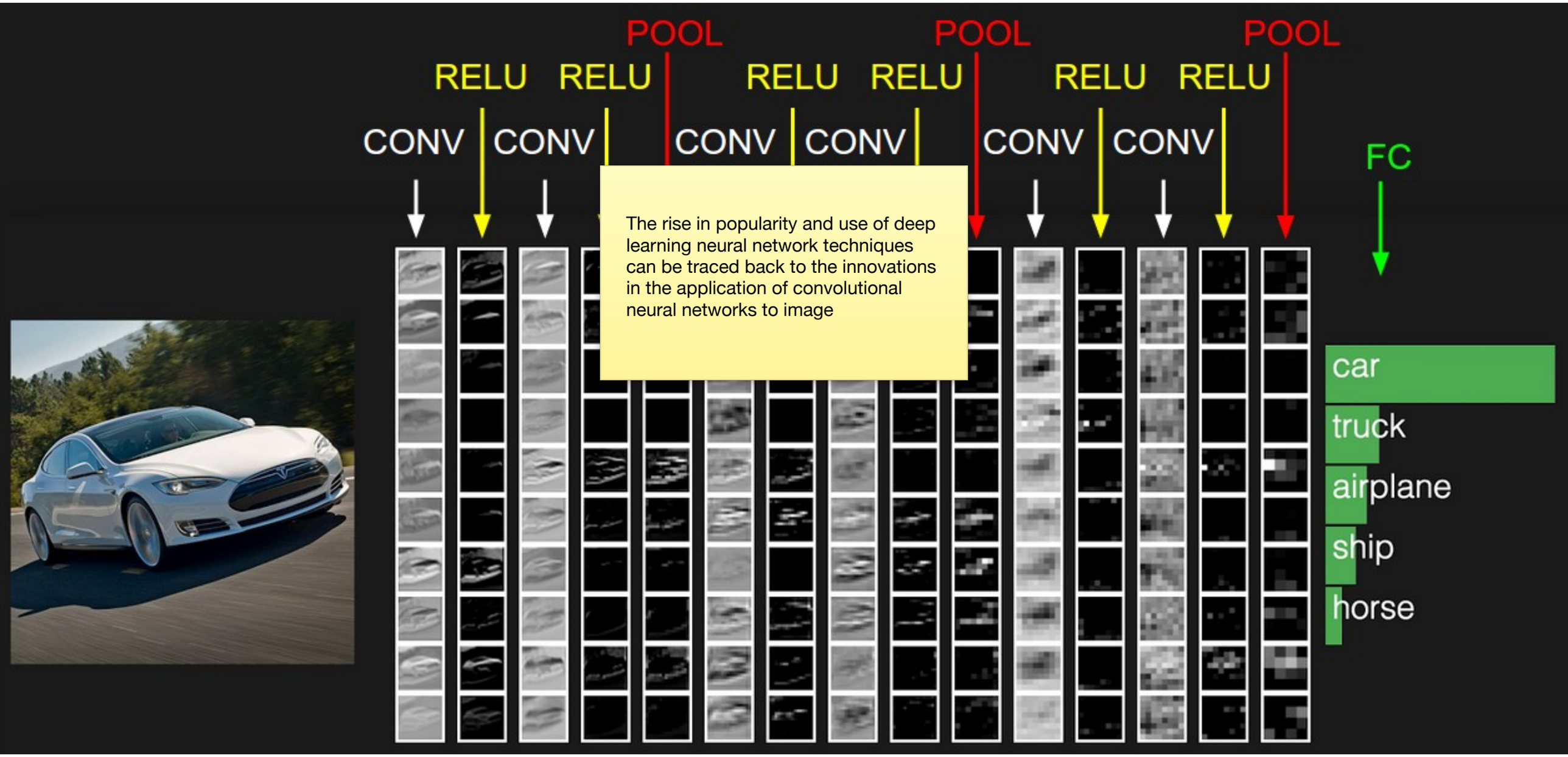
Image Classification Networks: classical architectures and common design patterns

Kate Farrahi

Vision, Learning and Control
University of Southampton

Motivation: Image Classification



The rise in popularity and use of deep learning neural network techniques can be traced back to the innovations in the application of convolutional neural networks to image

ImageNet Dataset

- ImageNet is a massive dataset of annotated photographs for computer vision research

- Collected by researchers at Stanford, led by Prof. Fei-Fei Li

- +14M images, +21K classes, +1M images with bounding box annotations

- Image annotations by humans using crowdsourcing (e.g. Amazon's Mechanical Turk)

- The database of annotations of third-party image URLs is freely available, though the images are not owned by ImageNet.

