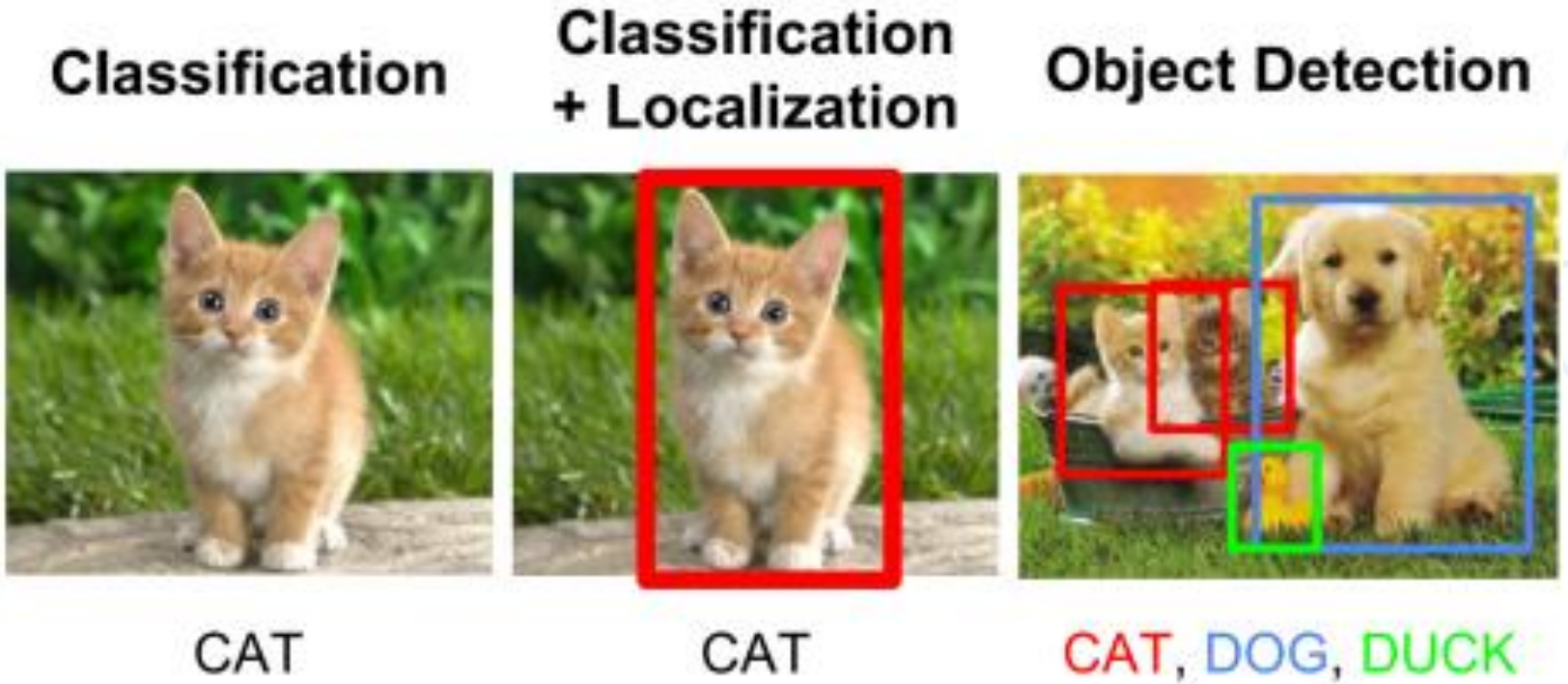
FORNAX

How to detect cats in the image? Deep learning for object detection

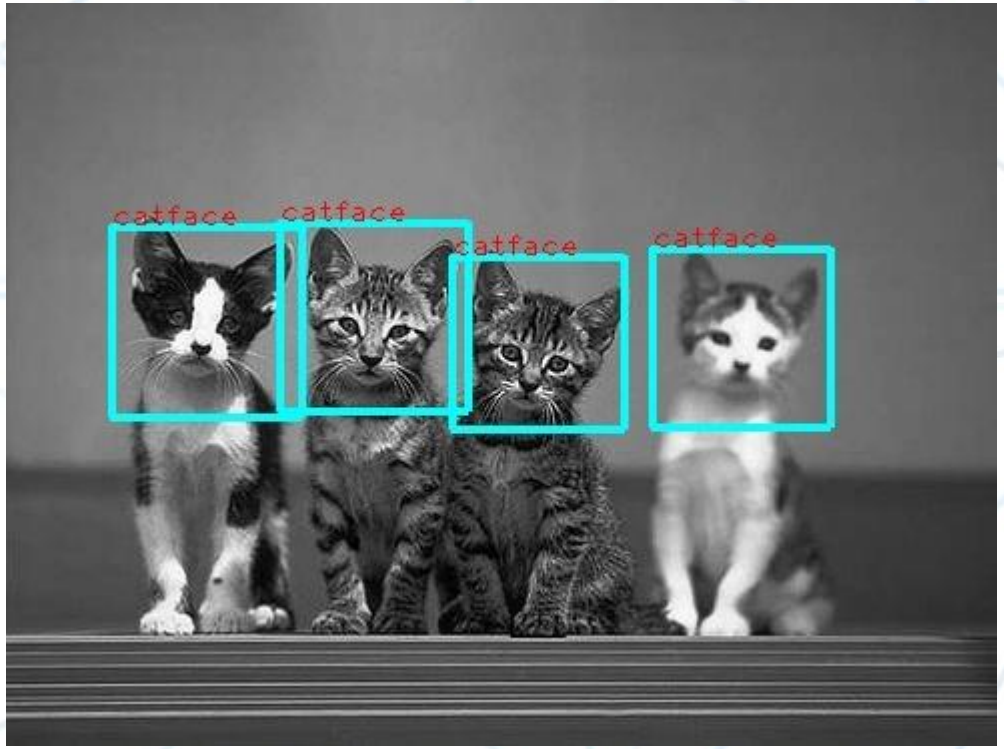
Rafał Cycoń
CTO

14.11.2017

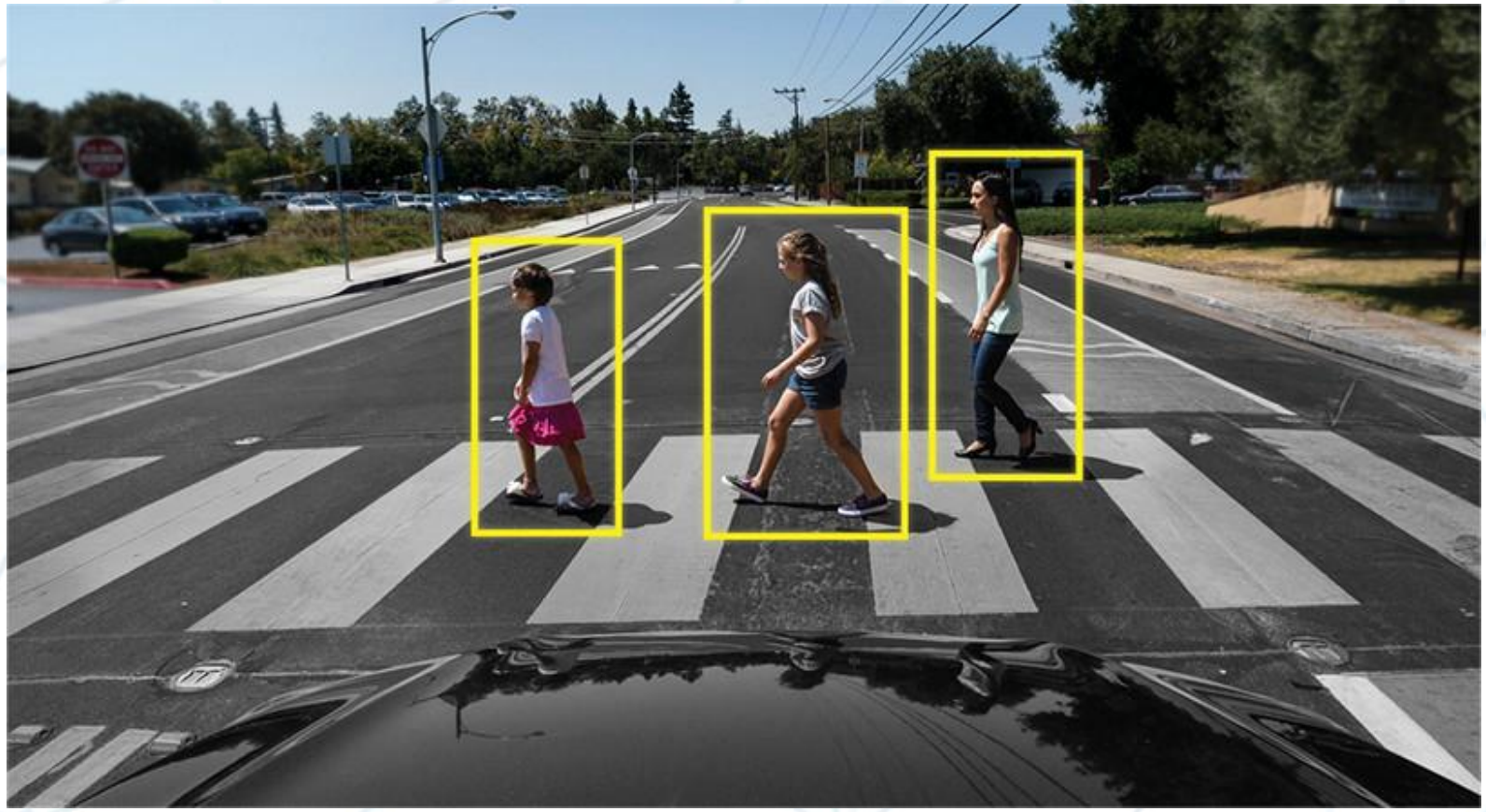
WWW.FORNAX.AI


Classification
Classification + Localization
Object Detection
CAT
CAT
CAT, DOG, DUCK

Detection!

