TTIC 31230, Fundamentals of Deep Learning, Winter 2018

David McAllester

Introduction and Historical Notes

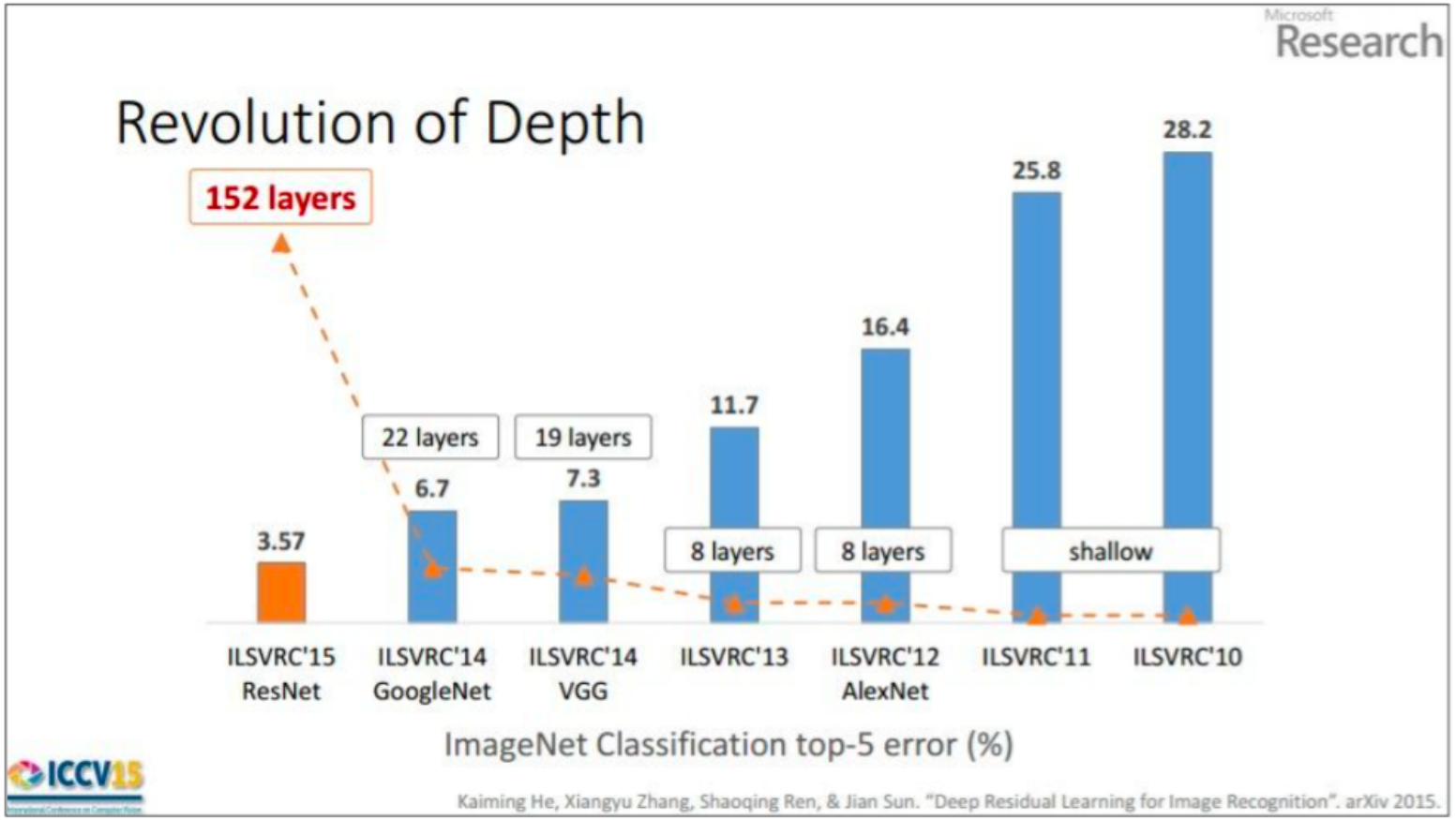
Deep Learning: A Moore's Law of AI?

