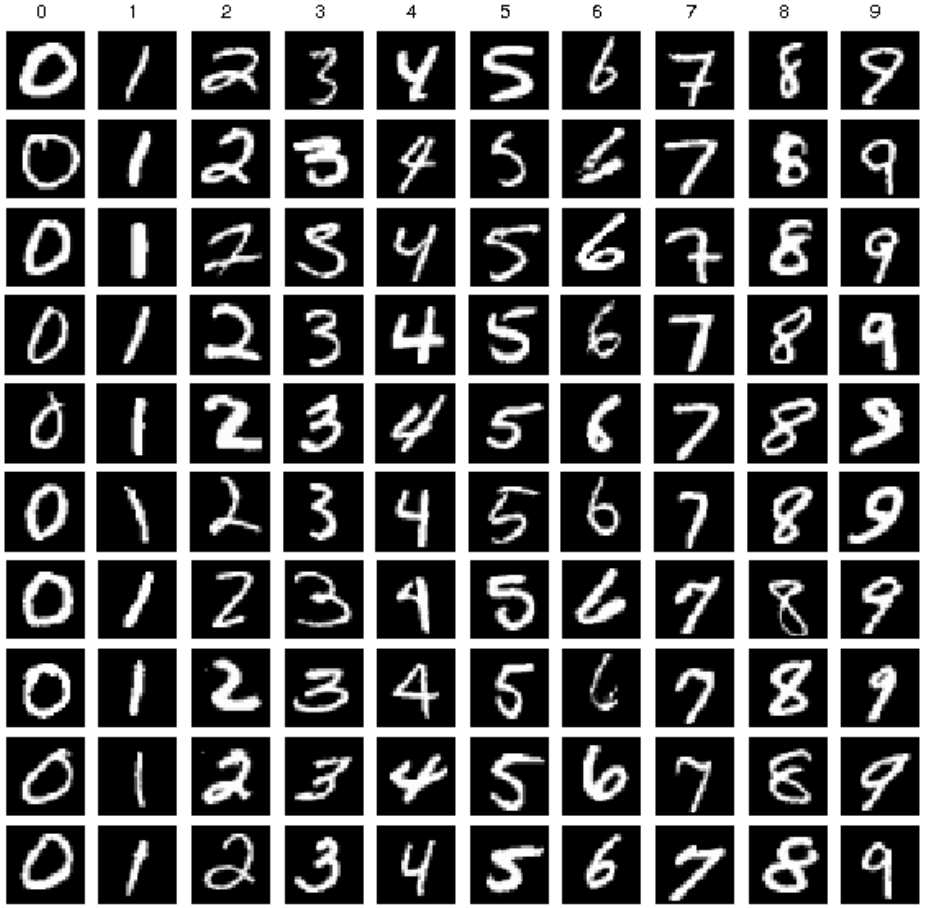
Image Classification

Image Classification

Which digit is presented in the image?

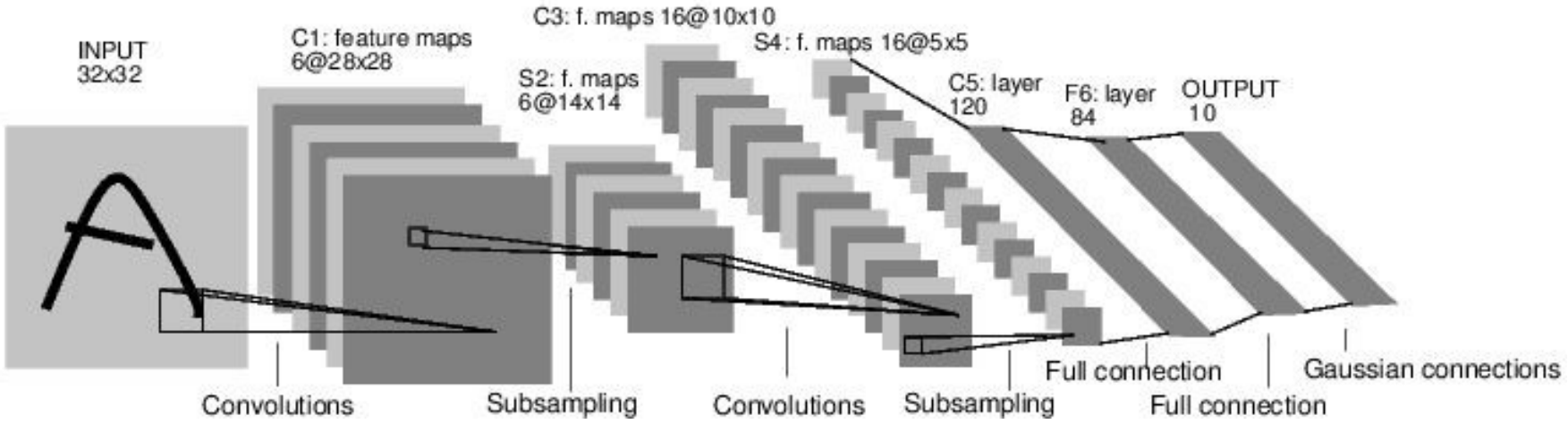


[LeCun, Cortes, Burges]

ConvNets for Image Classification

Case Study: LeNet-5

[LeCun et al., 1998]



Conv filters were 5x5, applied at stride 1
Subsampling (Pooling) layers were 2x2 applied at stride 2
i.e. architecture is [CONV-POOL-CONV-POOL-CONV-FC]

Image Classification

What object is presented in the image?

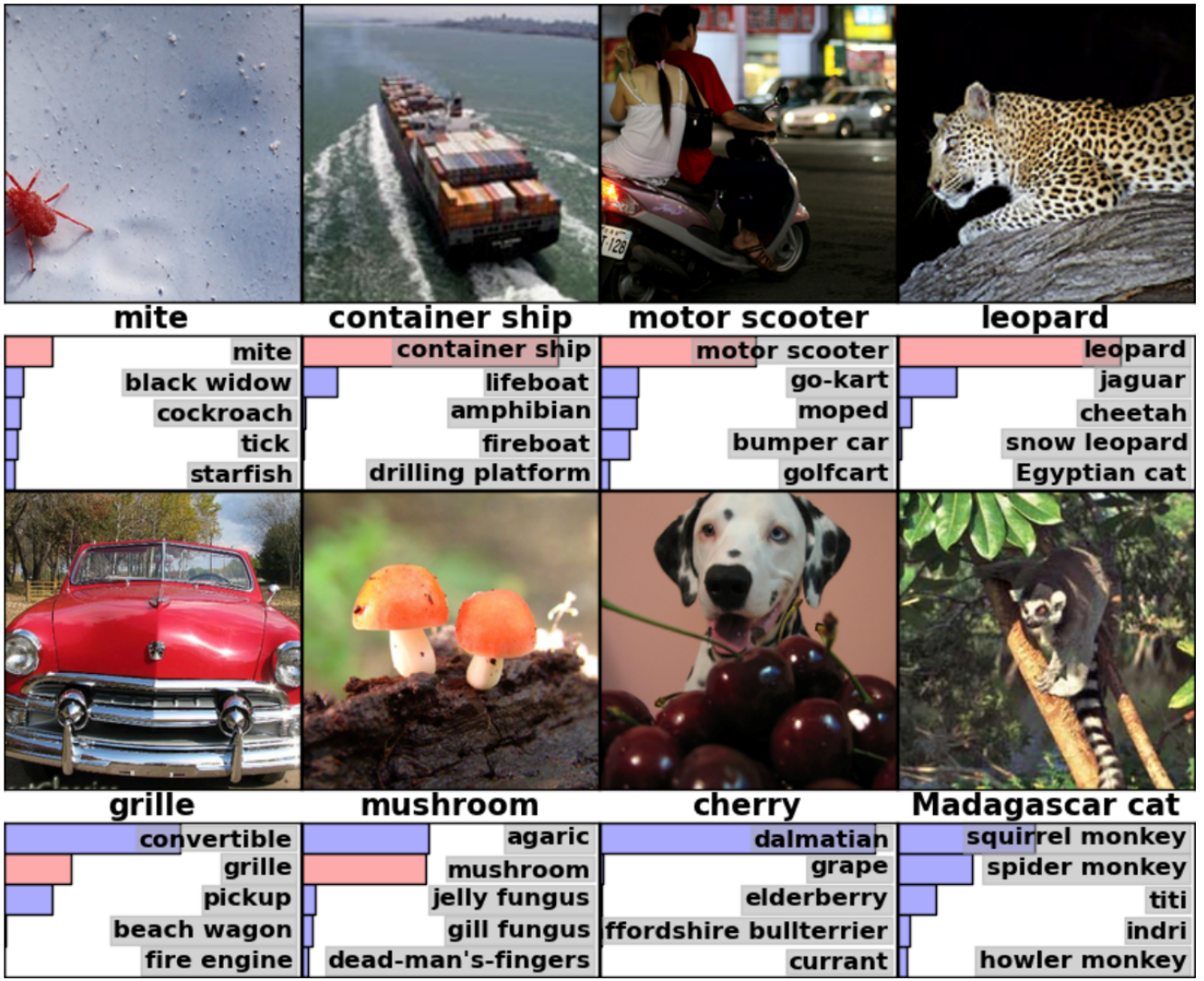
ImageNet: A Large-Scale Hierarchical Image Database,
Deng, Dong, Socher, Li, Li and Fei-Fei, *CVPR,* 2009

ImageNet Classification

What object is presented in the image?