History of ImageNet

- Fei-Fei Li began working on the idea for ImageNet in 2006. At that time, most AI research focused on models and algorithms.

- In 2007, Li met with Princeton professor Christiane Fellbaum, one of the creators of WordNet. As a result of this meeting, Li went on to build ImageNet starting from the word-database of WordNet and using many of its features.

- WordNet is a lexical database of semantic relations between words in more than 200 languages. Began in the mid 80s by George Miller.

- They presented the database for the first time as a poster at the 2009 CVPR Conference in Florida (paper).

https://en.wikipedia.org/wiki/ImageNet

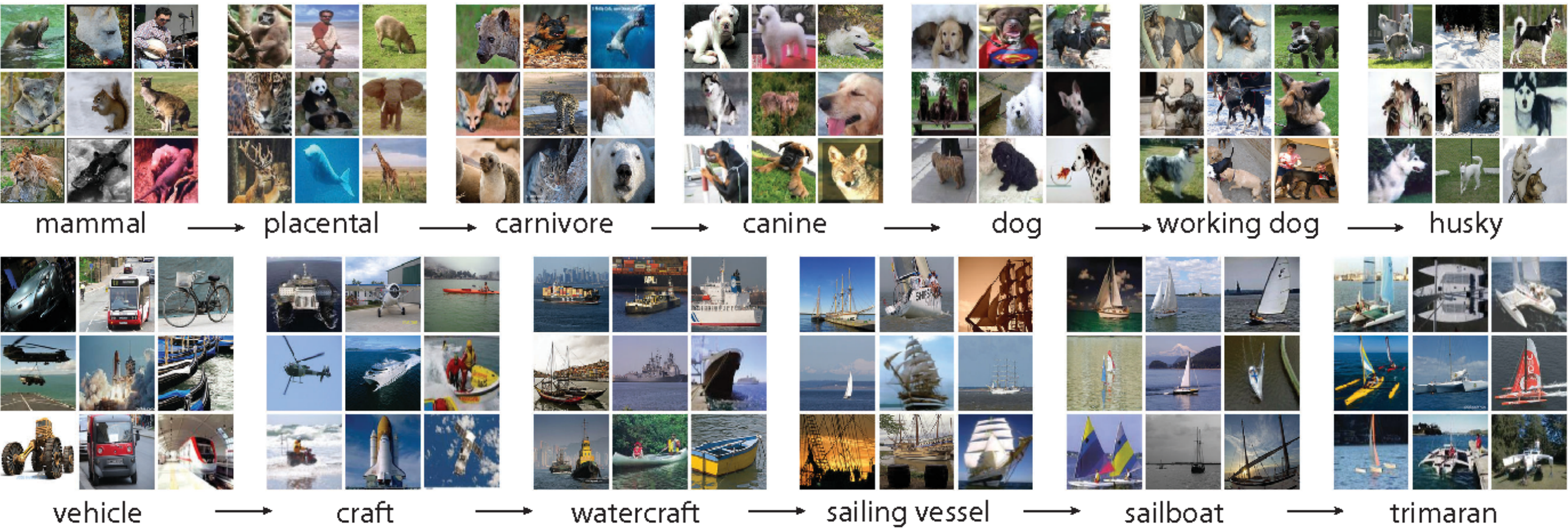


Image taken from the ImageNet paper CVPR 2009.

The ImageNet Challenge

- ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

- Annual challenge on a subset of ImageNet

- Designed to foster the development and benchmarking of state-of-the-art algorithms

- The challenge has led to milestone model architectures and techniques that are more widely used in deep learning

Download

Download ImageNet Data

The most highly-used subset of ImageNet is the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012-2017 image classification and localization dataset. This dataset spans 1000 object classes and contains 1,281,167 training images, 50,000 validation images and 100,000 test images. This subset is available on Kaggle.

For access to the full ImageNet dataset and other commonly used subsets, please login or request access. In doing so, you will need to agree to our terms of access.

