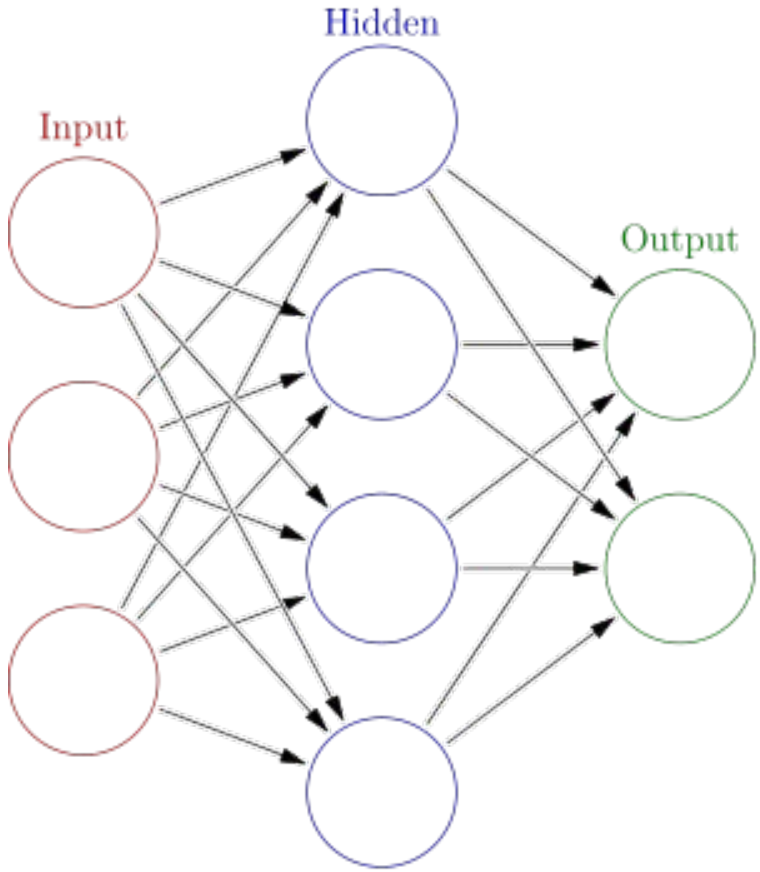
Convolutional Neural Networks

November 17, 2015

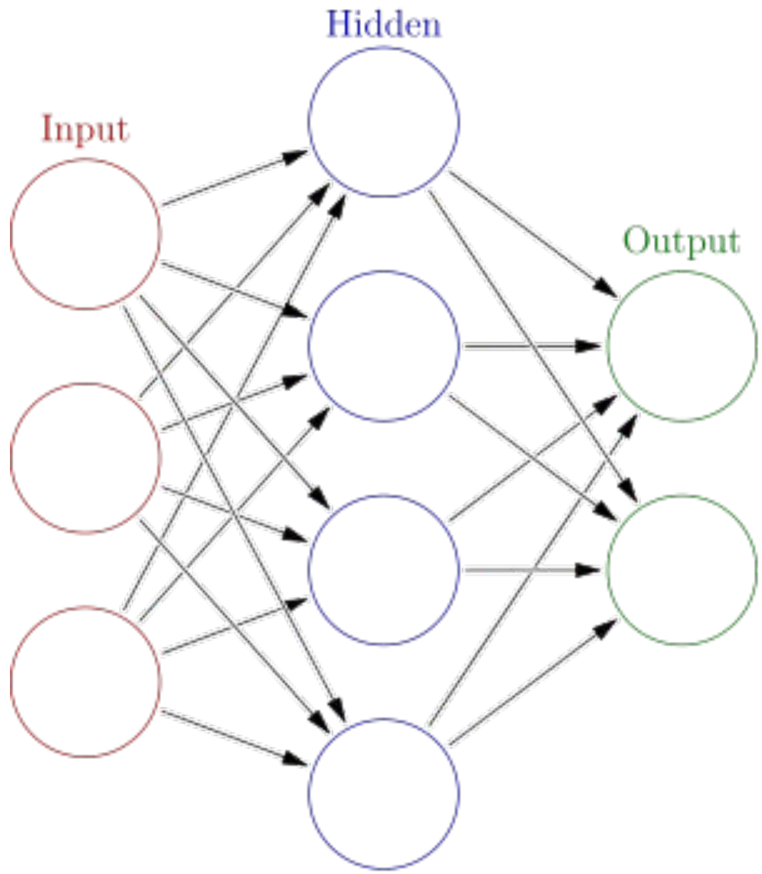
Artificial Neural Networks

- Feedforward neural networks



Artificial Neural Networks

- Feedforward, *fully-connected* neural networks



Artificial Neural Networks

- Feedforward, *fully-connected* neural networks
  - Large modeling capacity

Artificial Neural Networks

- Feedforward, *fully-connected* neural networks
  - Large modeling capacity
  - Require large amounts of data

Artificial Neural Networks

- Feedforward, *fully-connected* neural networks
  - Large modeling capacity
  - Require large amounts of data
  - Work fairly well for handwritten digits

Natural images? …not so much.

Natural Images

Natural Images

- Much more detail

Natural Images

- Much more detail
  - Intricate spatial relationships

Natural Images

- Much more detail
  - Intricate spatial relationships
- More variety within a class of examples

Natural Images

- Much more detail
  - Intricate spatial relationships
- More variety within a class of examples

Natural Images

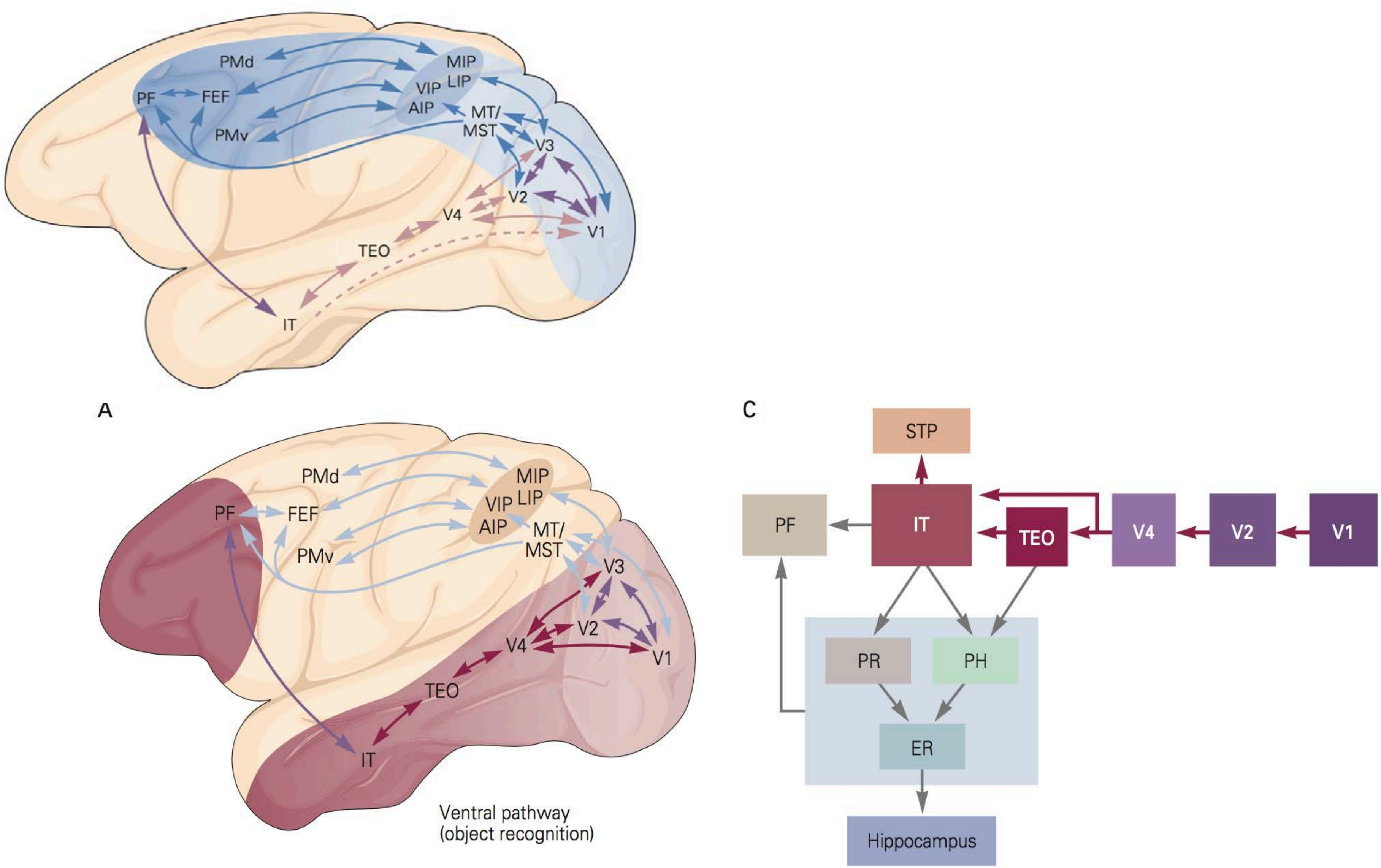
- Much more detail
  - Intricate spatial relationships
- More variety within a class of examples
  - Natural variations
  - Color
  - Viewing angle
  - Lighting
  - Size
  - Position

Can we build a better network?

Take inspiration from neuroscience

Biological Vision

Biological Vision



Biological Vision

- Hubel & Wiesel (1950s)

Biological Vision

- Hubel & Wiesel (1950s)

RECEPTIVE FIELDS OF SINGLE NEURONES IN THE CAT'S STRIATE CORTEX

By D. H. HUBEL* and T. N. WIESEL*

From the Wilmer Institute, The Johns Hopkins Hospital and University, Baltimore, Maryland, U.S.A.