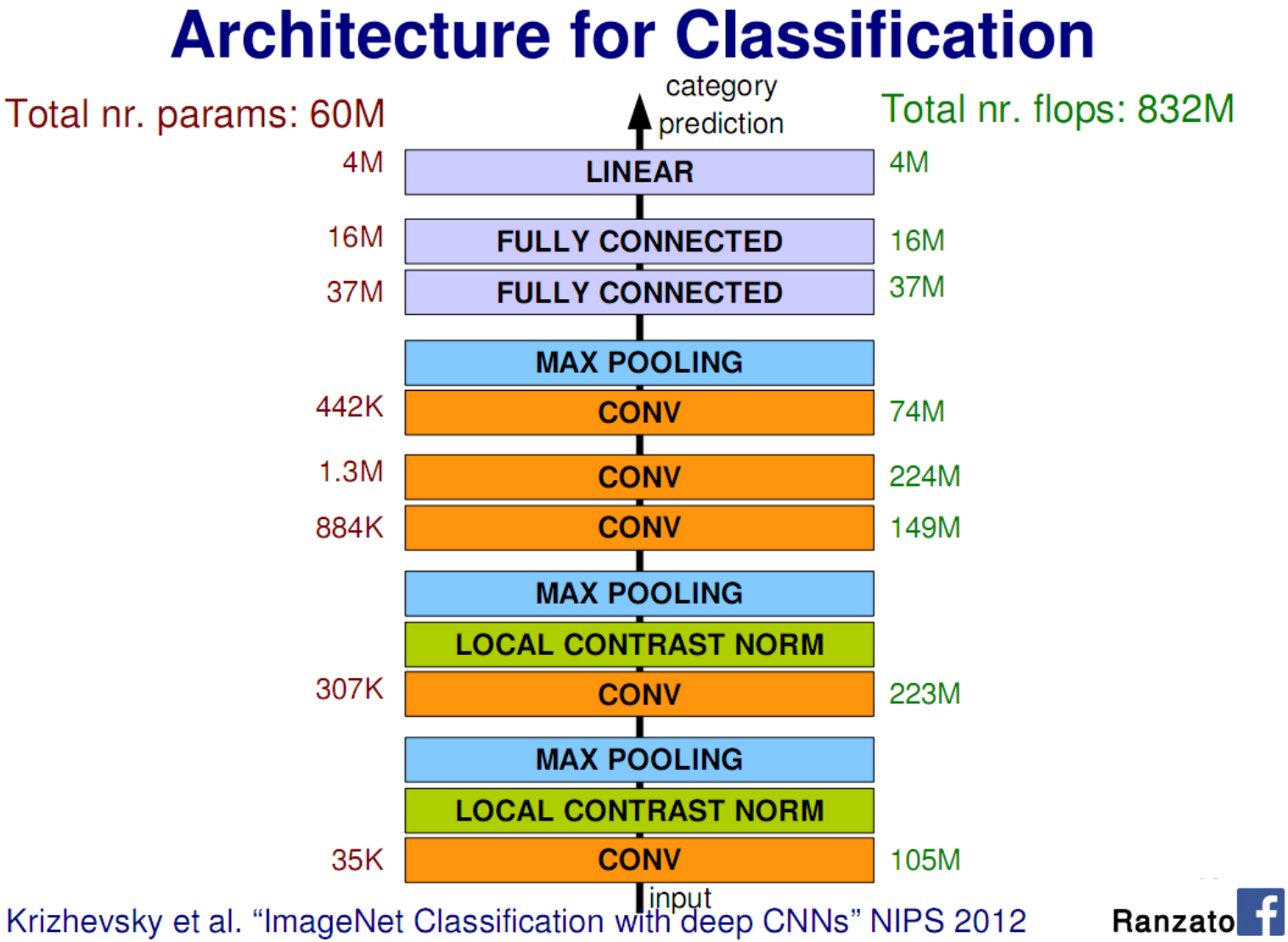
- Assume a single object of interest
- Target object usually lies around the center
- Somewhat random set of objects (e.g., dogs)
- Still interesting and very challenging!

ConvNets for Image Classification



"ImageNet Classification with Deep Convolutional Neural Networks",
Krizhevsky, Sutskever, Hinton, *NIPS,* 2012

ConvNets for Image Classification



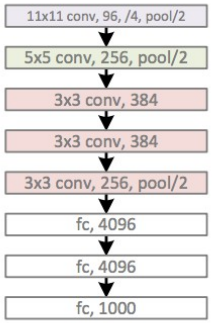
Krizhevsky et al. "ImageNet Classification with deep CNNs" NIPS 2012

Ranzato

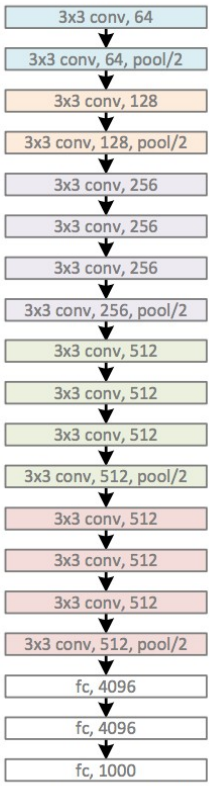
ConvNets for Image Classification

Revolution of Depth

AlexNet, 8 layers
(ILSVRC 2012)



VGG, 19 layers
(ILSVRC 2014)



GoogleNet, 22 layers
(ILSVRC 2014)

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Slide Credit: Kaiming He

ConvNets for Image Classification

Revolution of Depth

AlexNet, 8 layers
(ILSVRC 2012)

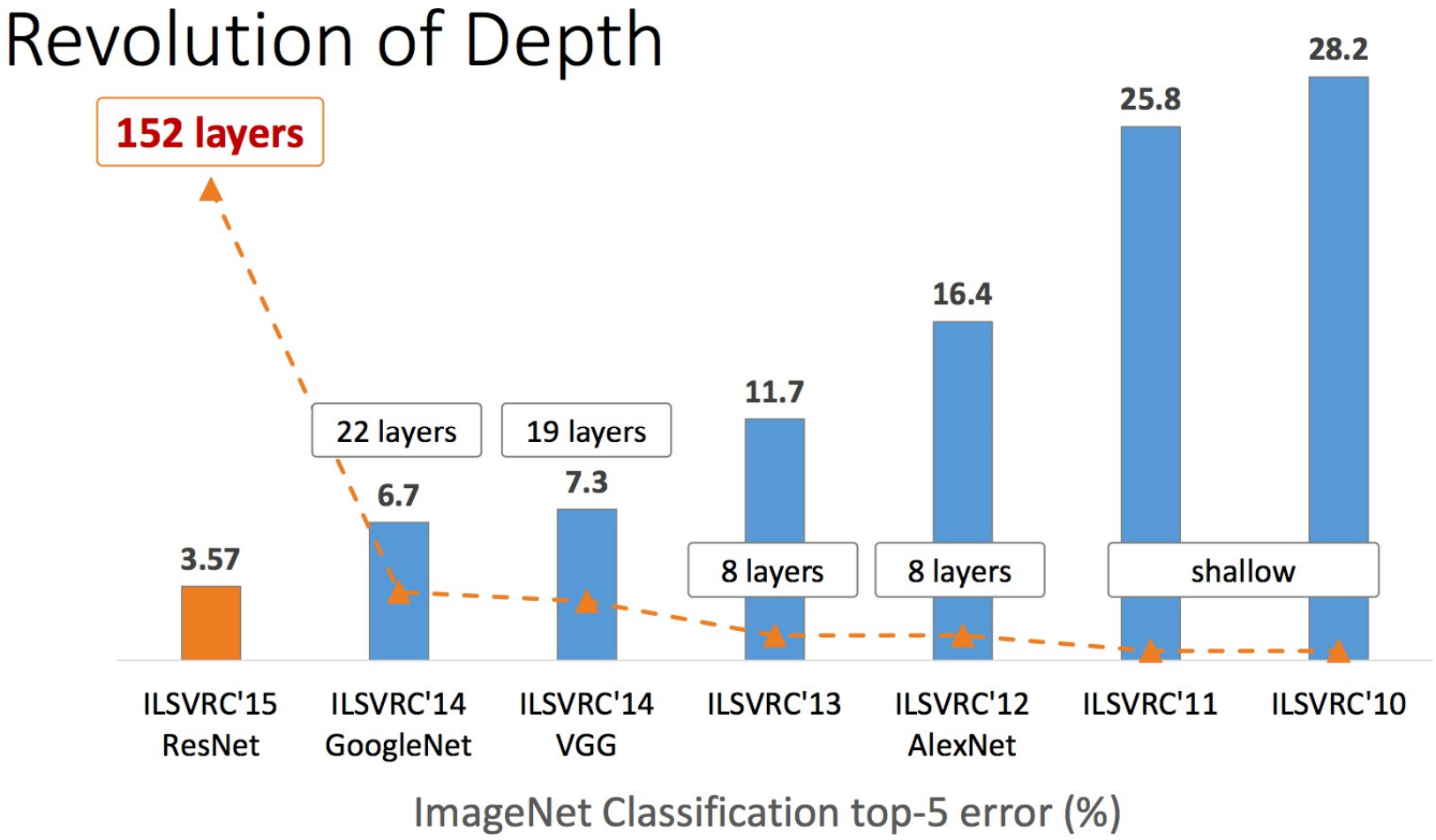
VGG, 19 layers
(ILSVRC 2014)