ILSVRC

- The annual competition was held between 2010 and 2017 using subsets of the ImageNet dataset

- Typically, the training dataset comprised of 1 million images, with 50,000 for a validation dataset and 150,000 for a test set (available to download here)

- The general challenge tasks for most years (Image classification, Single-object localization, Object detection)

- Publication: Int J Comput Vis paper link, TED talk

Image classification

Steel drum

Ground truth

Steel drum
Folding chair
Loudspeaker

Accuracy: 1

Scale
T-shirt
Steel drum
Drumstick
Mud turtle

Accuracy: 1

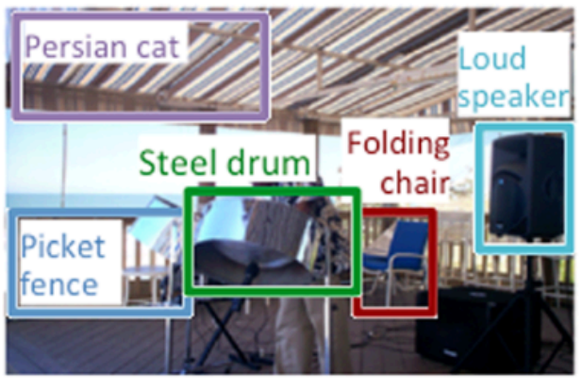
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