PASCAL VOC Object Detection

	bicycle	bus	car	motorbike	person	20 class average
2007	36.9	23.2	34.6	27.6	21.3	17.1
2008	42.0	23.2	32.0	38.6	42.0	22.9
2009	46.8	43.8	37.2	42.0	41.5	27.9
2010	54.3	54.2	49.1	51.6	47.5	36.8
2011	58.1	57.6	54.4	58.3	51.6	40.9
2012	54.5	57.1	49.3	59.4	46.1	41.1
2013 DNN	56.3	51.4	48.7	59.8	44.4	43.2
2014 DNN						63.8
2015 ResNet	88.4	86.3	87.8	89.6	90.9	83.8
2016 ResNet						86

Imagenet Classification

1000 kinds of objects.



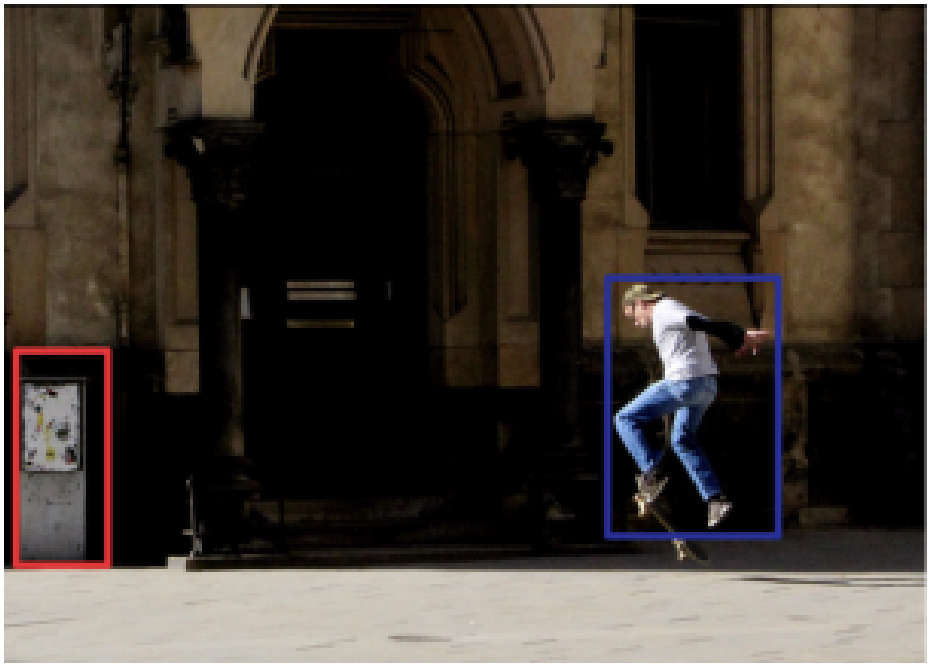
(slide from Kaiming He's recent presentation)