ResNet, 152 layers
(ILSVRC 2015)

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Slide Credit: Kaiming He

ConvNets for Image Classification

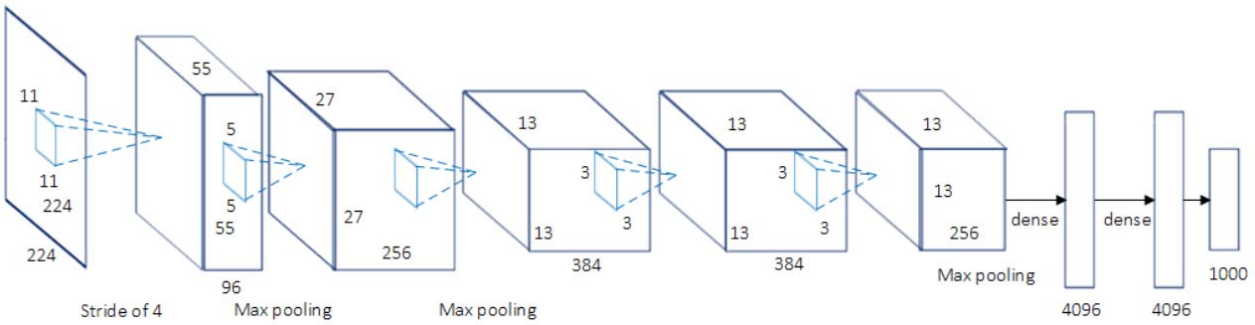


Slide Credit: Kaiming He

What do we know about these networks?

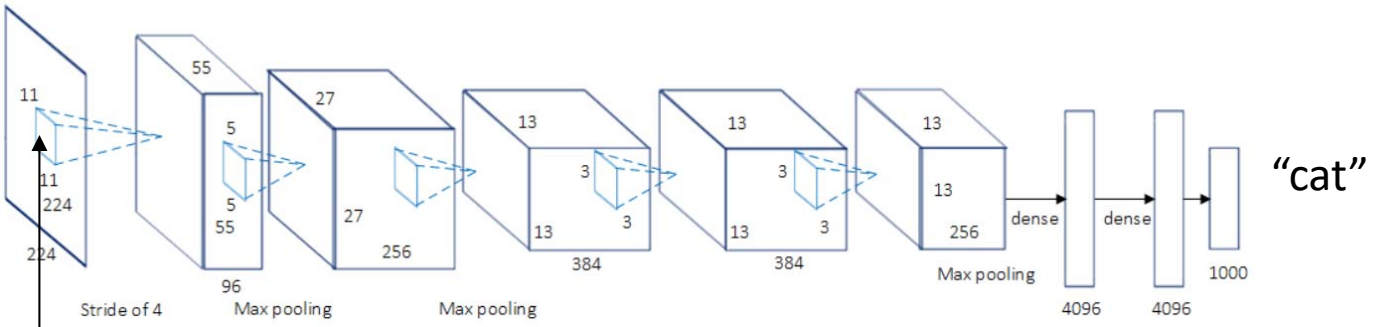
Many slides from Lana Lazebnik

What has been learned by ConvNets?

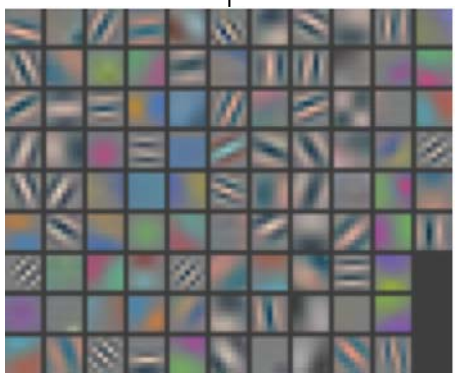


"cat"

What has been learned by ConvNets?



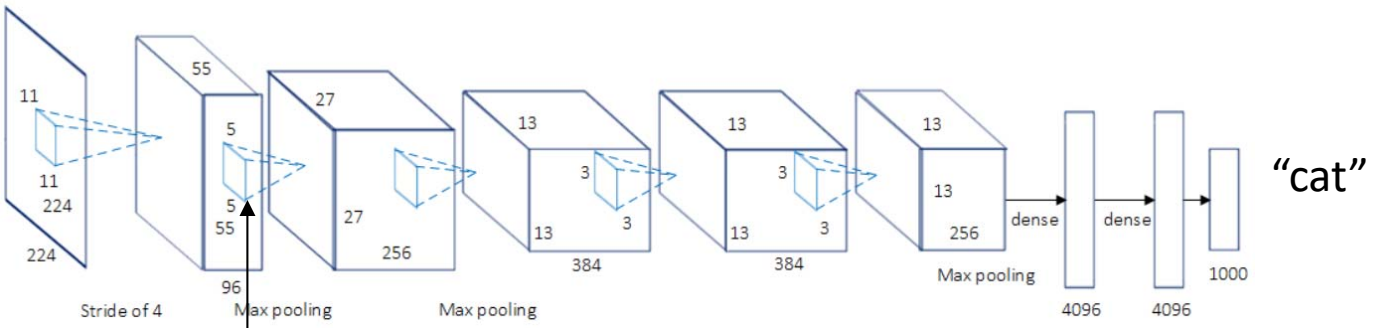
Visualize first-layer weights directly



Visualizing ConvNets

- “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.”

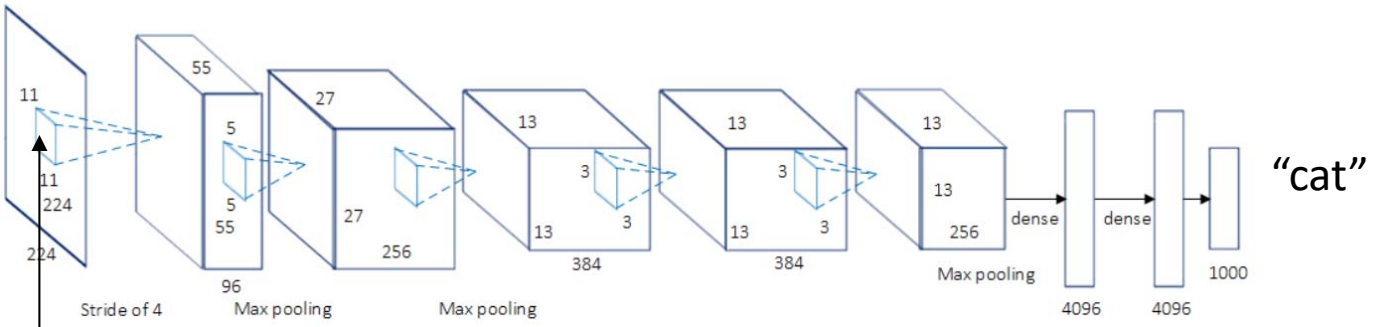
What has been learned by ConvNets?



Not too helpful for subsequent layers

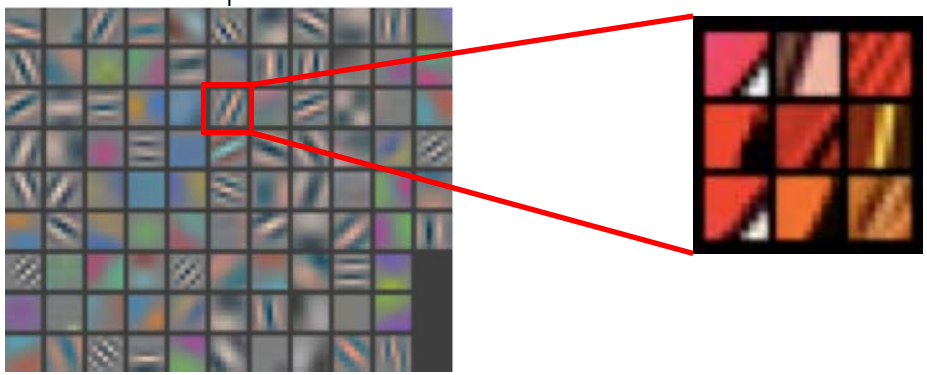
Features from a CIFAR10 network, via Stanford CS231n

What has been learned by ConvNets?