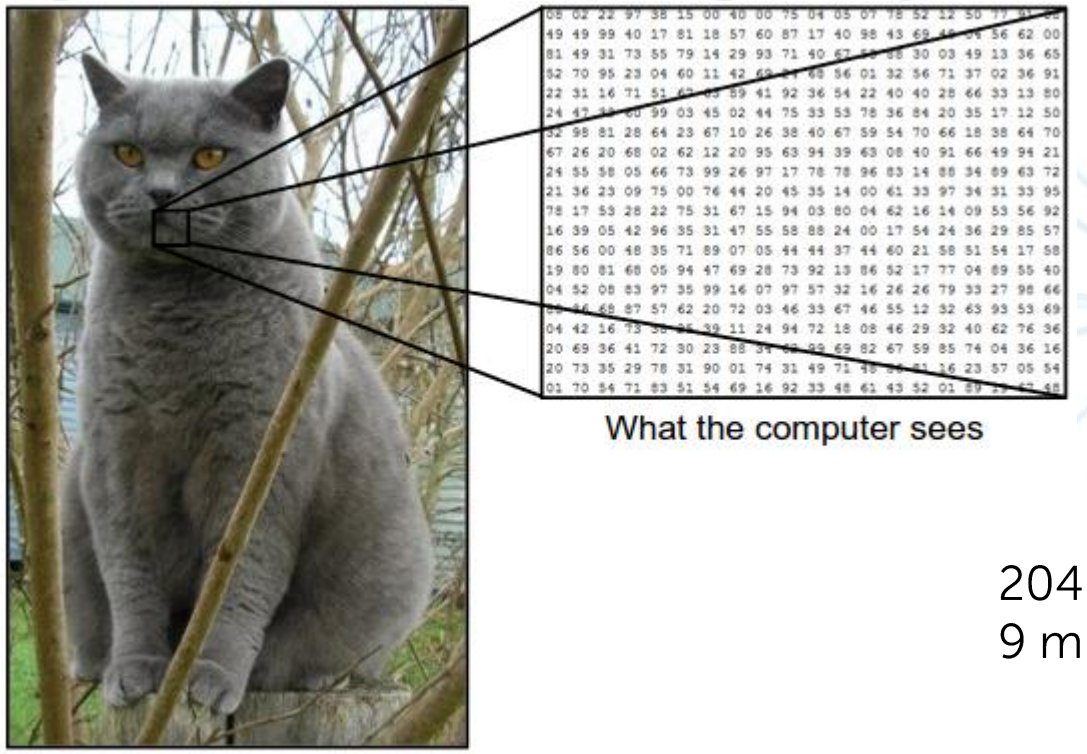


Self-driving cars



Analyze satellite imagery

Face detection



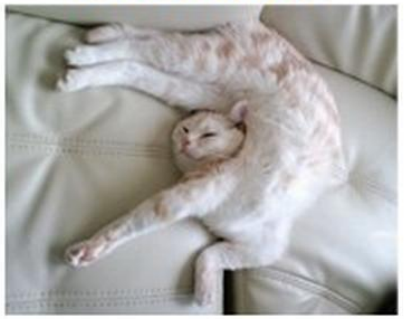
What the computer sees

2048x1536 = ~3 mln pixels
9 mln pixels for a color image

Viewpoint variation

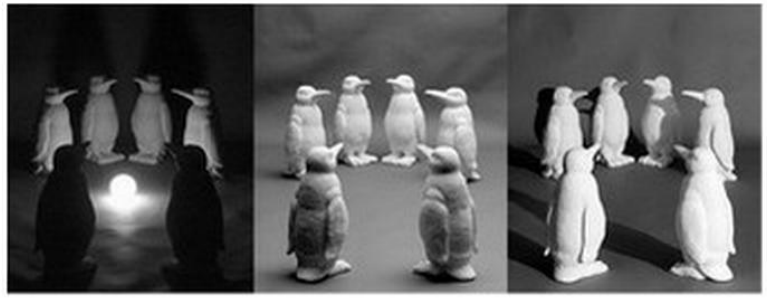
Scale variation

Deformation



Occlusion

Illumination conditions

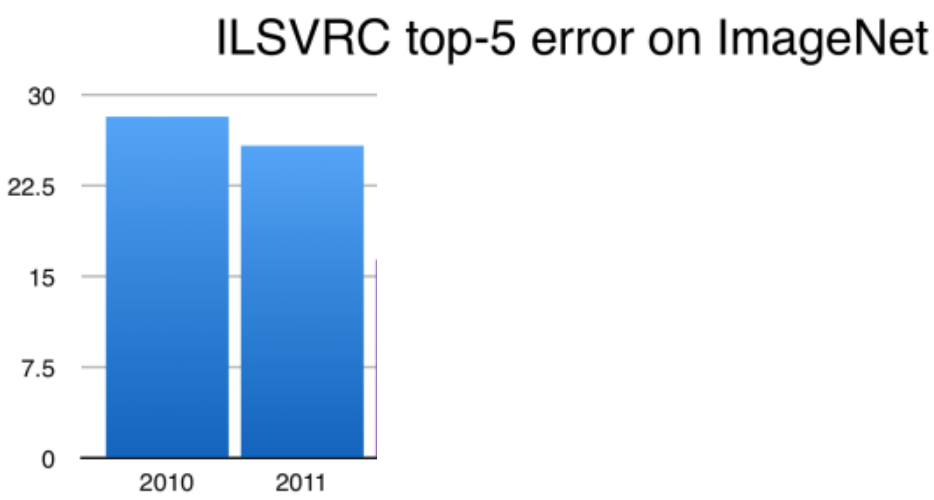


Background clutter

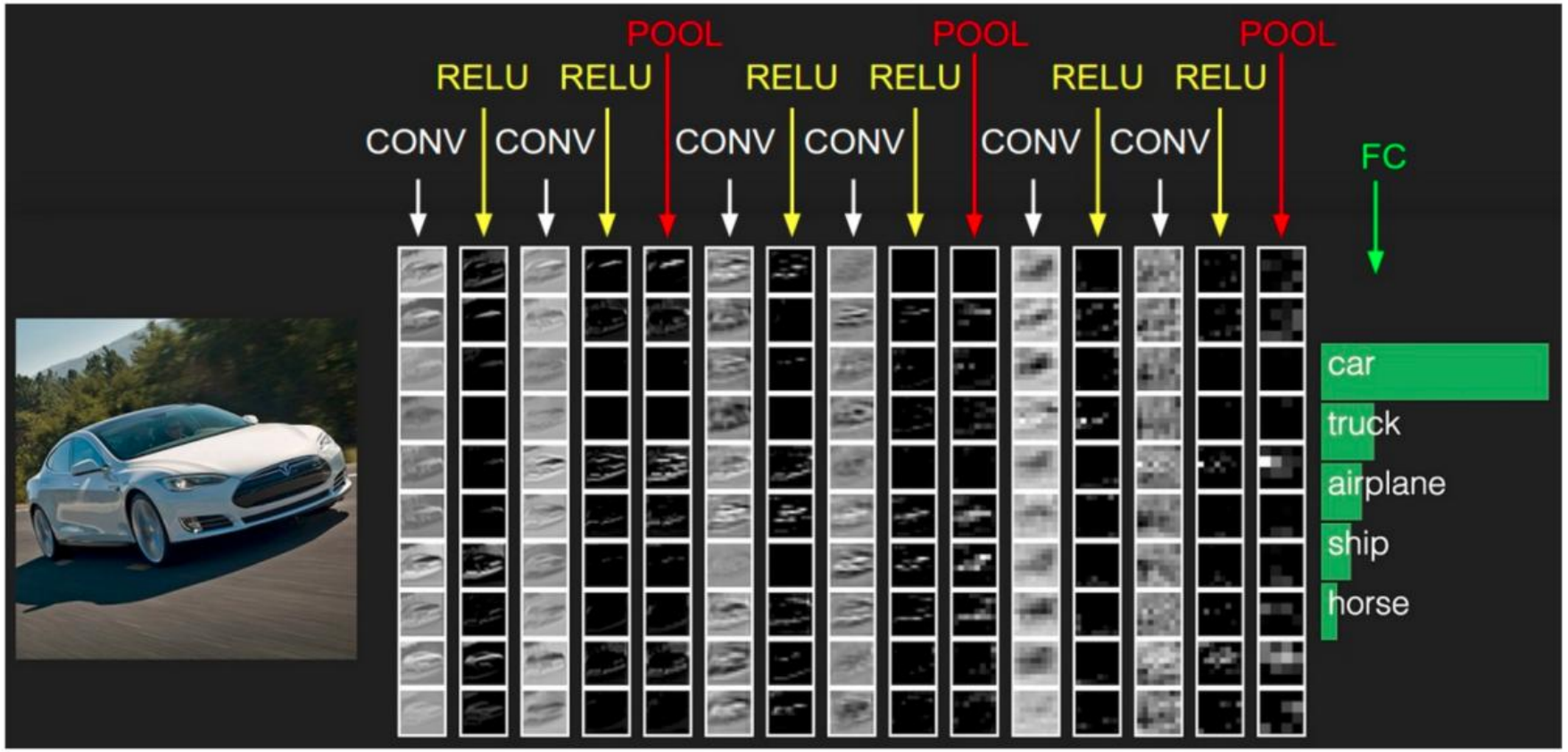
Intra-class variation

ImageNet

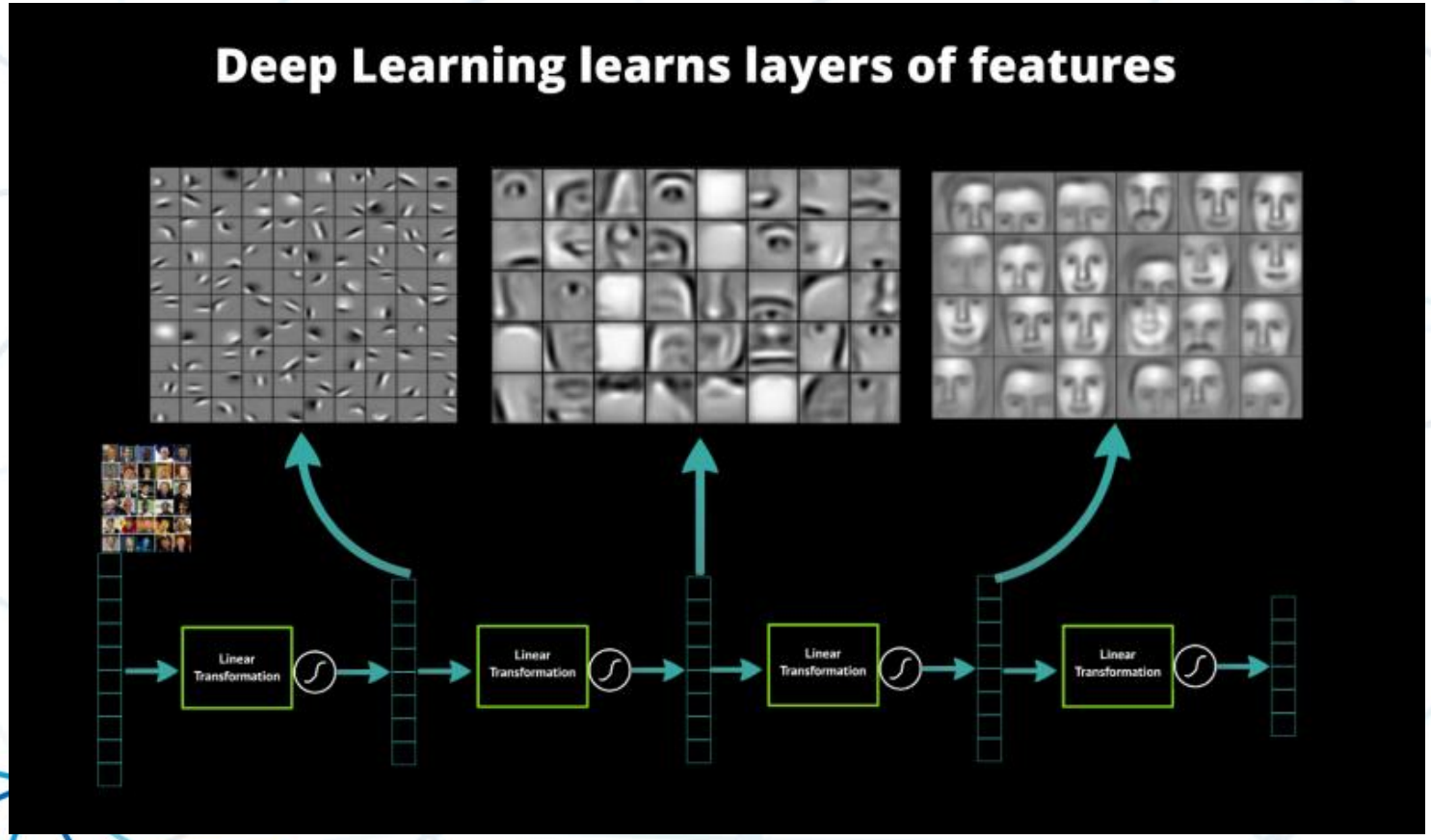

mite
container ship
motor scooter
leopard
mite
black widow
cockroach
tick
starfish
container ship
lifeboat
amphibian
fireboat
drilling platform
motor scooter
go-kart
moped
bumper car
golfcart
leopard
jaguar
cheetah
snow leopard
Egyptian cat


ILSVRC top-5 error on ImageNet
30
22.5
15
7.5
0
2010
2011

Enter Deep Learning!



