Deep Learning for Computer Vision

Lecture 10: An Imperfect Sampling of Image Datasets and ConvNets, and more…

Peter Belhumeur

Computer Science
Columbia University

1958 New York Times...

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

Yale Face Database A

15 subjects x 11 = 165 images (with expression + lighting + glasses/no glasses)

[Belhumeur et al. 1996]

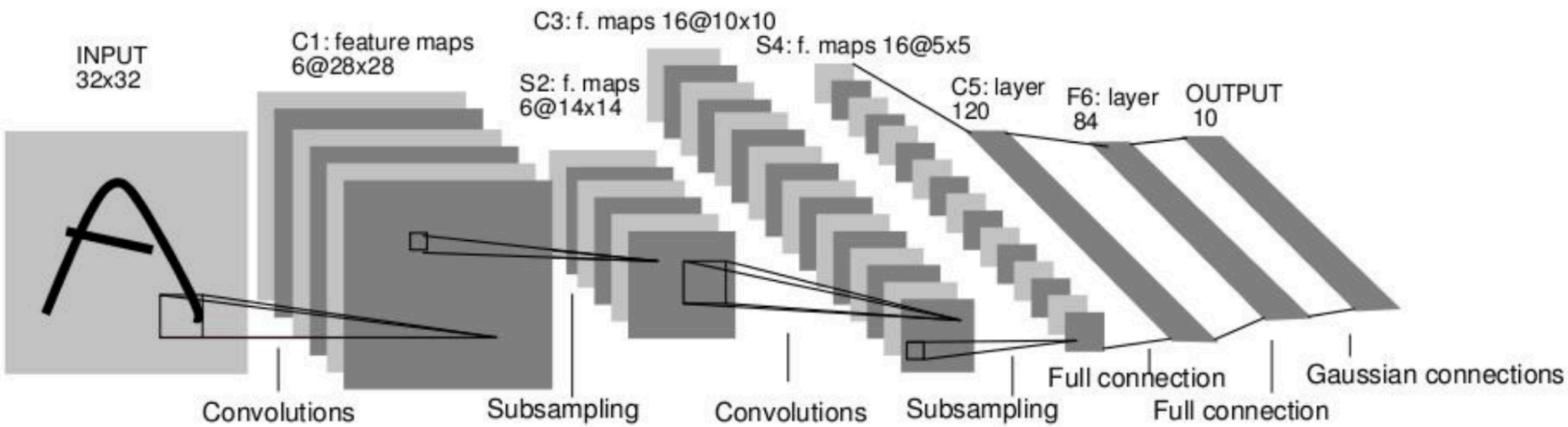
MNIST

10 digits x 7,000 instances = 70,000 images

[LeCun et al. 1998]

LeNet [LeCun et al., 1998]

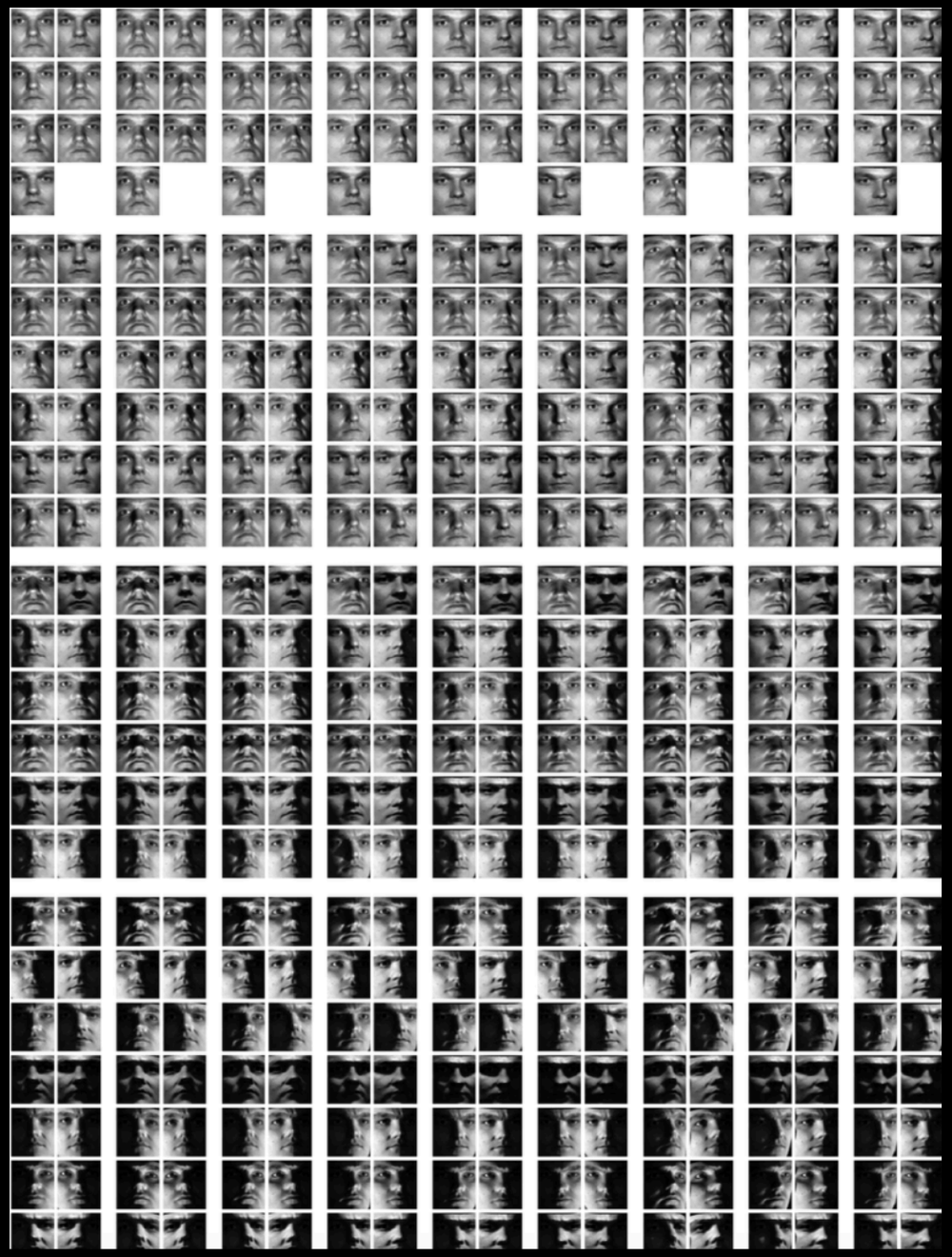
LeNet 5



Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner,
Gradient-based learning applied to document recognition, Proc. IEEE 86(11): 2278–2324, 1998.

Yale Face Database B

[Georghiades et al. 2000]



10 subjects x 9 poses x 65 lighting conditions = 5,850 images

Caltech 101

[Fei-Fei Li et al. 2004]