2016 error rate is 3.0% 2017 error rate is 2.25%

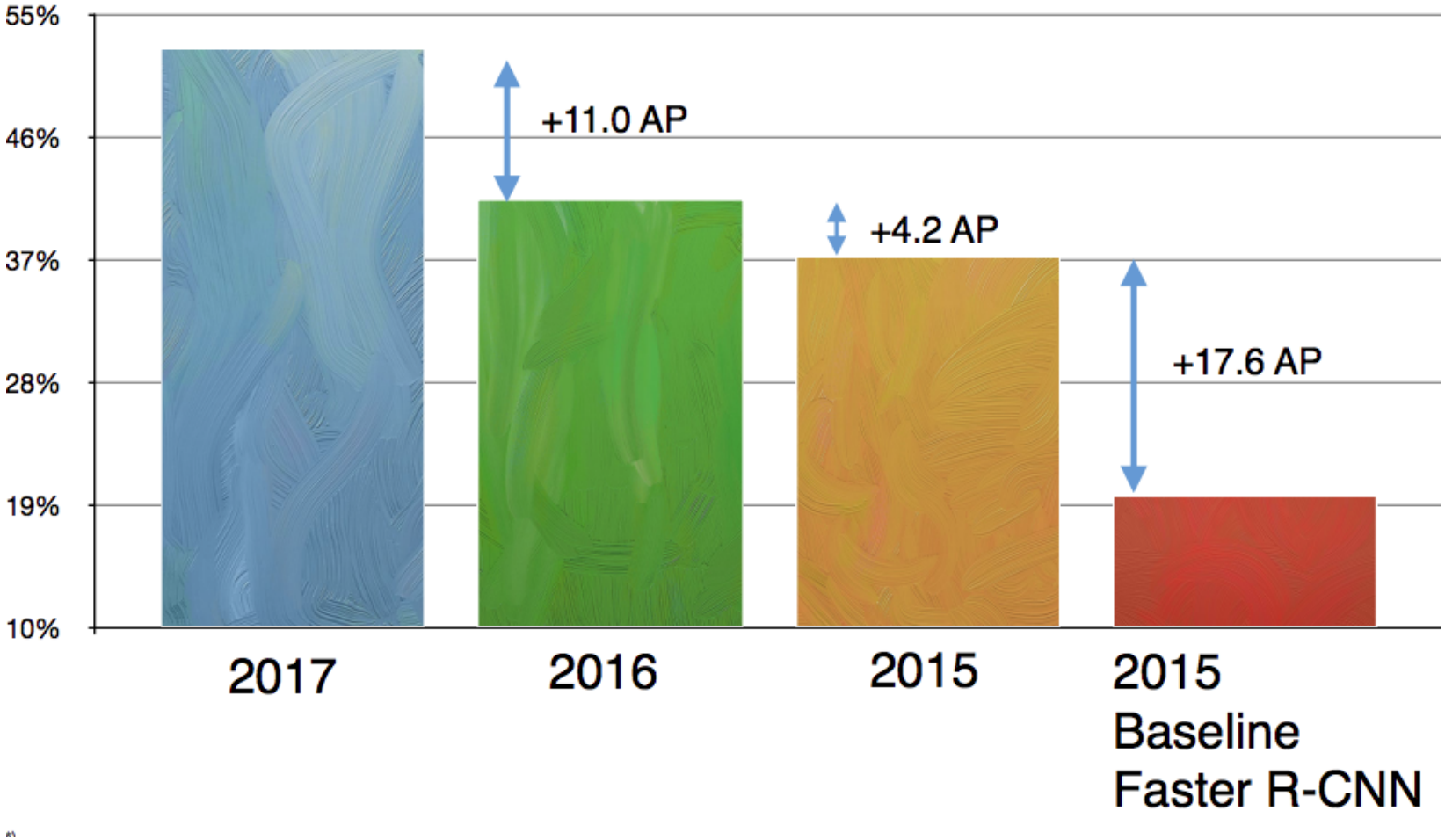
Coco Challenge 17

Detection

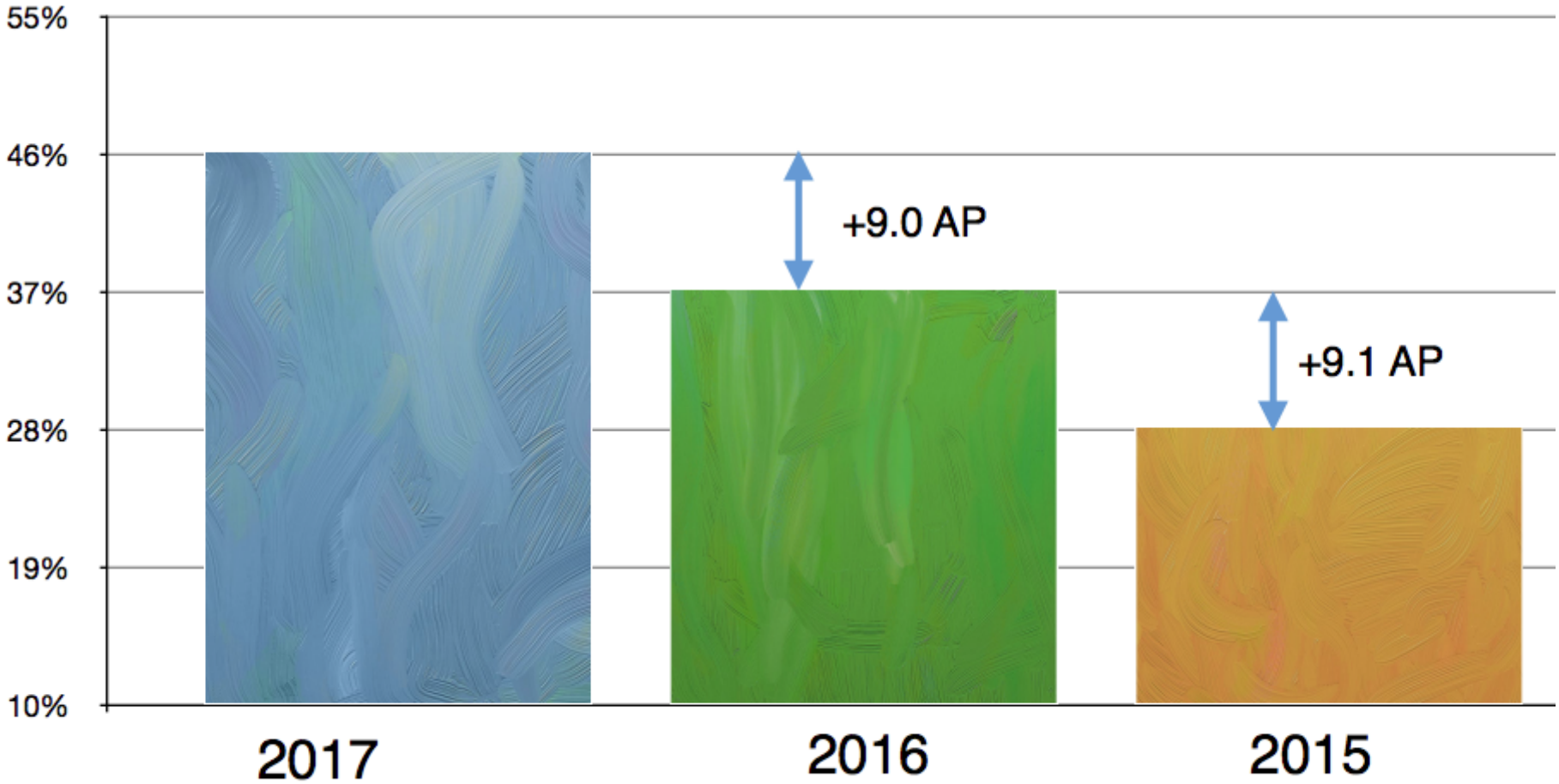