(*Received* 22 *April* 1959)

Biological Vision

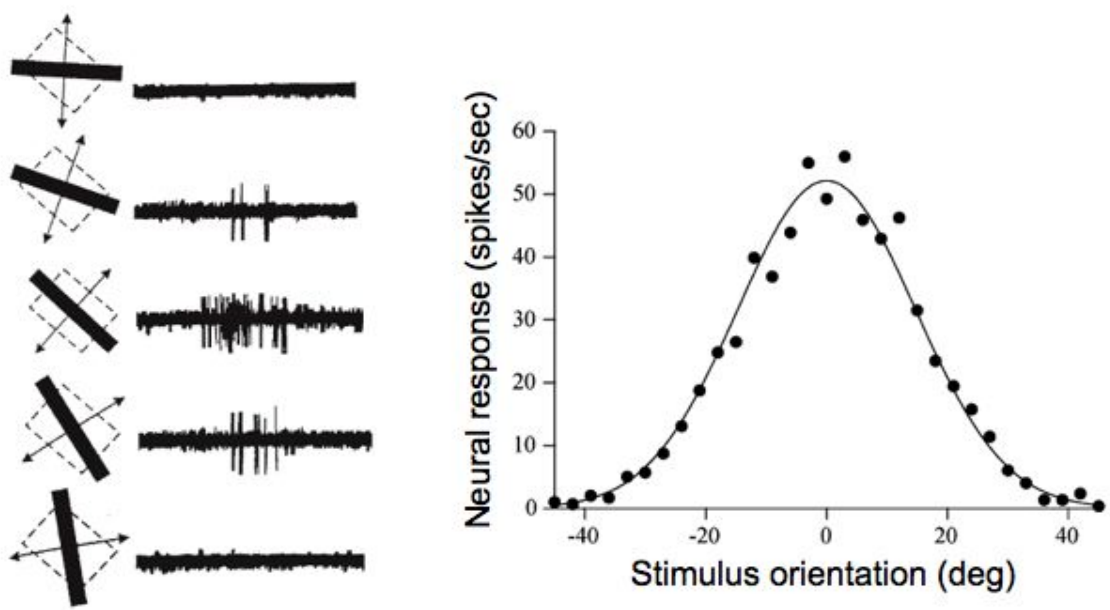
- Hubel & Wiesel (1950s)
  - Record from neurons in V1

Biological Vision

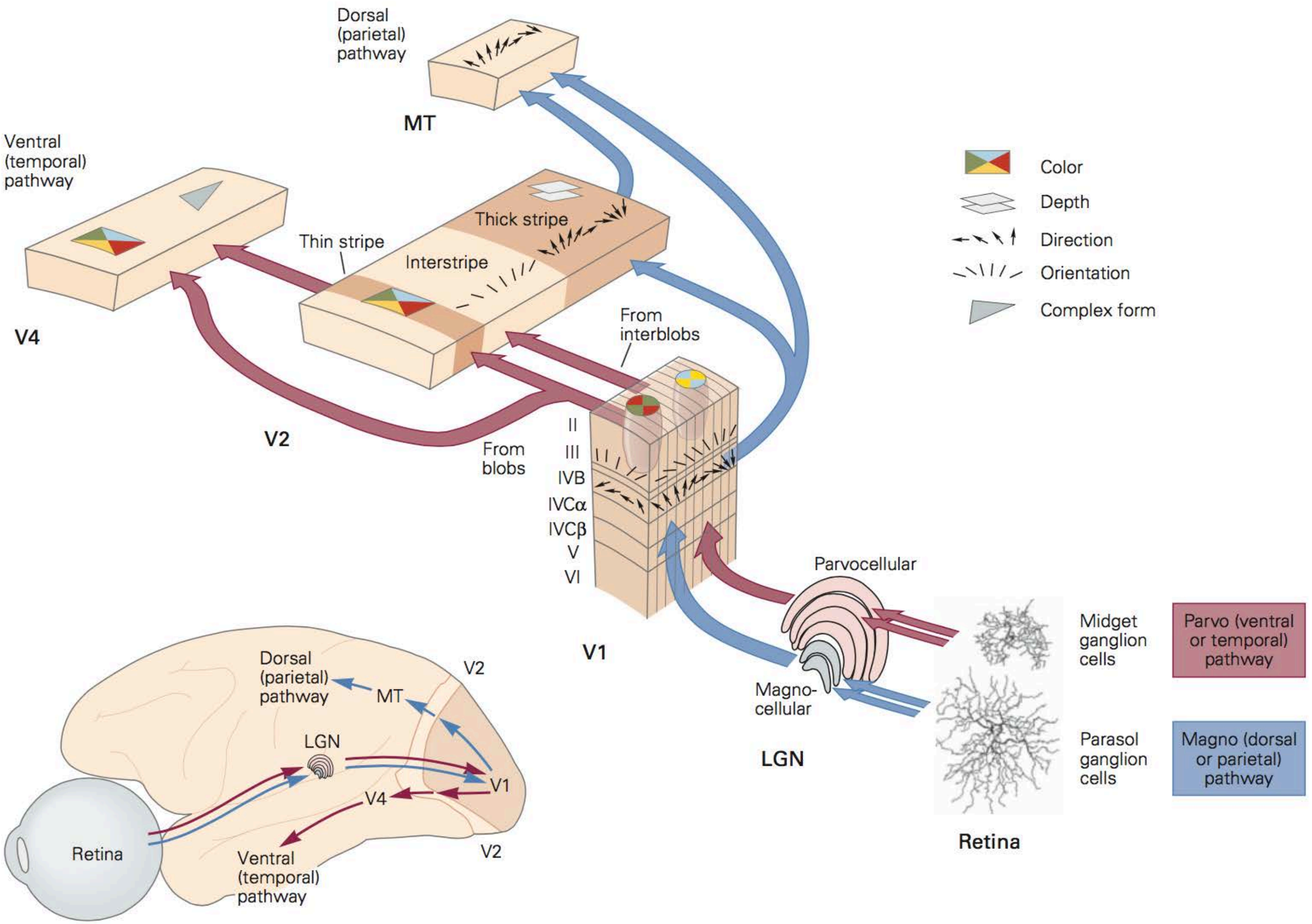
- Hubel & Wiesel (1950s)
  - Record from neurons in V1
  - Present moving gratings

Biological Vision

- Hubel & Wiesel (1950s)
  - Record from neurons in V1
  - Present moving gratings



Biological Vision

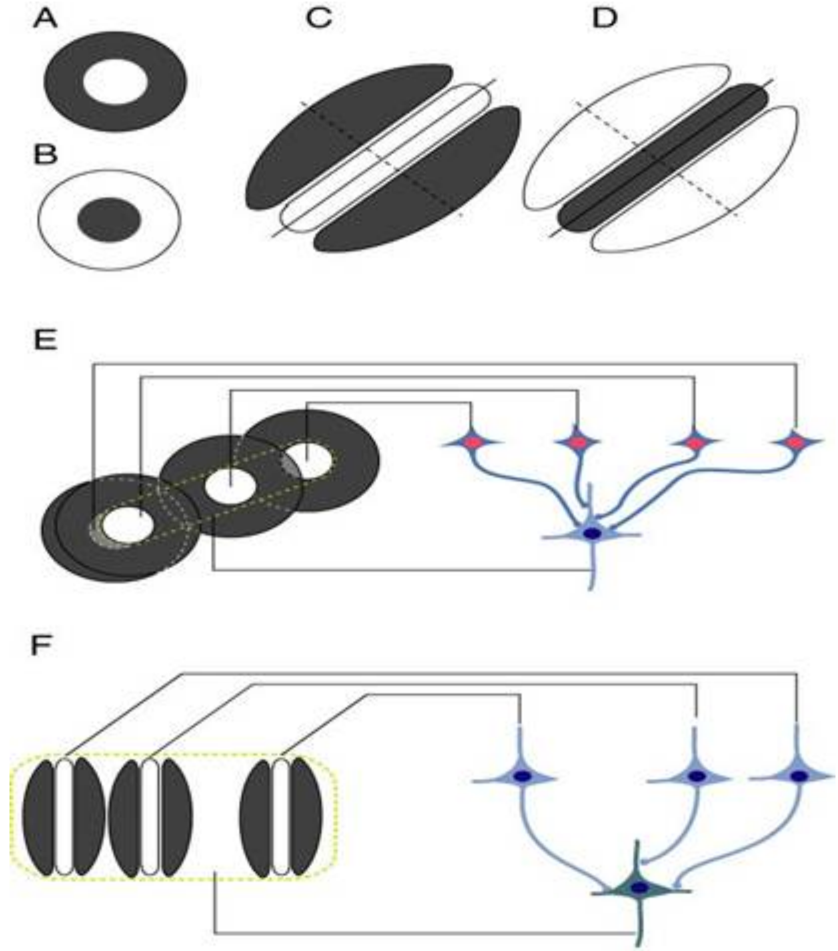


Biological Vision

- Simple and complex cells

Biological Vision

- Simple and complex cells



Biological Vision

- Higher visual areas

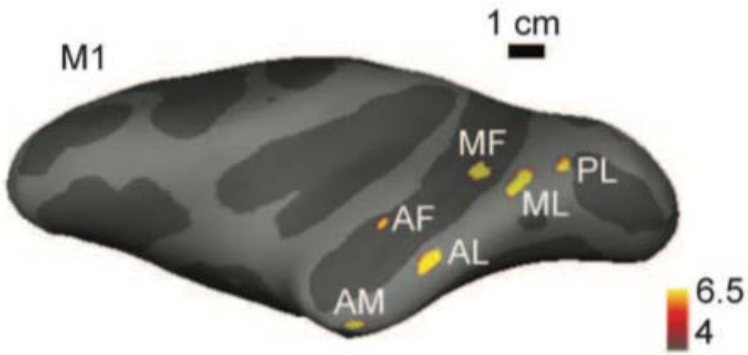
Biological Vision

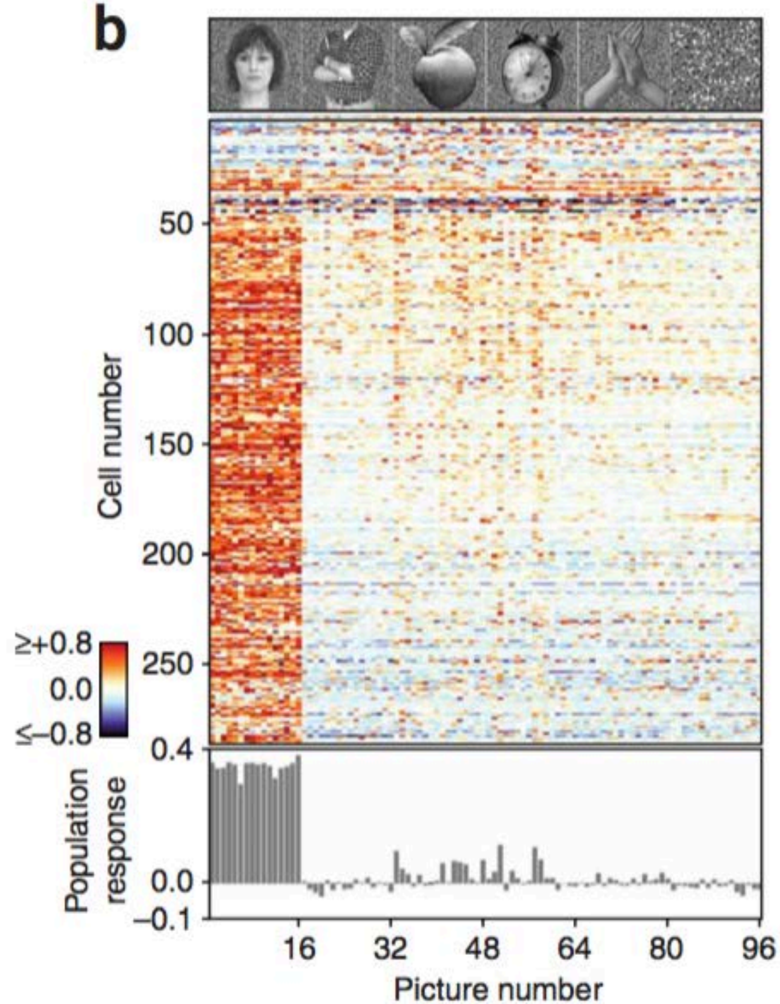
- Higher visual areas
  - Encode complex stimuli

Biological Vision

- Higher visual areas
  - Encode complex stimuli
    - Professor Doris Tsao, Caltech