101 categories x 40 - 800 instances = 50,000 images

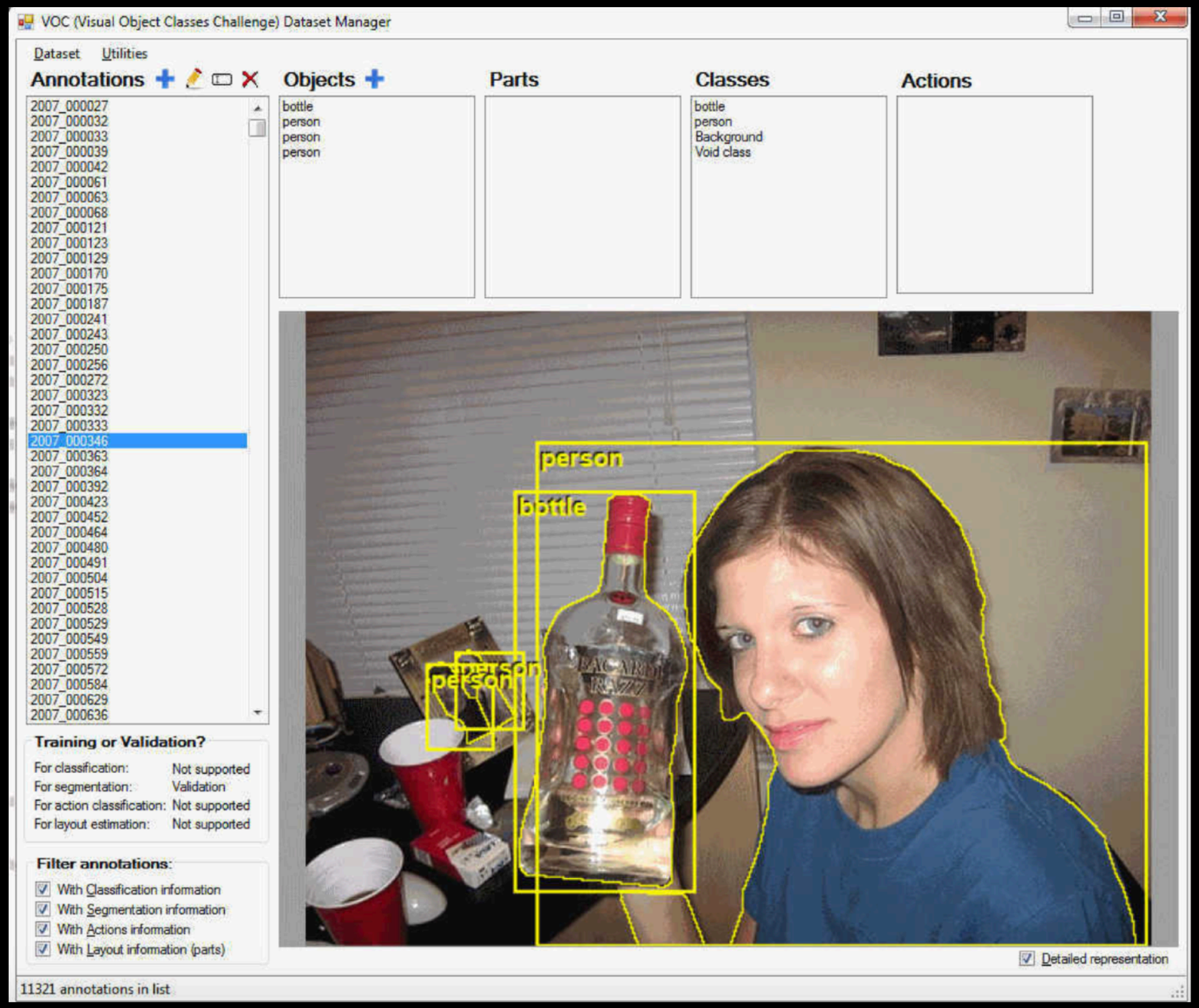
PASCAL VOC

Year	Statistics	New developments	Notes
2005	Only 4 classes: bicycles, cars, motorbikes, people. Train/validation/test: 1578 images containing 2209 annotated objects.	Two competitions: classification and detection	Images were largely taken from exising public datasets, and were not as challenging as the flickr images subsequently used. This dataset is obsolete.
2006	10 classes: bicycle, bus, car, cat, cow, dog, horse, motorbike, person, sheep. Train/validation/test: 2618 images containing 4754 annotated objects.	Images from flickr and from Microsoft Research Cambridge (MSRC) dataset	The MSRC images were easier than flickr as the photos often concentrated on the object of interest. This dataset is obsolete.
2007	20 classes: • *Person:* person • *Animal:* bird, cat, cow, dog, horse, sheep • *Vehicle:* aeroplane, bicycle, boat, bus, car, motorbike, train • *Indoor:* bottle, chair, dining table, potted plant, sofa, tv/monitor Train/validation/test: 9,963 images containing 24,640 annotated objects.	• Number of classes increased from 10 to 20 • Segmentation taster introduced • Person layout taster introduced • Truncation flag added to annotations • Evaluation measure for the classification challenge changed to Average Precision. Previously it had been ROC-AUC.	This year established the 20 classes, and these have been fixed since then. This was the final year that annotation was released for the testing data.
2008	20 classes. The data is split (as usual) around 50% train/val and 50% test. The train/val data has 4,340 images containing 10,363 annotated objects.	• Occlusion flag added to annotations • Test data annotation no longer made public. • The segmentation and person layout data sets include images from the corresponding VOC2007 sets.	
2009	20 classes. The train/val data has 7,054 images containing 17,218 ROI annotated objects and 3,211 segmentations.	• From now on the data for all tasks consists of the previous years' images augmented with new images. In earlier years an entirely new data set was released each year for the classification/detection tasks. • Augmenting allows the number of images to grow each year, and means that test results can be compared on the previous years' images. • Segmentation becomes a standard challenge (promoted from a taster)	• No difficult flags were provided for the additional images (an omission). • Test data annotation not made public.
2010	20 classes. The train/val data has 10,103 images containing 23,374 ROI annotated objects and 4,203 segmentations.	• Action Classification taster introduced. • Associated challenge on large scale classification introduced based on ImageNet. • Amazon Mechanical Turk used for early stages of the annotation.	• Method of computing AP changed. Now uses all data points rather than TREC style sampling. • Test data annotation not made public.
2011	20 classes. The train/val data has 11,530 images containing 27,450 ROI annotated objects and 5,034 segmentations.	• Action Classification taster extended to 10 classes + "other".	• Layout annotation is now not "complete": only people are annotated and some people may be unannotated.
2012	20 classes. The train/val data has 11,530 images containing 27,450 ROI annotated objects and 6,929 segmentations.	• Size of segmentation dataset substantially increased. • People in action classification dataset are additionally annotated with a reference point on the body.	• Datasets for classification, detection and person layout are the same as VOC2011.

20 categories in 11,530 images with 27,450 ROIs and 6,929 segmentations

[Everingham et al. 2005—2012]

PASCAL VOC


VOC (Visual Object Classes Challenge) Dataset Manager
Dataset
Utilities
Annotations
Objects
Parts
Classes
Actions
2007_000027
2007_000032
2007_000033
2007_000039
2007_000042
2007_000061
2007_000063
2007_000068
2007_000121
2007_000123
2007_000129
2007_000170
2007_000175
2007_000187
2007_000241
2007_000243
2007_000250
2007_000256
2007_000272
2007_000323
2007_000332
2007_000333
2007_000346
2007_000363
2007_000364
2007_000392
2007_000423
2007_000452
2007_000464
2007_000480
2007_000491
2007_000504
2007_000515
2007_000528
2007_000529
2007_000549
2007_000559
2007_000572
2007_000584
2007_000629
2007_000636
bottle
person
person
person
bottle
person
Background
Void class
Training or Validation?
For classification: Not supported
For segmentation: Validation
For action classification: Not supported
For layout estimation: Not supported
Filter annotations:
With Classification information
With Segmentation information
With Actions information
With Layout information (parts)
person
bottle
person
person
Detailed representation
11321 annotations in list

Object Detection: PASCAL VOC mean Average Precision (mAP)

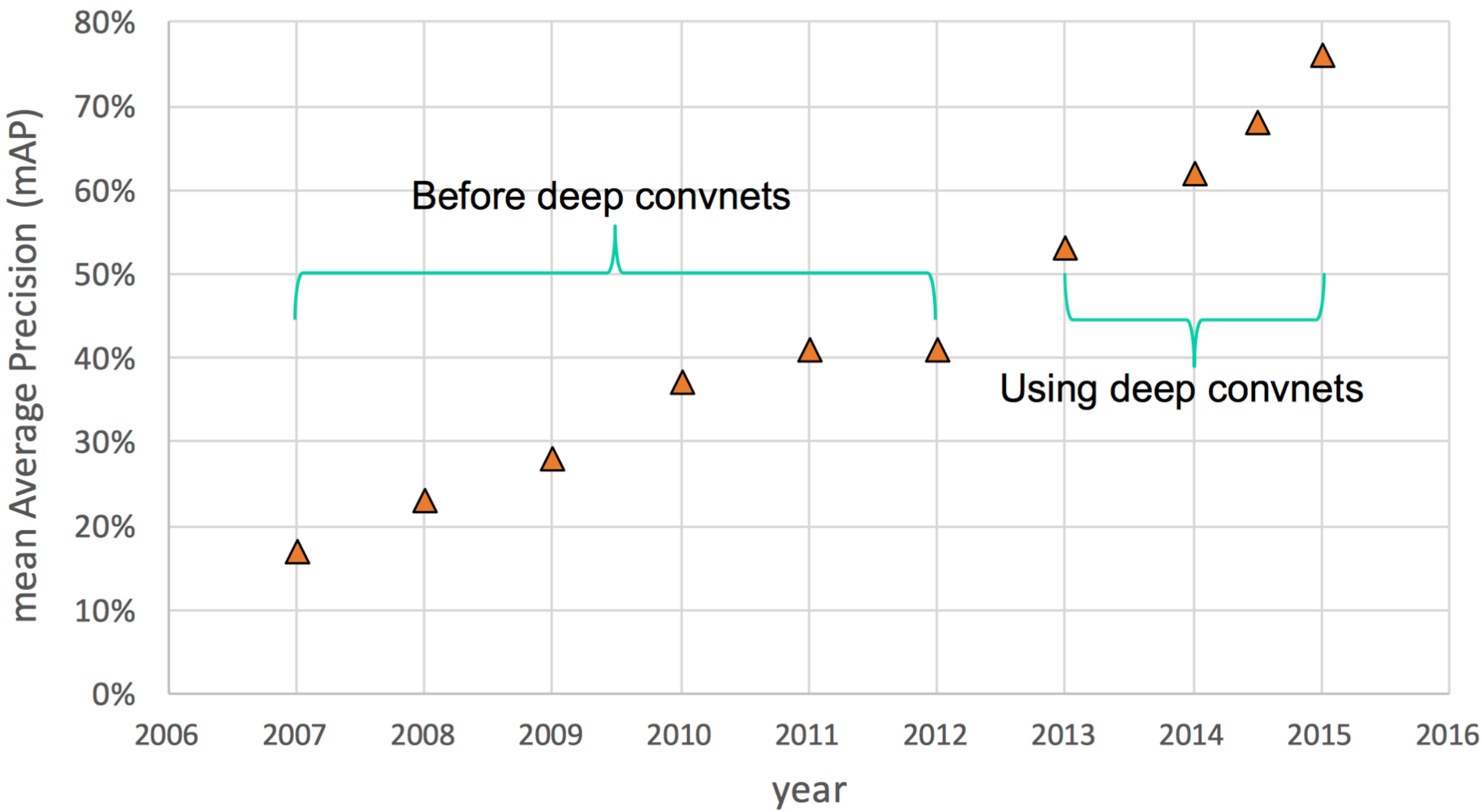


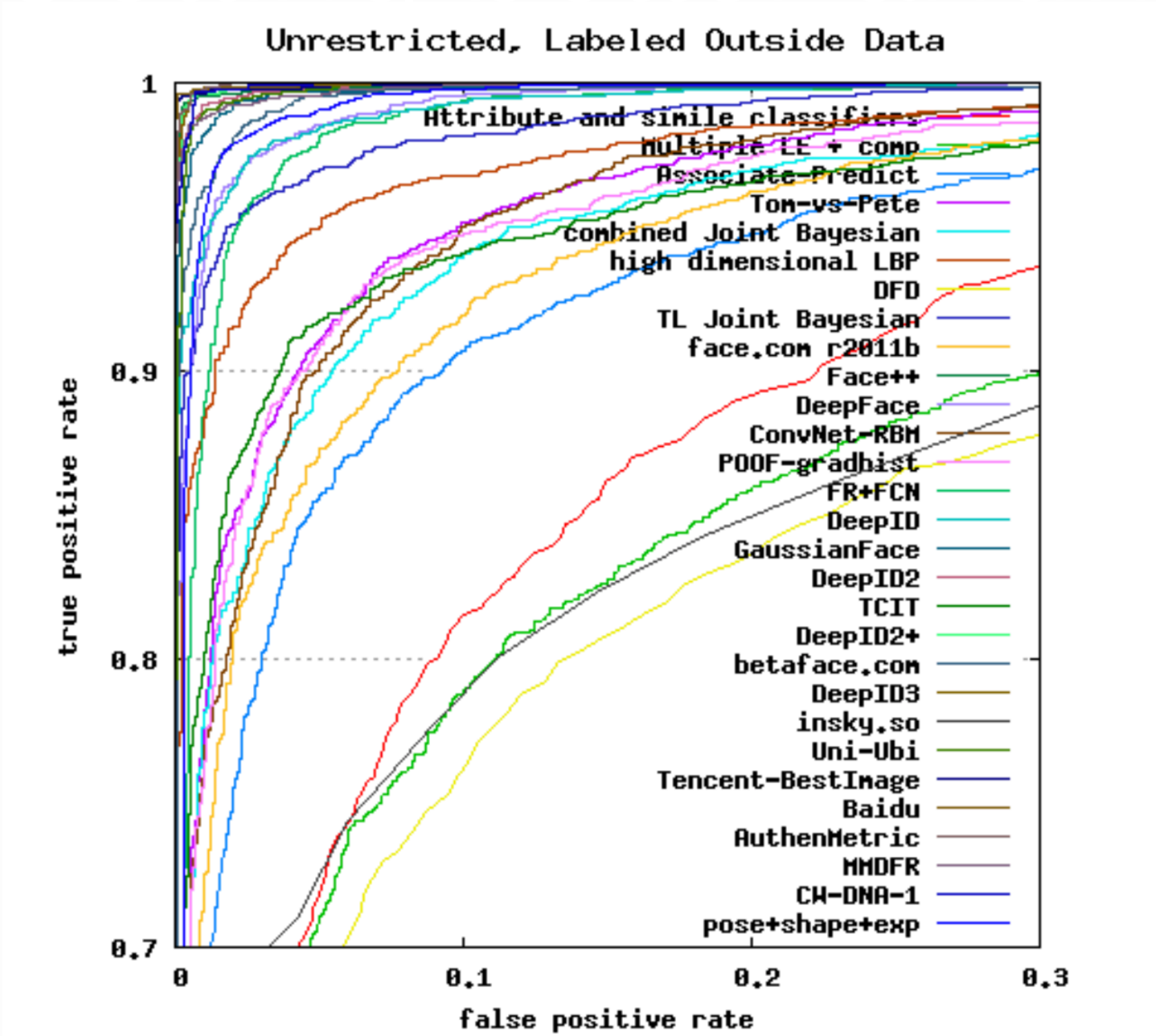
Figure source: Ross Girshick

Labeled Faces in the Wild (LFW)