Segmentation



Detection

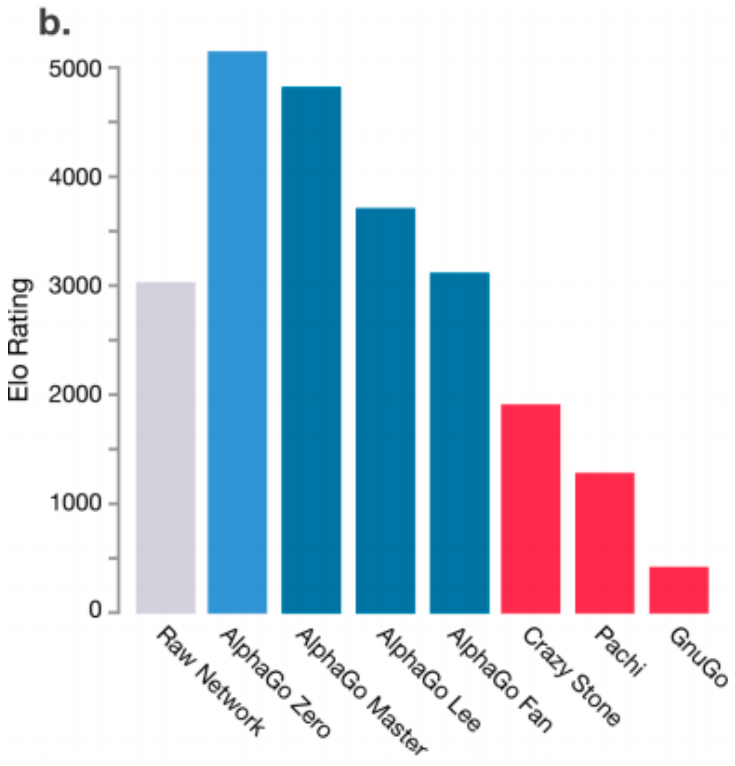

55%
46%
37%
28%
19%
10%
+11.0 AP
+4.2 AP
+17.6 AP
2017
2016
2015
2015
Baseline
Faster R-CNN

