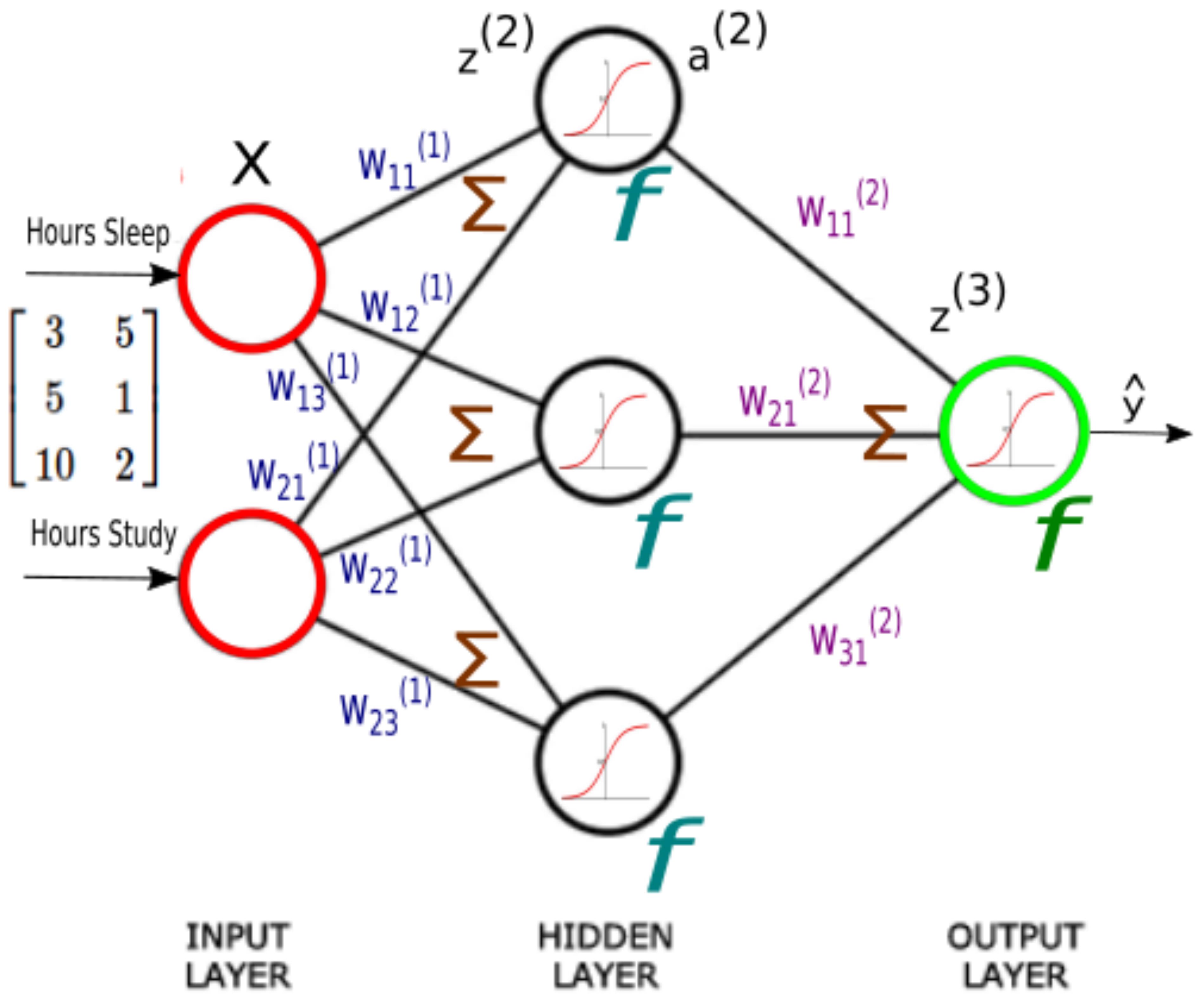
Deep Learning

What is it good for?