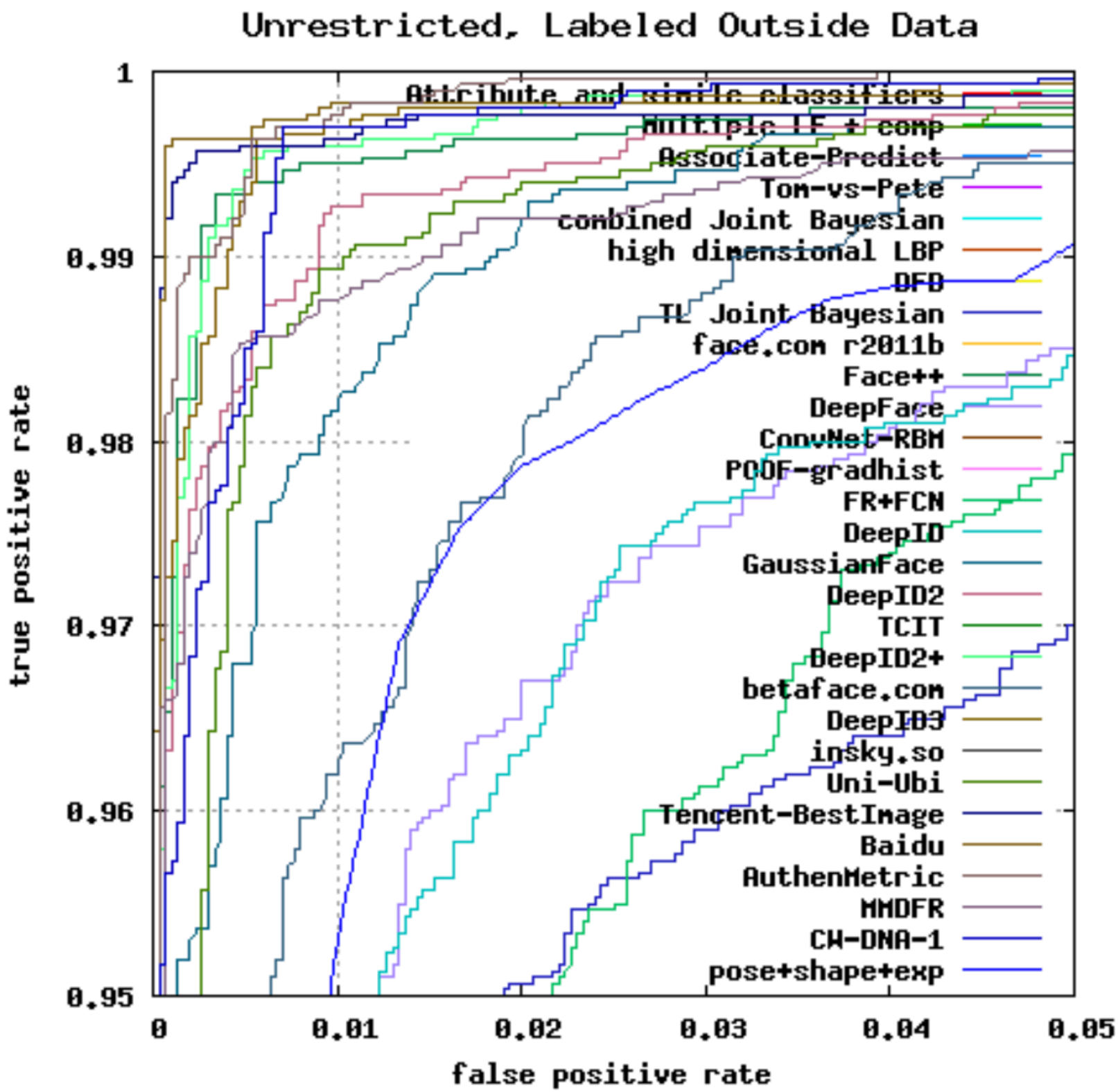
5,749 subjects x 1+ instances = 13,233 images taken in the “wild”

[Huang et al. 2007]

LFW Face Verification

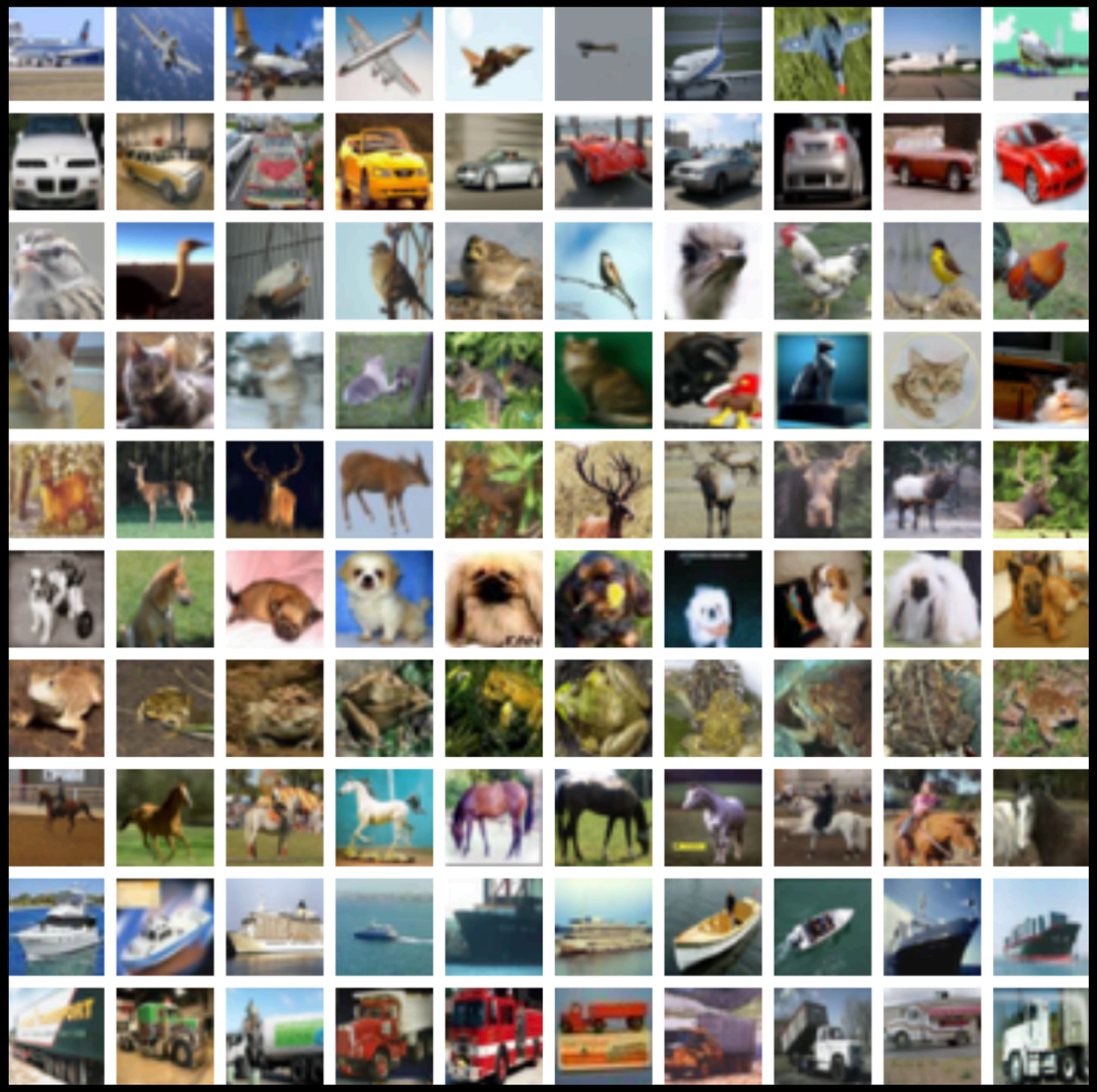


LFW Face Verification



CIFAR-10

[Krizhevsky 2009]