Segmentation




nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO

AlphaZero

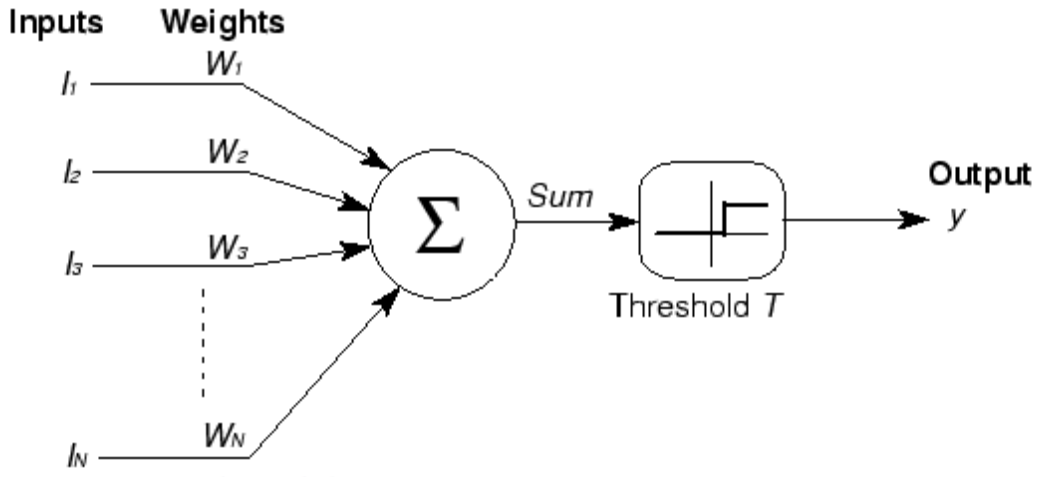

b.
Elo Rating
5000
4000
3000
2000
1000
0
Raw Network
AlphaGo Zero
AlphaGo Master
AlphaGo Lee
AlphaGo Fan
Crazy Stone
Pachi
GnuGo

The Deep Revolution is Everywhere

- Computer Vision
- Speech Recognition
- Machine Translation
- Computer Games
- Information Retrieval (Google Search)
- Computational Chemistry
- ...

Some History of Deep Learning

McCullock and Pitts 1943 — introduced the linear threshold "neuron".



Rosenblatt 1962 — Applied a "Hebbian" learning rule.

Novikoff 1962 — proved the perceptron convergence theorem.

Deep Winter I: late 60s through early 80s

Robinson 1965 — introduces resolution theorem proving.

Minsky 1969 — wins Turing Award for "promoting AI".

McCarthy and Hayes 1968 — introduced the situation calculus.

Minsky and Papert 1969 — published the book *Perceptrons*. They proved that many properties of images could not be determined by (single layer) perceptrons. Caused a decline of activity in neural network research.

McCarthy, 1971 — wins Turing Award.

Minsky 1974 — wrote "A Framework for Representing Knowledge".

McCarthy 1980 — introduces "non-monotonic logic".

Deep Resurgence I, late 80s

Fukushima 1980 — introduced the neocognitron (a form of CNN)

Hinton and Sejnowski 1985 — introduce the Boltzman machine

Rummelhart, Hinton and Williams 1986 — demonstrated empirical success with backpropagation (itself dating back to 1961).

Deep Winter II: Late 90s' and 00's

Valiant 1984 — introduces the formal definition of PAC learnability. Credited with starting learning theory as a branch of computer science. Turing Award, 2010.

Pearl 1995 — publishes *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Credited with driving the "statistical revolution" in AI. Turing Award, 2011.

Convex optimization and convex relaxations (the marginal polytope of a graphical model)

Nonparametric Bayesian inference (Dirichlet processes).

Submodular optimization

Deep Learning in Winter II

Schmidhuber et al. 1997 — introduces LSTMs

LeCun 1998 — introduces convolutional neural networks (CNNs) (LeNet).

Bengio 2003 — introduced neural language modeling

Deep Learning Explodes in 2012

Alexnet dominates the 2012 Imagenet challenge.

Google speech recognition converts to deep learning.

Both developments were driven by Hinton's group at the University of Toronto.

The Four Horsemen of Deep Learning

 Geoff Hinton, 70, H index = 136

 Yann LeCun, 57, H index = 98

 Yoshua Bengio, 53, H index = 105

 Juergen Schmidhuber, 54, H index = 80

END