Biological Vision





Friewald, 2009 & 2010

Biological Vision

Biological Vision

- Hierarchical representation

Biological Vision

- Hierarchical representation
- Map of visual space at lower levels

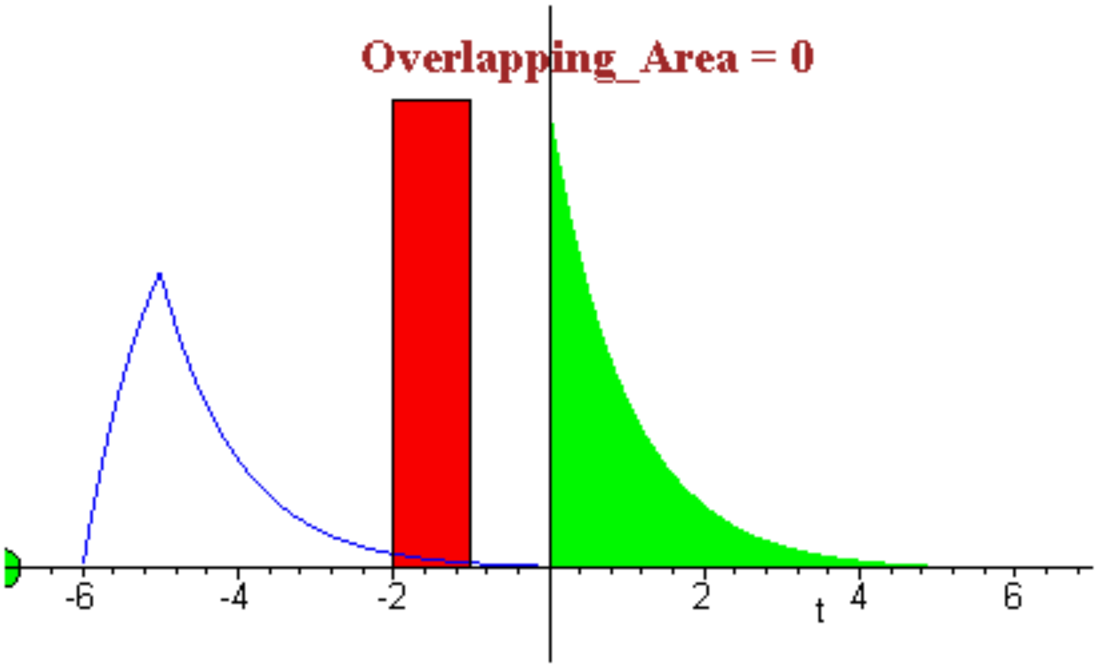
Biological Vision

- Hierarchical representation
- Map of visual space at lower levels
- Highly connected at upper levels of the hierarchy

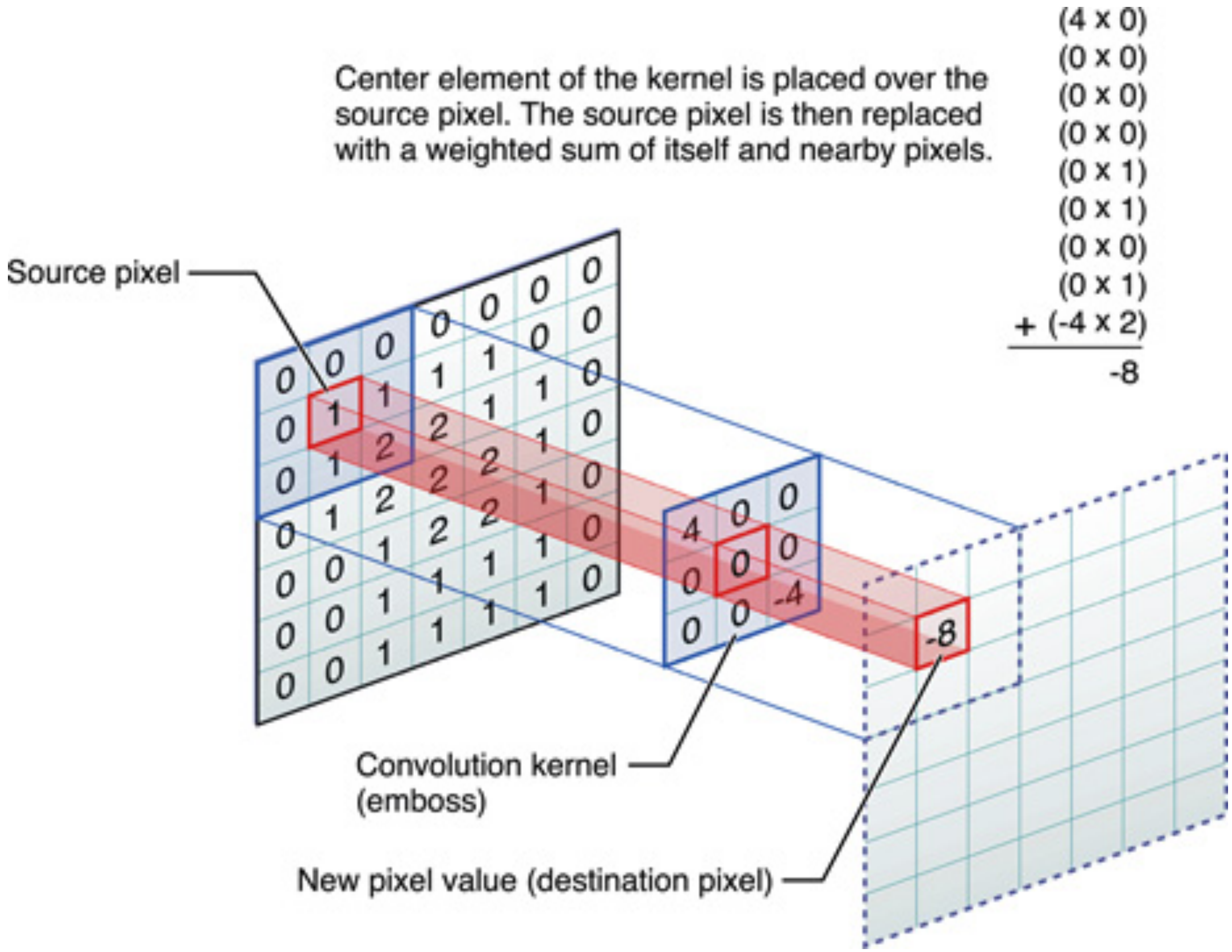
How do we turn this into a model?

Convolution & Pooling

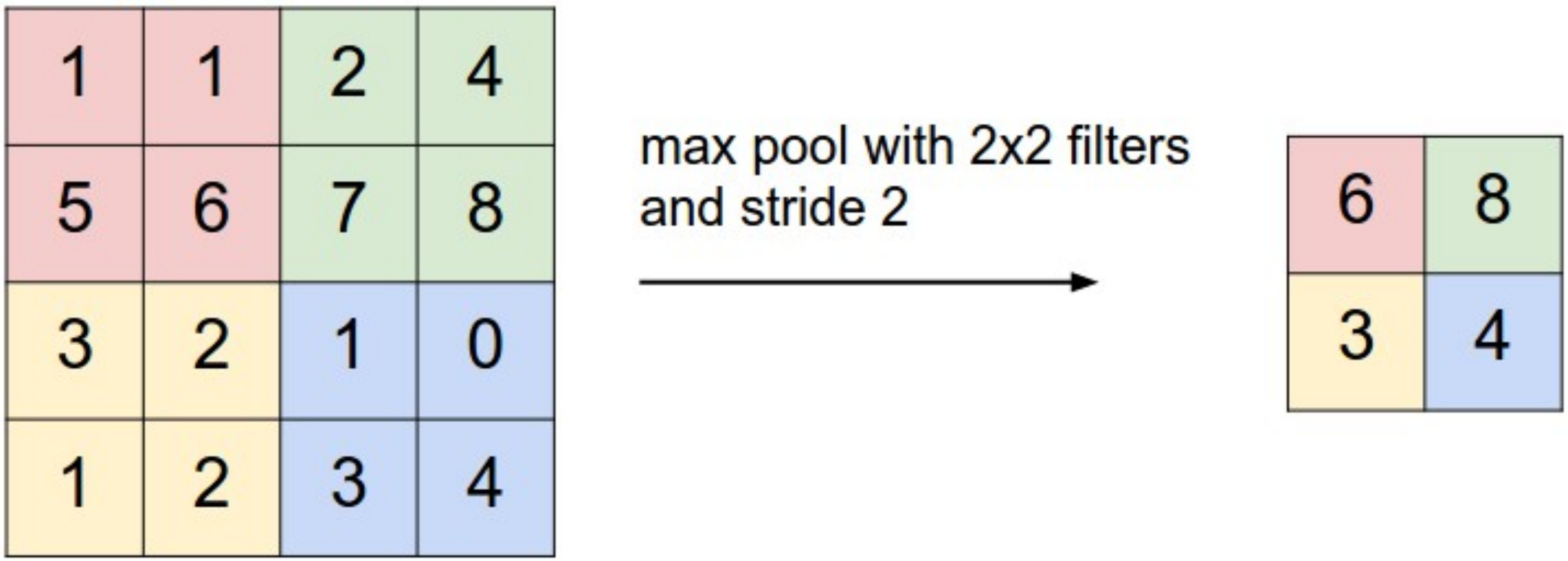
Convolutional Operation



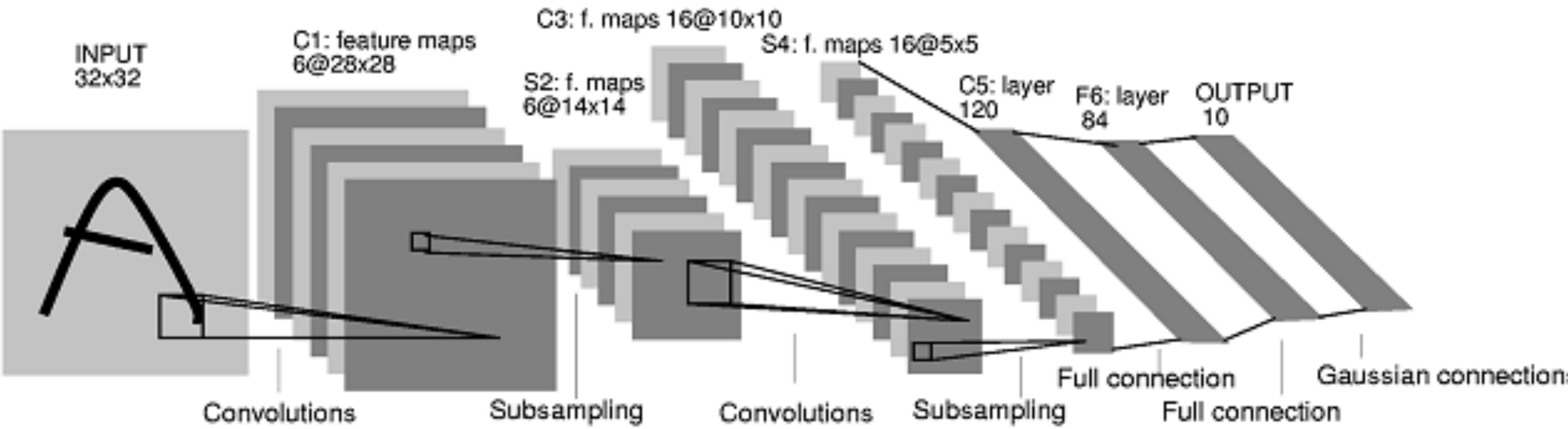
Convolutional Operation



Pooling Operation



LeNet



LeCun, 1989

AI Winter

AI Winter

AI Winter

- Convolutional neural networks are great, but…

AI Winter

- Convolutional neural networks are great, but…
  - They are hard to train

AI Winter

- Convolutional neural networks are great, but…
  - They are hard to train
  - They take a long time to train

AI Winter

- Convolutional neural networks are great, but…
  - They are hard to train
  - They take a long time to train
  - We don't have enough data to train them

