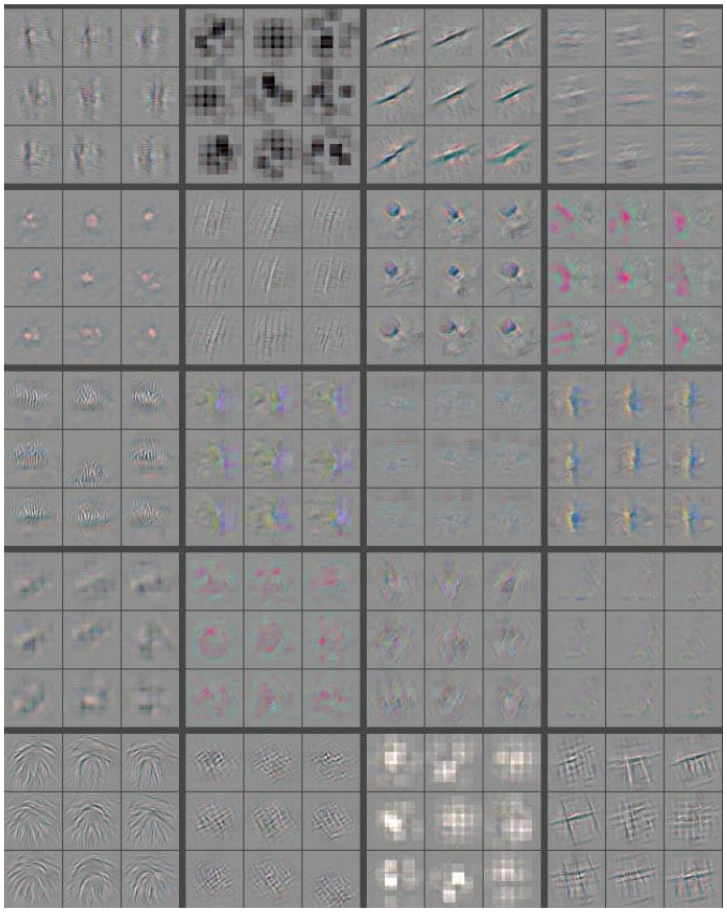
One Problem


3D Max Pooling (P)
Pooled Map p
Switches s
-0.86 -0.55
0.71 0.53
(Max abs values)
2,1,2 1,2,1
2,2,1 1,1,2
(Locations x,y,k)
Unpooling (U_s)
Feature Maps z
0.45 0 -0.55 0
-0.21 0 0 0.12
0 0.71 -0.45 0
-0.01 0 0.13 0.23
0.15 0.07 0 0.07
0 -0.86 0 -0.01
0 0.04 0 0
-0.51 0 0.53 0
y
x
k=1
k=2
Unpooled Feature Maps $\hat{z}$
0 0 -0.55 0
0 0 0 0
0 0.71 0 0
0 0 0 0
0 0 0 0
0 -0.86 0 0
0 0 0 0
0 0 0.53 0
k=1
k=2

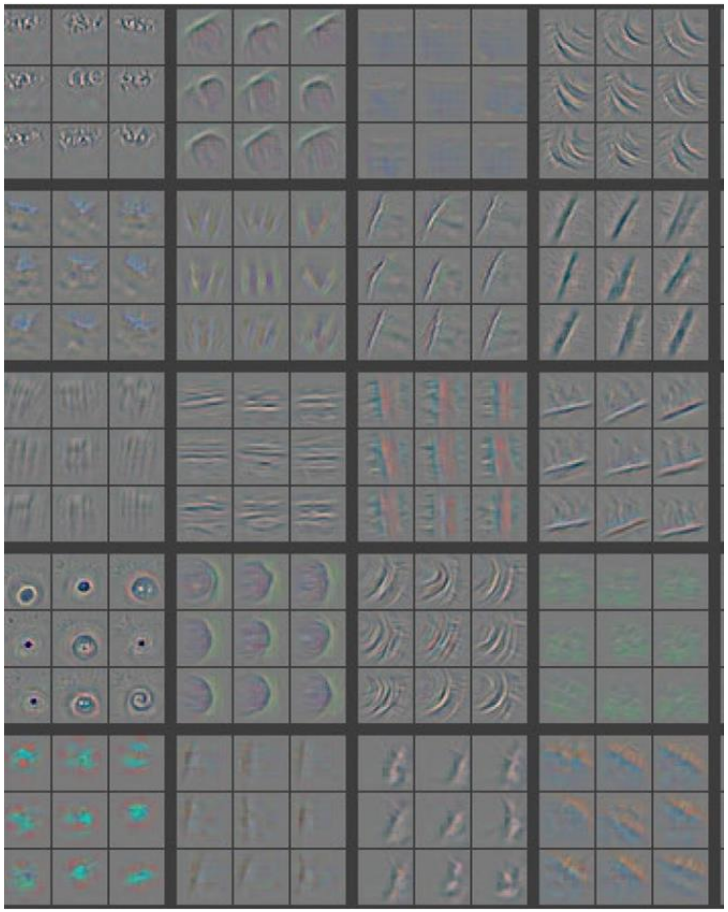
Tour through the Network

- Tour Through the Network

Is This Useful?



Alexnet



Zeiler and Fergus
+1.7% Accuracy