Accuracy: 0

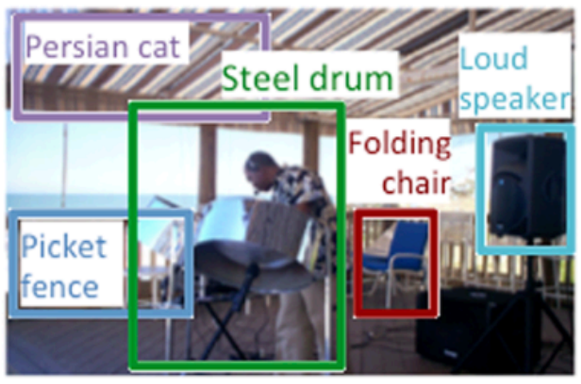
Single-object localization

Steel drum

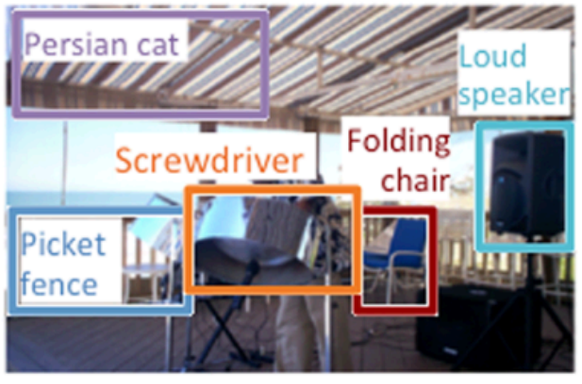
Ground truth



Accuracy: 1

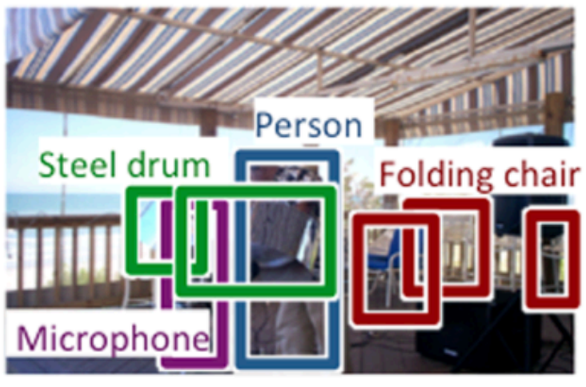


Accuracy: 0

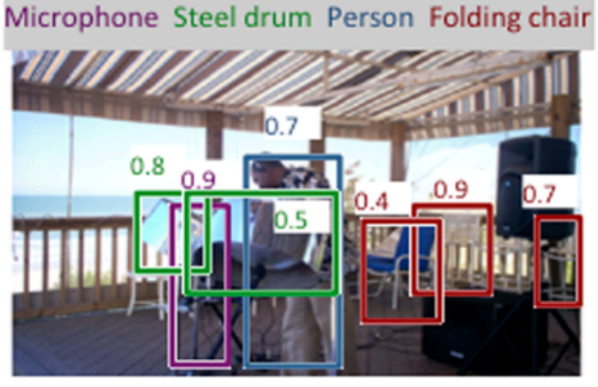


Accuracy: 0

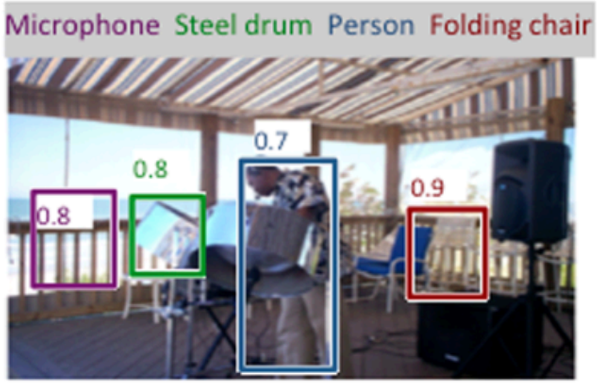
Object detection



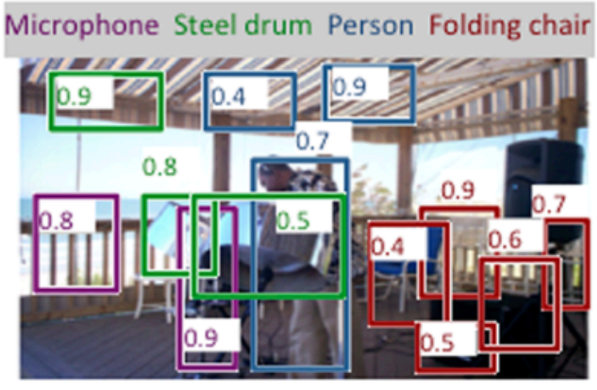
Ground truth