Enter Deep Learning!



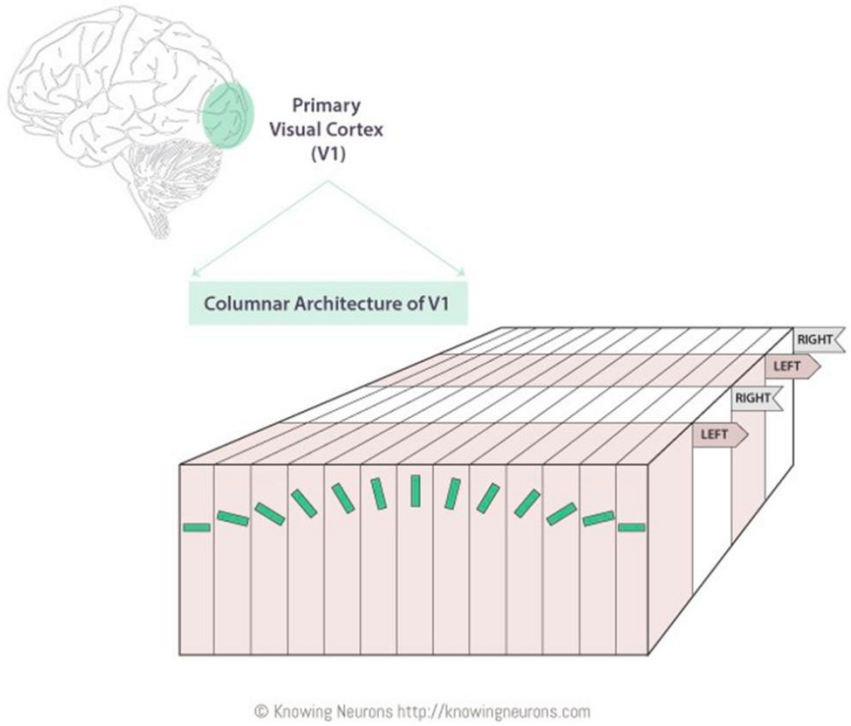
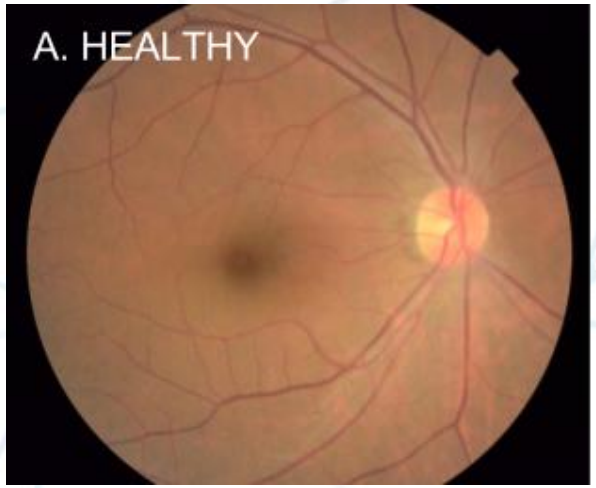
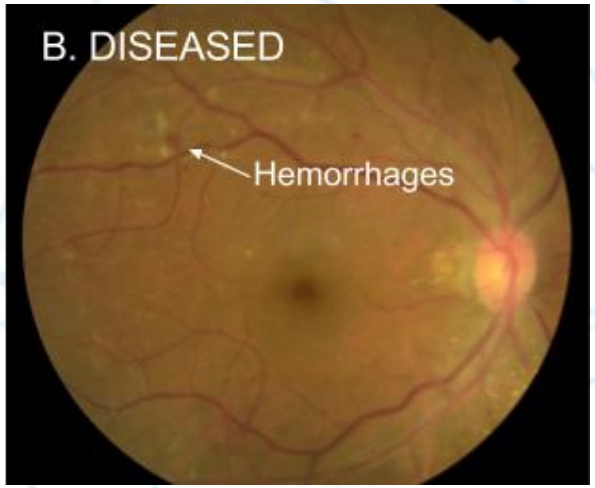