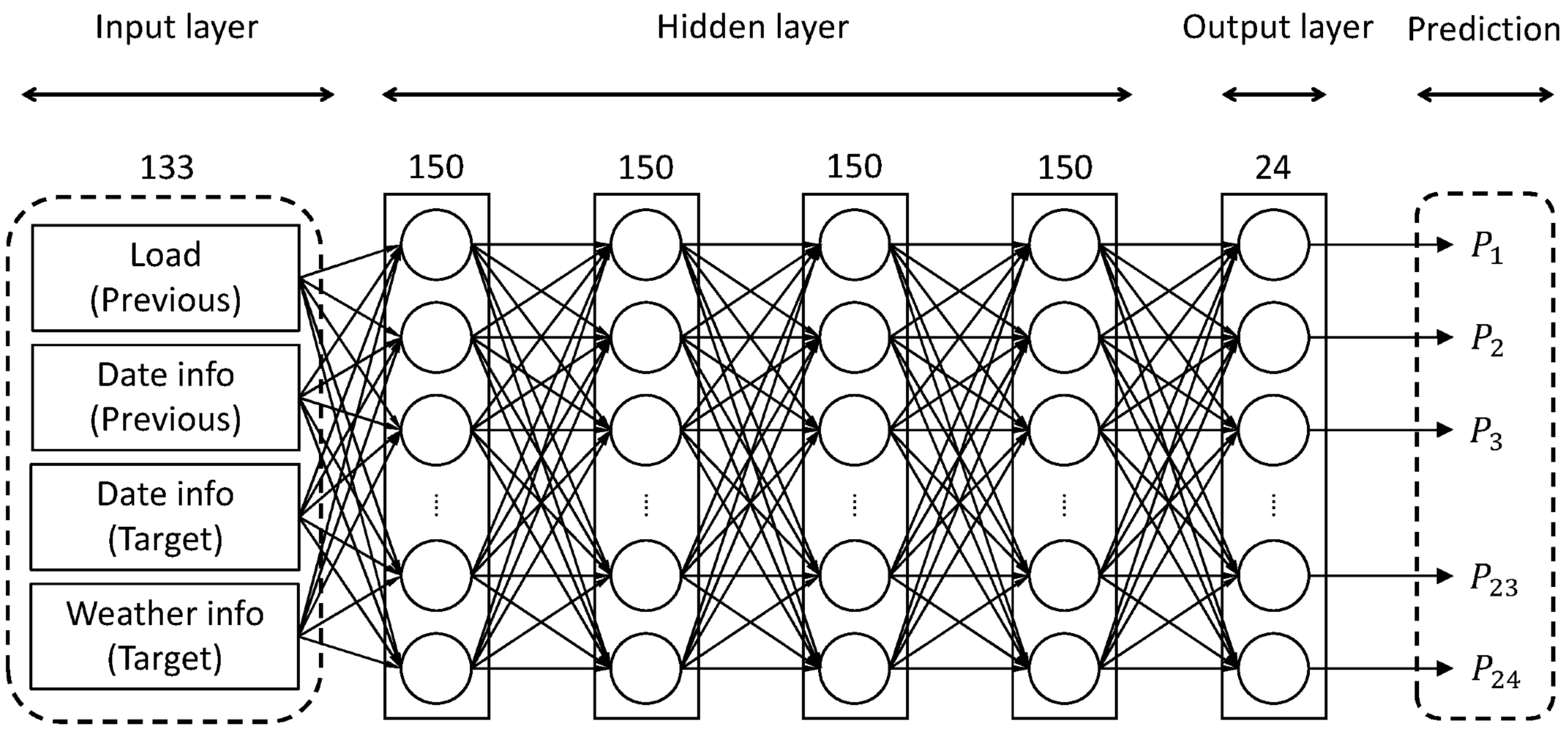
A short history of NN

- All information taken from “Wikipedia/Artificial Neural Networks”
- **1943 McCulloch and Pitts**, the first credible mathematical model of a Neuron.
- **1940’s Hebb rule**: a model of learning through neural plasticity.
- **1969 “Perceptron”** by Minsky and Pappert: showed that single NN cannot compute parity. NN research slowed until computers achieved far greater processing power.
- **1975 Webros invents BackPropagation** - renews interest in computational NN.
- **1986 Parallel Distributed Processing (PDP), Connectomics,** are made popular by **Rumelhard and Mclleland**.
- **1989 Yann Le-Cun** (Currently Vice President, Chief AI Scientist in FaceBook) uses Convolutional NN for zip code recognition.
- **1990’s Vanishing Gradient problem**
- **2006 Geoff Hinton** (Currently in Google and U. of Toronto) Suggests using hand-designed Deep NN.
- **2010 GPUs** (invented and mostly used for games) make it possible to learn from millions of examples.