GPUs

GPU

- Graphics Processing Unit

GPU

- Graphics Processing Unit
  - Rendering images is computationally intensive

GPU

- Graphics Processing Unit
  - Rendering images is computationally intensive
  - Parallel processing architecture to handle this task

GPU

- Graphics Processing Unit
  - Rendering images is computationally intensive
  - Parallel processing architecture to handle this task
- Can also handle matrix multiplication operations

GPU

- Graphics Processing Unit
  – Rendering images is computationally intensive
  – Parallel processing architecture to handle this task
- Can also handle matrix multiplication operations

Big Data

Big Data

- Cameras

Big Data

- Cameras
  - Digital cameras, smartphones

Big Data

- Cameras
  - Digital cameras, smartphones
- Internet

Big Data

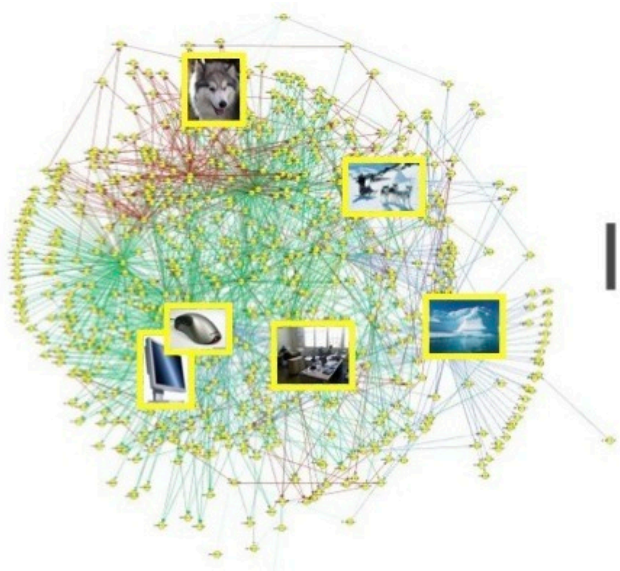
- Cameras
  - Digital cameras, smartphones
- Internet
  - Anyone can upload a picture

Big Data

- Cameras
  - Digital cameras, smartphones
- Internet
  - Anyone can upload a picture
  - Crowdsourcing

Big Data

- Cameras
  - Digital cameras, smartphones
- Internet
  - Anyone can upload a picture
  - Crowdsourcing
- ImageNet



IMAGENET

ImageNet Large Scale Visual Recognition Challenge

ImageNet Large Scale Visual Recognition Challenge

- Object recognition task

ImageNet Large Scale Visual Recognition Challenge

- Object recognition task
  - 1.2 million images

ImageNet Large Scale Visual Recognition Challenge

- Object recognition task
  - 1.2 million images
  - 1,000 classes of objects

ILSVRC 2012

ILSVRC 2012

- Krizhevsky, et al. use a deep convolutional network

ILSVRC 2012

- Krizhevsky, et al. use a deep convolutional network
  - Nearly halve the best error rate of the previous year

ILSVRC 2012

- Krizhevsky, et al. use a deep convolutional network
  - Nearly halve the best error rate of the previous year
  - Trained using GPUs and a few other tricks

Rectified Linear Units (ReLUs)

Rectified Linear Units (ReLUs)

- Researchers had primarily been using sigmoid non-linearities

Rectified Linear Units (ReLUs)

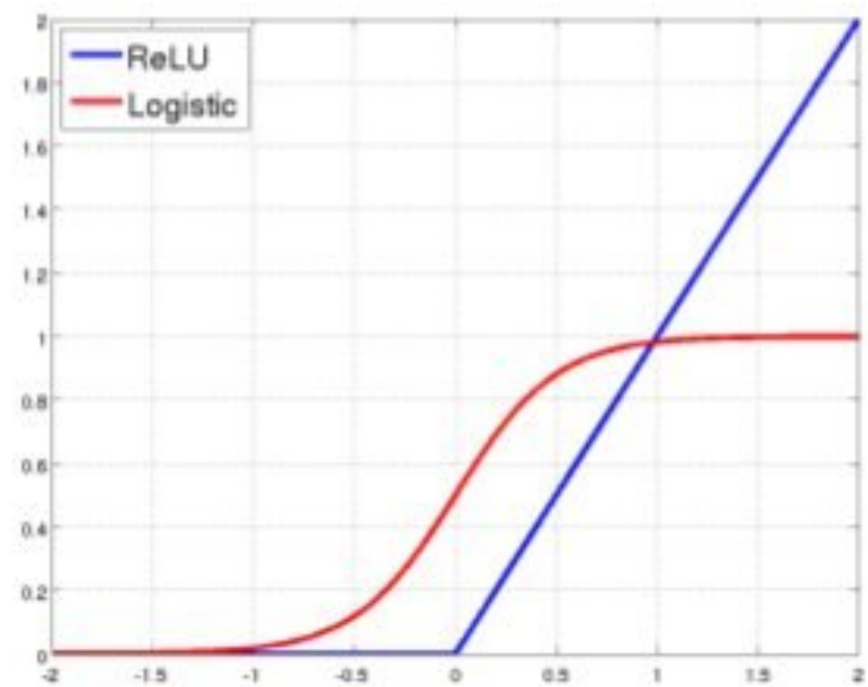
- Researchers had primarily been using sigmoid non-linearities
  - Vanishing gradient, saturation

Rectified Linear Units (ReLUs)

- Researchers had primarily been using sigmoid non-linearities
  - Vanishing gradient, saturation
- Instead, use ReLU

Rectified Linear Units (ReLUs)

- Researchers had primarily been using sigmoid non-linearities
  - Vanishing gradient, saturation
- Instead, use ReLU
  - Works much better!



Dropout

Dropout

- Unreliable connections between layers

Dropout

- Unreliable connections between layers
  - Randomly have connections ‘drop out’

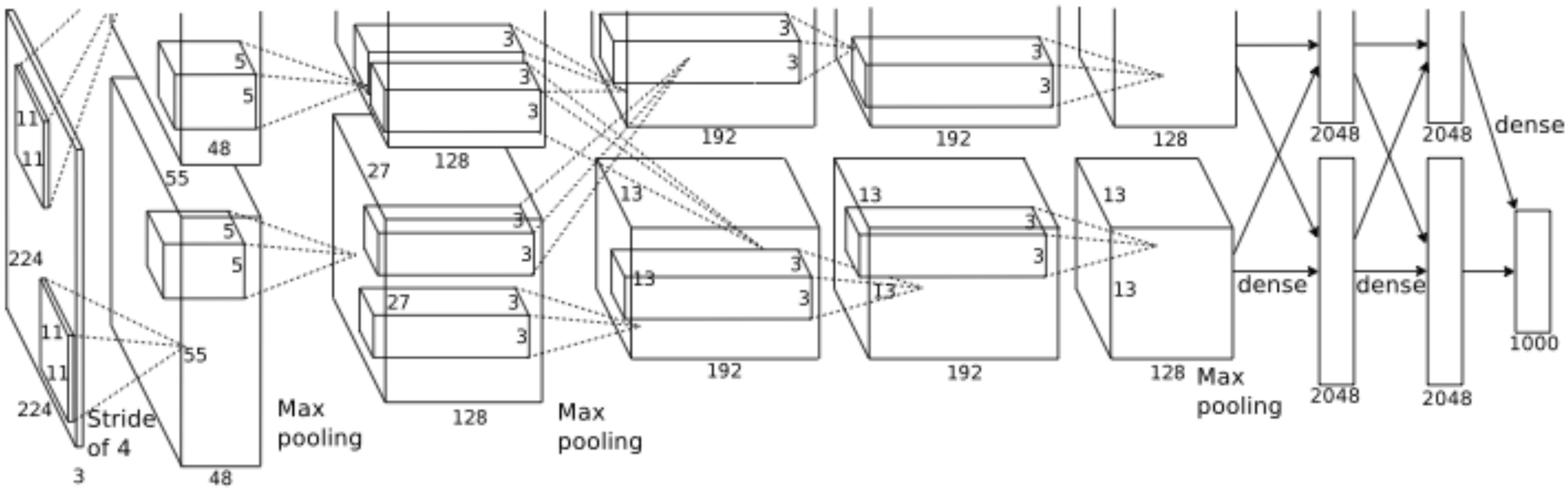
Dropout

- Unreliable connections between layers
  - Randomly have connections ‘drop out’
- Acts as a regularizer

Dropout

- Unreliable connections between layers
  - Randomly have connections ‘drop out’
- Acts as a regularizer
  - Forces the network to learn general features