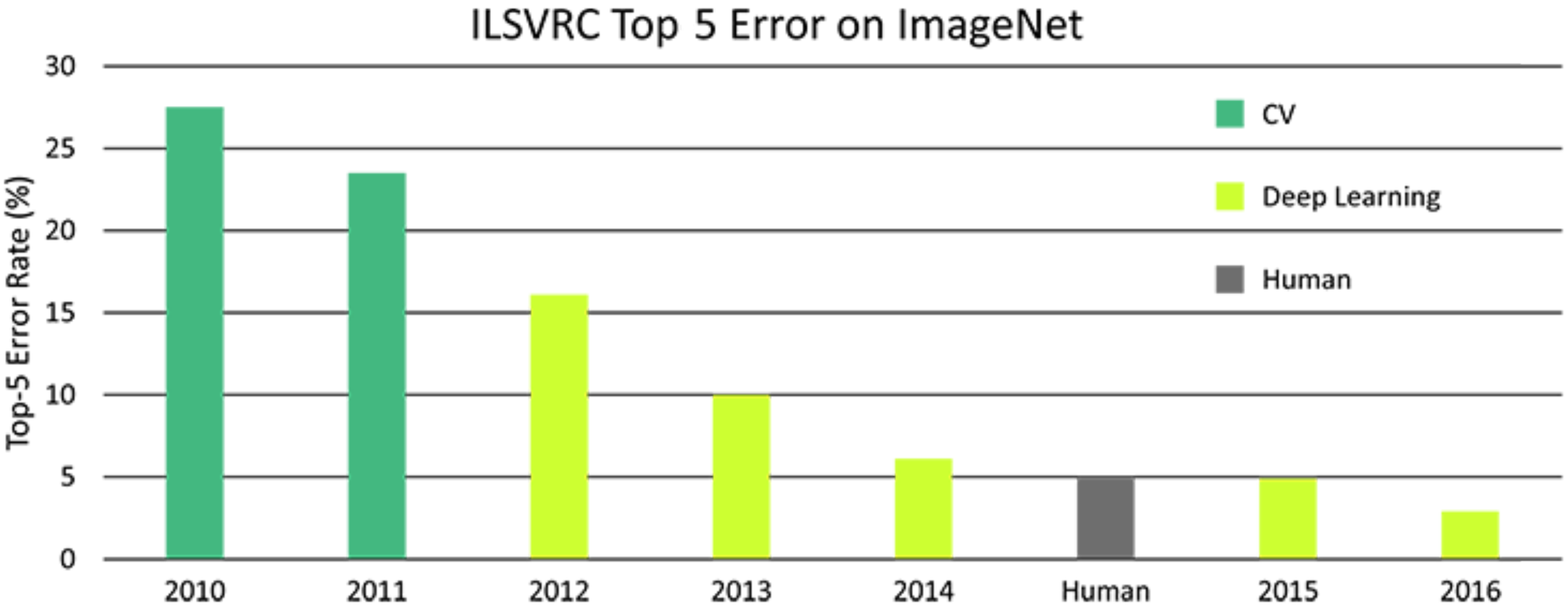
AP: **1.0** **1.0** **1.0** **1.0**

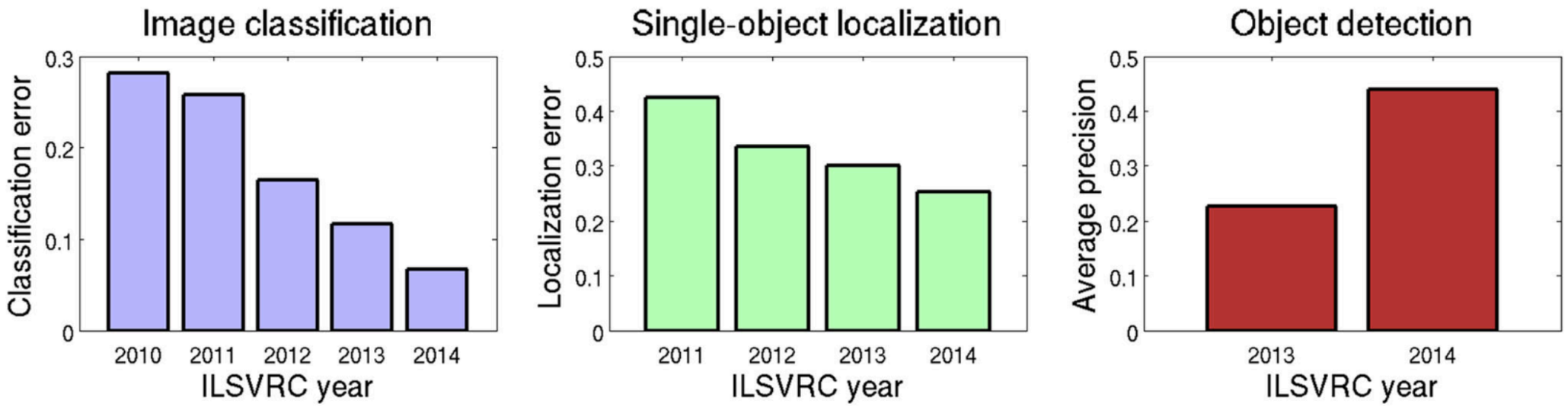


AP: **0.0** **0.5** **1.0** **0.3**



AP: **1.0** **0.7** **0.5** **0.9**


ILSVRC Top 5 Error on ImageNet
Top-5 Error Rate (%)
30
25
20
15
10
5
0
2010
2011
2012
2013
2014
Human
2015
2016
CV
Deep Learning
Human



Taken from paper ImageNet IJCV 2015 (link)