A Neural Network



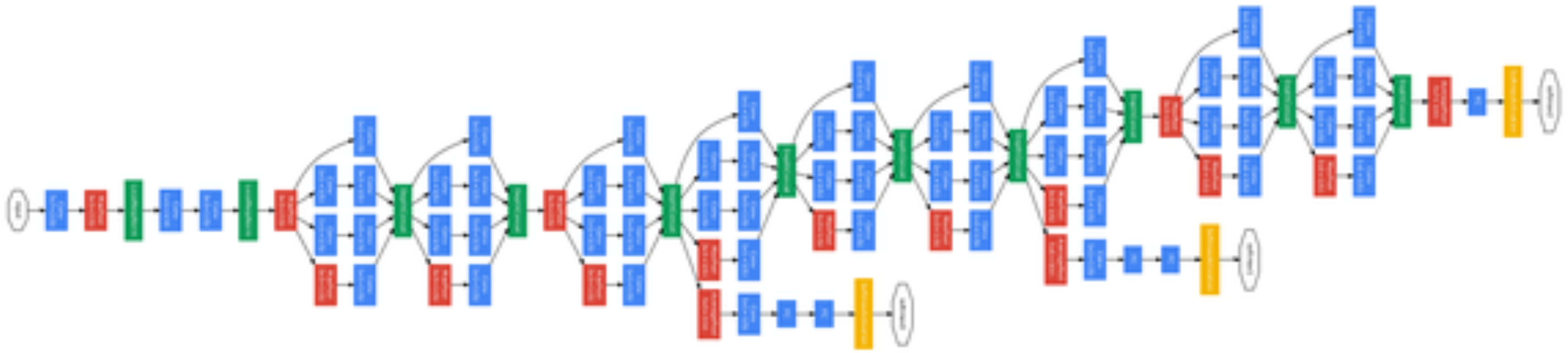
Deep Neural Networks



- More than 3 layers counts as "Deep"
- Equivalently: more than one hidden layer.

A really deep NN

- GoogleNet is a 22 layer network used for computer vision.



Why DNNs?

- It is exciting to be inspired by the brain!
  - These days, very few neuroscientists think of ANN as a useful model of the brain.
- Neural networks are very flexible - can represent almost any input-output relationship!
  - Representing ANY function <-> overfitting.
  - Might be better to encode assumptions directly.
- Deep is better than shallow!
  - Clearly true for people and relationships, not so clear why true for ANN.
- Best of all, you don’t need to know math!!
  - Or anything else!

Be a Skeptic

- Some NN based analysis is excellent, but the majority cannot be trusted.
- Same is true for **any** analysis method. But NN are particularly vulnerable because they are complex and there is barely any theory.
- To be a good analyst, you must be a skeptic. Don’t believe things just because an authority states them as good.
- Replicate what others did and compare results with other methods, including very simple ones (logistic regression).

examples

1. Excellent: AlphaGo and AlphaGo zero.
2. Debatable : Deep NN for Classification of Skin Cancer.
3. Good: YOLO
4. Embarrassing: Siraj Raval

Excellent: AlphaGo

Mastering the game of Go without human knowledge

David Silver[1]*, Julian Schrittwieser[1]*, Karen Simonyan[1]*, Ioannis Antonoglou[1], Aja Huang[1], Arthur Guez[1], Thomas Hubert[1], Lucas Baker[1], Matthew Lai[1], Adrian Bolton[1], Yutian Chen[1], Timothy Lillicrap[1], Fan Hui[1], Laurent Sifre[1], George van den Driessche[1], Thore Graepel[1] & Demis Hassabis[1]

[1]DeepMind, 5 New Street Square, London EC4A 3TW, UK.
*These authors contributed equally to this work.

354 | NATURE | VOL 550 | 19 OCTOBER 2017

- March 2016, Alpha go wins 4-1 against Lee Sedol , a 9-dan professional player

- May 2017, AlphaGo beats Ke Jie, ranked number 1, in a 3-game match. AlphaGo is given the 9-dan rank.

- October 2017: AlphaGo Zero, trained by only playing against itself. AlphaGo Zero wins 100:0 against AlphaGo Lee.

- Machines have clearly surpassed Humans in the game of Go.

Configuration and strength[60]

Versions	Hardware	Elo rating	Matches
AlphaGo Fan	176 GPUs,[51] distributed	3,144[50]	5:0 against Fan Hui
AlphaGo Lee	48 TPUs,[51] distributed	3,739[50]	4:1 against Lee Sedol
AlphaGo Master	4 TPUs,[51] single machine	4,858[50]	60:0 against professional players; Future of Go Summit
AlphaGo Zero	4 TPUs,[51] single machine	5,185[50]	100:0 against AlphaGo Lee 89:11 against AlphaGo Master
AlphaZero	4 TPUs, single machine	N/A	60:40 against AlphaGo Zero