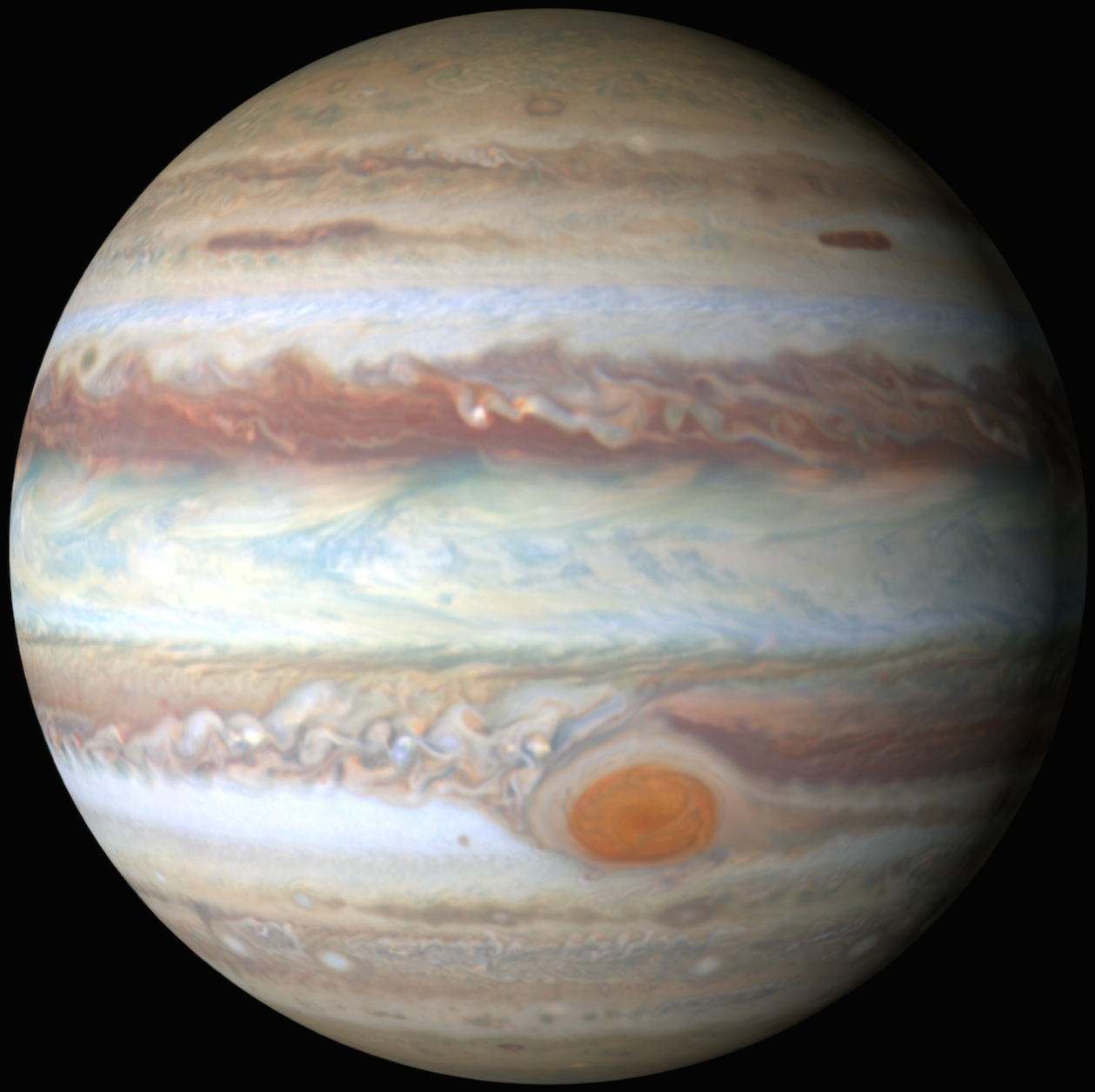


10 categories x 6,000 instances = 60,000 images (32x32x3)

The same simple CNN made 3 ways

- The CIFAR-10 Dataset has 60,000 32x32x3 thumbnail images of objects in 10 categories.
- We recreate a simple CNN to classify these using three separate wrappers/frameworks for building deep nets:

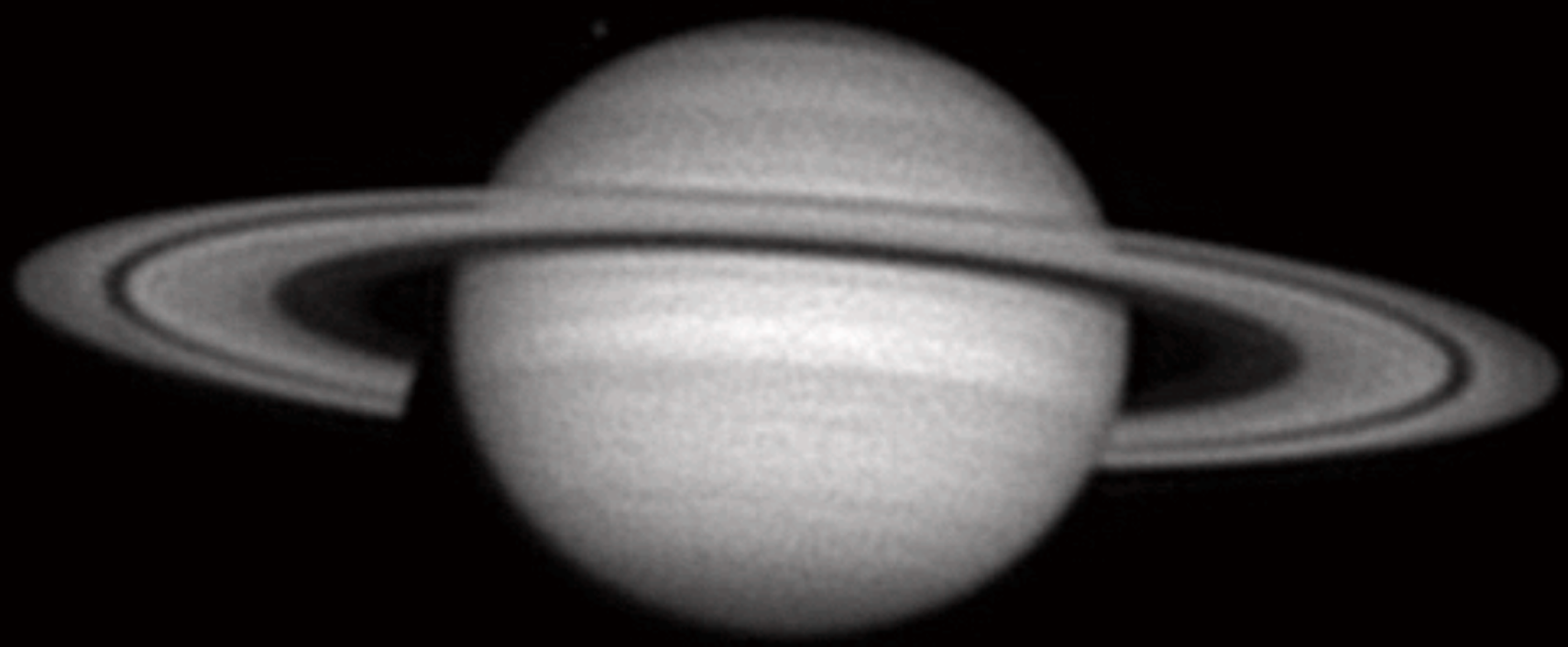
1. Keras: A python based wrapper that wraps both Theano and Tensorflow. This is recommended by many!
2. TFLearn: A simple tensor flow wrapper that may lack flexibility but is easy to use, especially for more timid coders.
3. Tensorflow: Google's API for making computational graphs, deep nets, and the like. It is getting better everyday, and becoming the standard for building deep nets in production. Great support for multi-GPU training, analysis, mobile platforms. And lots of pre-trained models for public use!



ImageNet

[Deng et al. 2009]

20,000+ categories x ~1000 instances = 14,000,000+ images



http://image-net.org

Classification: ImageNet Challenge top-5 error

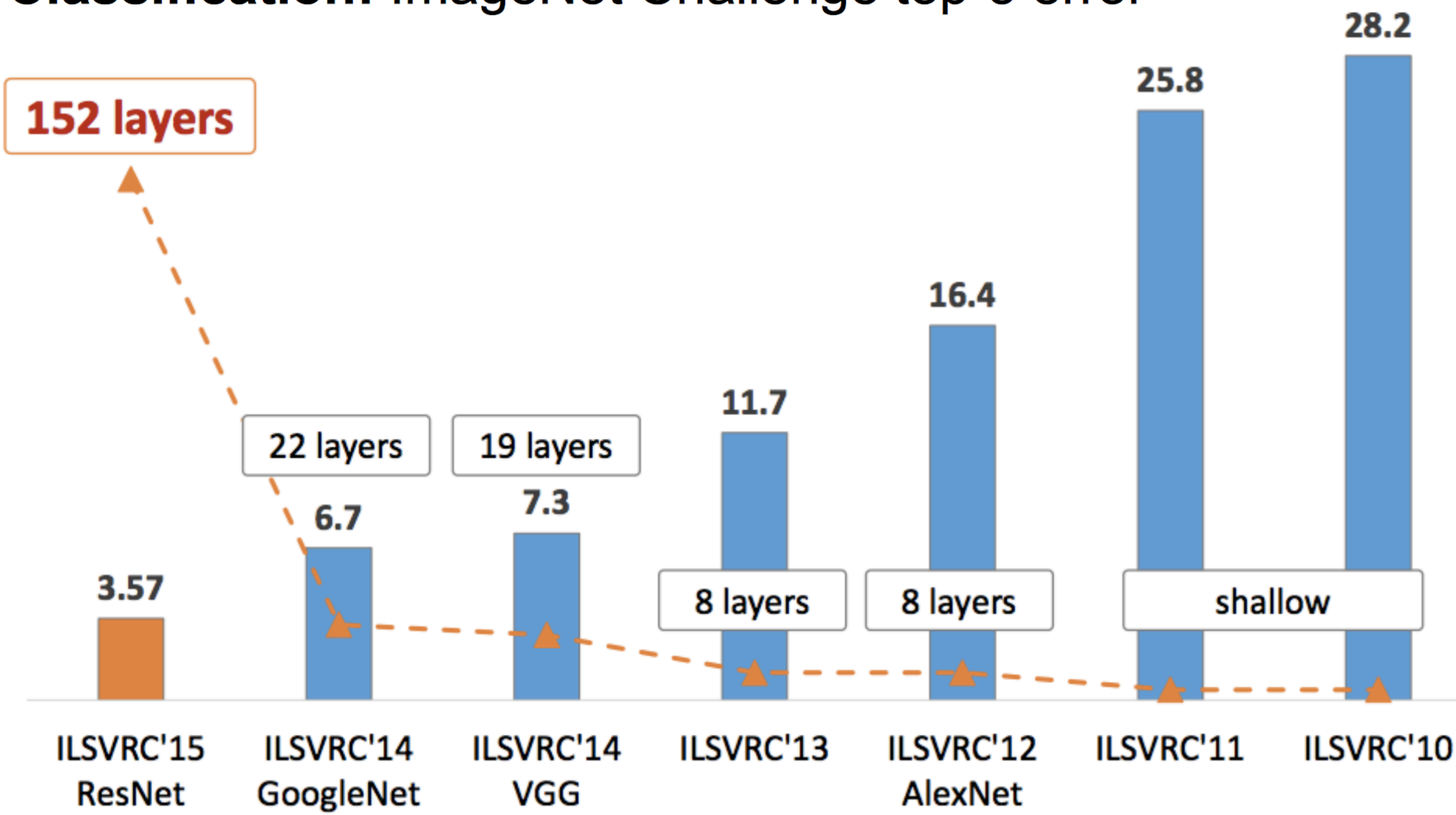
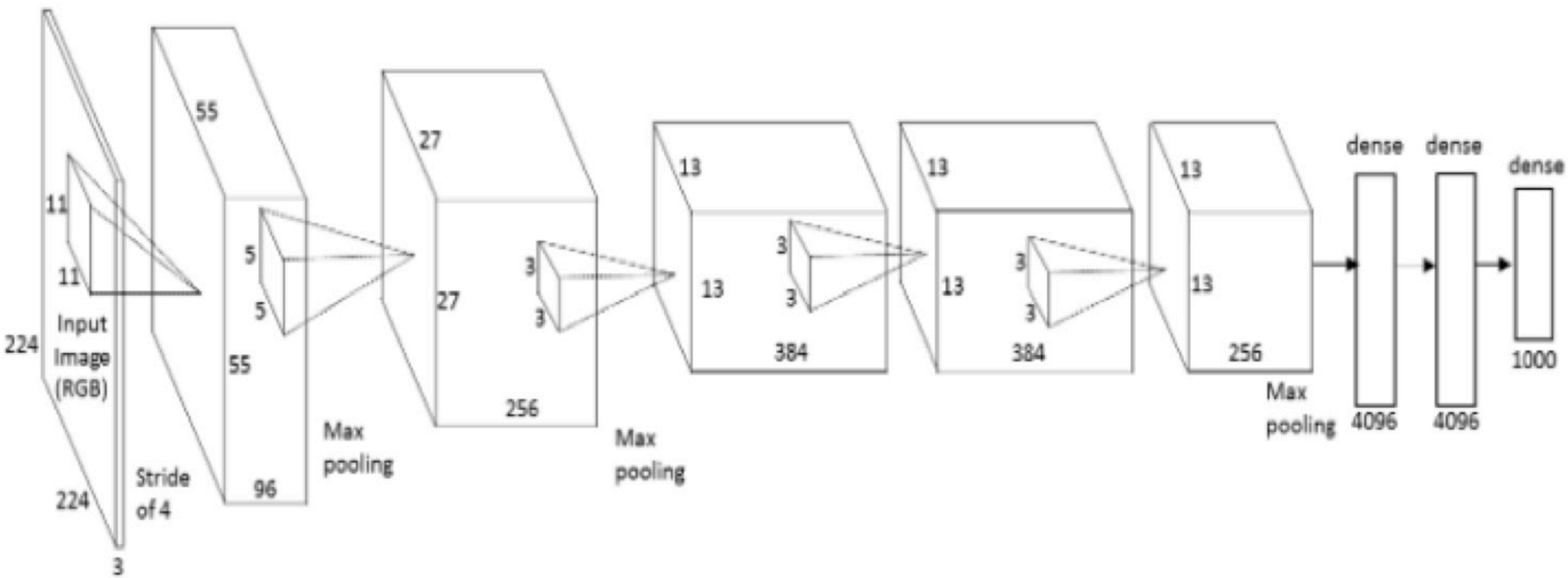