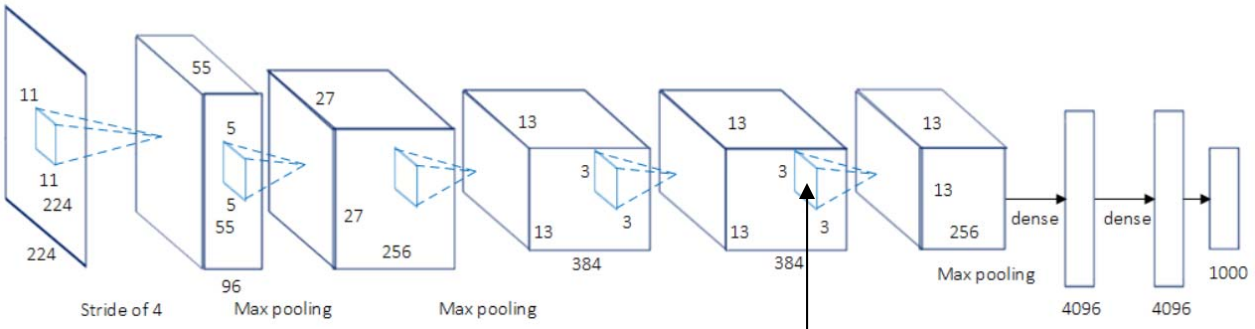


Visualize maximally activating patches:

pick a unit; run many images through the network;
visualize patches that produce the highest output values

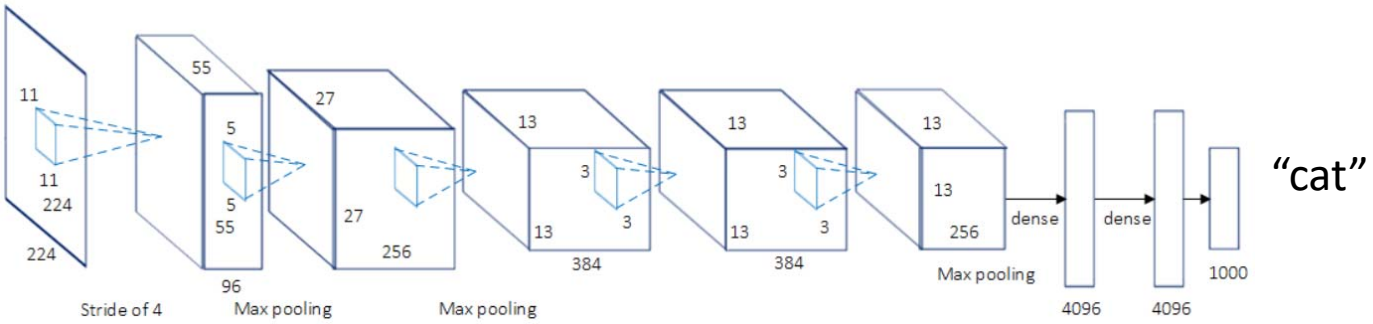


What has been learned by ConvNets?



Visualize maximally activating patches

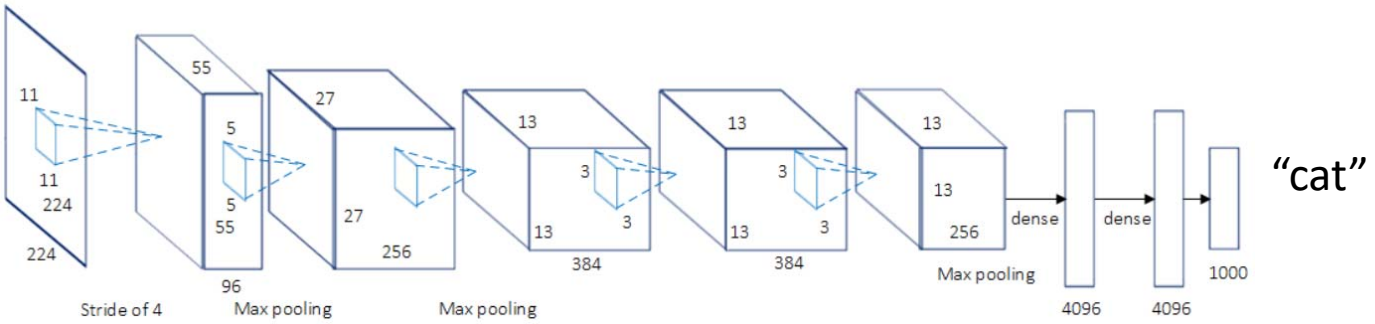
What has been learned by ConvNets?

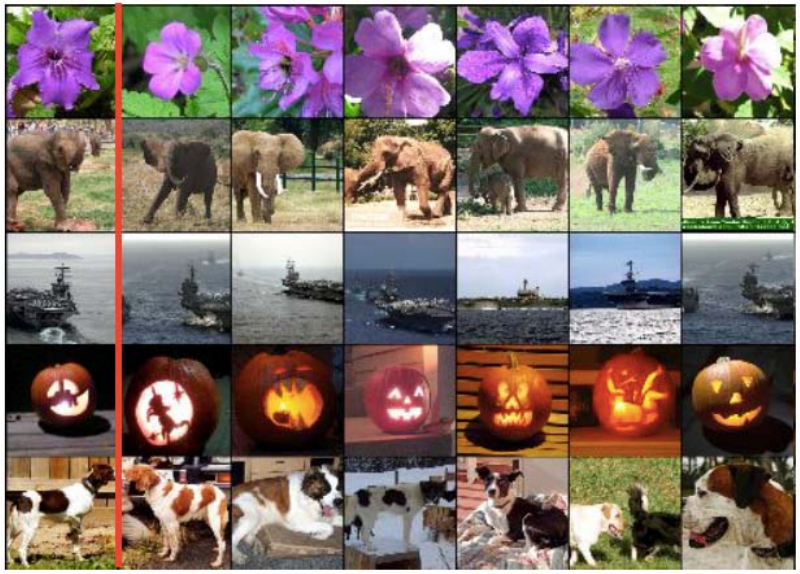


[A tour through the network!]

Visualization: Thanks to David Fouhey

What has been learned by ConvNets?



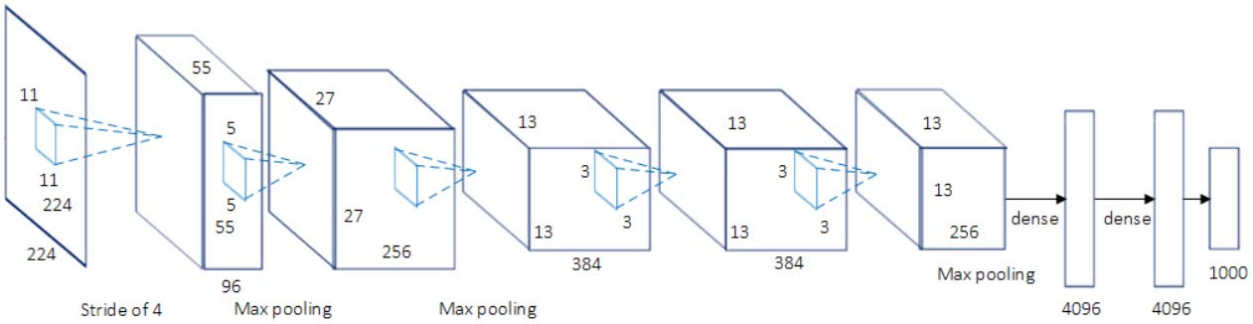


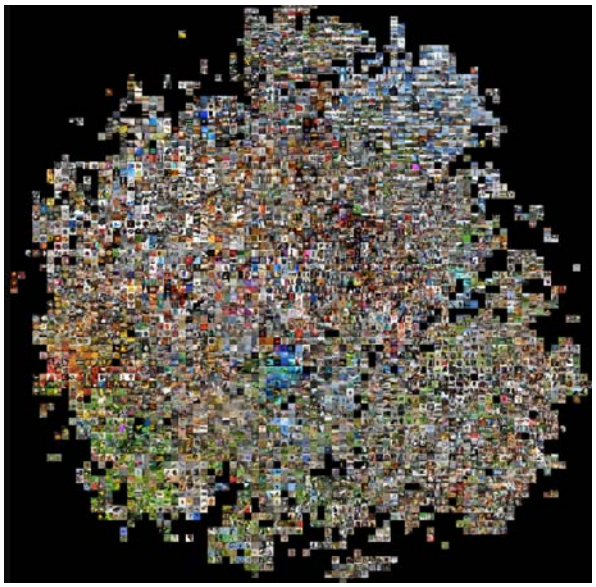
FC layers?