AlexNet



Image

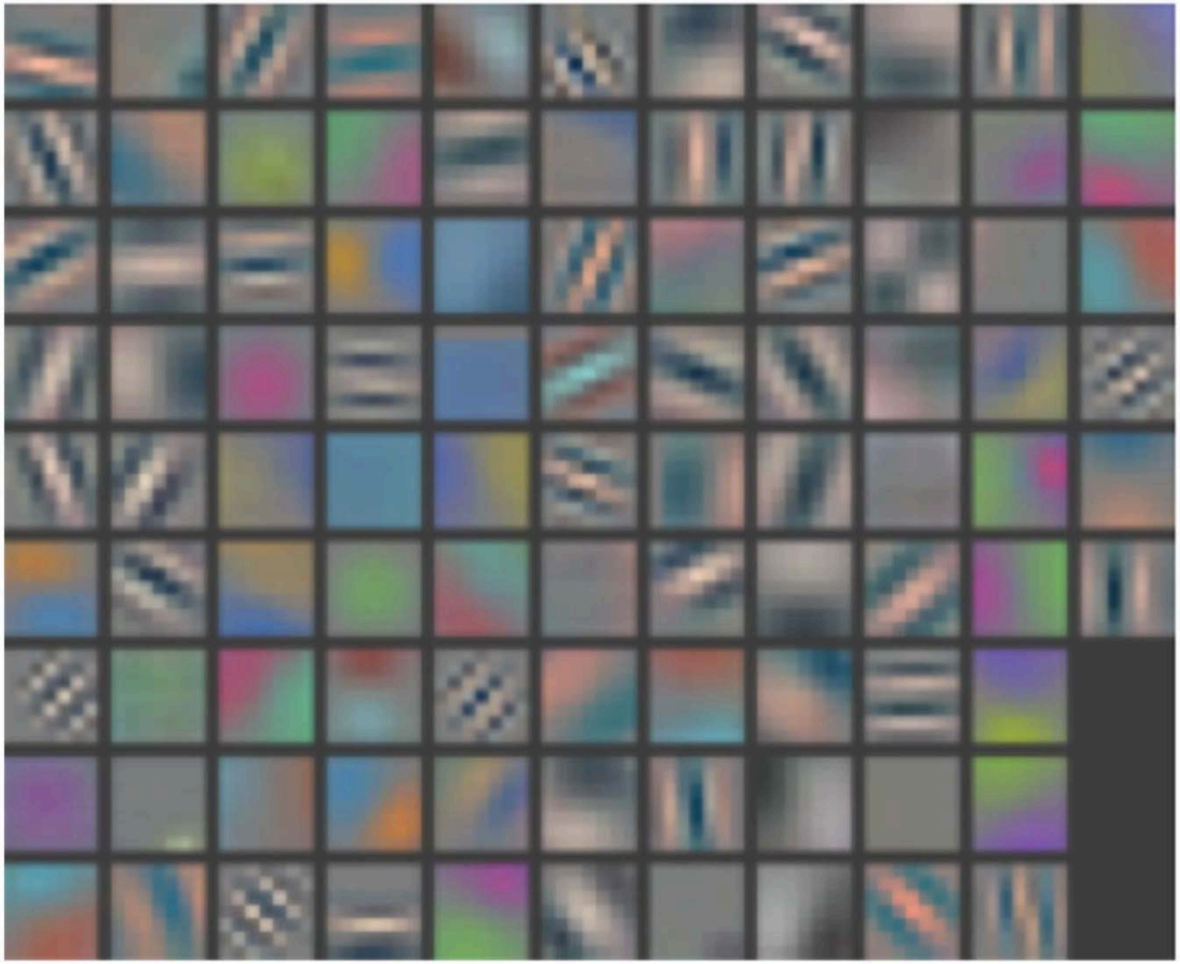
Convolution and Max Pooling Layers

Fully Connected
Layers

Features

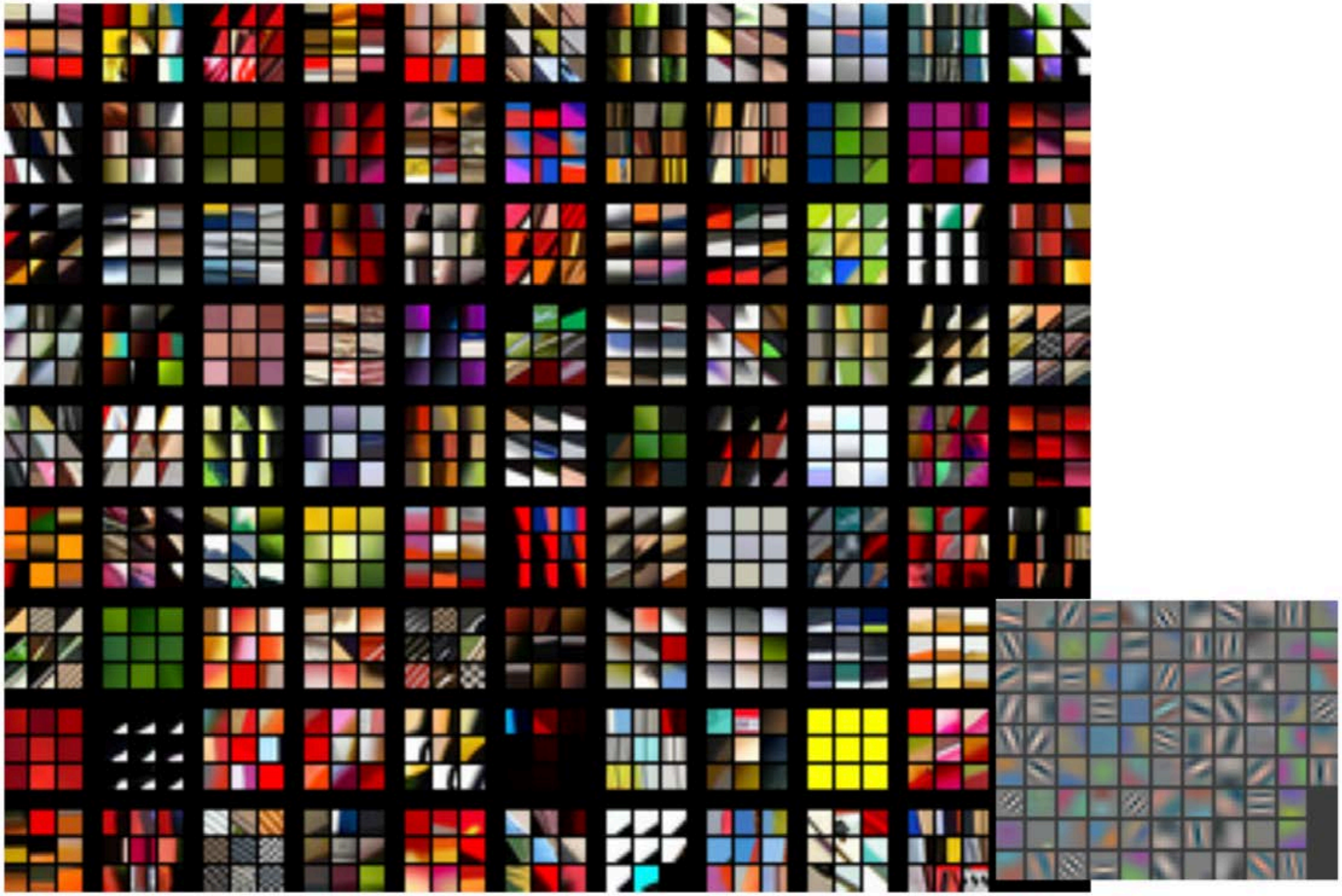
Features

- Conv1

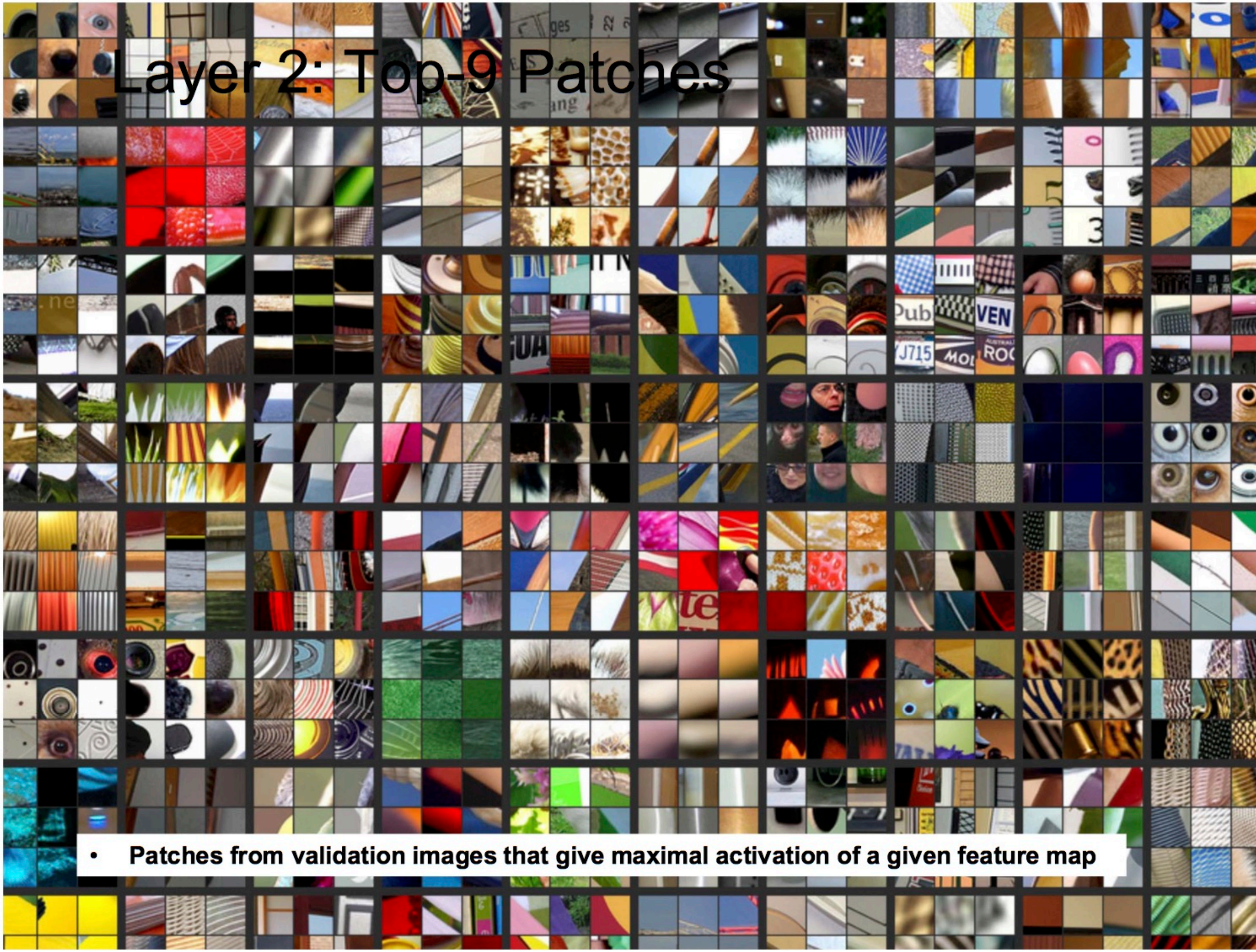