Debatable: detecting skin cancer

LETTER

doi:10.1038/nature21056

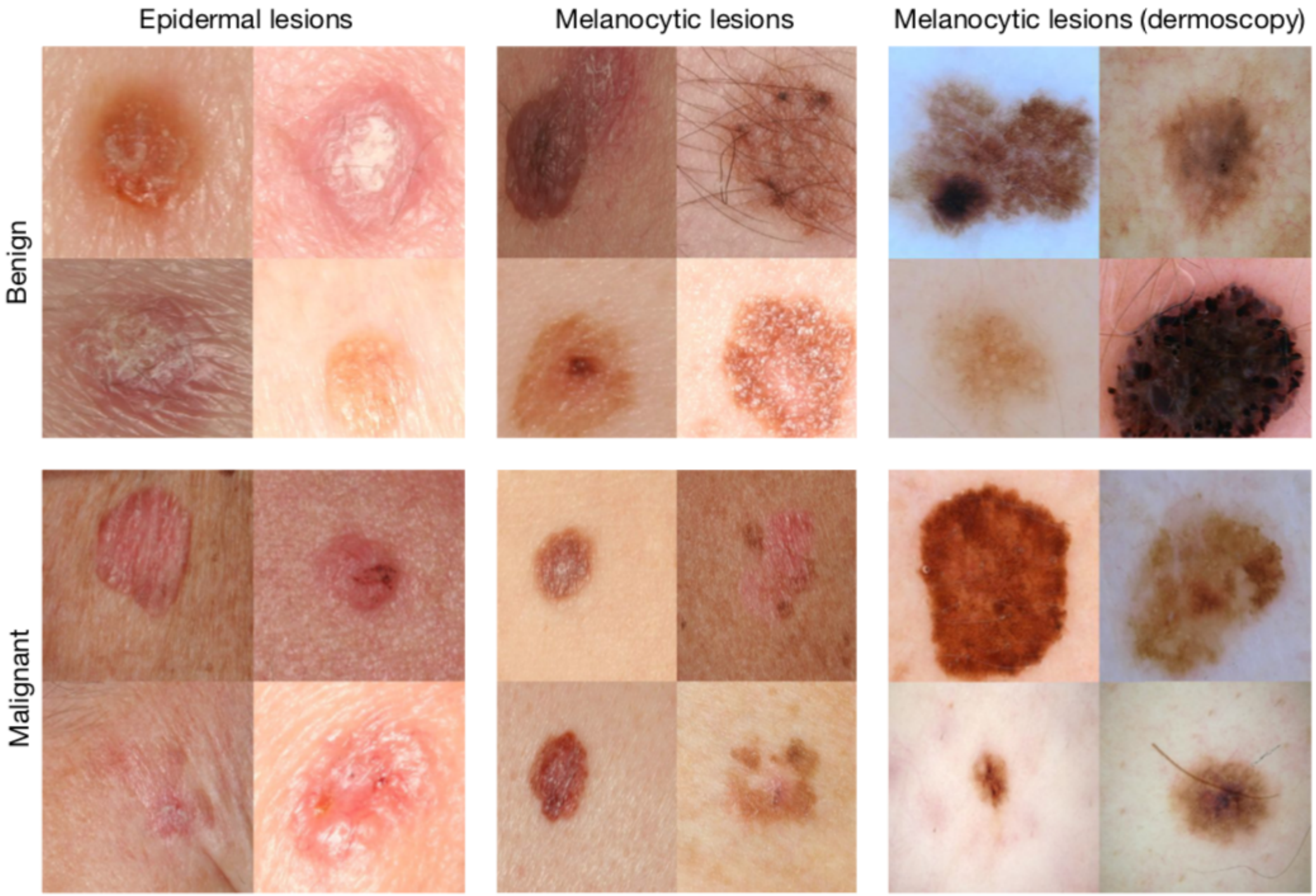
Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva[1]*, Brett Kuprel[1]*, Roberto A. Novoa[2,3], Justin Ko[2], Susan M. Swetter[2,4], Helen M. Blau[5] & Sebastian Thrun[6]

[1]Department of Electrical Engineering, Stanford University, Stanford, California, USA. [2]Department of Dermatology, Stanford University, Stanford, California, USA. [3]Department of Pathology, Stanford University, Stanford, California, USA. [4]Dermatology Service, Veterans Affairs Palo Alto Health Care System, Palo Alto, California, USA. [5]Baxter Laboratory for Stem Cell Biology, Department of Microbiology and Immunology, Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, California, USA. [6]Department of Computer Science, Stanford University, Stanford, California, USA.

*These authors contributed equally to this work.