Visualize nearest neighbors according to activation vectors

Source: Stanford CS231n

What has been learned by ConvNets?



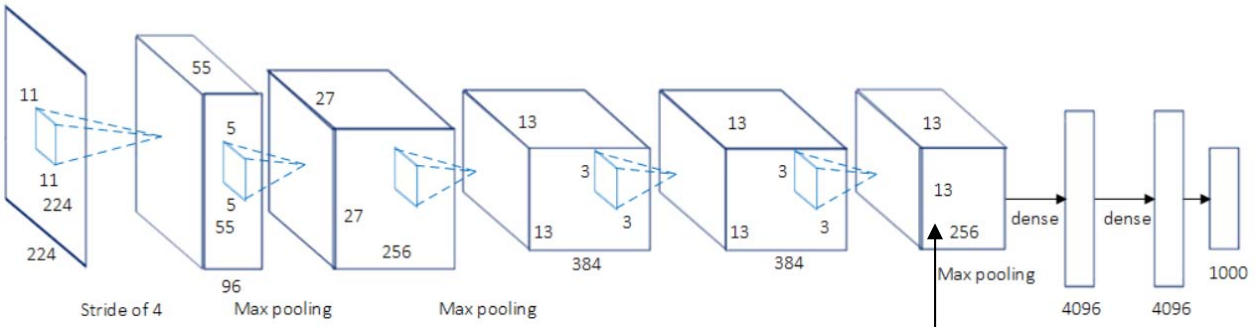


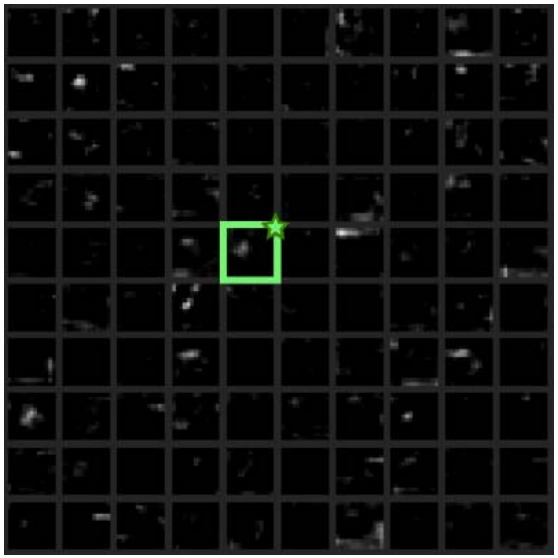
FC layers?

Fancy dimensionality reduction, e.g., t-SNE

Source: Andrej Karpathy

What has been learned by ConvNets?

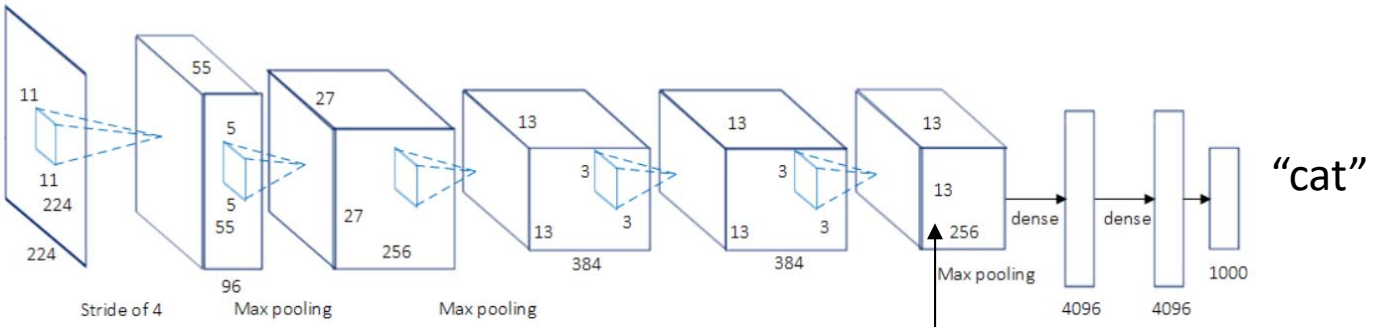




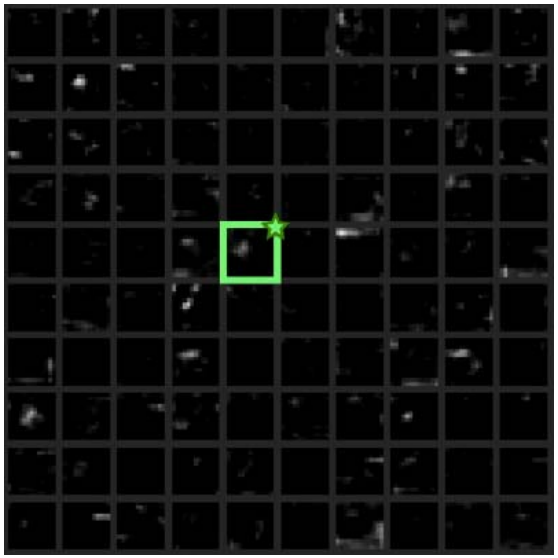
Visualize activations for an image

Source

What has been learned by ConvNets?



Going back to image?