Image classification

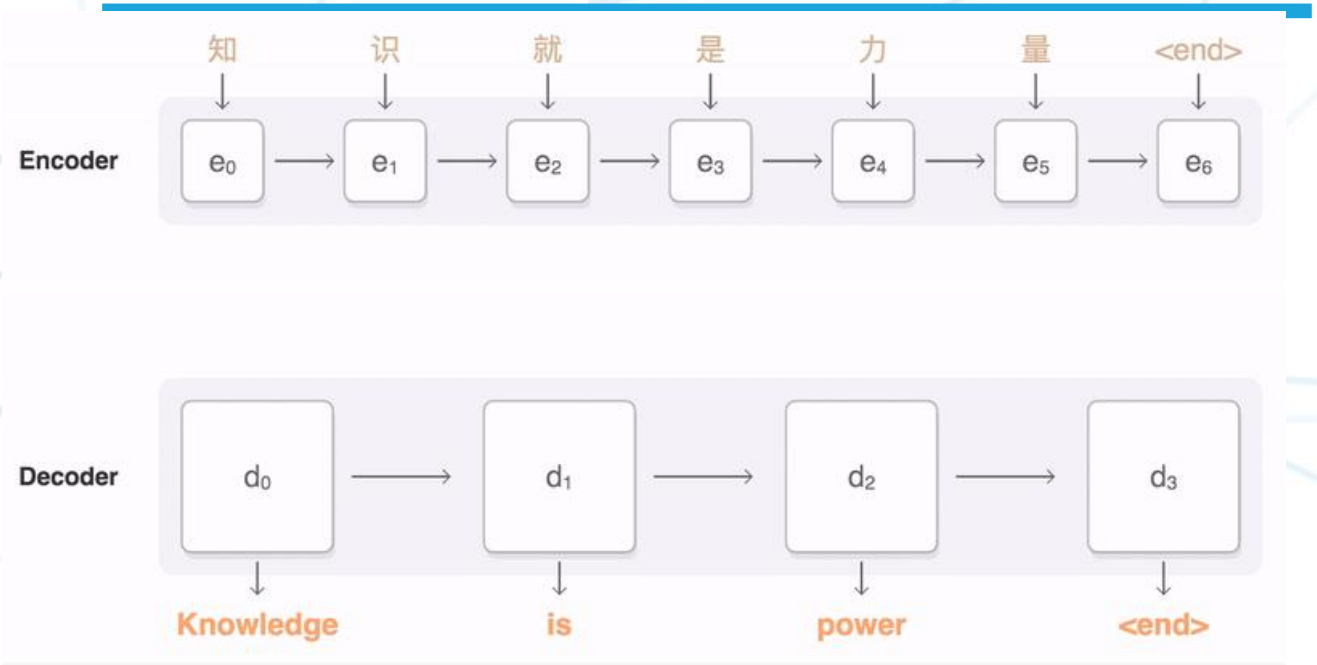
Image classification

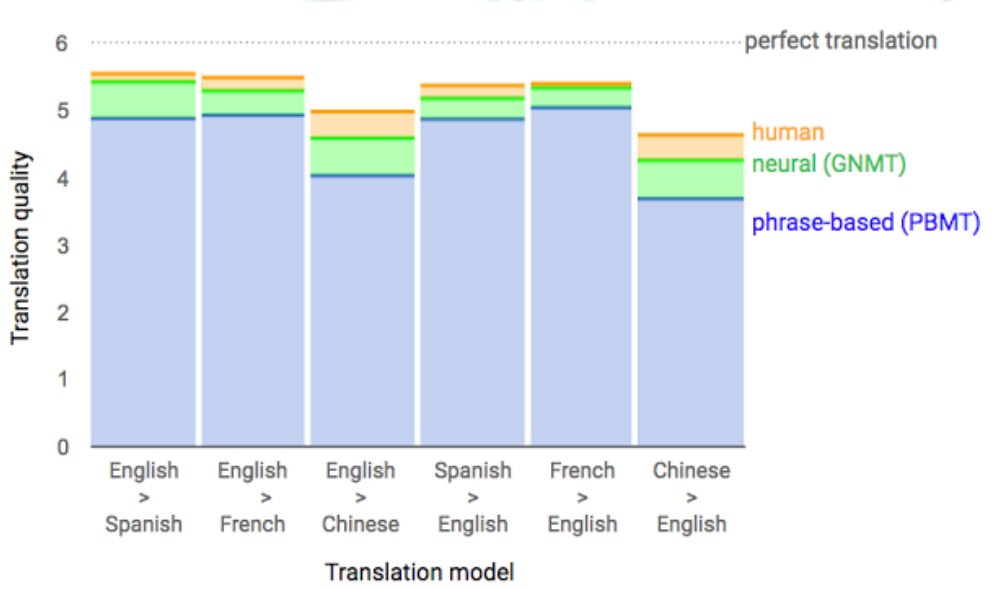
Diabetic retinopathy





Translation

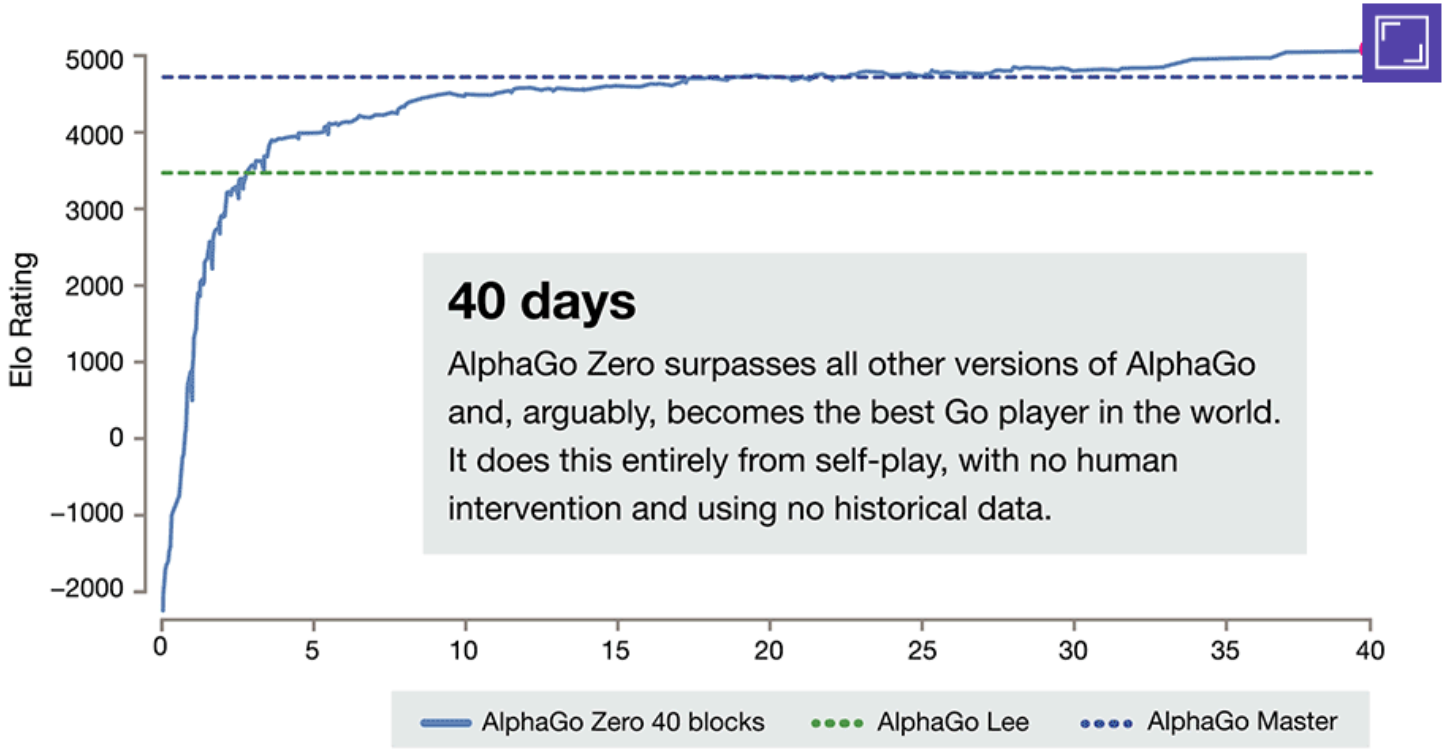




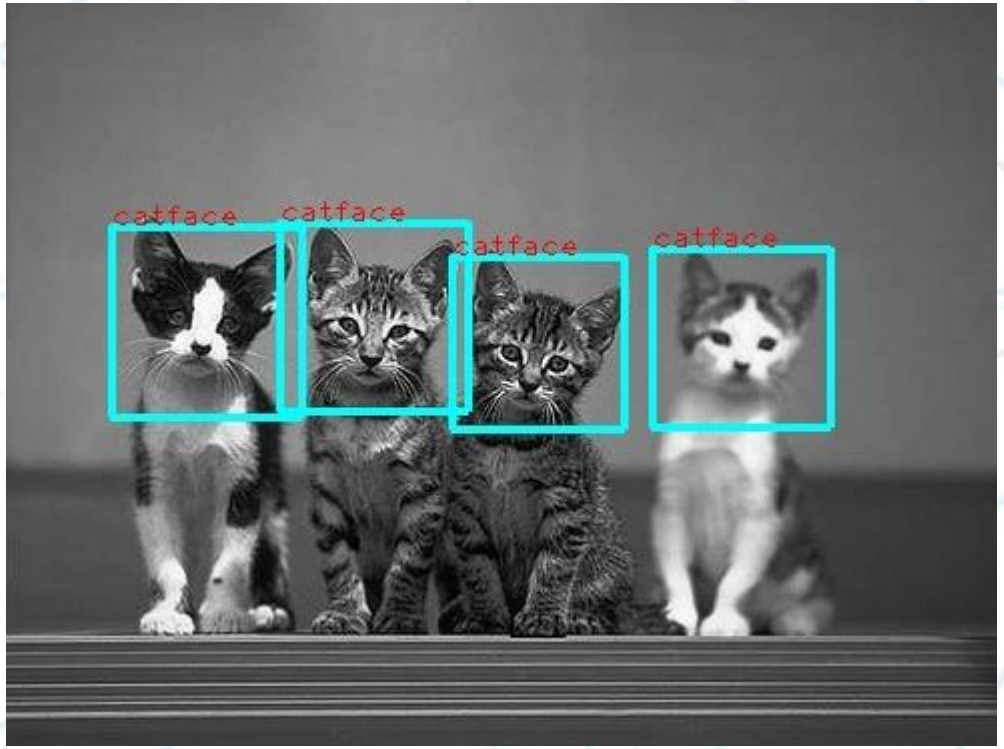
Photorealistic style transfer

AlphaGo Zero

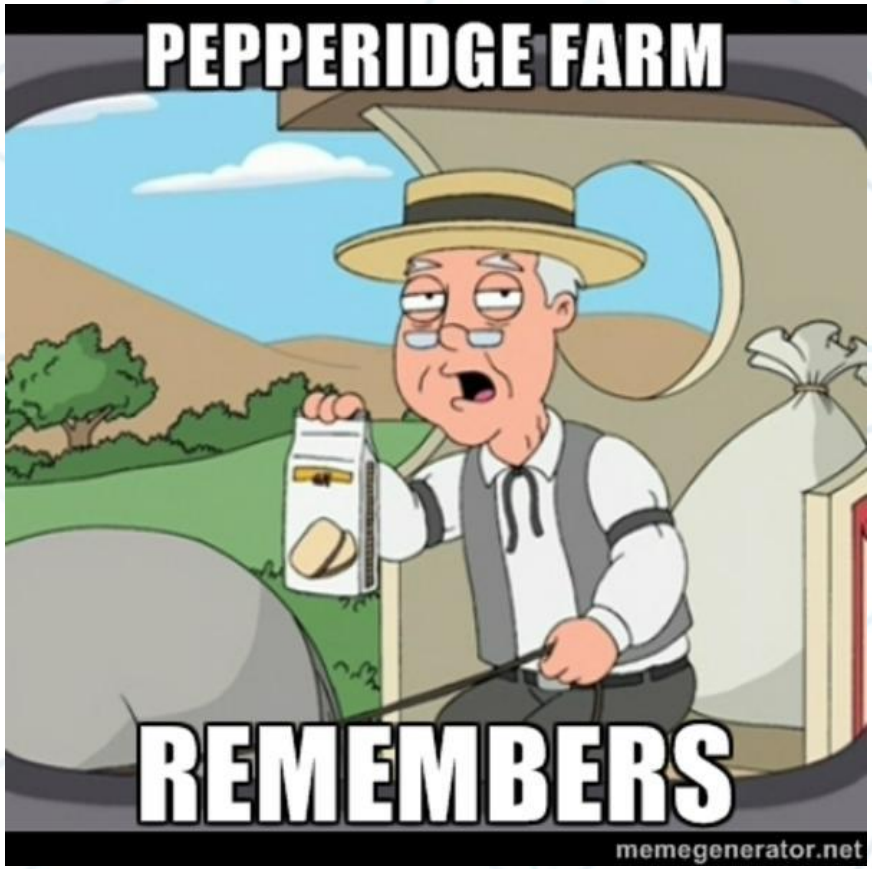
https://deepmind.com/blog/alphago-zero-learning-scratch/



Detection!



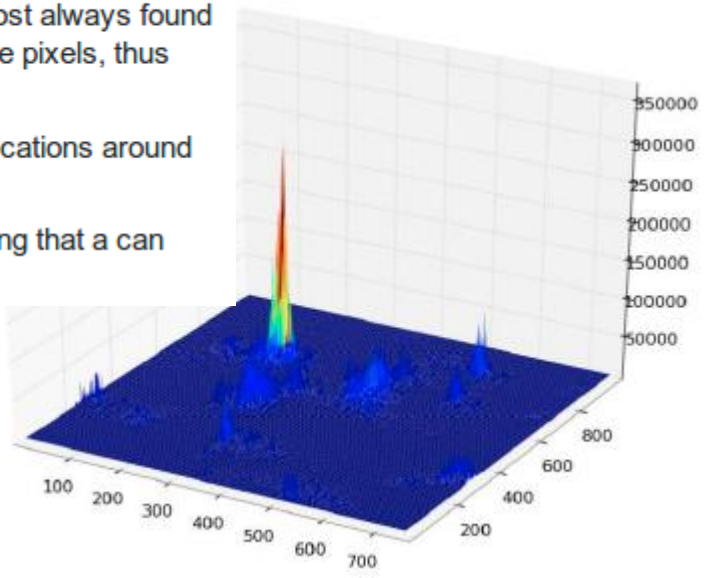
Detection - The Old Ways



Algorithm Improvement for 'Coca-Cola Can' Recognition

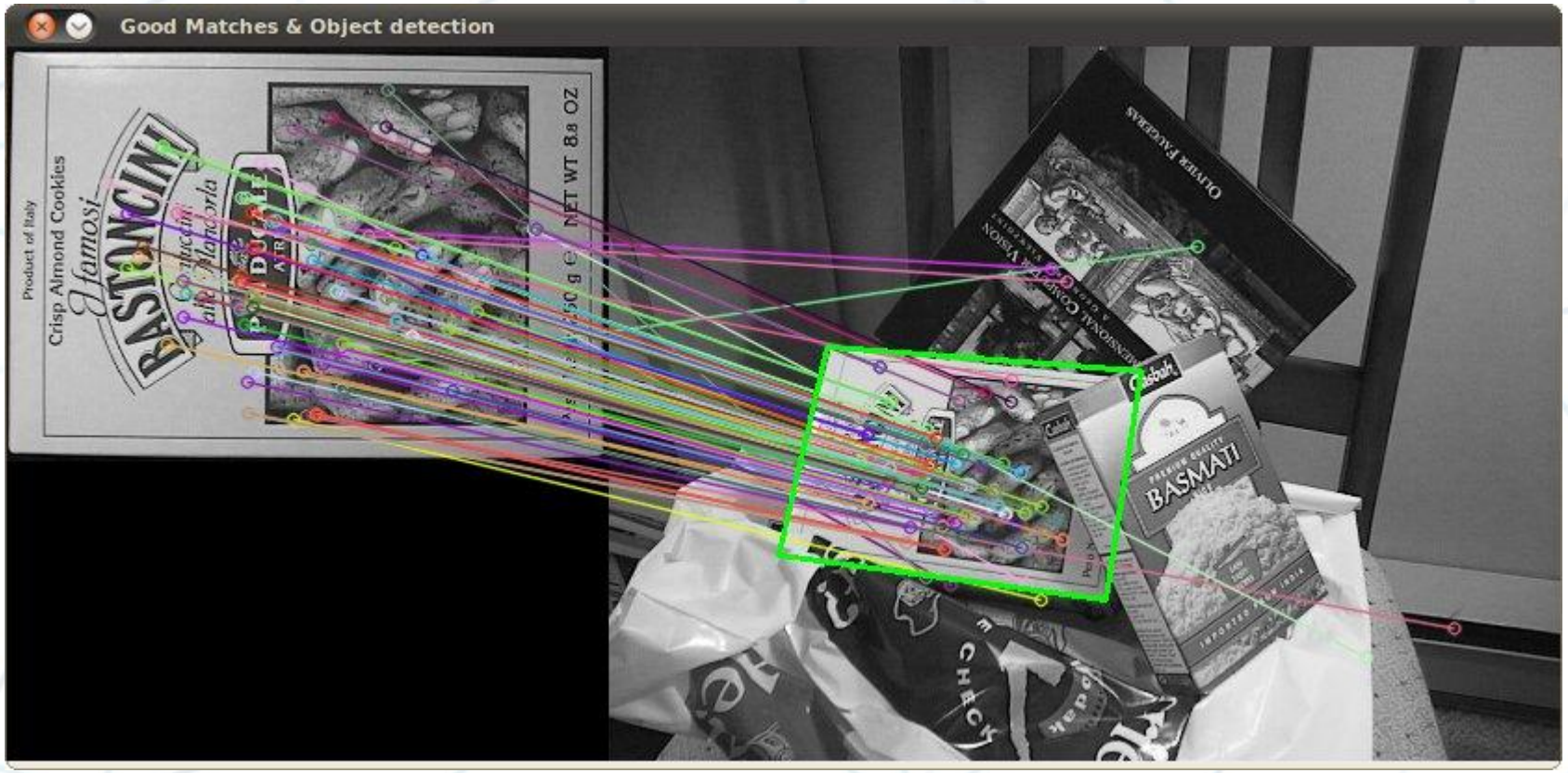
https://stackoverflow.com/questions/10168686/image-processing-algorithm-improvement-for-coca-cola-can-recognition