Figure source: Kaiming He

AlexNet [Krizhevsky et al., 2012]



Classification: ImageNet Challenge top-5 error

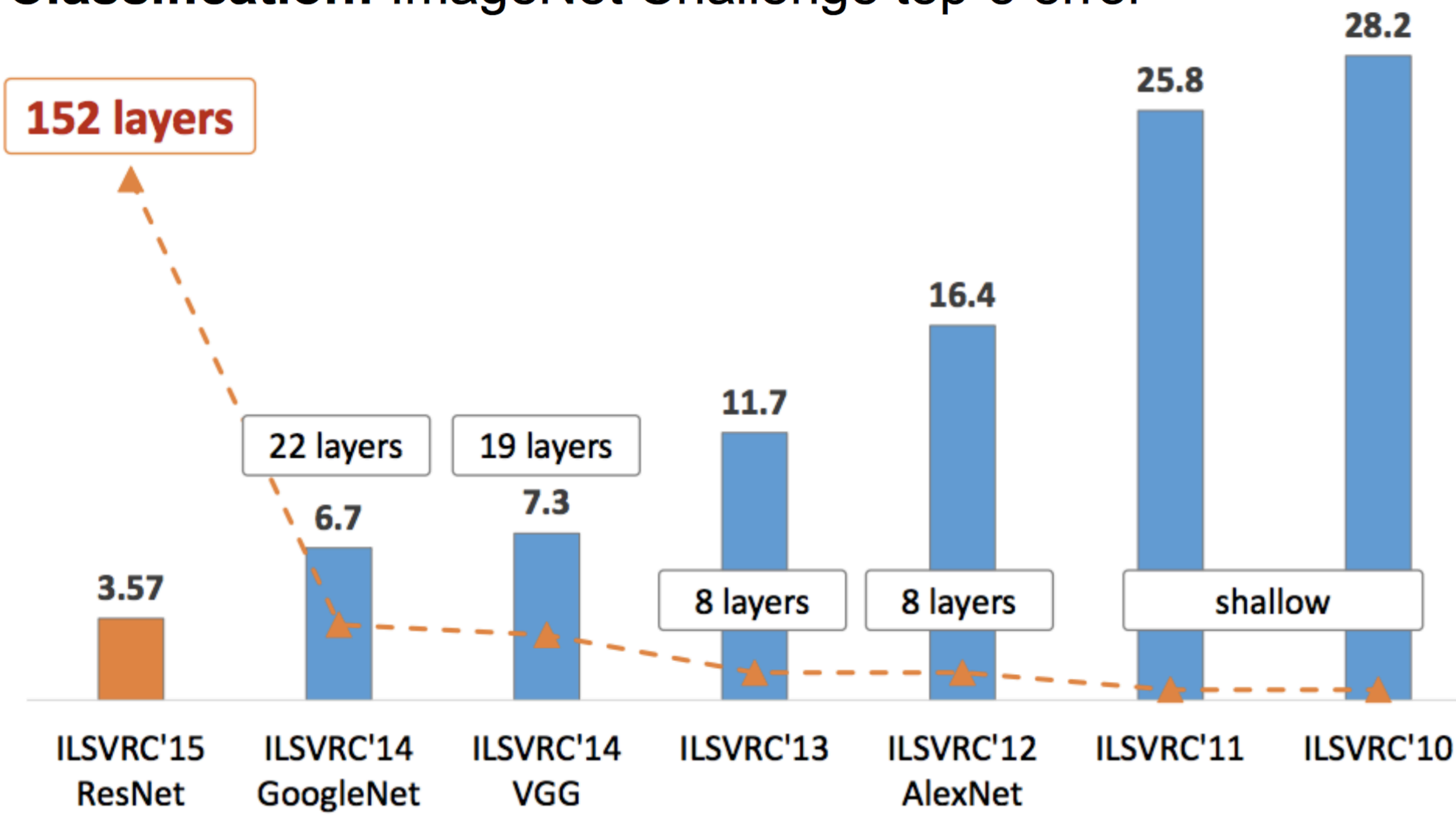
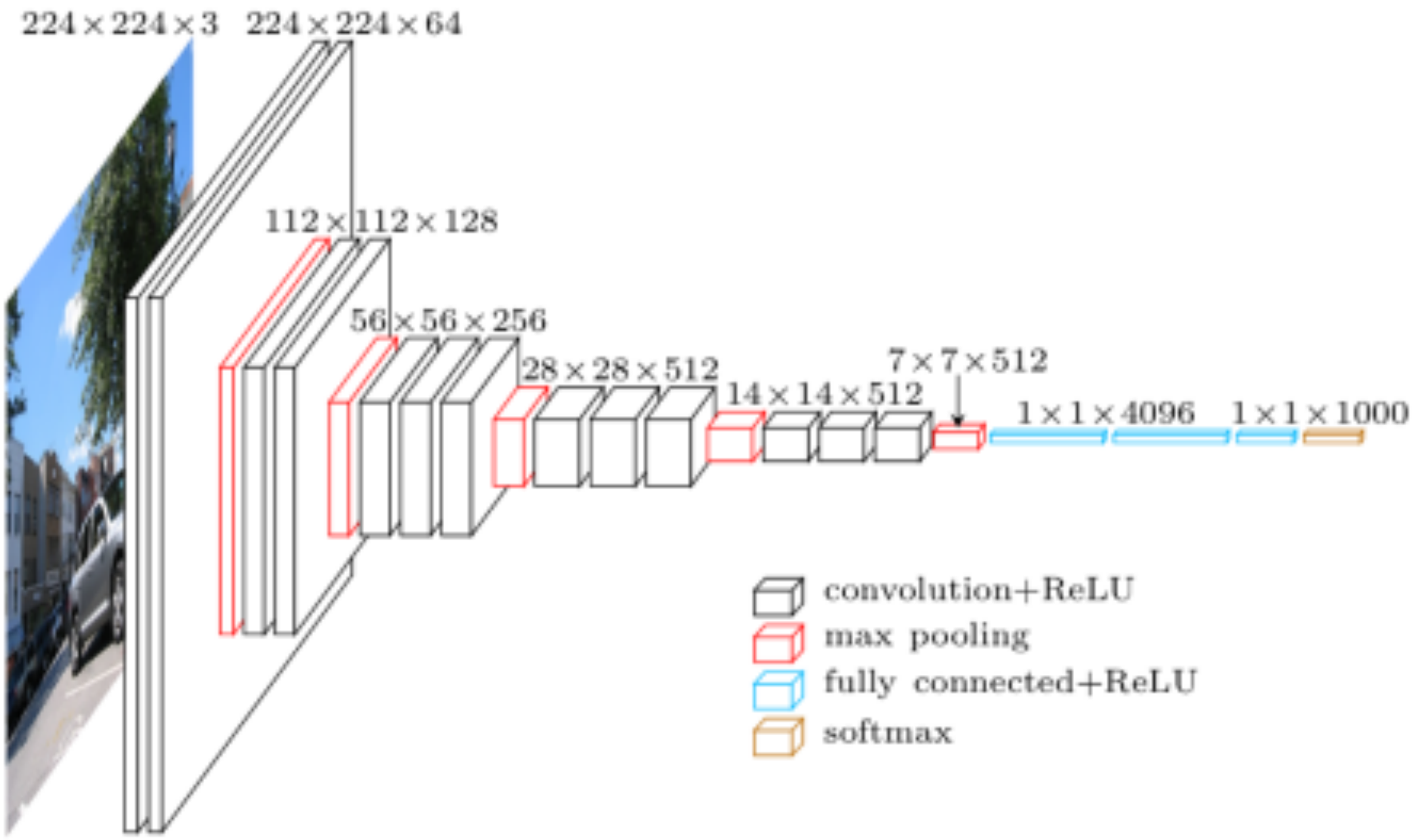


Figure source: Kaiming He

VGG16 [Simonyan and Zisserman, 2015]