Examples of the task



Comparing human and DNN performance

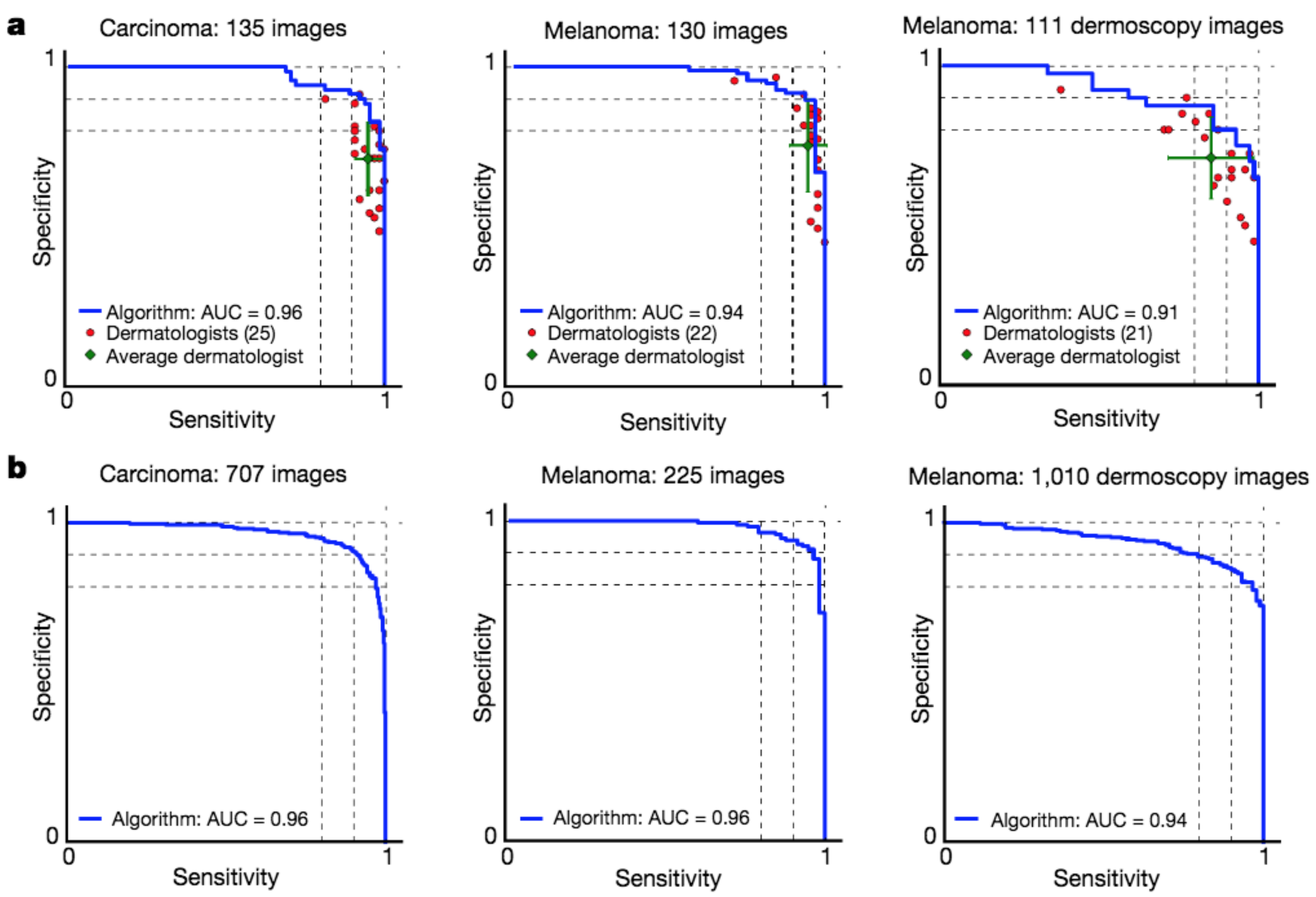


Figure 3 | Skin cancer classification performance of the CNN and dermatologists. a, The deep learning CNN outperforms the average of the dermatologists at skin cancer classification using photographic and dermoscopic images. Our CNN is tested against at least 21 dermatologists at keratinocyte carcinoma and melanoma recognition. For each test, previously unseen, biopsy-proven images of lesions are displayed, and dermatologists are asked if they would: biopsy/treat the lesion or reassure the patient. Sensitivity, the true positive rate, and specificity, the true negative rate, measure performance. A dermatologist outputs a single prediction per image and is thus represented by a single red point. The green points are the average of the dermatologists for each task, with error bars denoting one standard deviation (calculated from $n = 25$, 22 and 21 tested dermatologists for keratinocyte carcinoma, melanoma and melanoma under dermoscopy, respectively). The CNN outputs a malignancy probability P per image. We fix a threshold probability t such that the prediction $\hat{y}$ for any image is $\hat{y} = P \geq t$, and the blue curve is drawn by sweeping t in the interval 0–1. The AUC is the CNN's measure of performance, with a maximum value of 1. The CNN achieves superior performance to a dermatologist if the sensitivity–specificity point of the dermatologist lies below the blue curve, which most do. Epidermal test: 65 keratinocyte carcinomas and 70 benign seborrheic keratoses. Melanocytic test: 33 malignant melanomas and 97 benign nevi. A second melanocytic test using dermoscopic images is displayed for comparison: 71 malignant and 40 benign. The slight performance decrease reflects differences in the difficulty of the images tested rather than the diagnostic accuracies of visual versus dermoscopic examination. **b,** The deep learning CNN exhibits reliable cancer classification when tested on a larger dataset. We tested the CNN on more images to demonstrate robust and reliable cancer classification. The CNN's curves are smoother owing to the larger test set.

Critiques

- Label from pathological analysis of biopsy: depends on pathologist's judgement, might itself be incorrect.

- In the test set, the number of benign and malignant are similar, in reality most are benign. The Pathologist has this bias built in.

- A pathologist will compare with other skin areas and change lighting to see better.