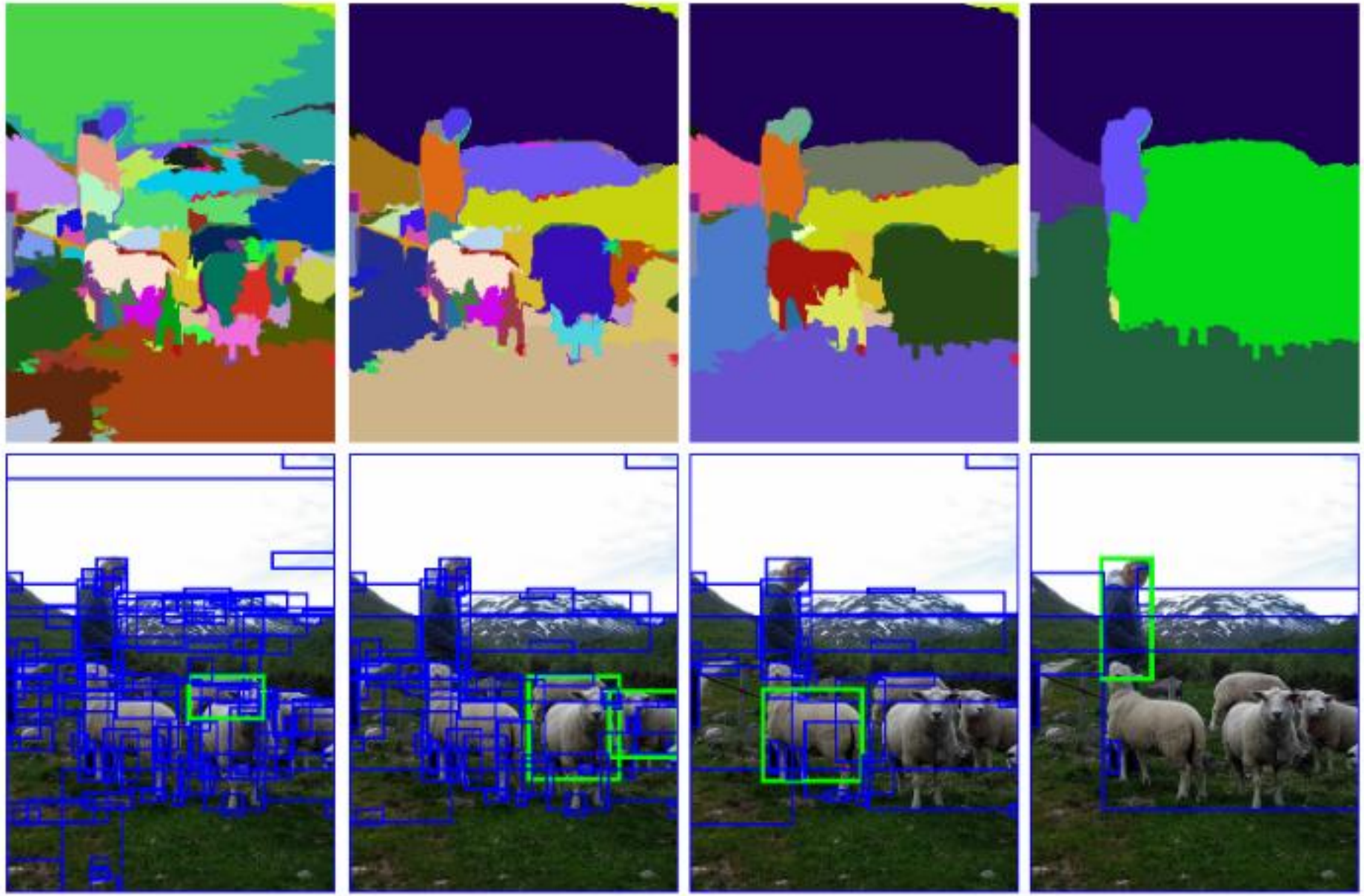
- It is **extremely slow**! I'm not stressing this enough. Almost a full day was needed to process the 30 test images, obviously because I had a very high scaling factor for rotation and translation, since some of the cans were very small.
- It was completely lost when bottles were in the image, and for some reason almost always found the bottle instead of the can (perhaps because bottles were bigger, thus had more pixels, thus more votes)
- Fuzzy images were also no good, since the votes ended up in pixel at random locations around the center, thus ending with a very noisy heat map.
- In-variance in translation and rotation was achieved, but not in orientation, meaning that a can that was not directly facing the camera objective wasn't recognized.

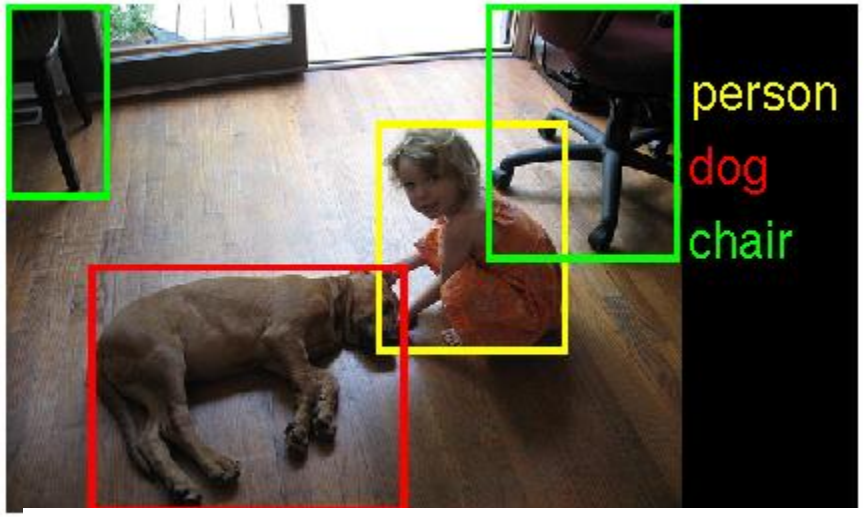


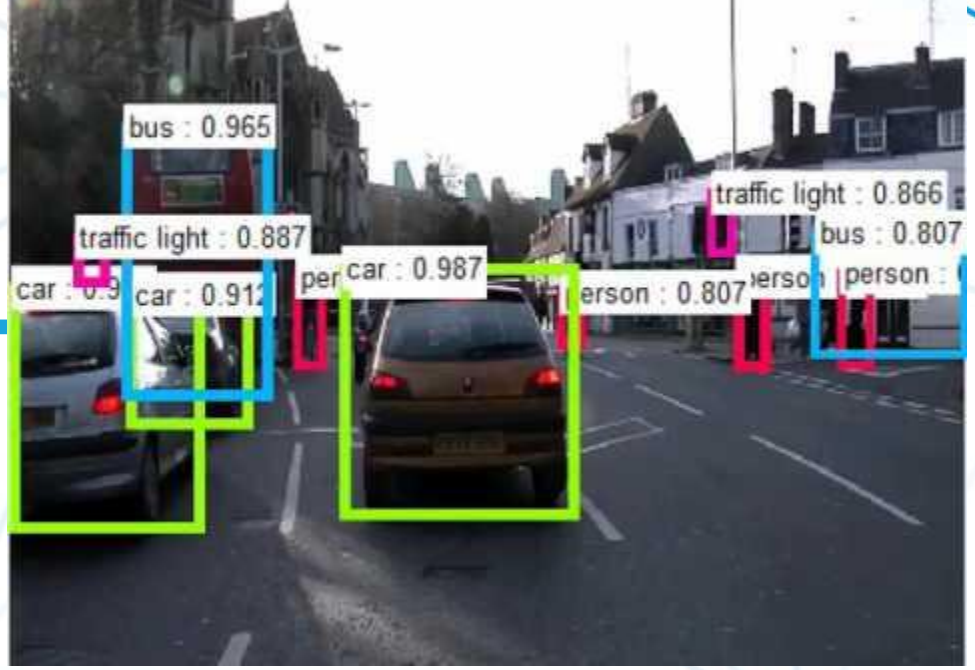
Algorithm Improvement for 'Coca-Cola Can' Recognition

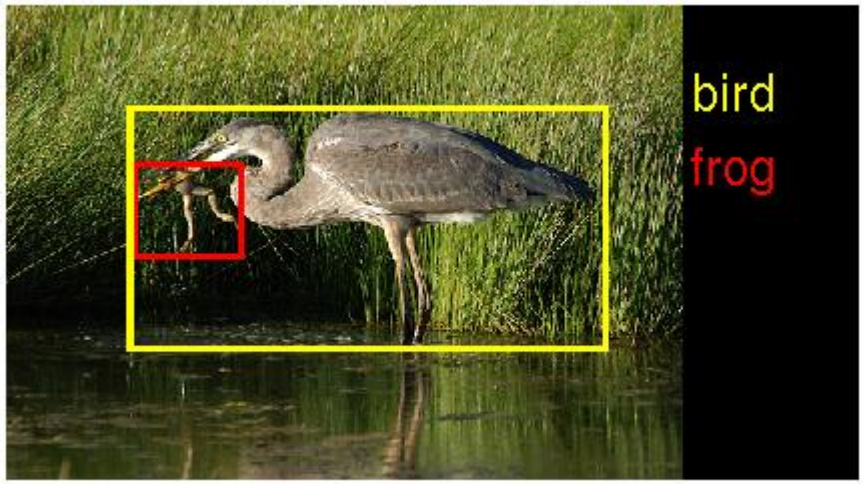
https://stackoverflow.com/questions/10168686/image-processing-algorithm-improvement-for-coca-cola-can-recognition