Features

- Top Image Patches

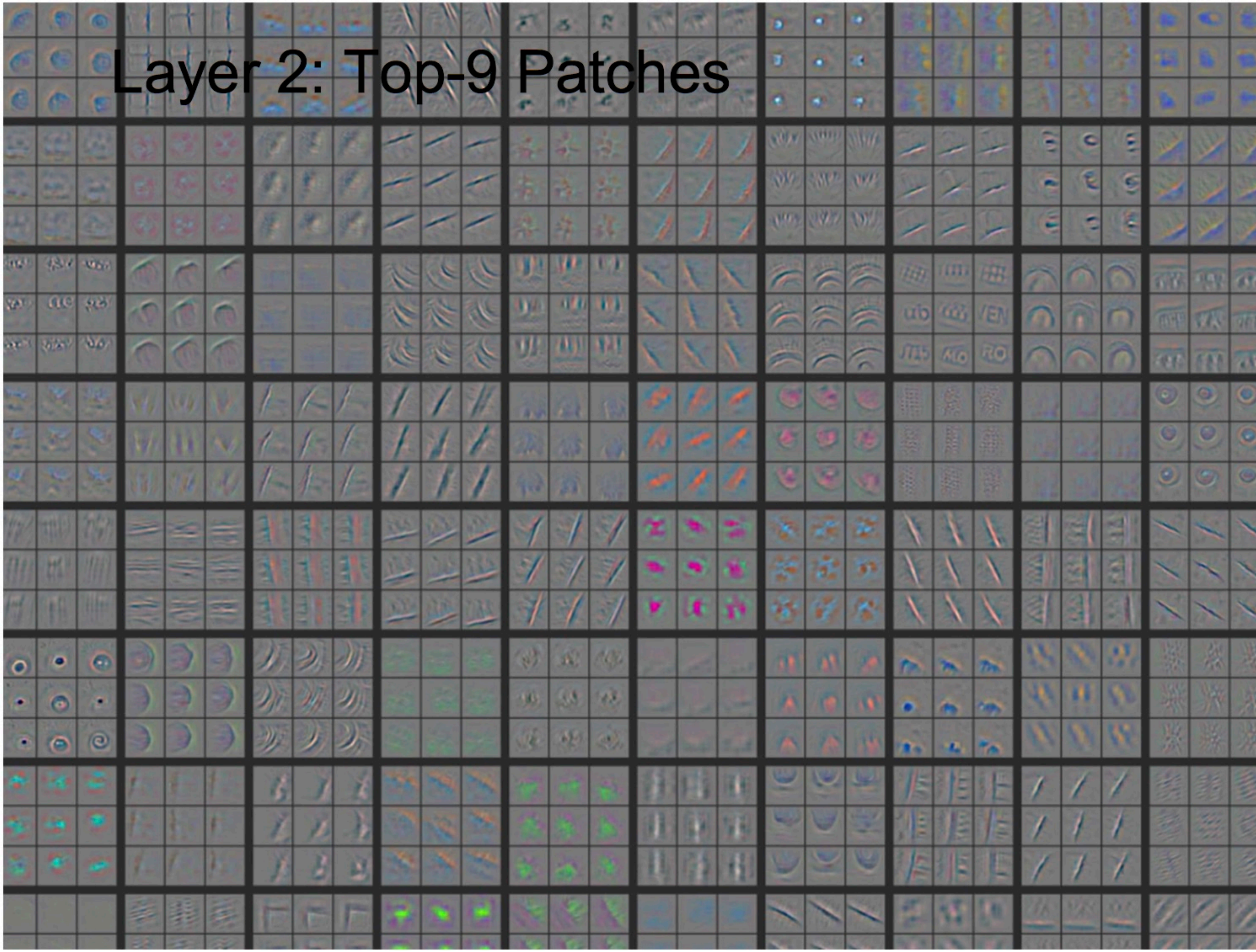


Layer 2: Top-9 Patches

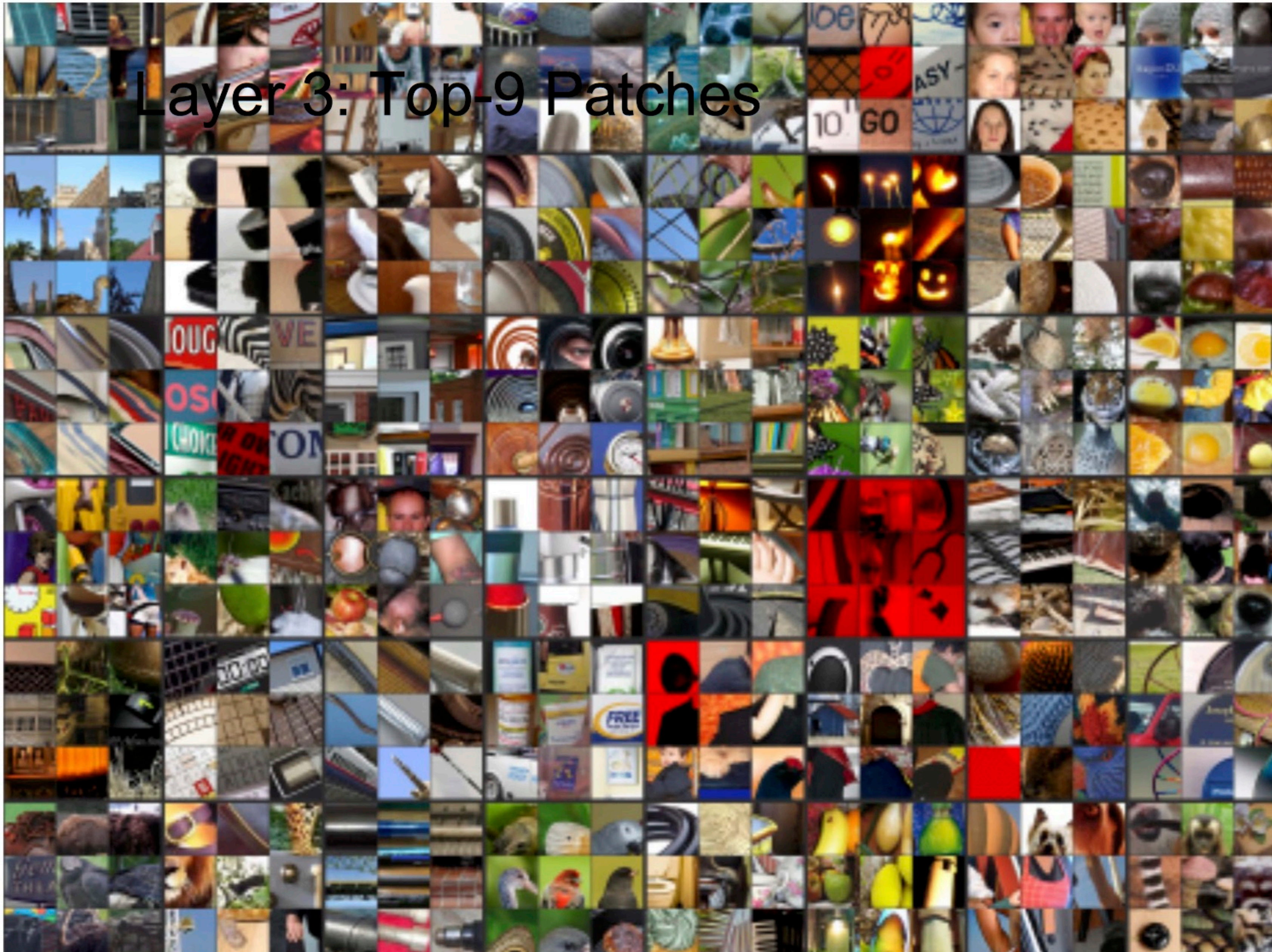


- **Patches from validation images that give maximal activation of a given feature map**