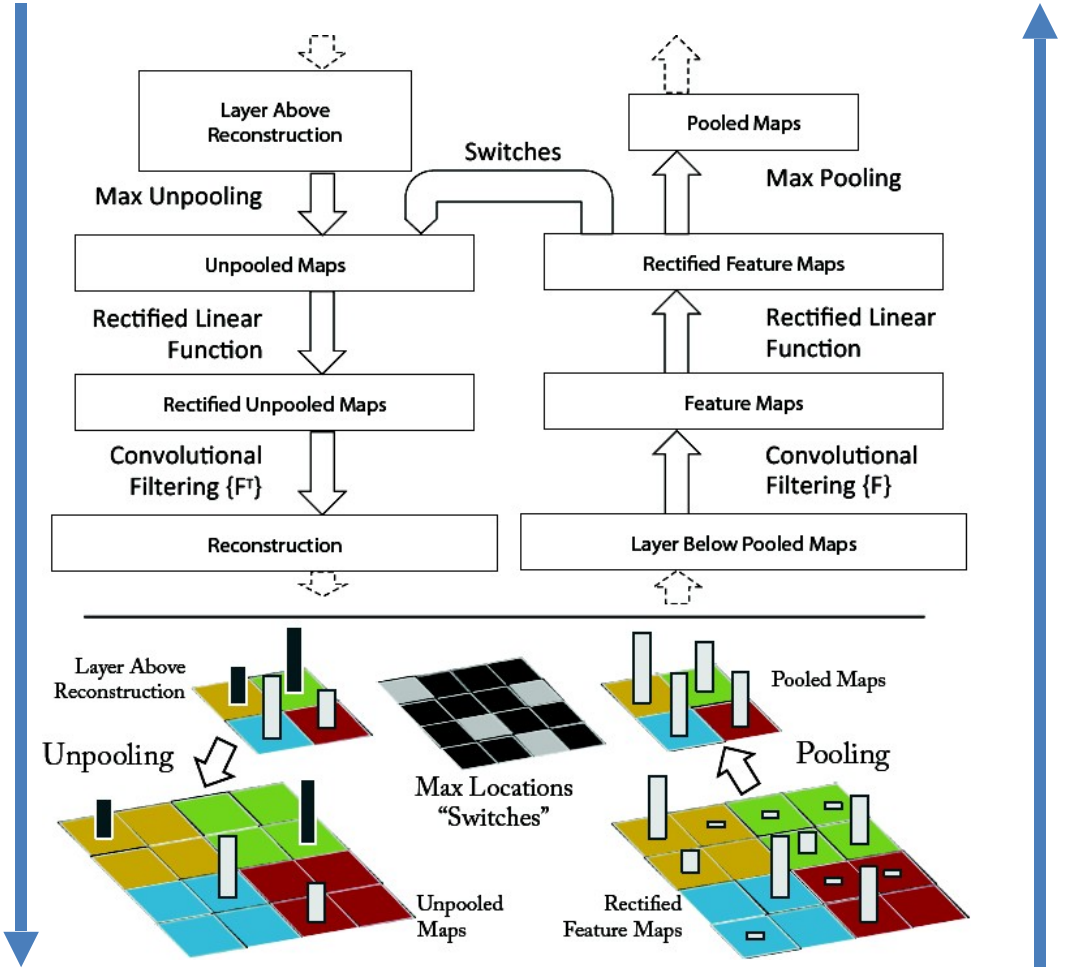


Source

Visualize activations for an image

Visualizing ConvNets: Going back to images

Towards Image



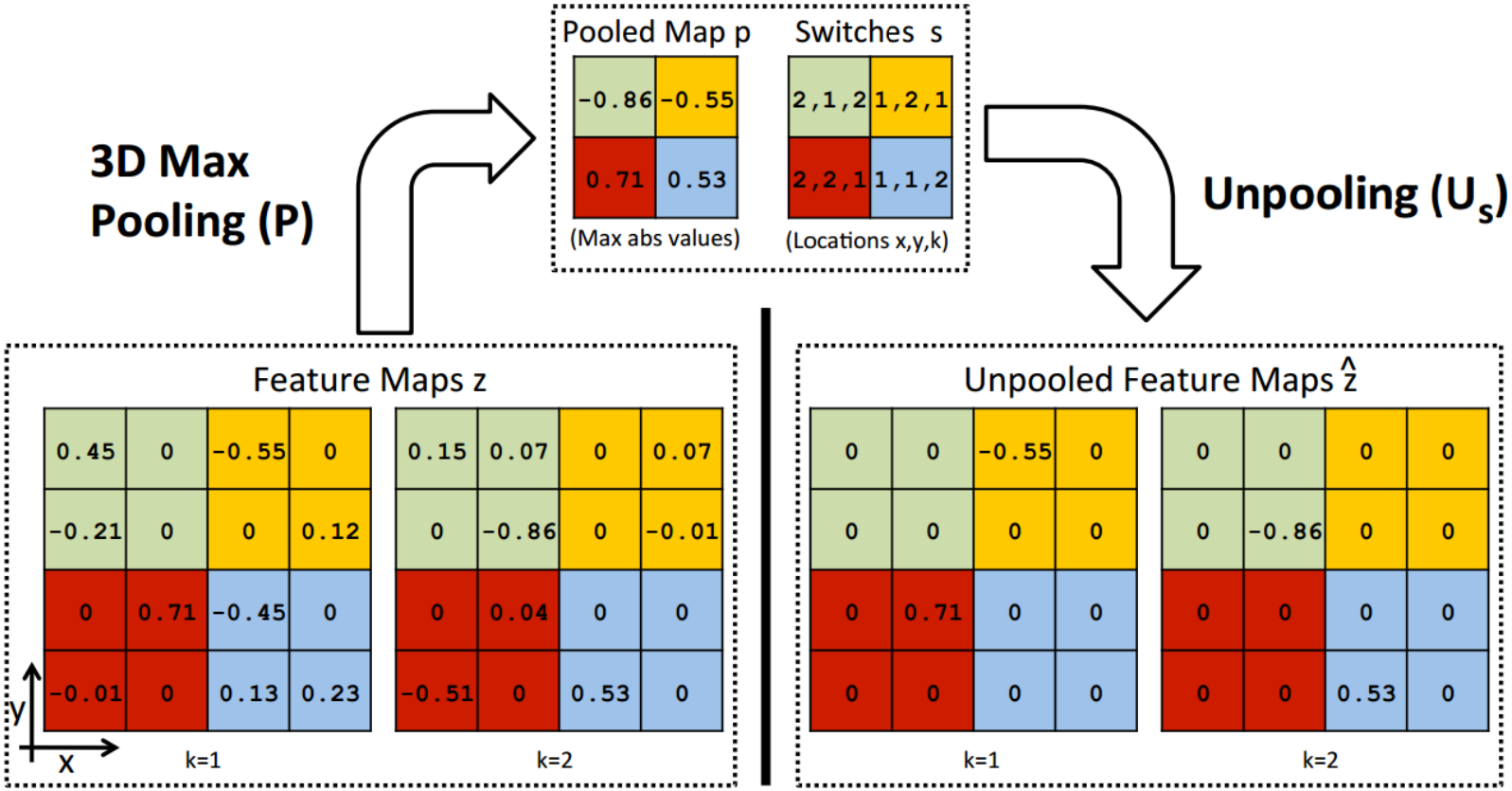
Towards Predictions

[Zeiler and Fergus]

Visualizing ConvNets: Things to invert

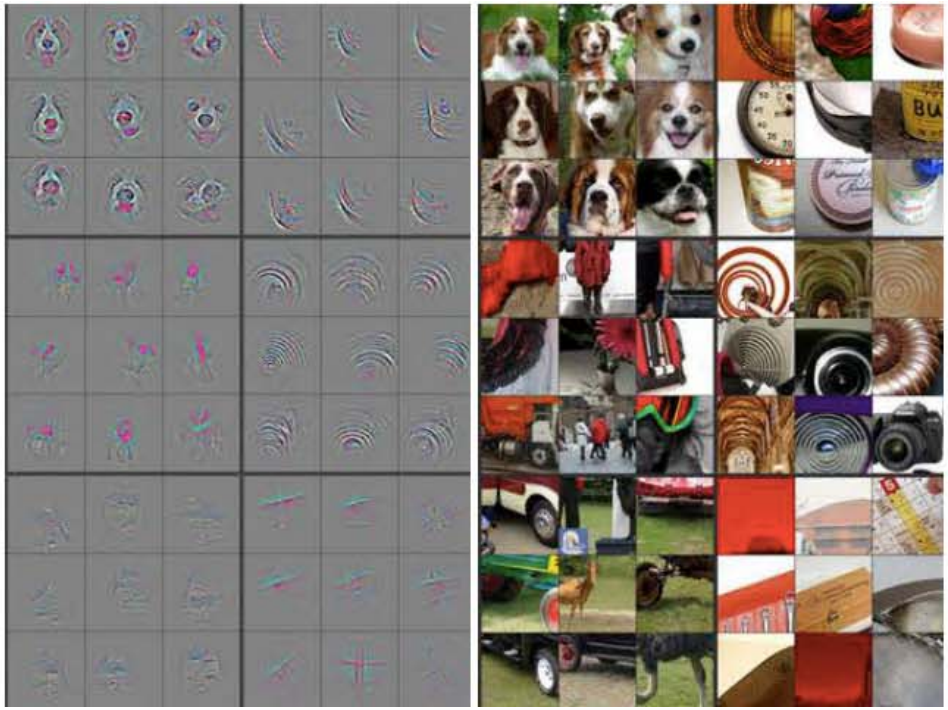
- Convolutions/Filtering
- Rectification/Non-linearity
- Pooling