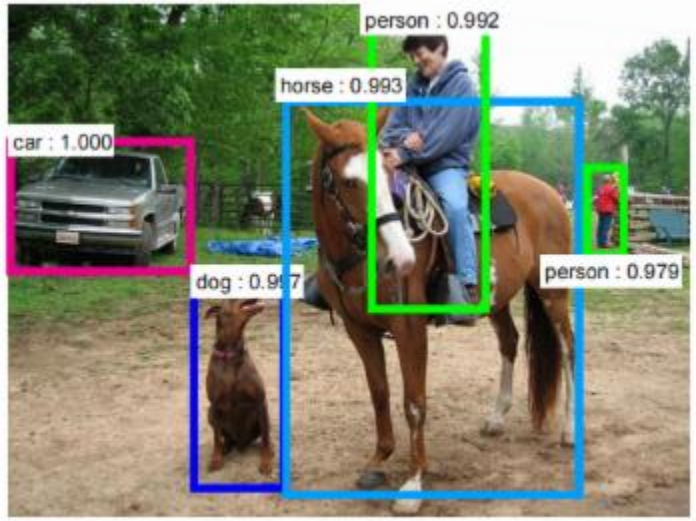


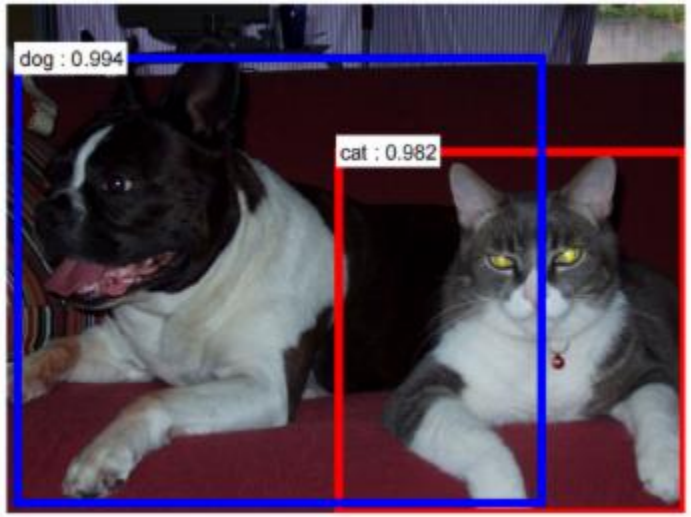



person
dog
chair

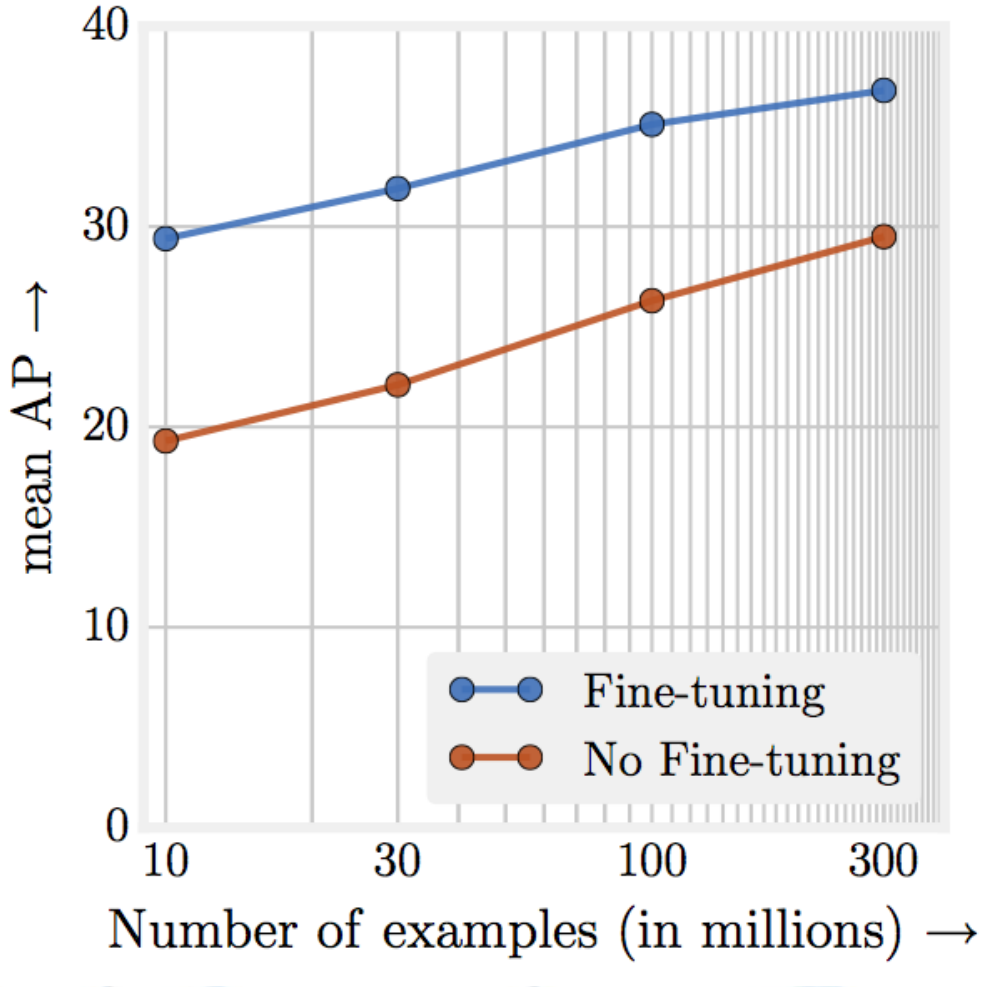

bus : 0.965
traffic light : 0.887
car : 0.912
car : 0.987
traffic light : 0.866
bus : 0.807
person : 0.807