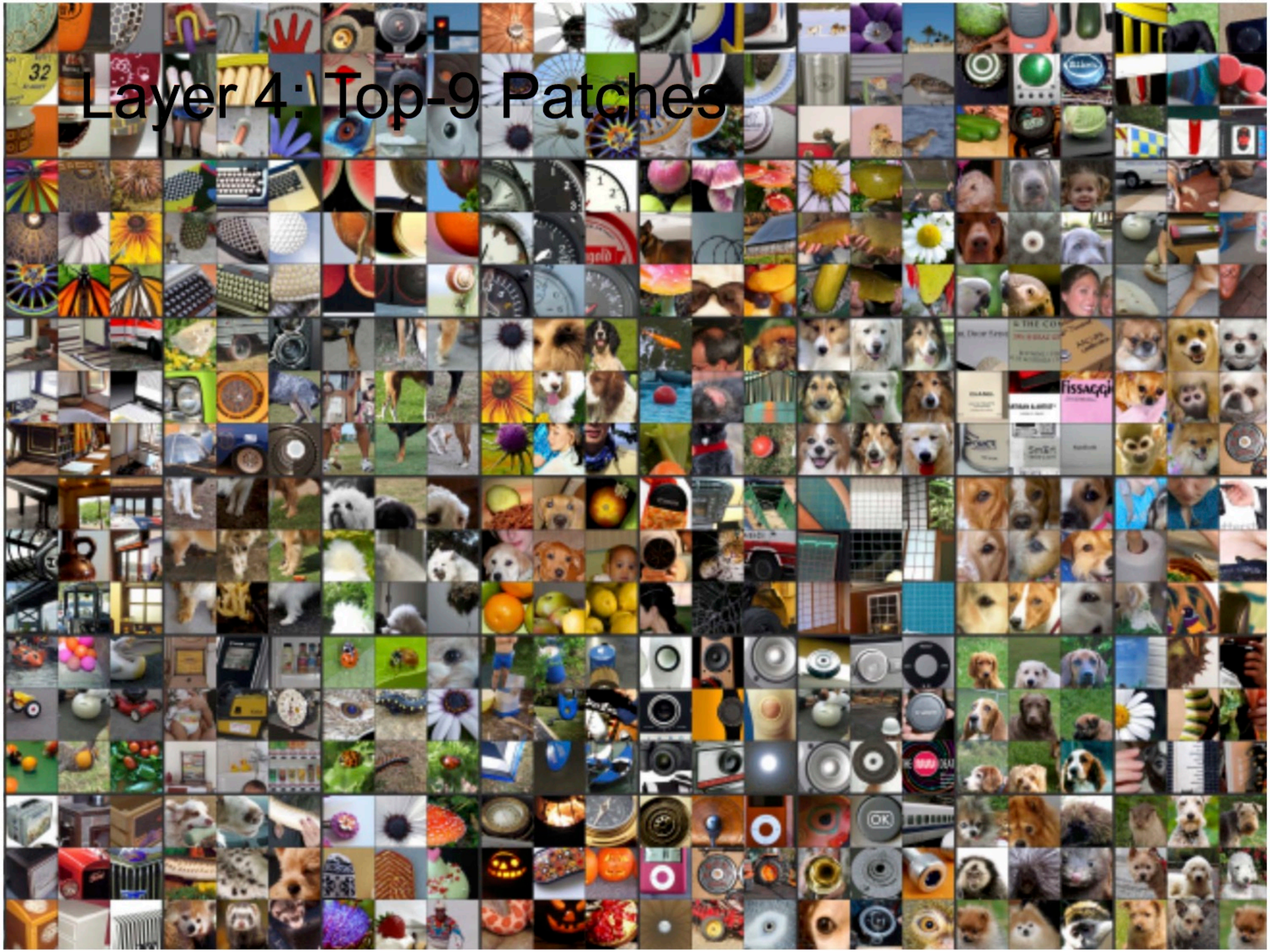
Layer 2: Top-9 Patches



Layer 3: Top-9 Patches

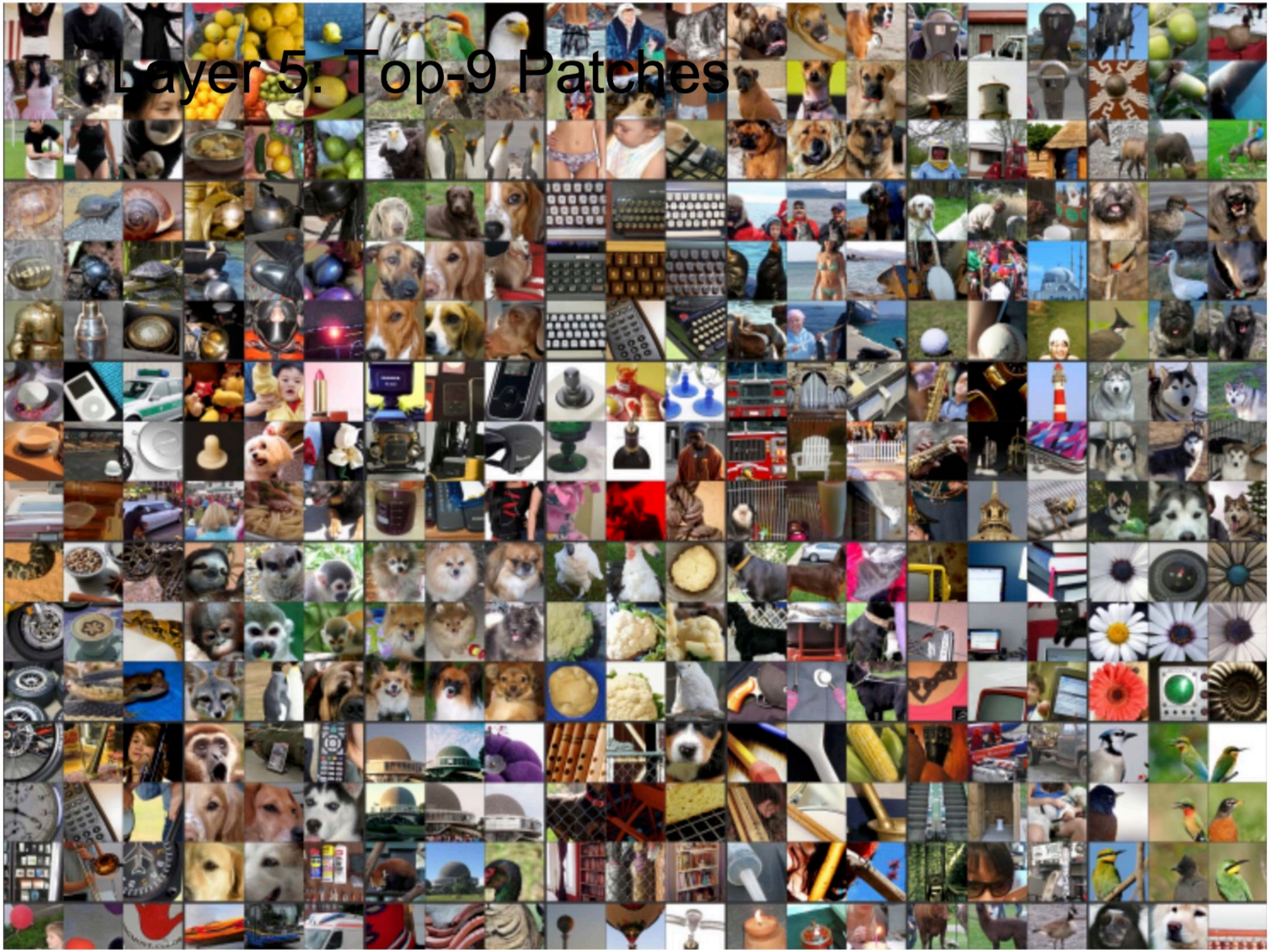


Layer 3: Top-9 Patches


Layer 4: Top-9 Patches

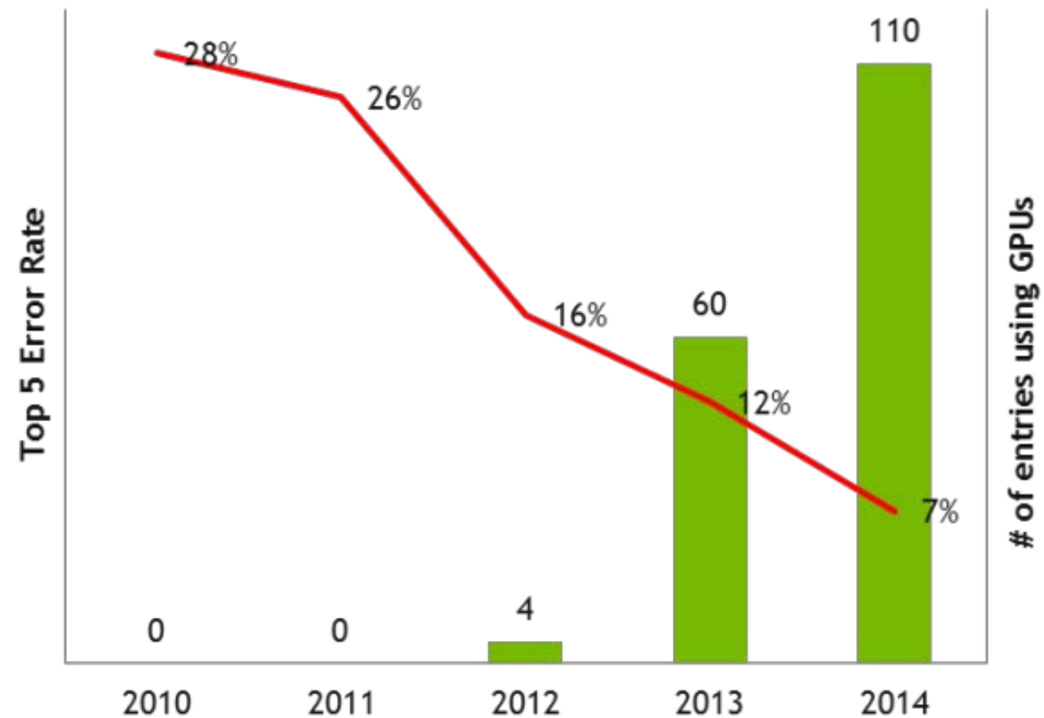
Layer 4: Top-9 Patches

Layer 5: Top-9 Patches



Layer 5: Top-9 Patches

IMAGENET


Top 5 Error Rate
# of entries using GPUs
28%
26%
16%
12%
7%
0
0
4
60
110
2010
2011
2012
2013
2014