- A pathologist will consult external information: age, sun exposure, medical history…

- Possible that other methods / smaller networks would work as well.

- Positioning: hard to argue this can replace the dermatologist, might be useful as home diagnosis tool, to be followed up with dermatologist.

How to be truthful and misleading at the same time

- The authors acknowledge many of these limitations in the text of the article. But not in the title or abstract.
- The title is **“Dermatologist-level classification of skin cancer with deep neural networks”**
- Might be interpreted by some readers to mean “Deep neural networks can detect cancer better than humans”
- No comparison is made to the performance achievable by simpler classifiers.

Good:You Only Look Once

arXiv.org > cs > arXiv:1506.02640

Search or Art

(Help | Advanced

Computer Science > Computer Vision and Pattern Recognition

You Only Look Once: Unified, Real-Time Object Detection

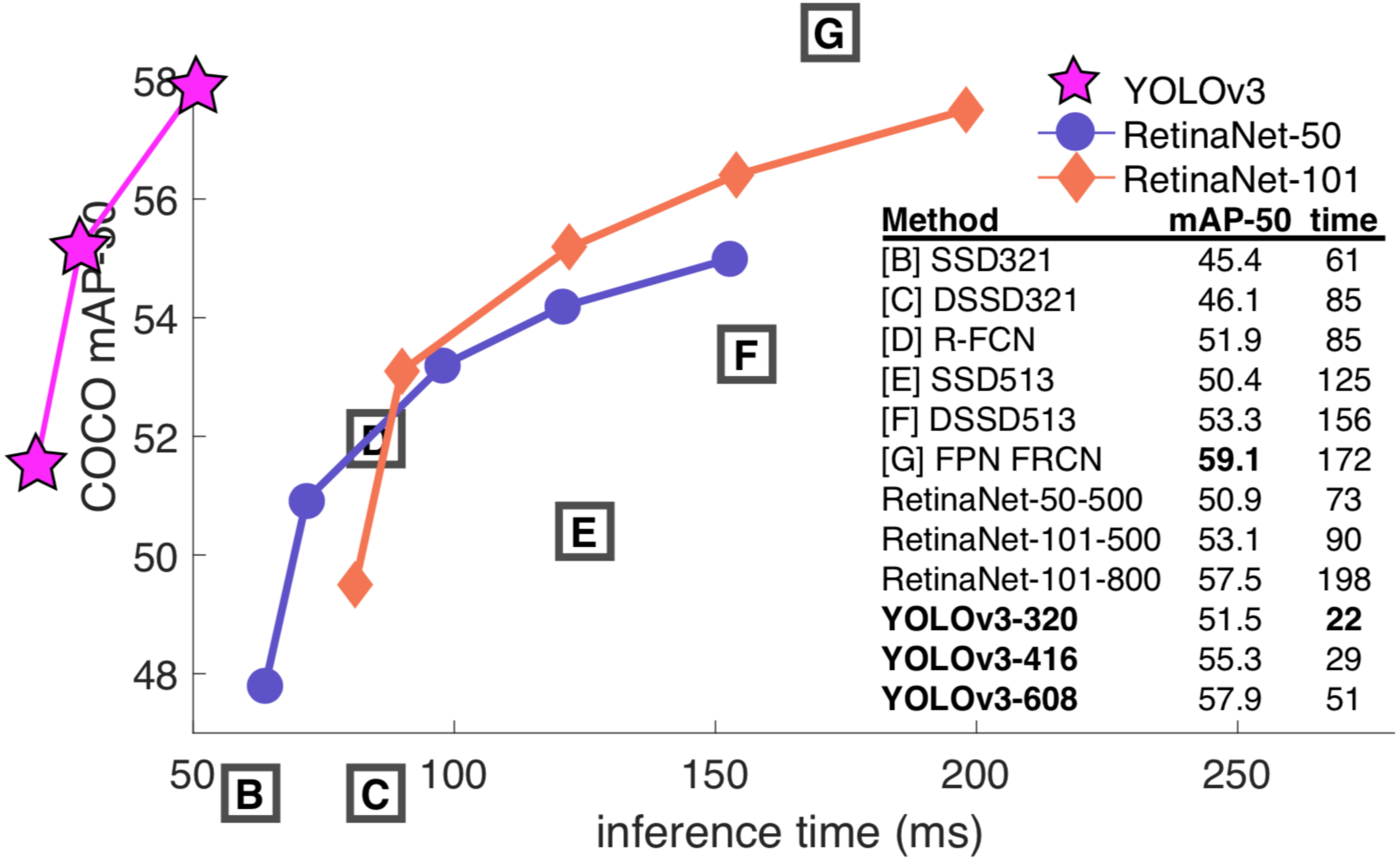
Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi

(Submitted on 8 Jun 2015 (v1), last revised 9 May 2016 (this version, v5))

We present YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, we frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance.
Our unified architecture is extremely fast. Our base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is far less likely to predict false detections where nothing exists. Finally, YOLO learns very general representations of objects. It outperforms all other detection methods, including DPM and R-CNN, by a wide margin when generalizing from natural images to artwork on both the Picasso Dataset and the People-Art Dataset.