bird
frog

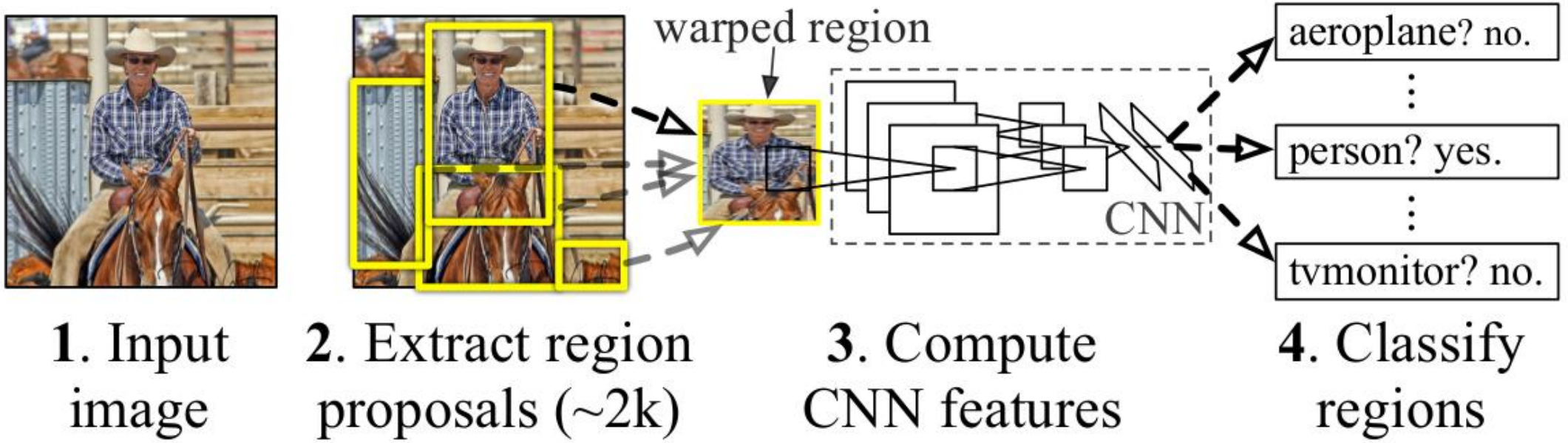

person : 0.992
horse : 0.993
car : 1.000
person : 0.979


dog : 0.994
cat : 0.982

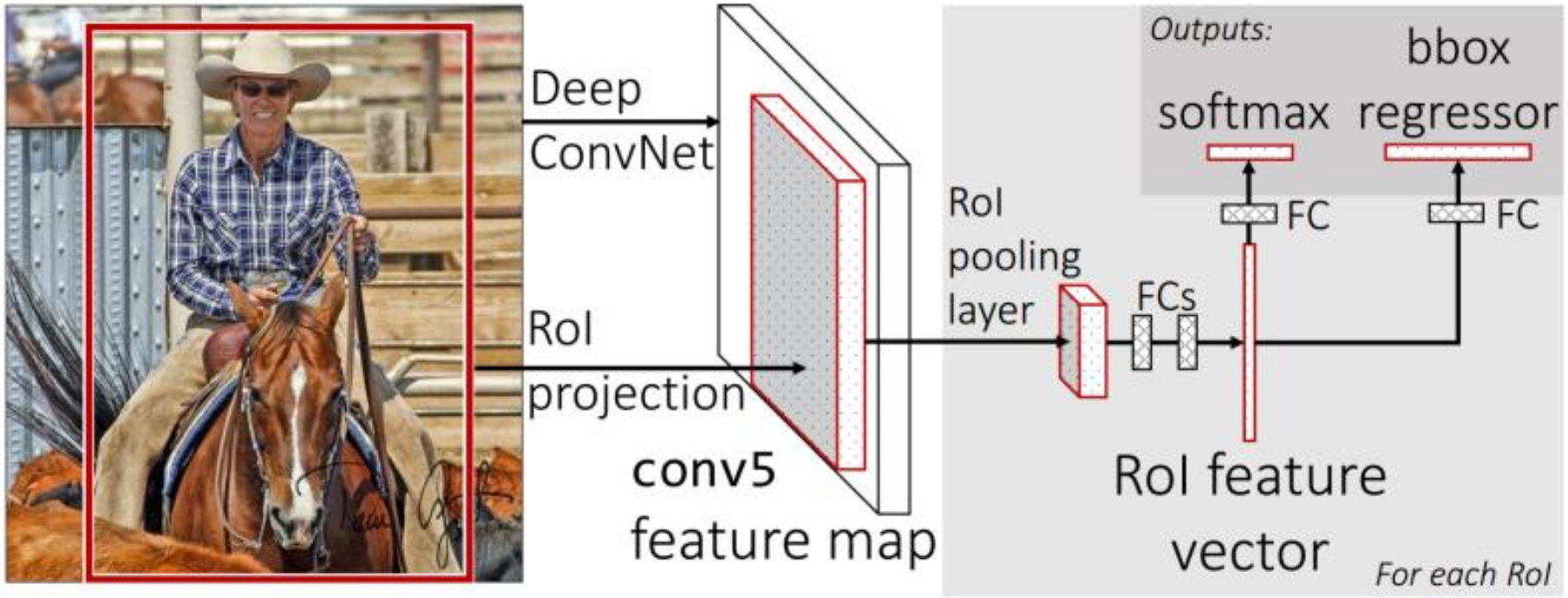
https://arxiv.org/abs/1707.02968



RCNN: Regions with Convolutional Neural Network



Fast RCNN



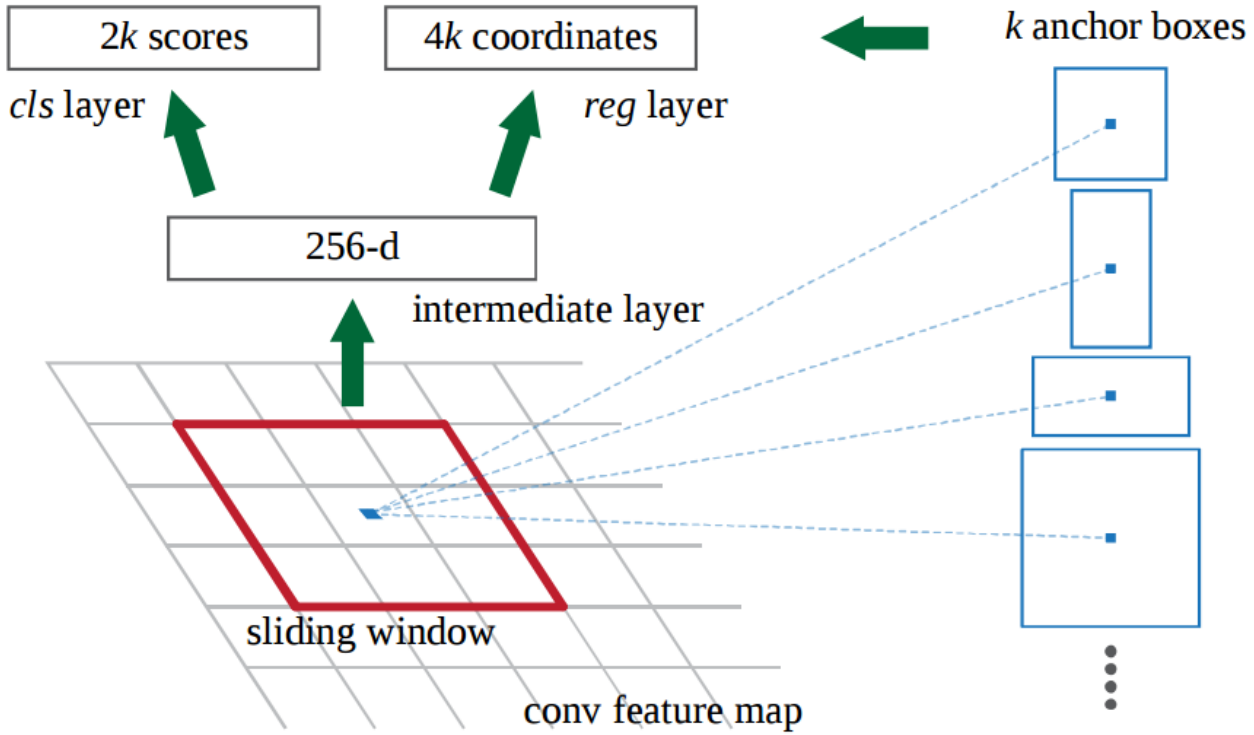
9x faster in training

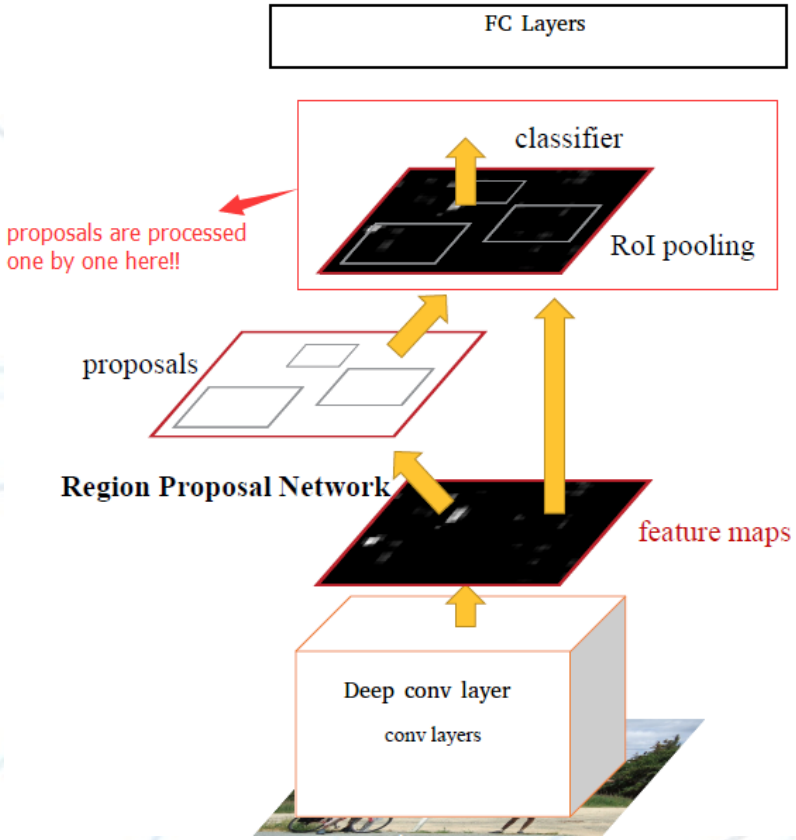
213x faster in inference

More accurate


YOU FOOL
THIS ISN'T EVEN MY FINAL FORM
quickmeme.com

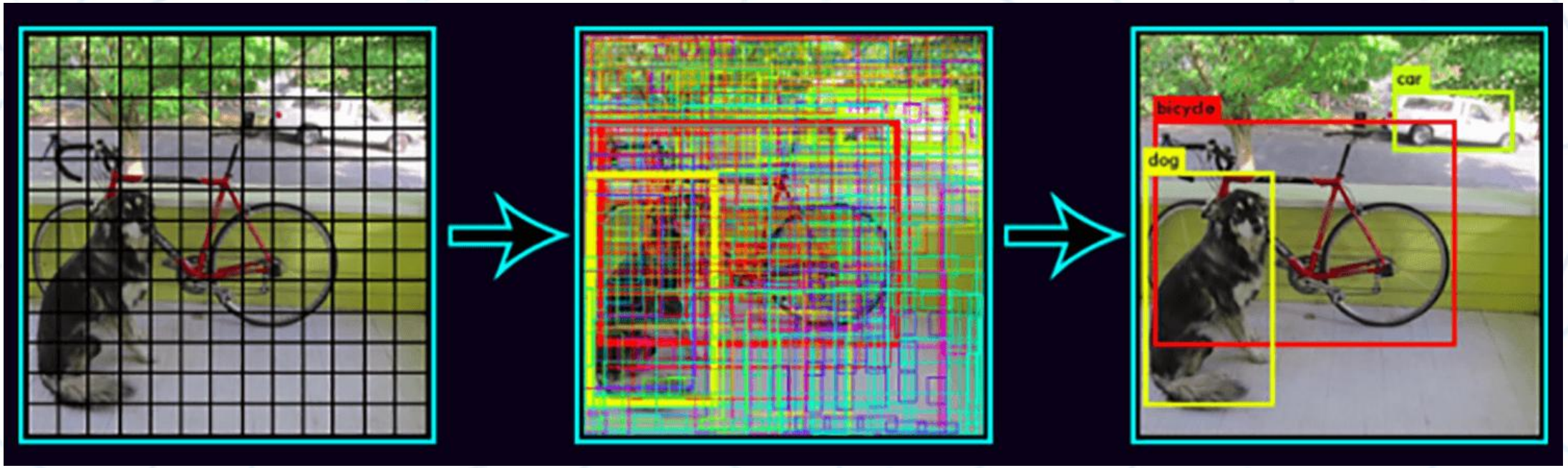
Faster RCNN
Region Proposal Network





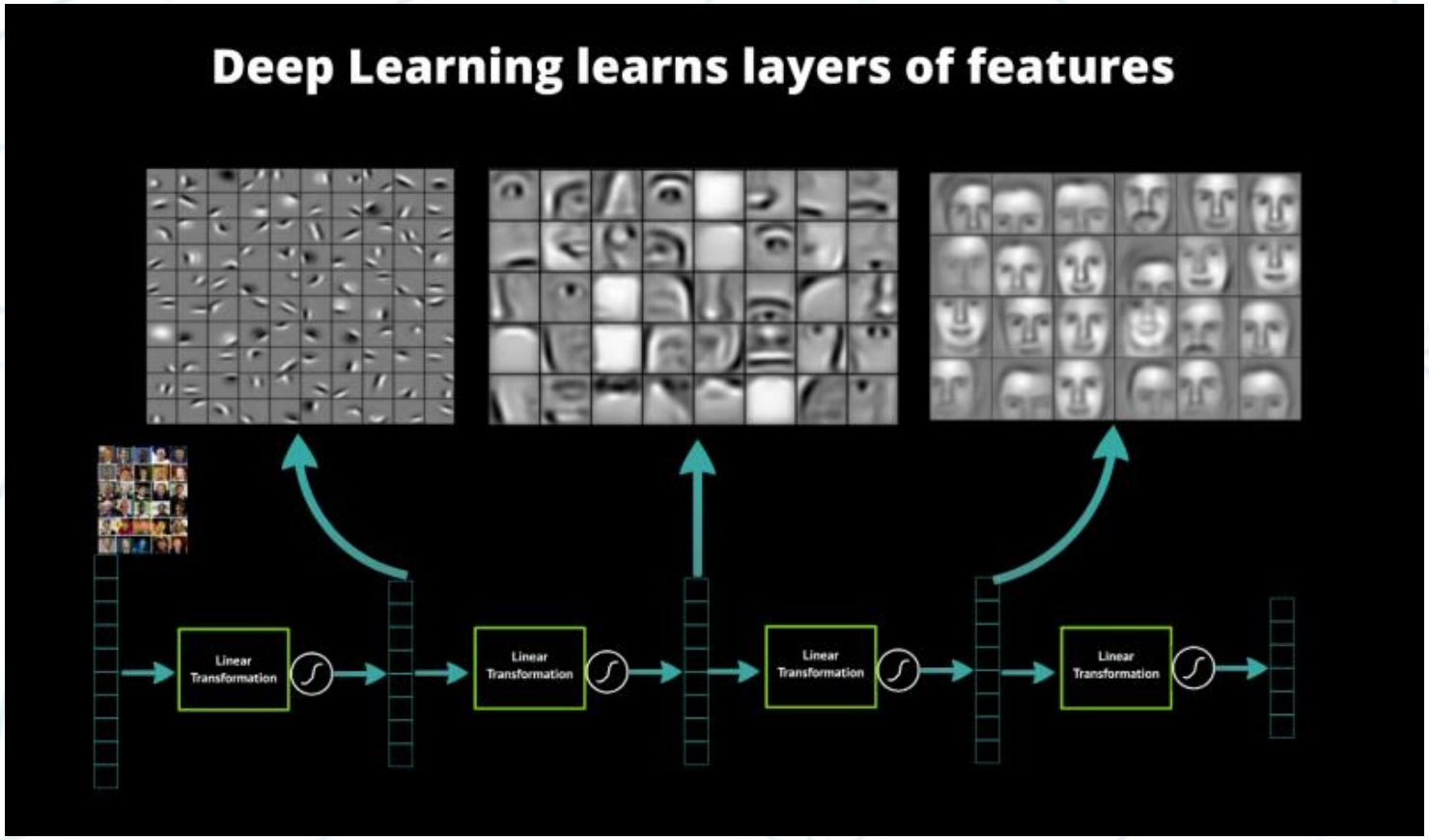
~5 FPS

YOLO & YOLOv2 (YOLO9000)