Image Classification on ImageNet

Leaderboard Dataset

View Top 5 Accuracy by Date for All models

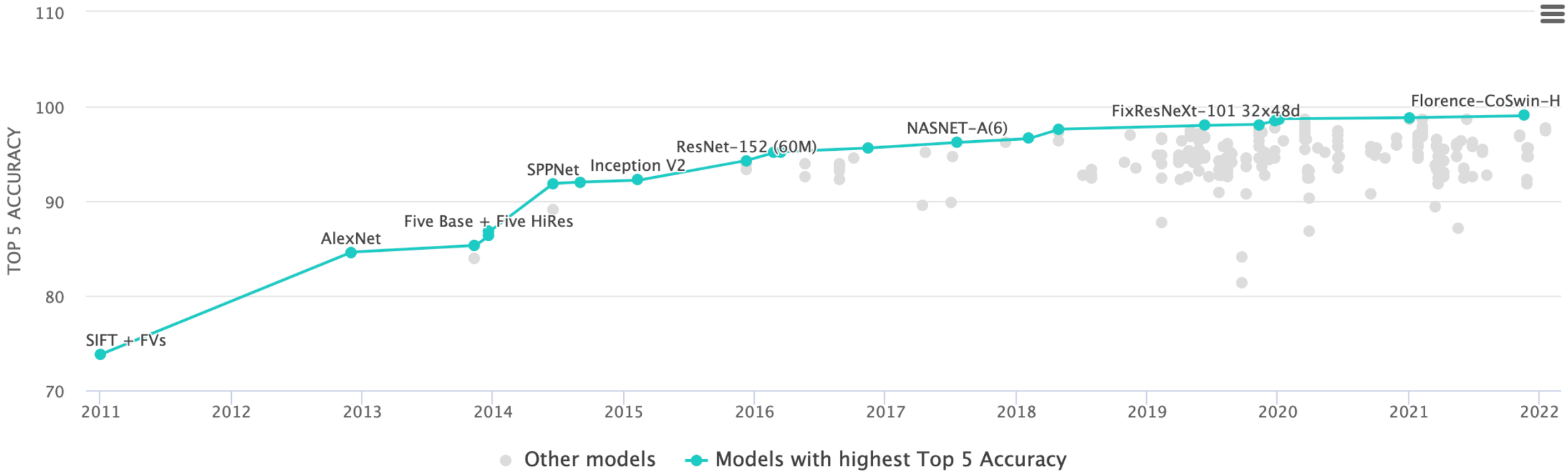


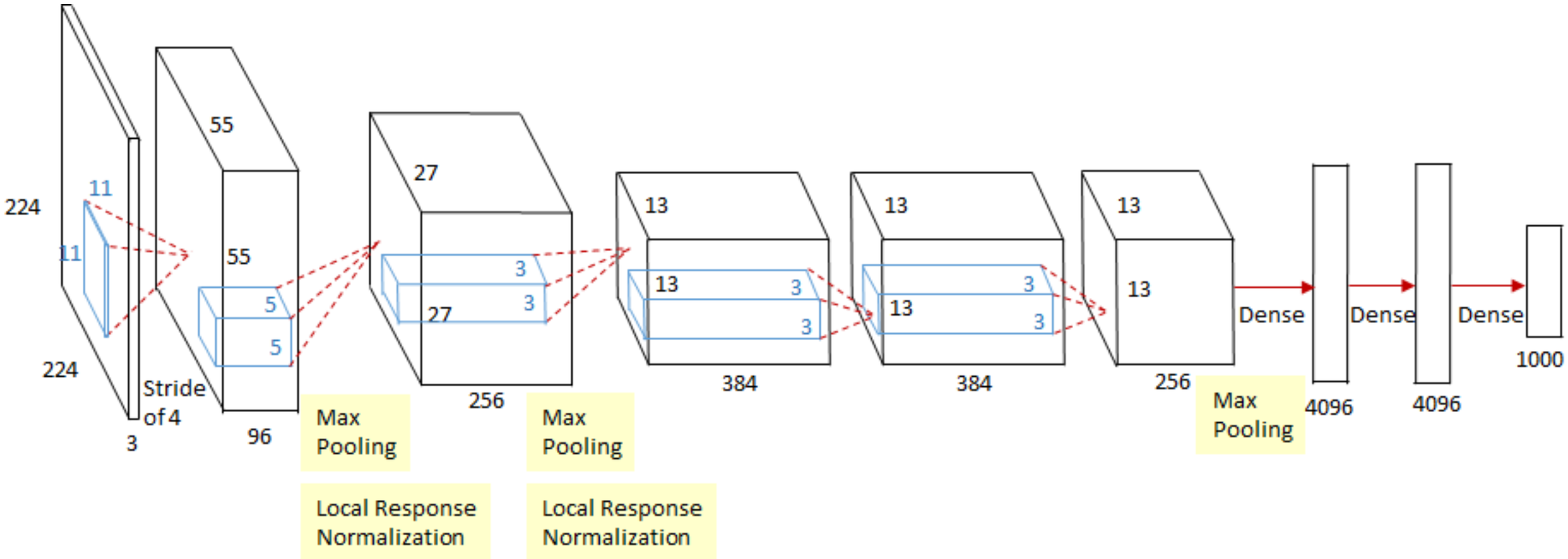
Image taken from here