Solid Performance

mAP-50: Correct detection/classification = ground truth and detection for same object type intersect, and area of intersection / area of union > 50%
* It is easier to achieve high mAP50 if the objects are large.



Method	mAP-50	time
[B] SSD321	45.4	61
[C] DSSD321	46.1	85
[D] R-FCN	51.9	85
[E] SSD513	50.4	125
[F] DSSD513	53.3	156
[G] FPN FRCN	**59.1**	172
RetinaNet-50-500	50.9	73
RetinaNet-101-500	53.1	90
RetinaNet-101-800	57.5	198
YOLOv3-320	51.5	**22**
YOLOv3-416	55.3	29
YOLOv3-608	57.9	51

Siraj Raval Fake News

On paper, seems like a solid guy:

Education

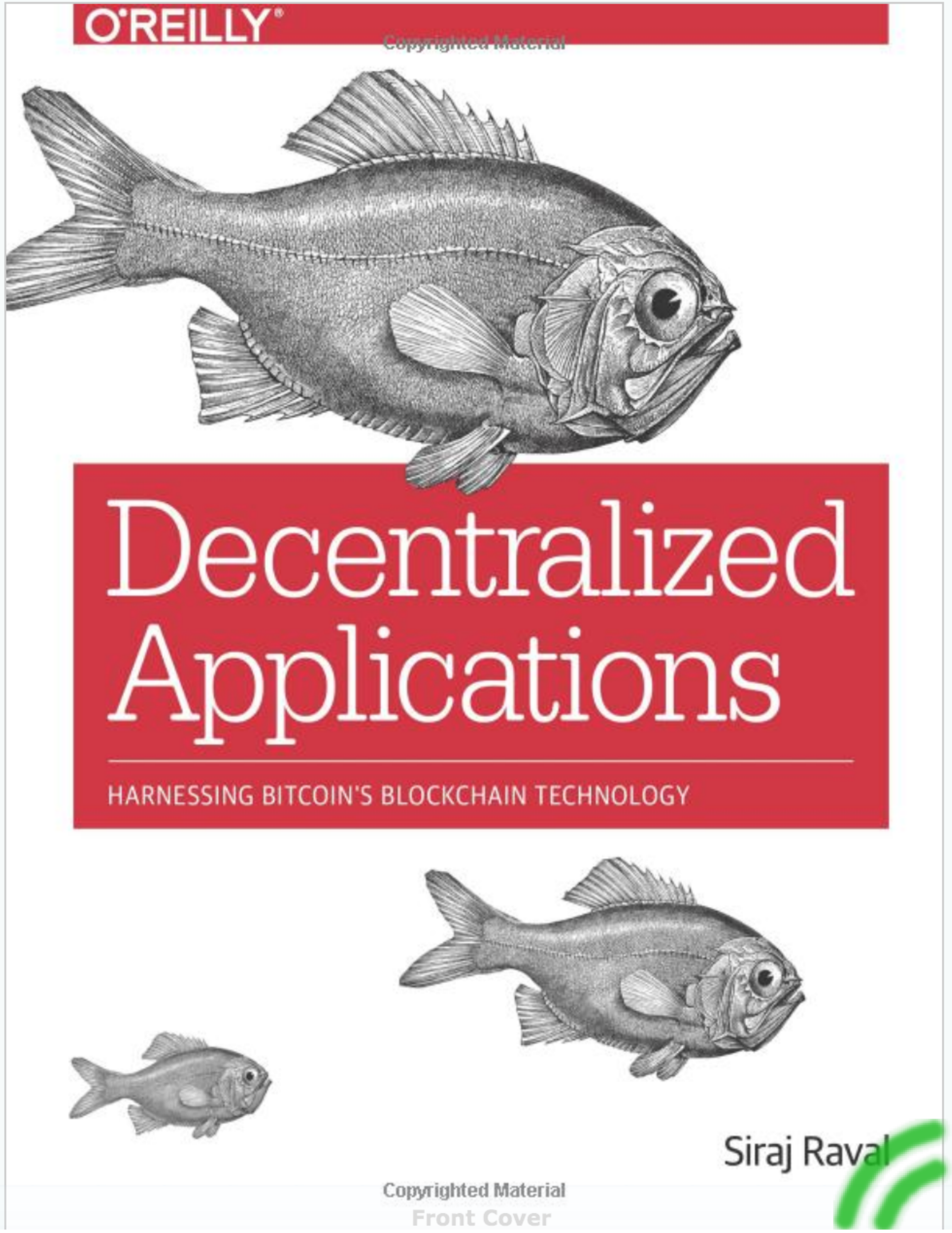
Columbia University in the City of New York
Computer Science
2009 – 2012