Classification: ImageNet Challenge top-5 error

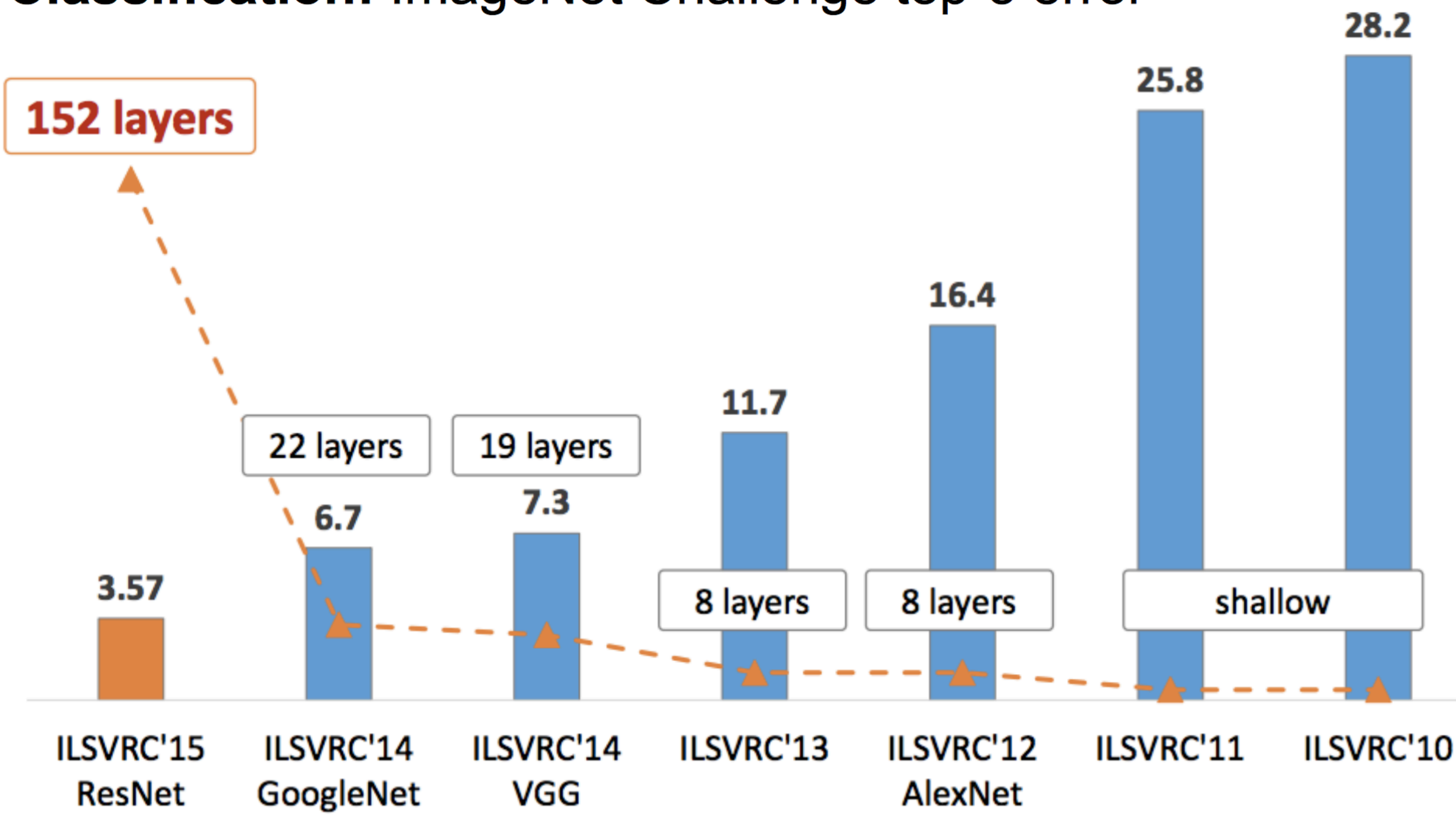
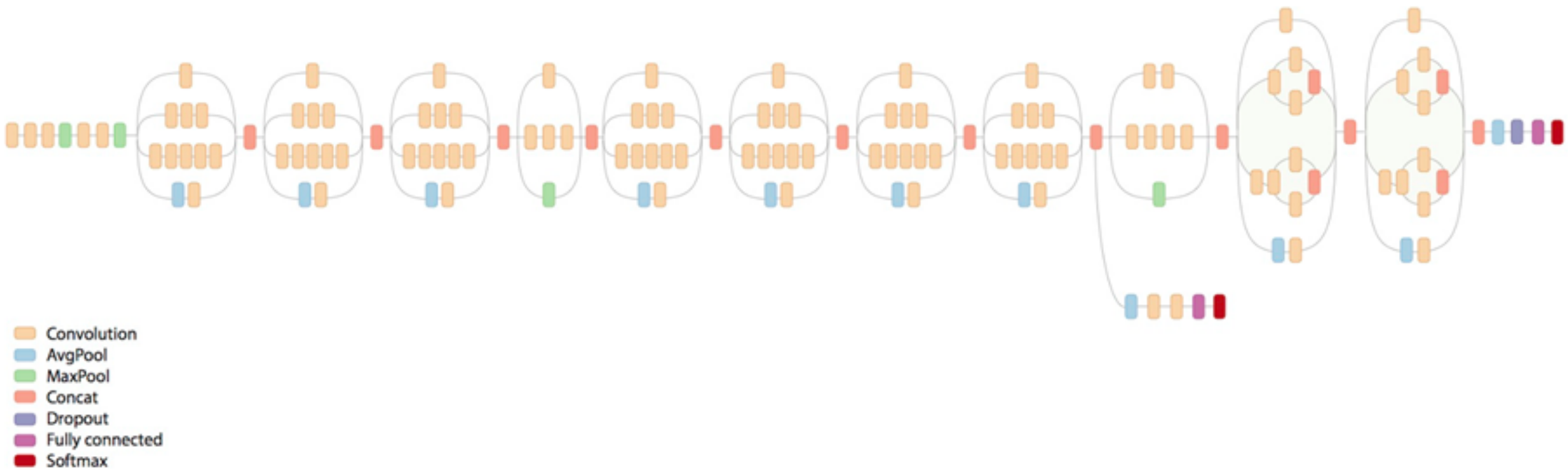


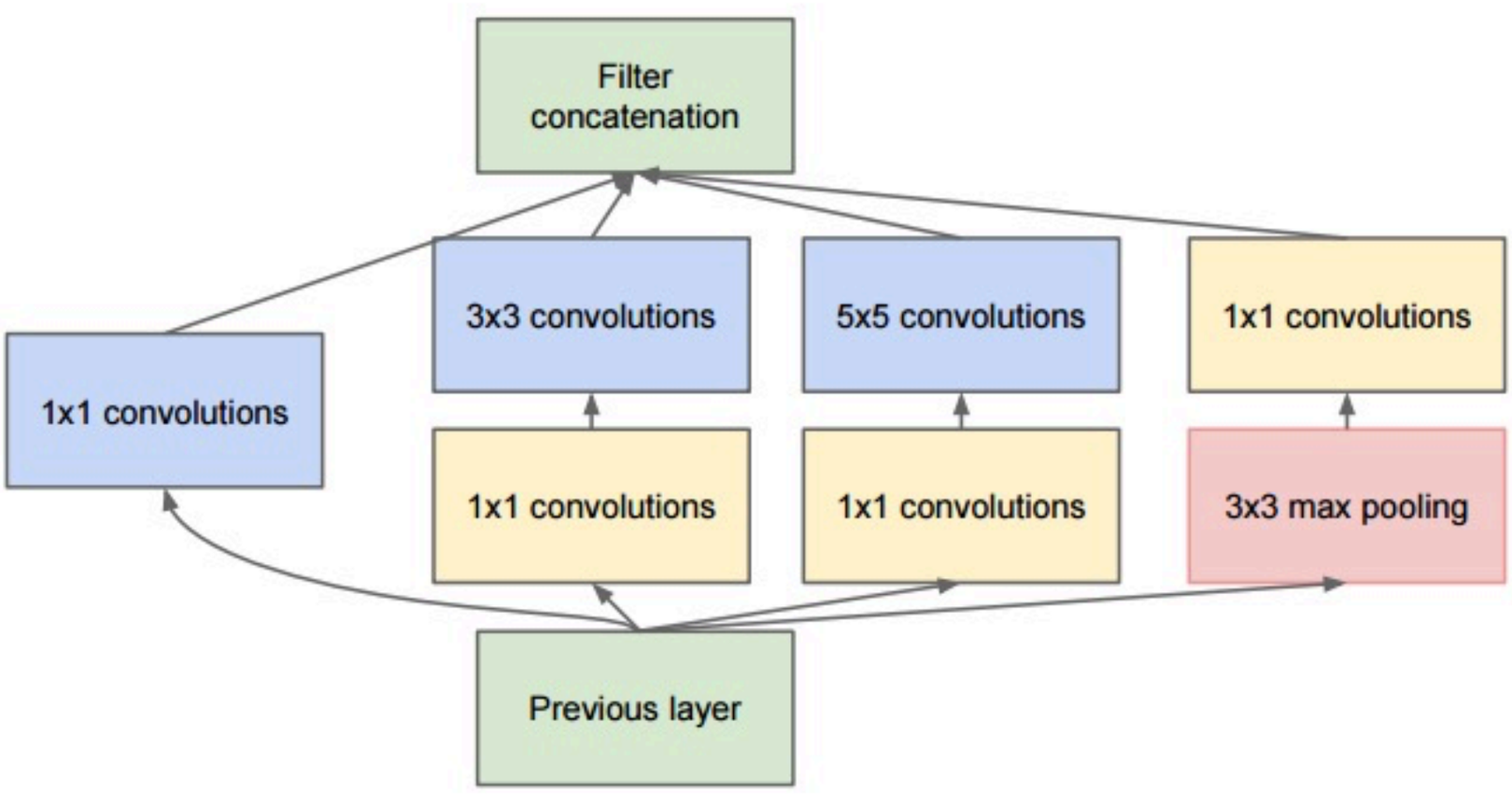
Figure source: Kaiming He

GoogLeNet [Szegedy et al., 2014]



Another view of GoogLeNet's architecture.