Visualizing ConvNets: Things to invert

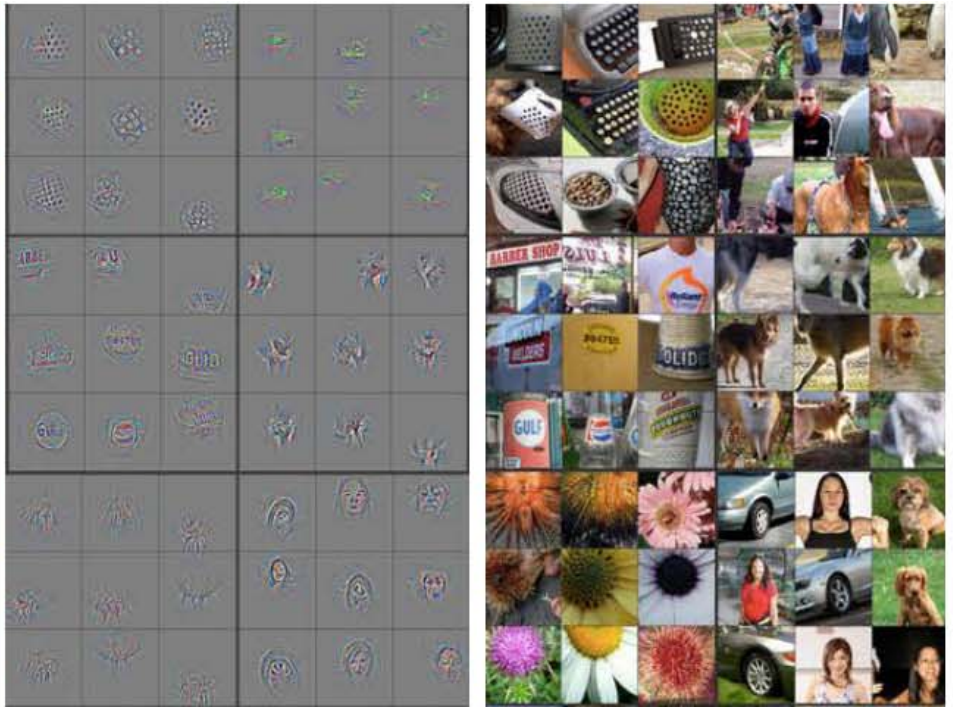


Visualizing ConvNets: Going back to images

AlexNet Layer 4



AlexNet Layer 5



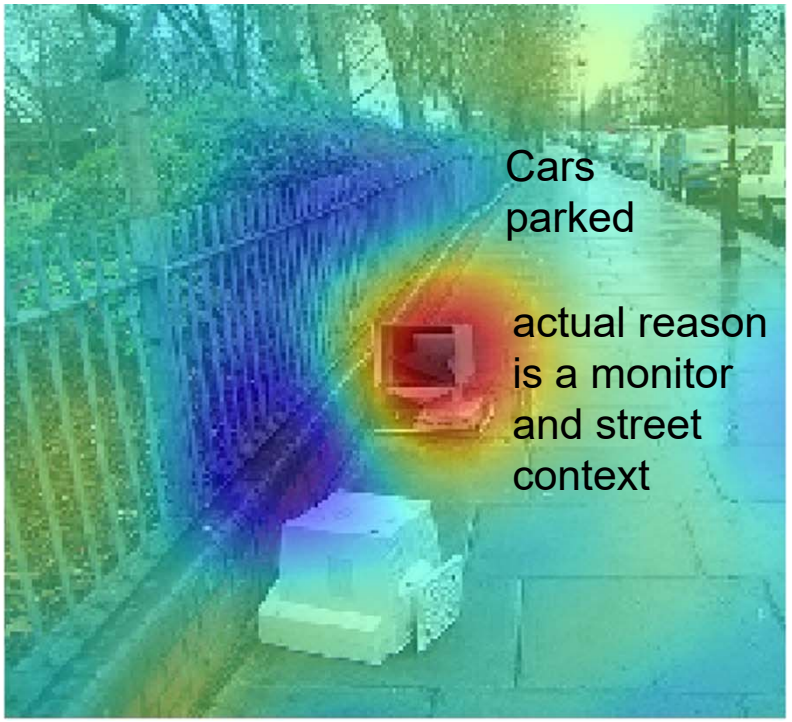
M. Zeiler and R. Fergus,
Visualizing and Understanding Convolutional Networks, ECCV 2014

Visualizing ConvNets: Saliency Maps

Which parts of the image played the most important role in the network's decision?

Prediction: "car" 64%

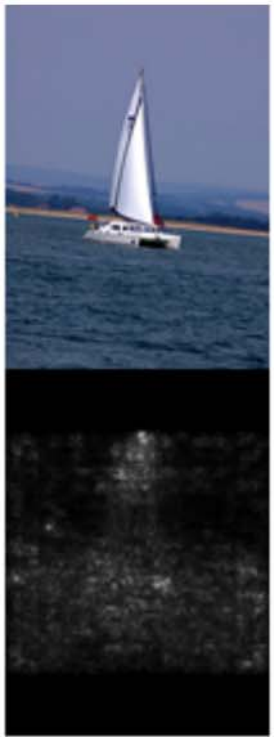
Explanation for "car"

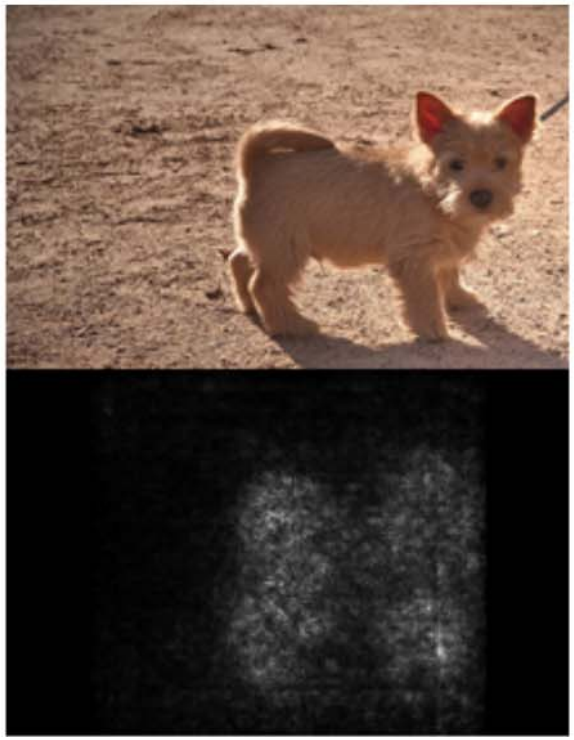


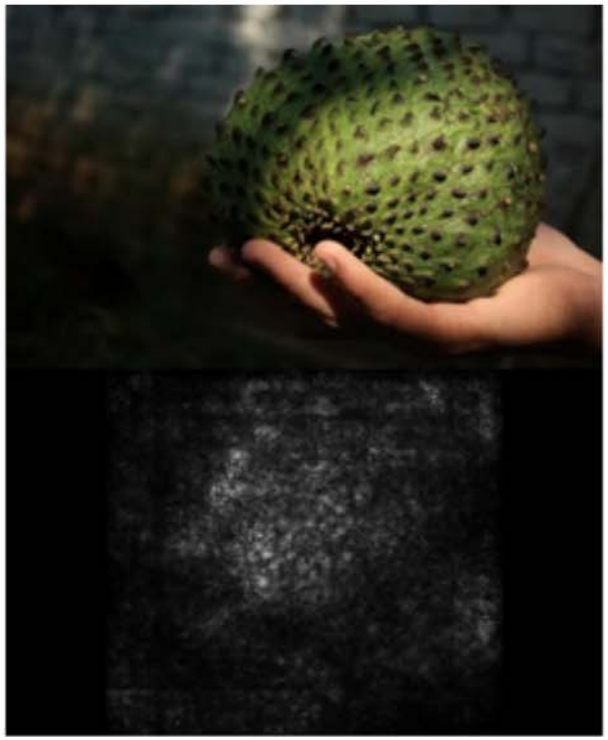
Source: K. Saenko

“White Box” Saliency: Using Gradients

Compute $-\frac{\partial L(x;\theta)}{x}$ and display the max of absolution values across three channels



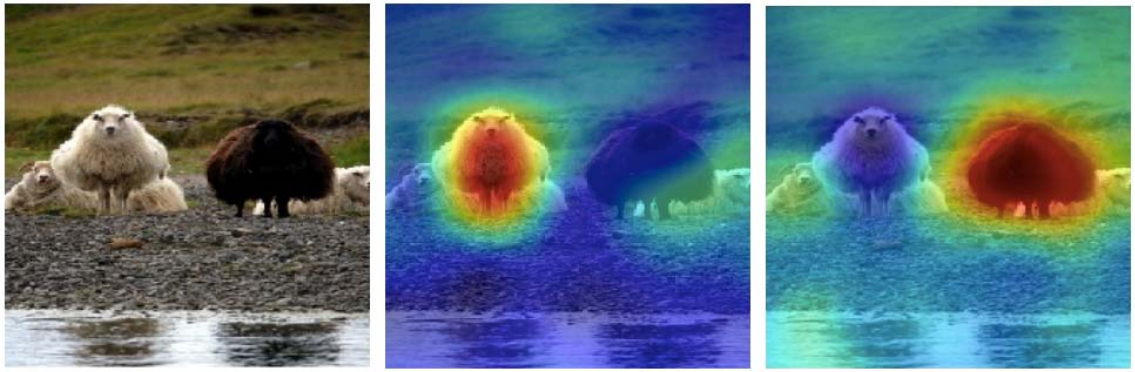




K. Simonyan, A. Vedaldi, and A. Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ICLR 2014

“Black Box” Saliency: Via Masking

Pixel saliency is class dependent



(a) Sheep - 26%, Cow - 17% (b) Importance map of ‘*sheep*’ (c) Importance map of ‘*cow*’

(d) Bird - 100%, Person - 39% (e) Importance map of ‘*bird*’ (f) Importance map of ‘*person*’

V. Petsiuk, A. Das, K. Saenko, RISE: Randomized Input Sampling for Explanation of Black-box Models, BMVC 2018

Quantify Interpretability of Units

- From the very beginning, people have observed that many units in higher layers seem to fire on semantically meaningful concepts
- But how can we quantify this?