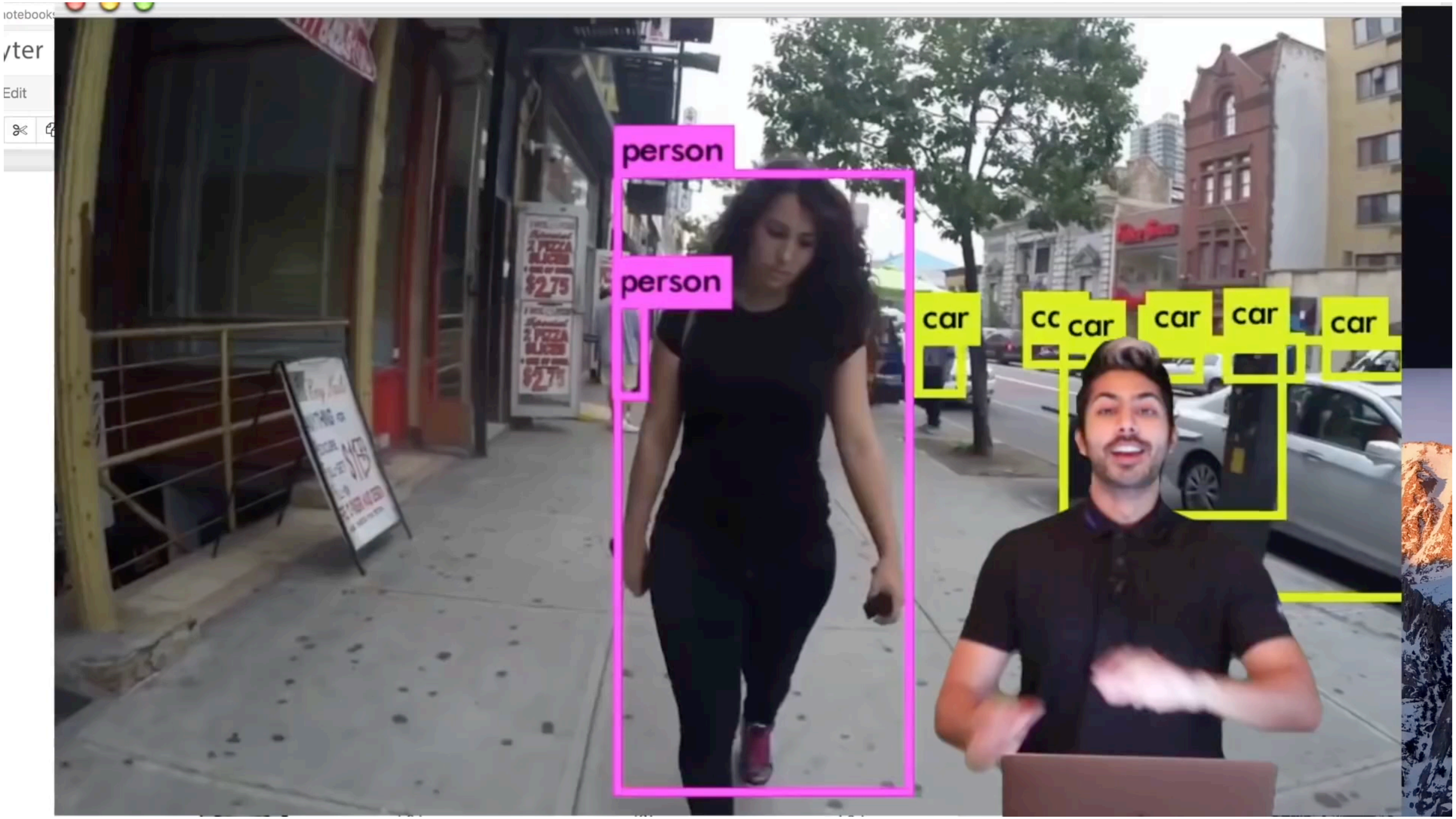
I found him by searching for “yolo object detection”on YouTube

YOLO Object Detection (TensorFlow tutorial)

301,208 views

Siraj Raval
Published on Nov 15, 2017




Edit
person
person
car
car
car
car
car

What I think of NN

- The name is misleading, there is little relation to what we know about the brain.

- It is a flexible representation. Might be too flexible.

- It is easy to try out things, and with GPUs you can do it fast!

- Popularity is not a good predictor of performance.

Recommendations

- There is a great infrastructure for learning DNN, (Tensorflow, GPUs,…) take advantage of it!
- Use NN, but not exclusively. Compare with methods previously suggested for the problem, try also SVM, Decision trees, boosting…
- Make sure that the measure of quality of your classifier in a way that corresponds to your overall goals.
- Think about how you can help people, not how you can replace them.