Inception Module



Classification: ImageNet Challenge top-5 error

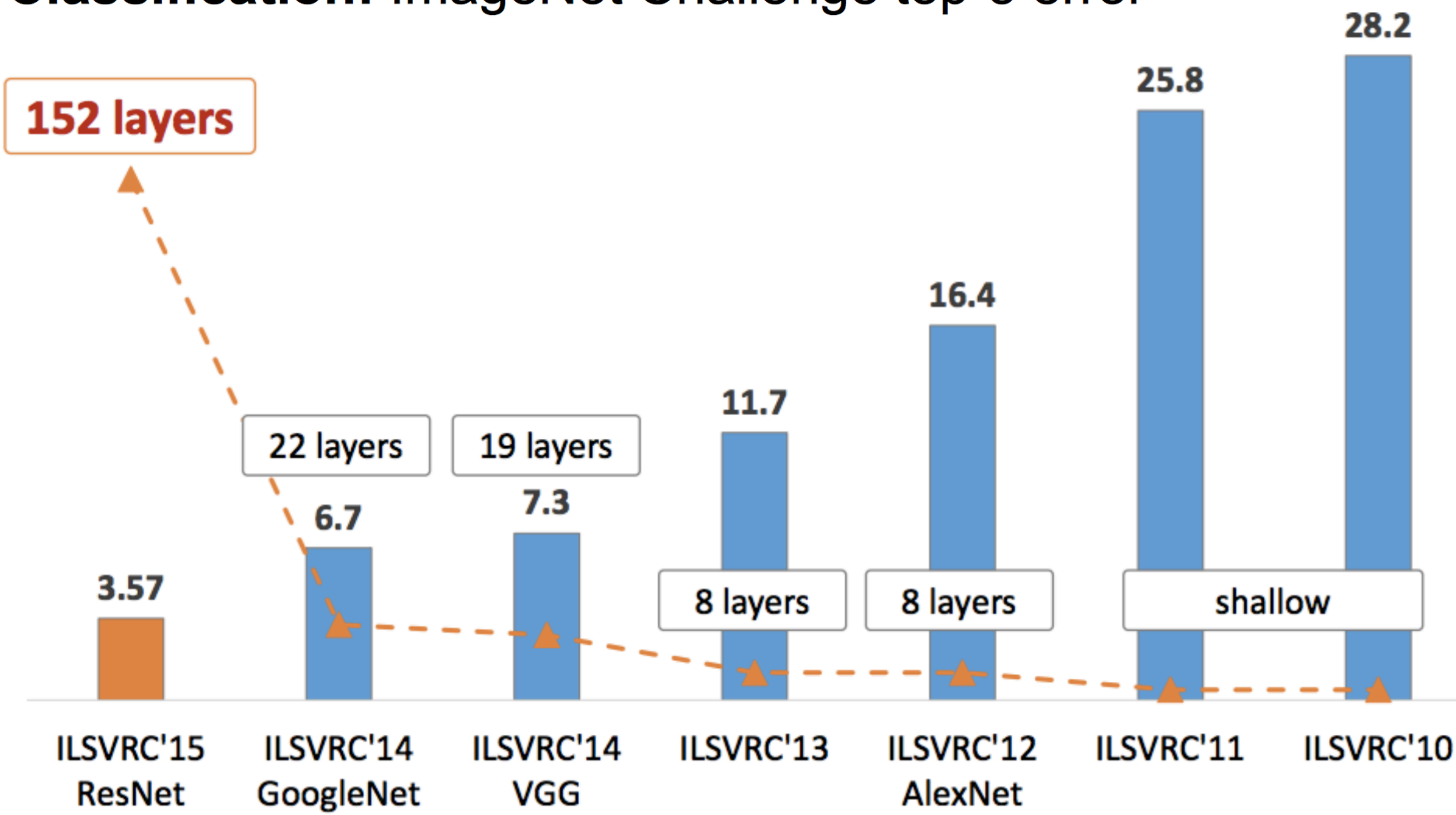
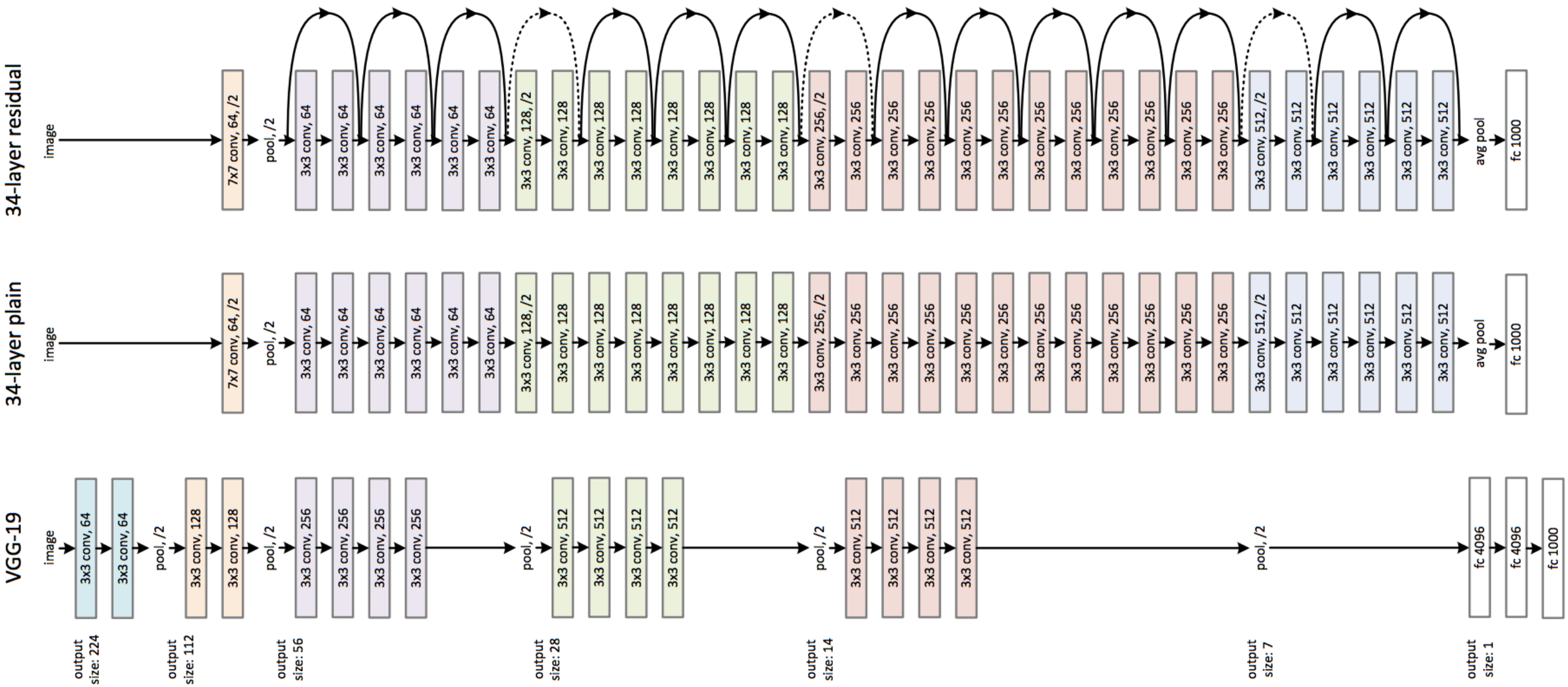
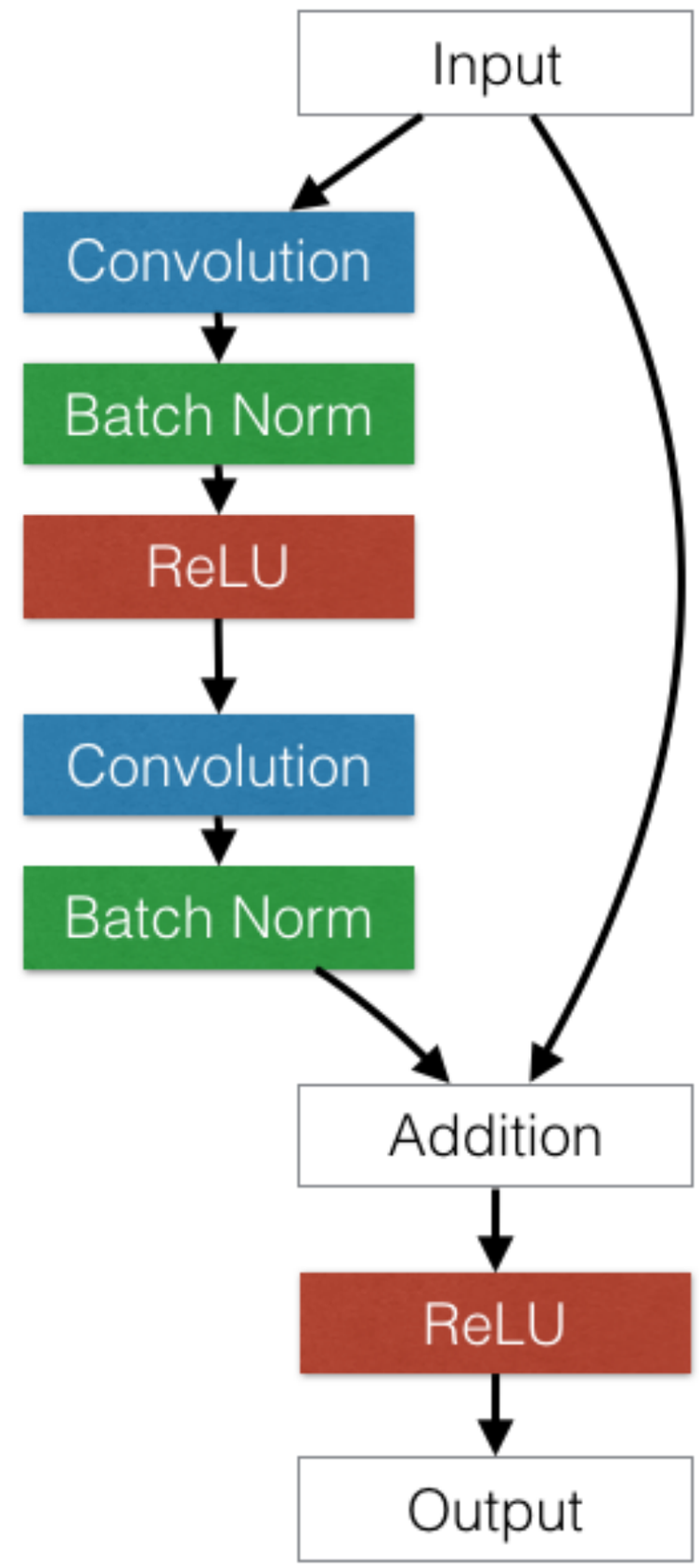


Figure source: Kaiming He

ResNet [He et al., 2016]