Classic Architectures

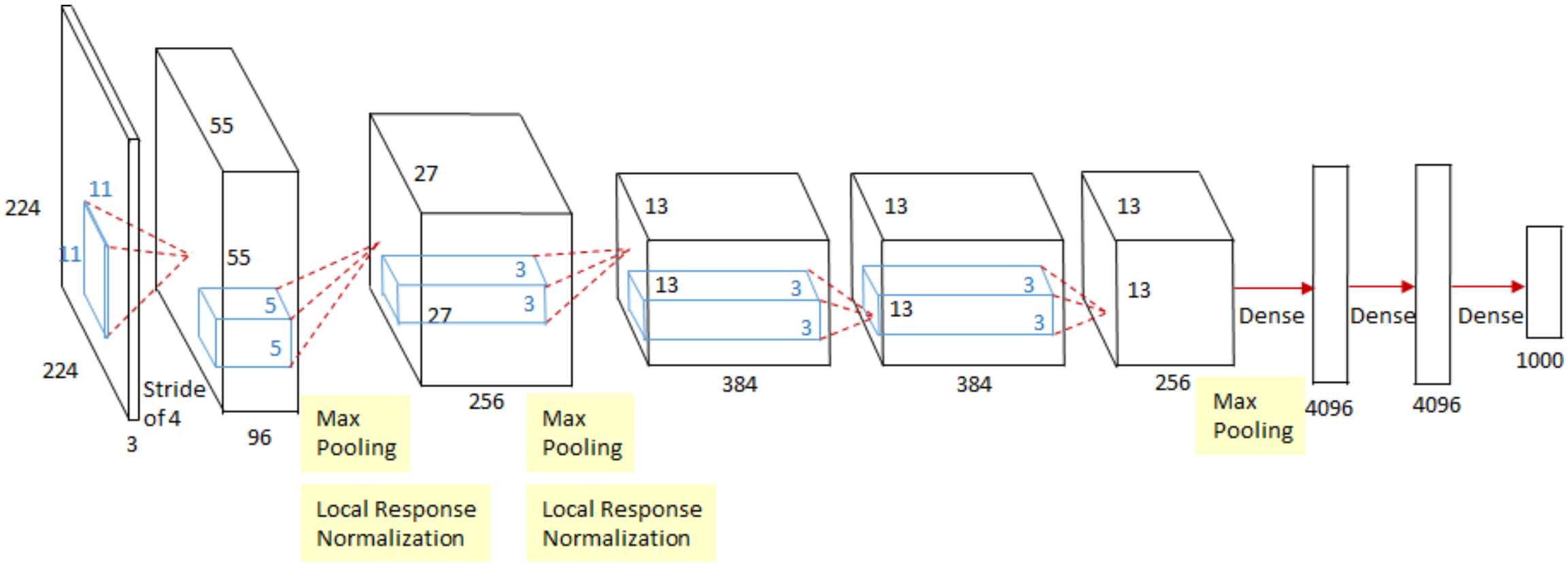
AlexNet

ImageNet Classification with Deep Convolutional Neural Networks. https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