AlexNet Layer 4

AlexNet Layer 5

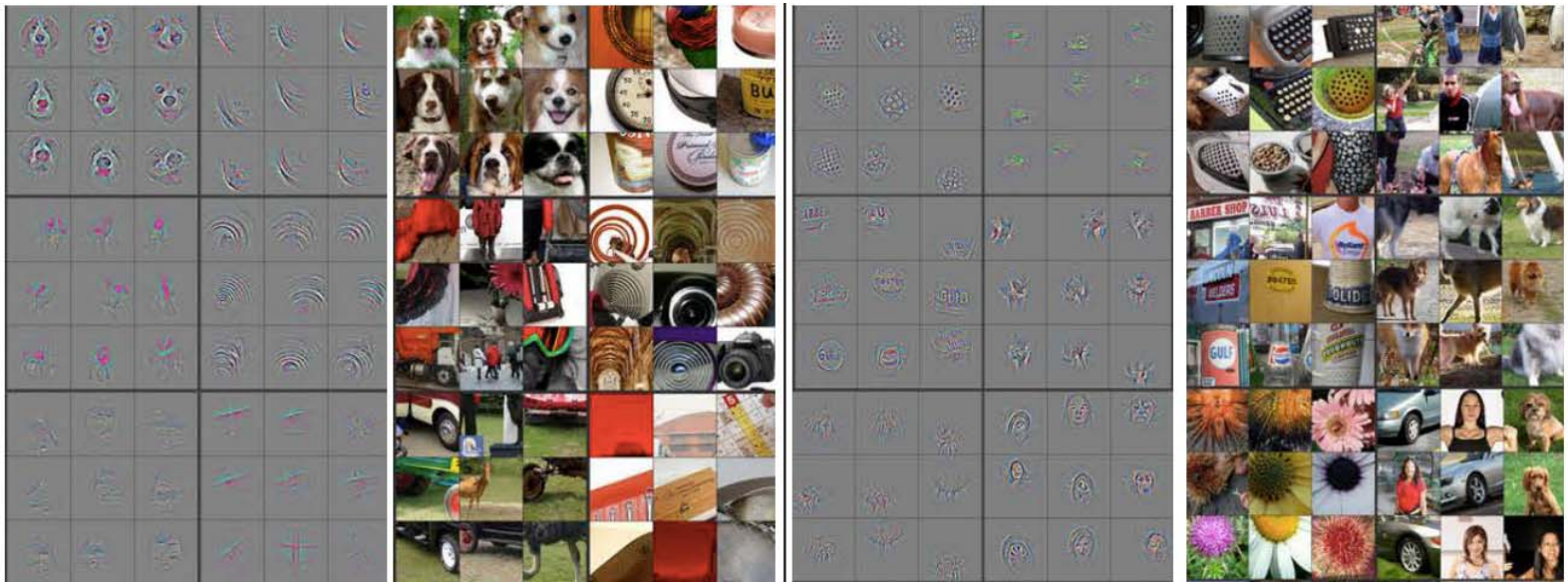
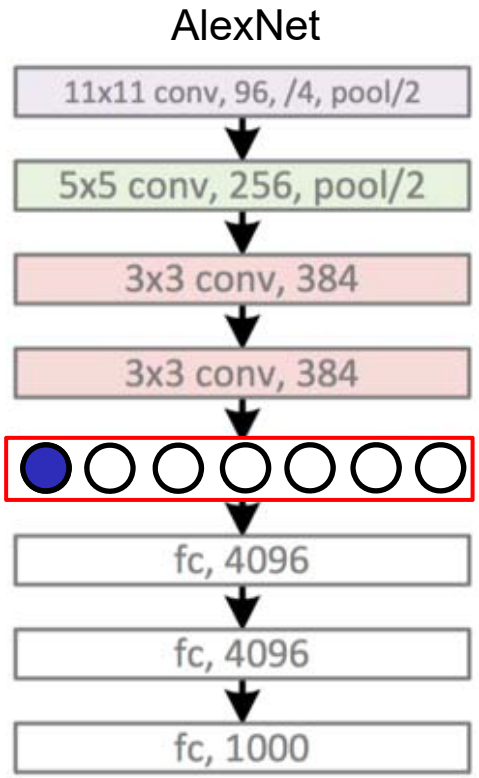
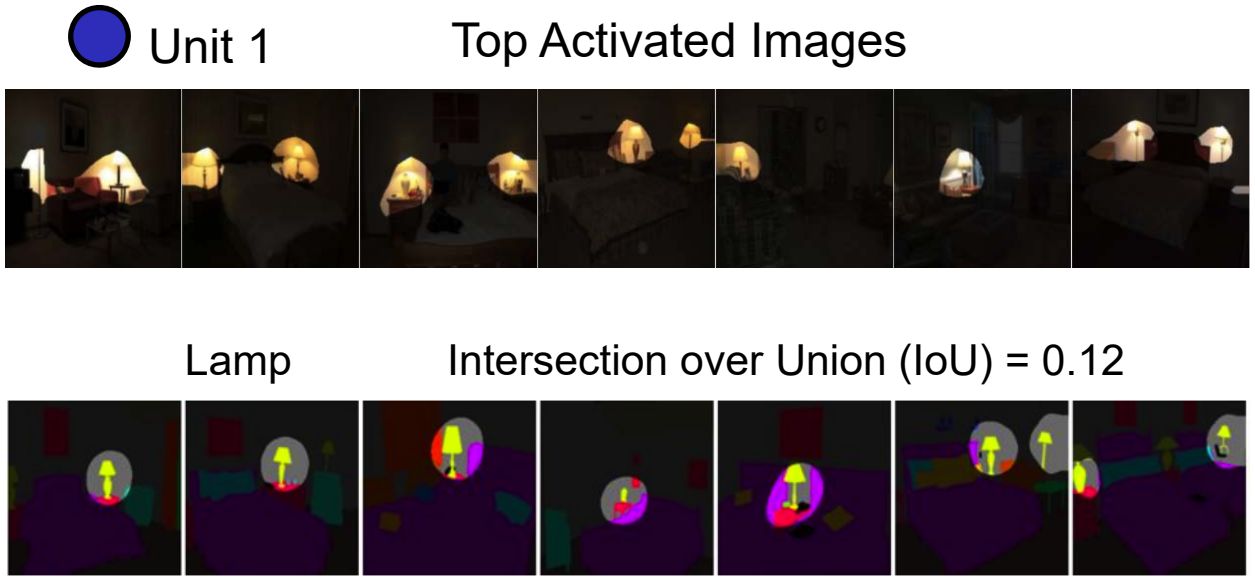


Figure: Zeiler & Fergus

Quantify Interpretability of Units

For a given unit, measure how much areas of high activation overlap semantic segmentations for a large set of visual concepts

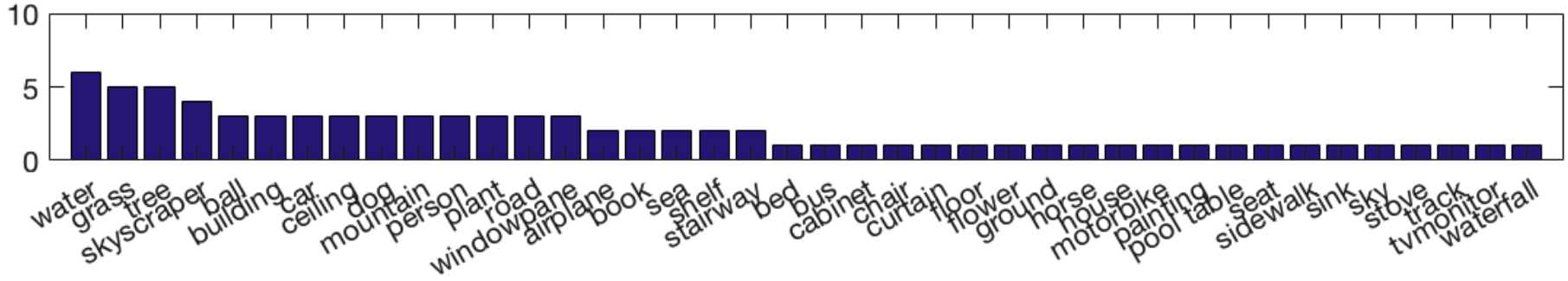




D. Bau, B. Zhou, A. Khosla, A. Oliva, A. Torralba, Network Dissection: Quantifying Interpretability of Deep Visual Representations, CVPR 2017

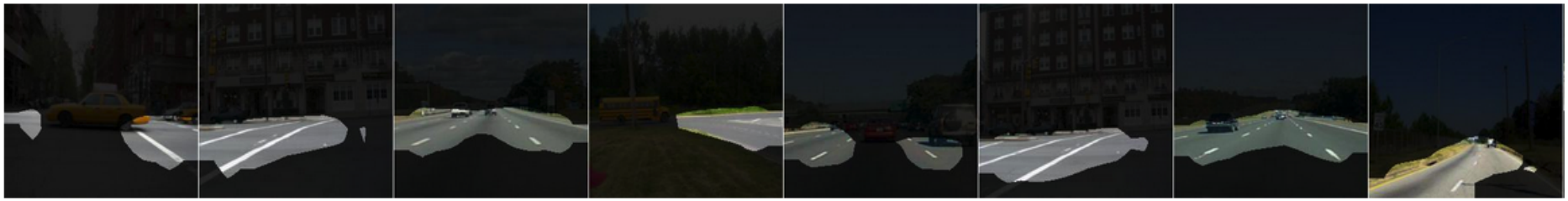
Quantify Interpretability of Units

Histogram of object detectors for Places AlexNet conv5 units
81/256 units with IoU > 0.04



conv5 unit 79 car (object) IoU=0.13

conv5 unit 107 road (object) IoU=0.15



D. Bau, B. Zhou, A. Khosla, A. Oliva, A. Torralba, Network Dissection: Quantifying Interpretability of Deep Visual Representations, CVPR 2017

ConvNet Visualizations: Recap

- Basic visualization techniques
  - showing weights
  - top activated patches
  - nearest neighbors
- Mapping activations back to the image
  - Deconvolution
- Saliency maps
  - “White box” vs. “black box”
- Explainability/interpretability
  - Explaining network decisions
  - Quantifying interpretability of intermediate units