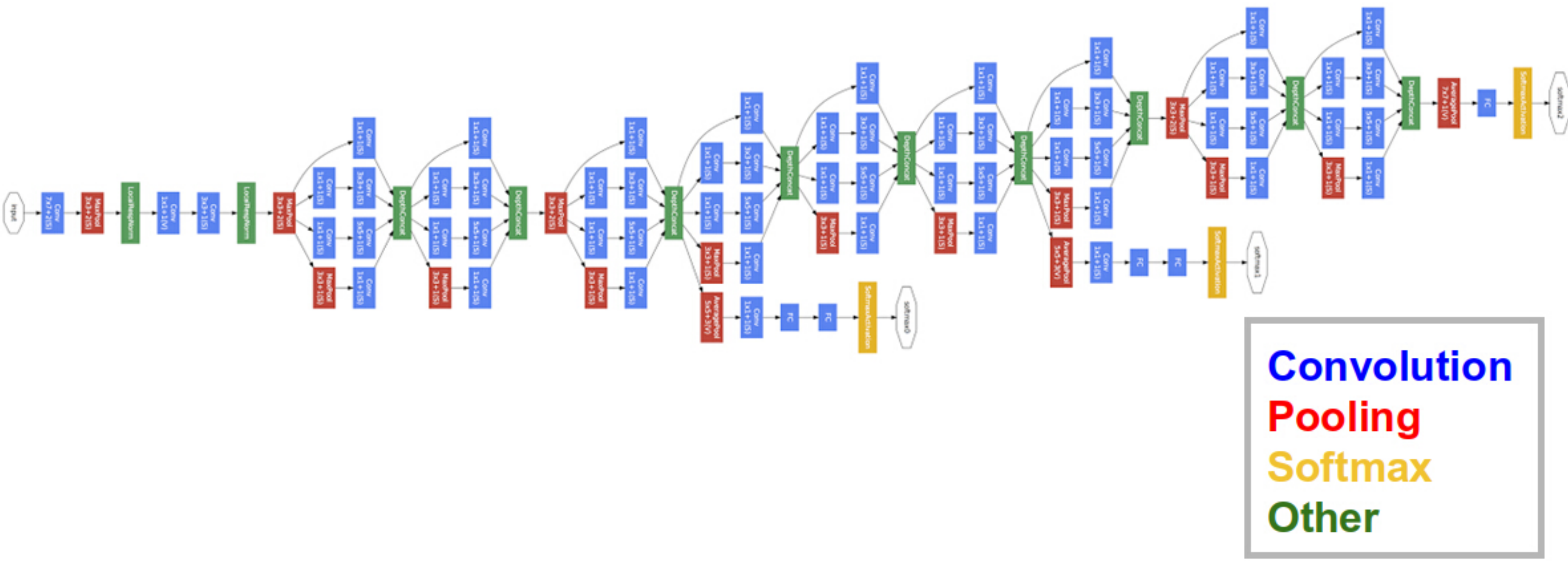
2014

2014

- GoogLeNet

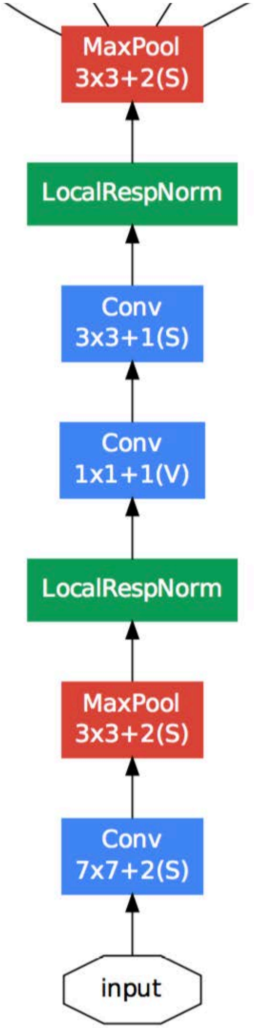
2014

- GoogLeNet



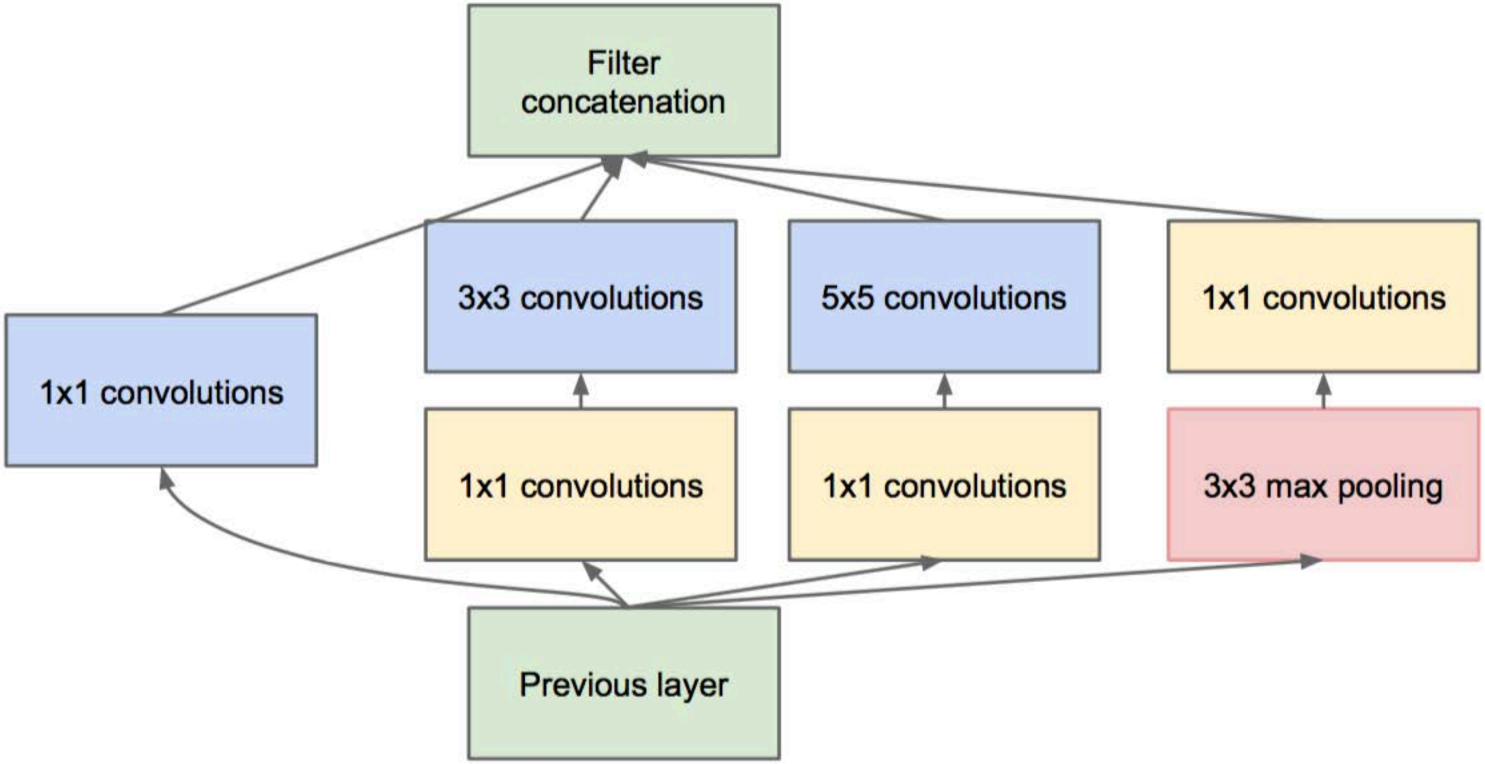
2014

- GoogLeNet



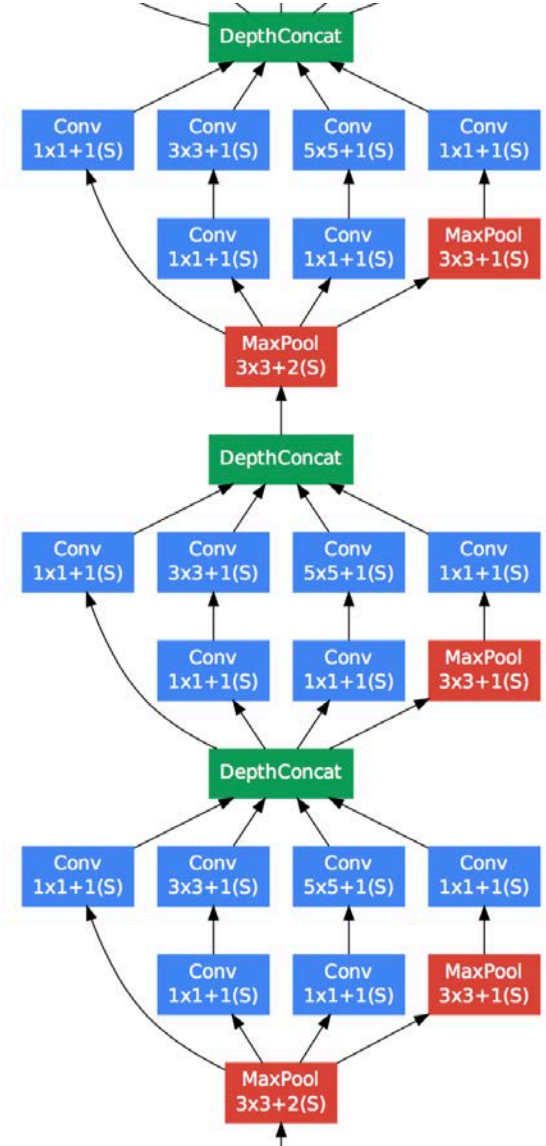
2014

- GoogLeNet



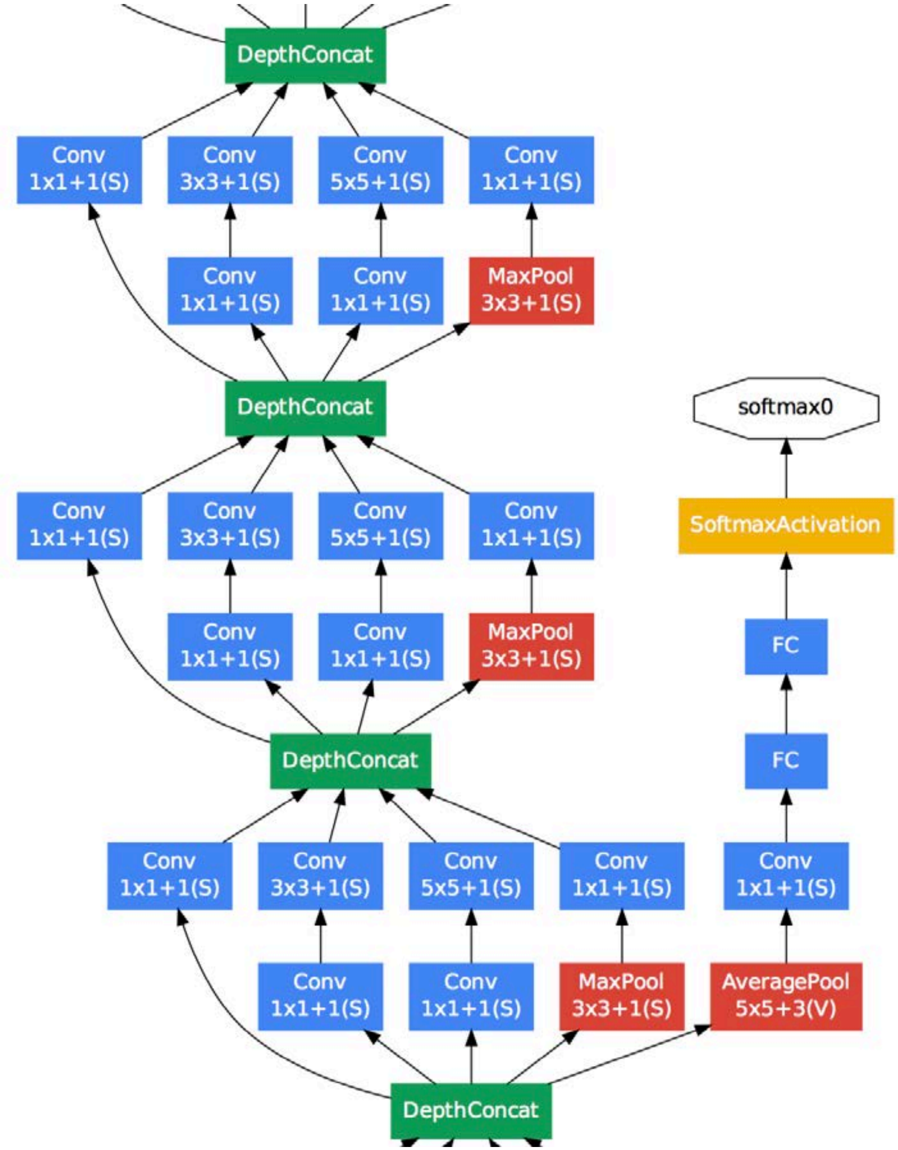
2014

- GoogLeNet



2014

- GoogLeNet



2014

- GoogLeNet
  - 7% top-5 error

2015

- Microsoft

2015

- Microsoft
  - 5% top-5 accuracy

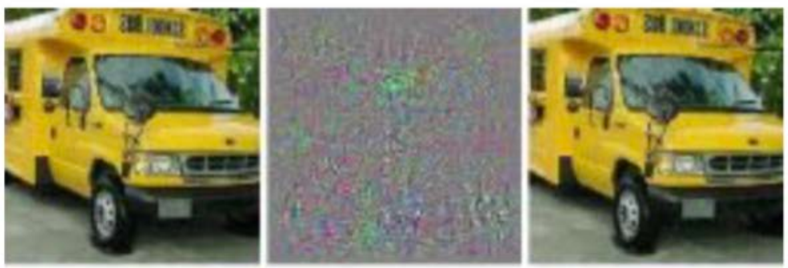
2015

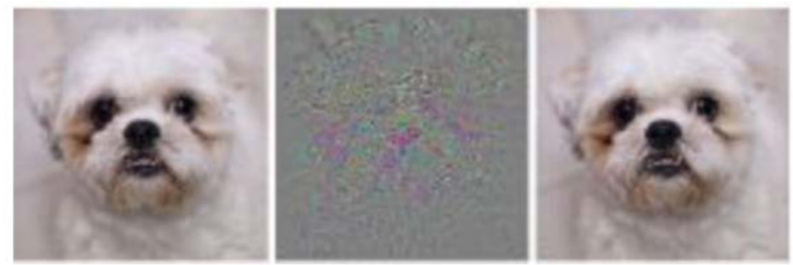
- Microsoft
  - 5% top-5 accuracy
  - Surpassed human level performance

Issues

Issues

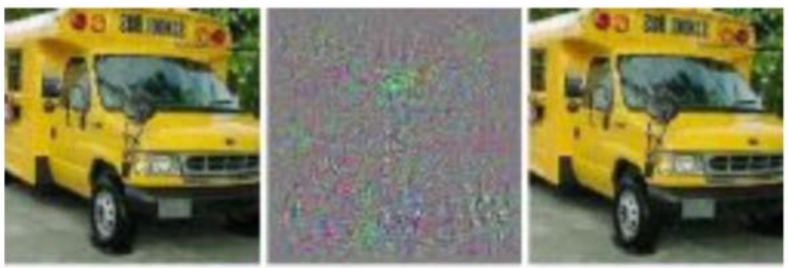
- Adversarial examples

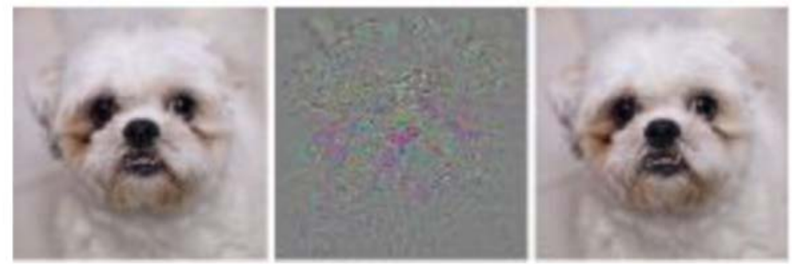




Issues

- Adversarial examples

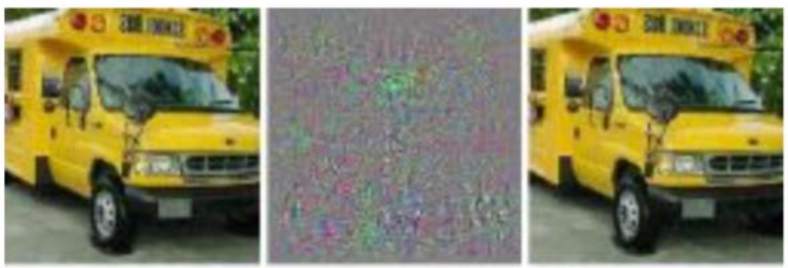


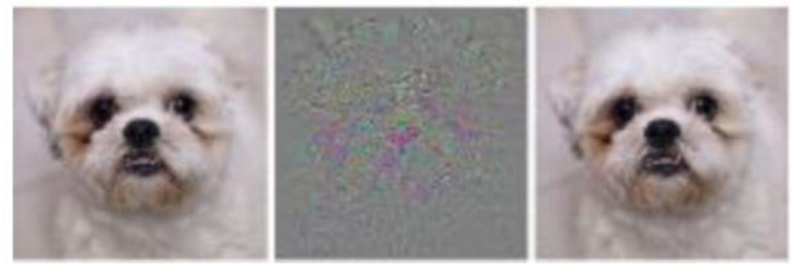


- Lacking a theoretical understanding of these models

Issues

- Adversarial examples





- Lacking a theoretical understanding of these models
- Learning is dependent on class labels. Unsupervised deep learning is less developed.

Software Packages

- Caffe - https://github.com/BVLC/caffe
- Torch - https://github.com/torch/torch7
- Theano - https://github.com/Theano/Theano
- Neon - https://github.com/NervanaSystems/neon
- TensorFlow - https://github.com/tensorflow/tensorflow

Resources

LeNet: Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. *Advances in Neural Information Processing Systems.* 1990.

ImageNet: Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) ImageNet Large Scale Visual Recognition Challenge. *arXiv:1409.0575,* 2014.

AlexNet: Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

Network Visualization: Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." *Computer Vision–ECCV 2014*. Springer International Publishing, 2014. 818-833.

GoogLeNet: Szegedy, Christian, et al. "Going deeper with convolutions." *arXiv preprint arXiv:1409.4842* (2014).

Microsoft Network: He, Kaiming, et al. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." *arXiv preprint arXiv:1502.01852* (2015).

Adversarial Examples: Szegedy, Christian, et al. "Intriguing properties of neural networks." *arXiv preprint arXiv:1312.6199* (2013).