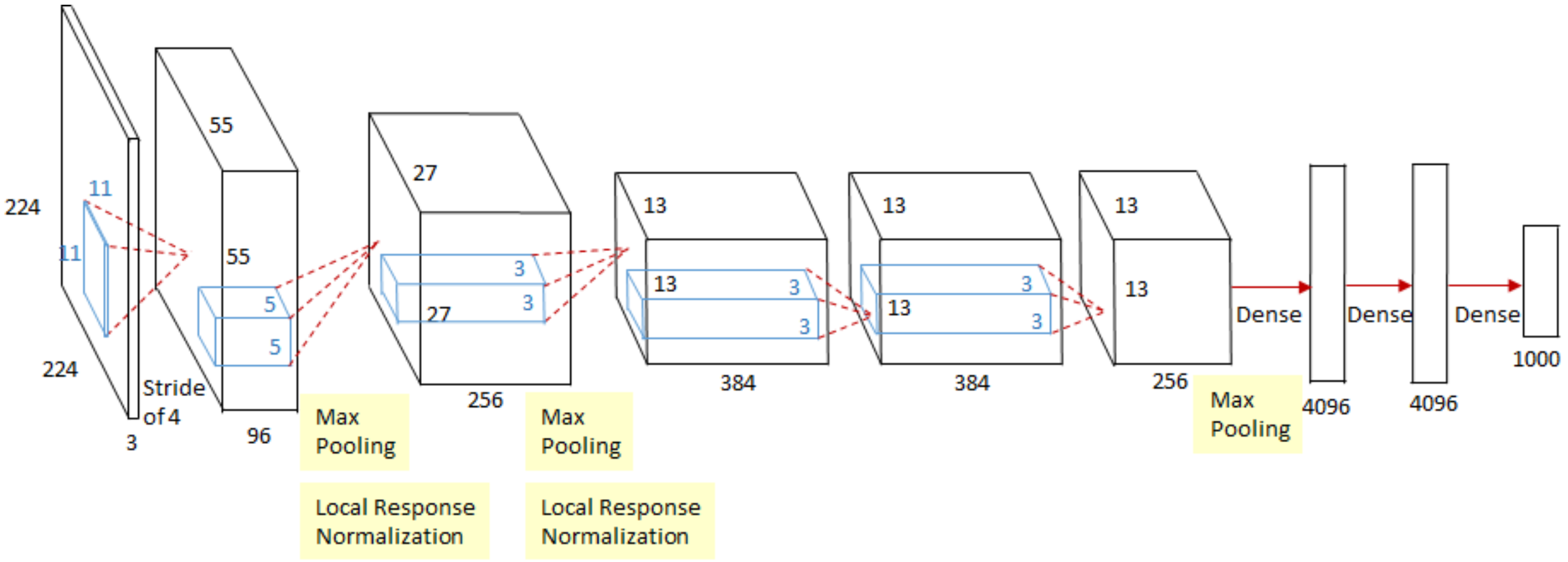
CaffeNet


224
11
11
224
3
Stride
of 4
55
55
96
5
5
Max
Pooling
Local Response
Normalization
27
27
256
3
3
Max
Pooling
Local Response
Normalization
13
13
384
3
3
13
13
384
3
3
13
13
256
Max
Pooling
Dense
4096
Dense
4096
Dense
1000

CaffeNet - How many parameters?



CaffeNet - Effective Receptive Field Sizes



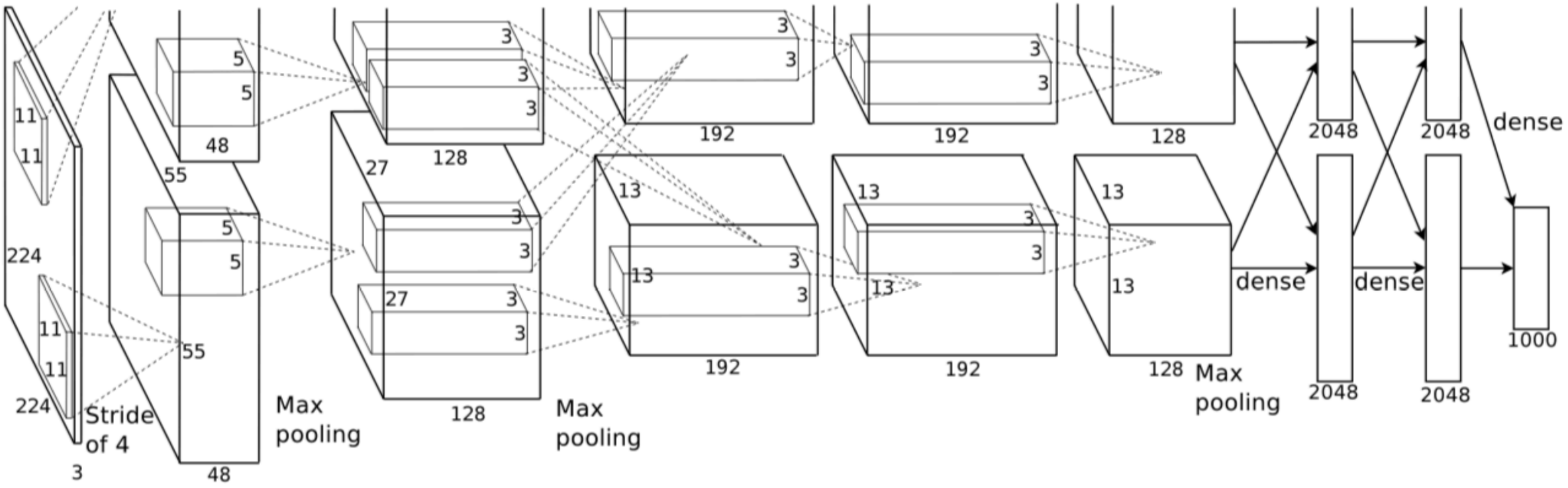


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input 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.