ResNet Module



Classification: ImageNet Challenge top-5 error

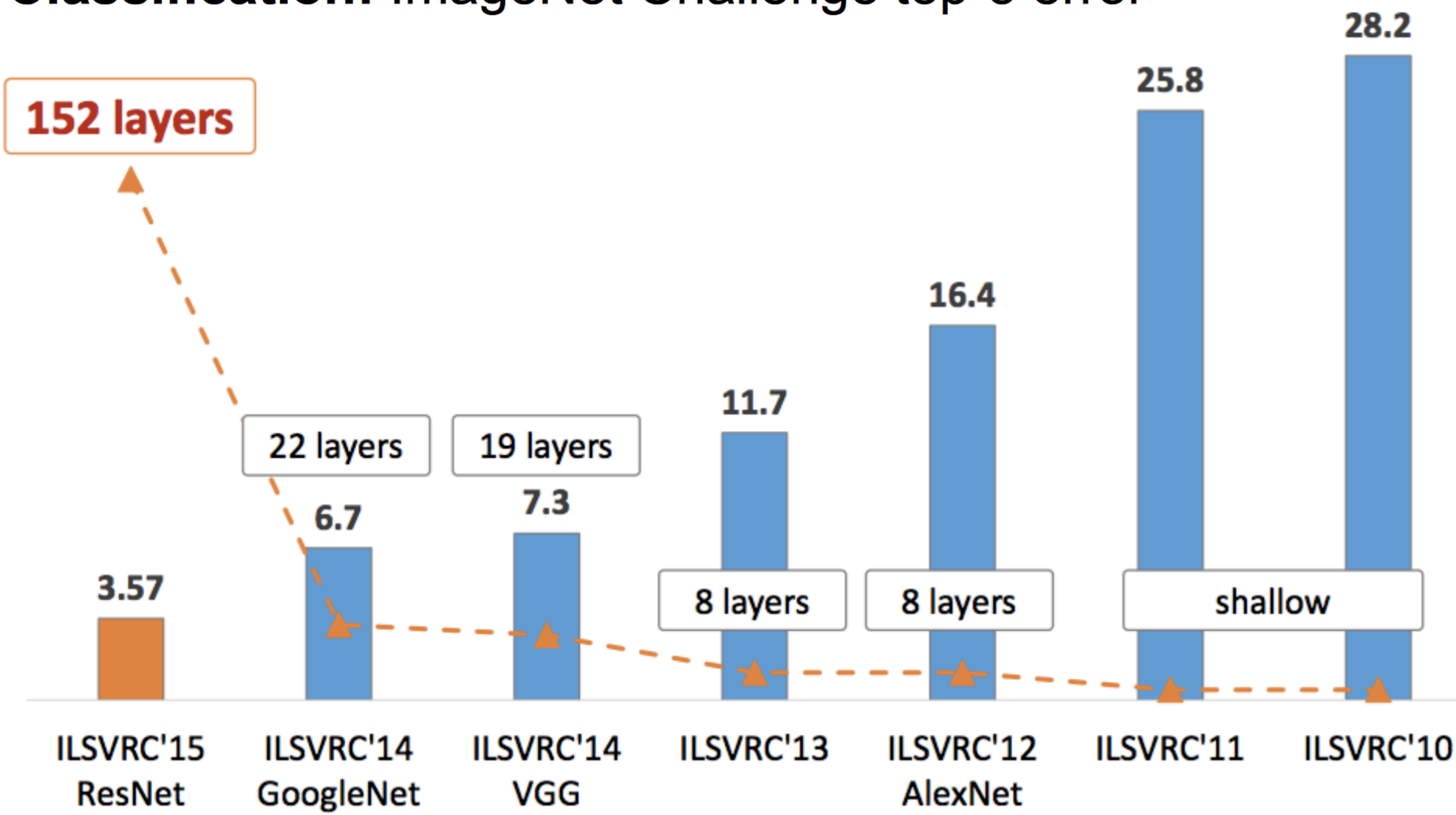
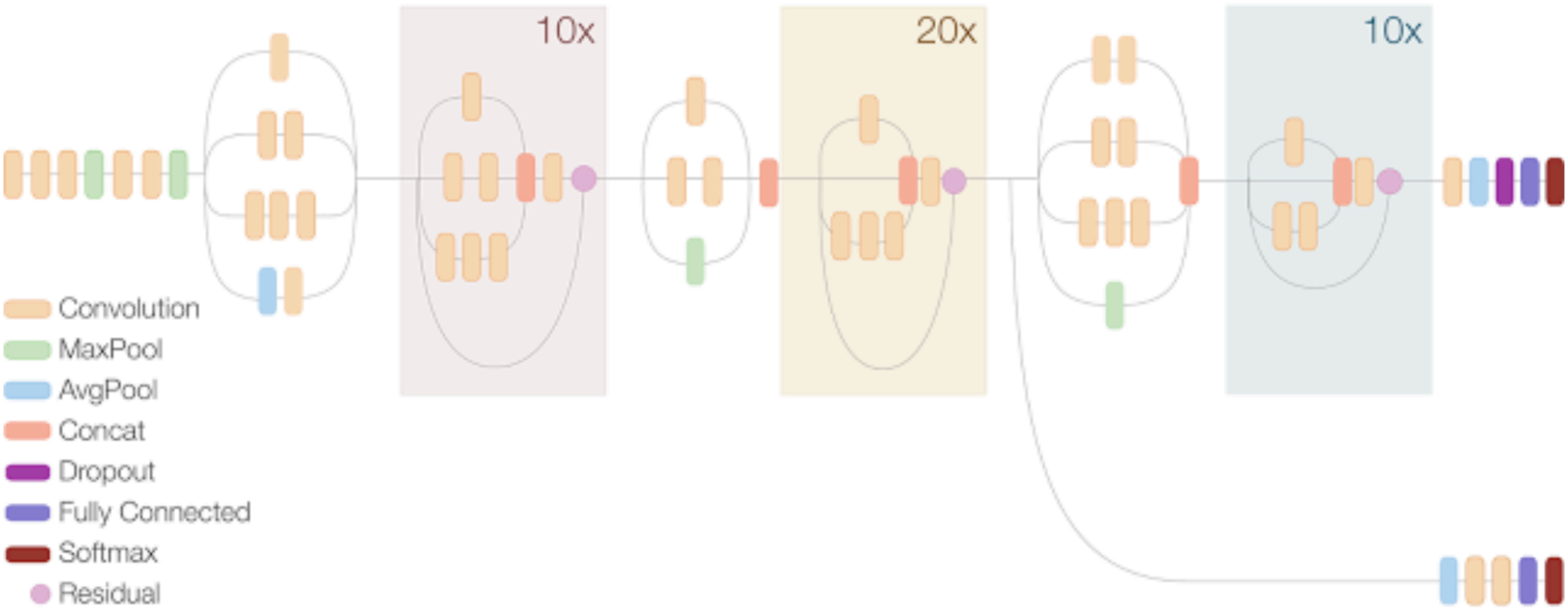


Figure source: Kaiming He

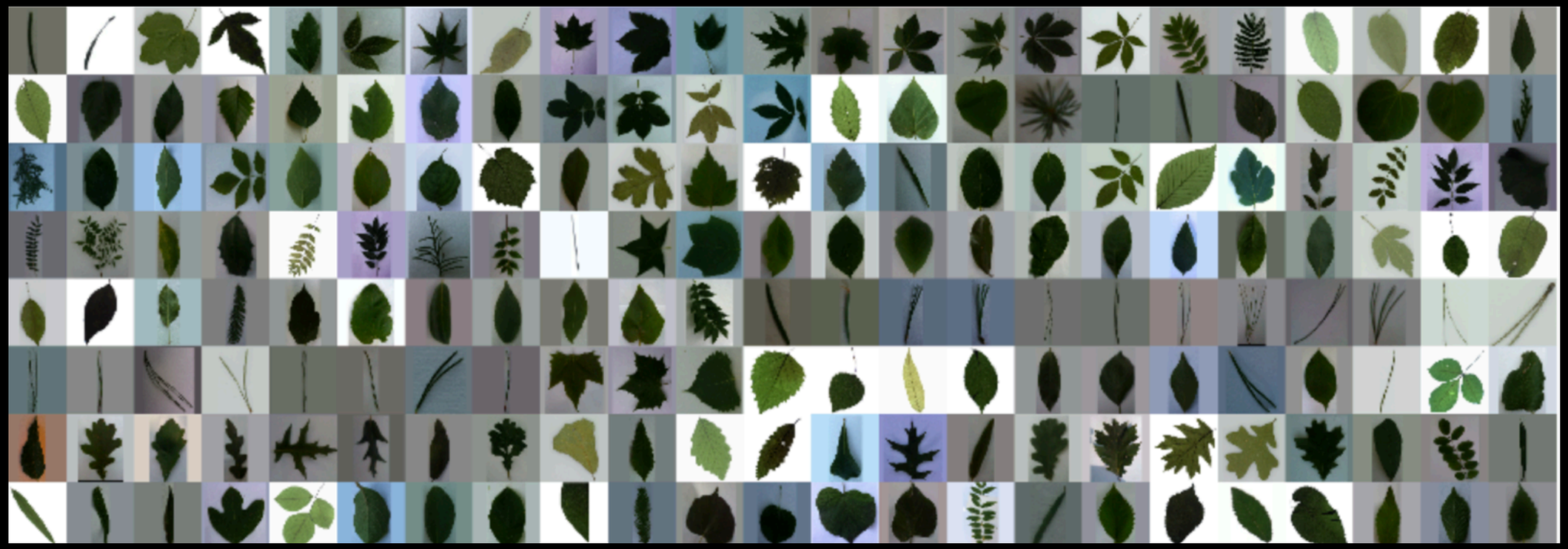
Inception Resnet V2 Network

Compressed View



Schematic diagram of Inception-ResNet-v2

Leafsnap Dataset



184 tree species x 25+ instances = 5,000+ iPhone images

[Kumar et al. 2010]

leafsnap

Home | Species | Dataset | About

Leafsnap: An Electronic Field Guide

Leafsnap is a series of electronic field guides being developed by researchers from Columbia University, the University of Maryland, and the Smithsonian Institution. The free mobile apps use visual recognition software to help identify tree species from photographs of their leaves. They contain beautiful high-resolution images of leaves, flowers, fruits, petioles, seeds and bark.

The original Leafsnap currently includes trees found in the Northeastern United States and Canada, and will soon grow to include the trees of the entire continental United States. The high-resolution images in the original app were created by the conservation organization. Finding Species.

This website shows the tree species included in Leafsnap, the collections of its users, and the team of research volunteers working to produce it.

The Leafsnap UK app includes trees from the United Kingdom with species information and imagery provided by the Natural History Museum in London. More information can be found on the Natural History Museum website.

The City College of New York developed and tested curricular materials that use the Leafsnap app to help middle school students notice, group, and contextualize street trees in the patterns of evolution. Curricular guide and other educational materials are available from here.

Leafsnap

Free for iPhone: Available on the App Store **and iPad:** Available on the iPad App Store

Leafsnap UK

Free for iPhone Available on the App Store

Flower of the Japanese Pagoda Tree

You Tube

The New York Times

CUB-200

200 bird species x ~30 instances = 6,033 images

[Welinder et al. 2010]

bird snap

USA Eastern Western Backyard Local

Birds of the United States

Bird Wheel | Bird List | Bird Lab | About

Sort by: Tree of Life | Alphabetical

Visual Recognition | Text Search

Order: *Anseriformes*
Family: *Anatidae*
Subfamily: *Anatinae*
Genus: *Anas*
Species: *A. rubripes*

by Bill Evans

★ - visually similar
- arriving
- departing
- migrating through

Some recordings include other species

View original image