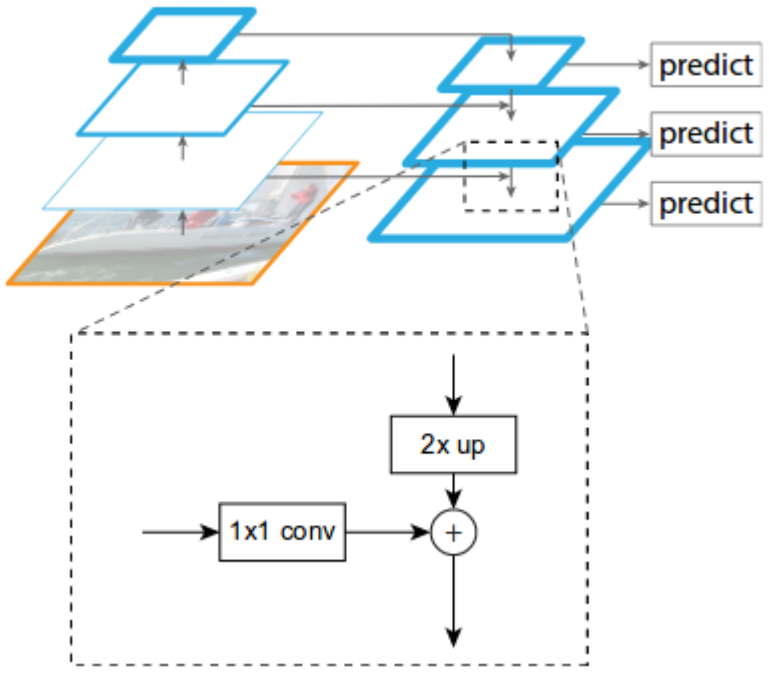


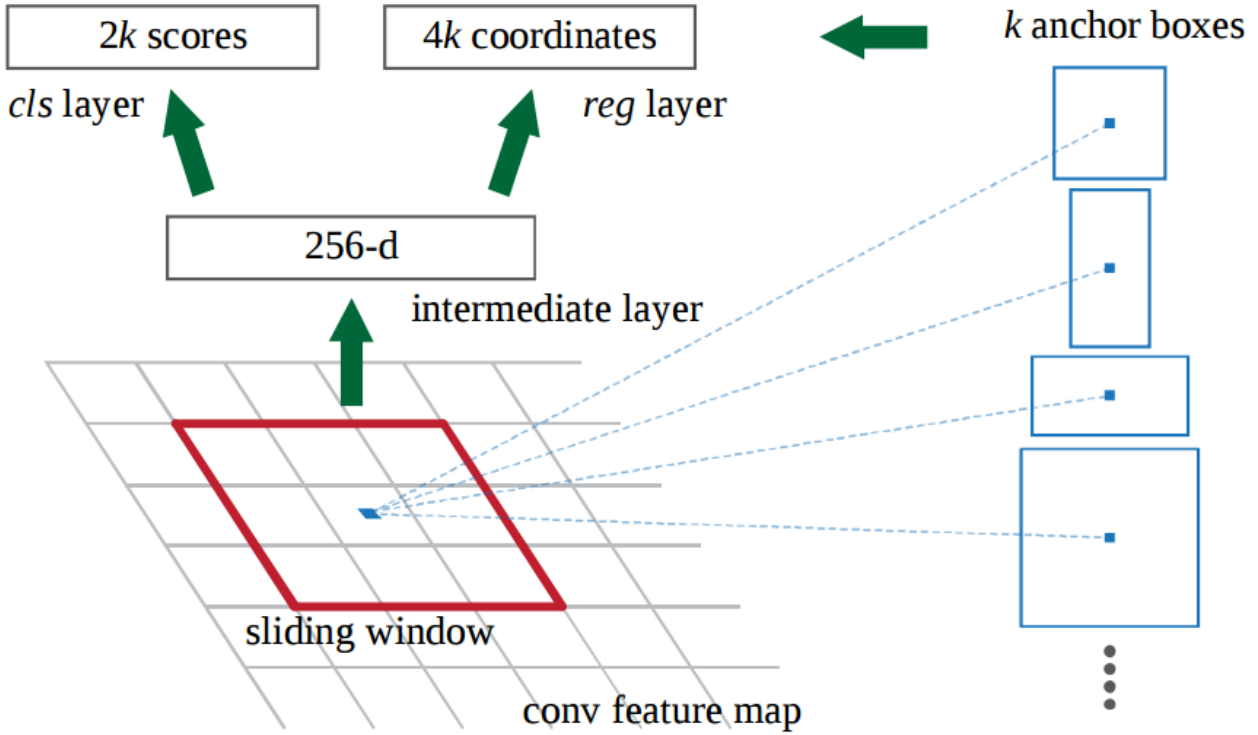
>40 FPS

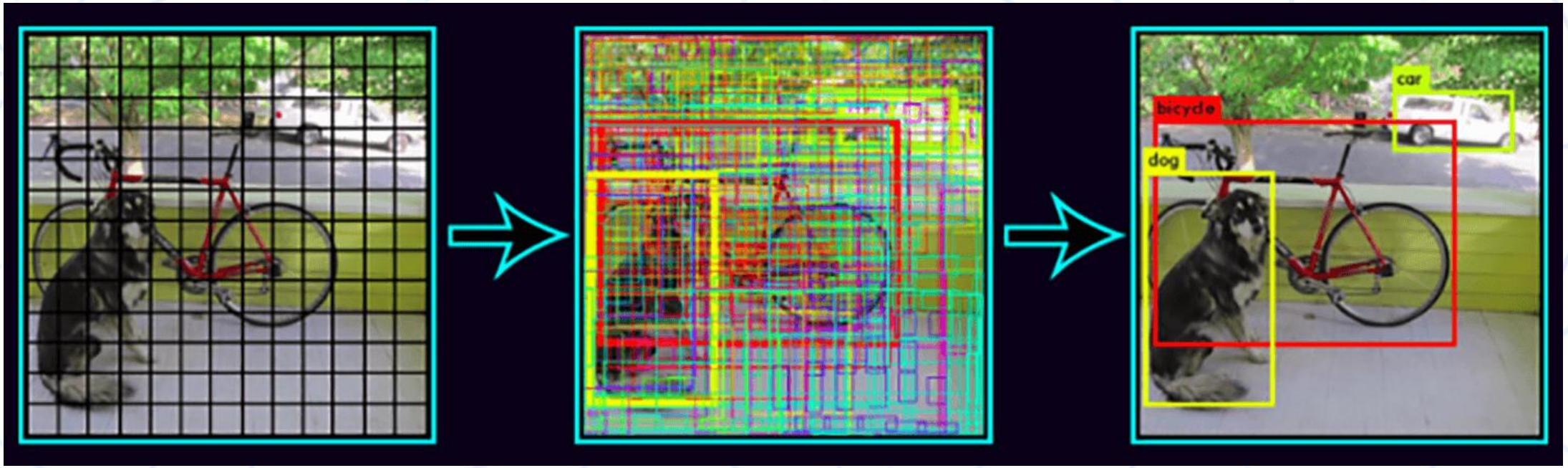
Representations

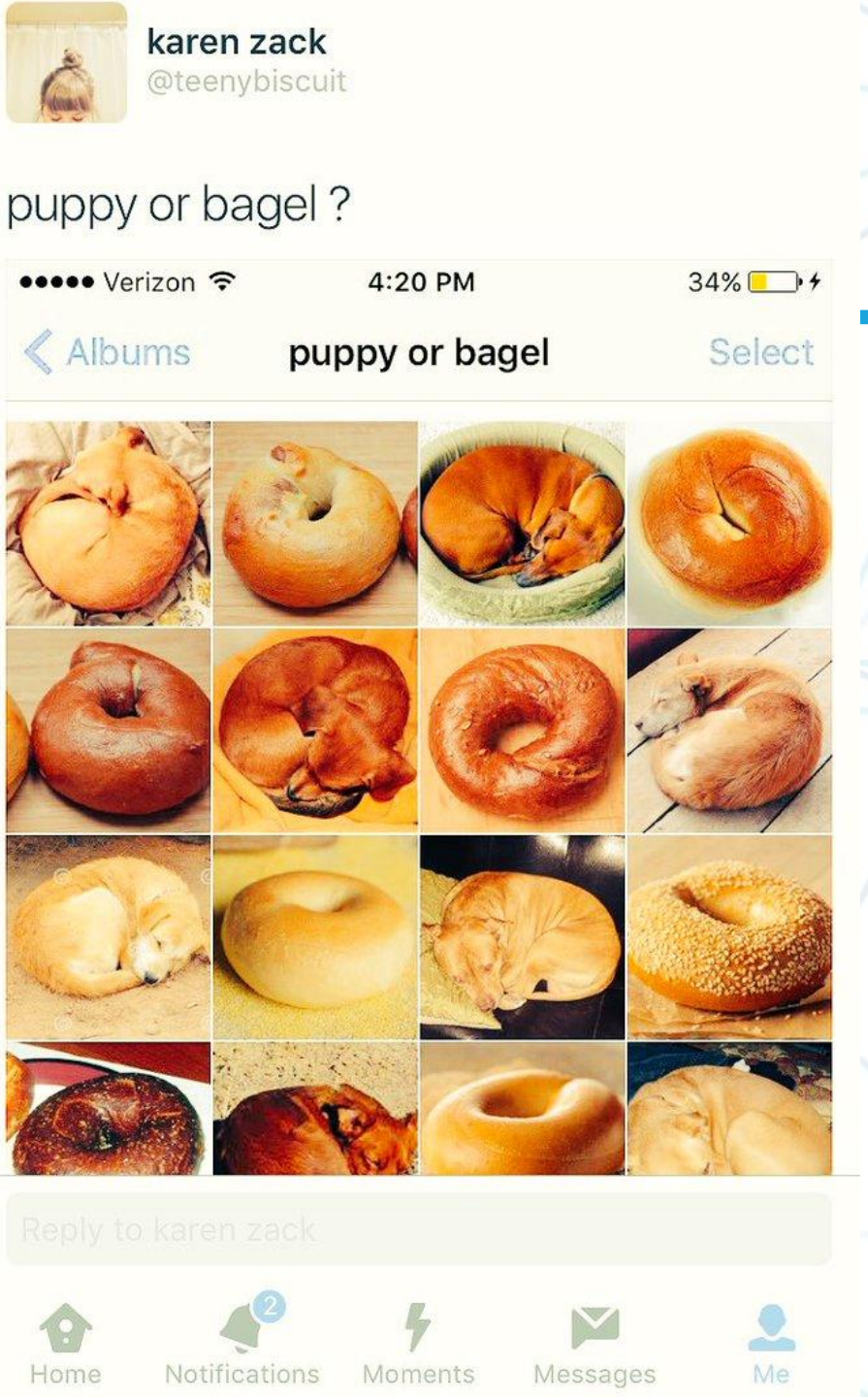


Feature Pyramid Network






bicyde
car
dog


karen zack
@teenybiscuit
puppy or bagel ?
Verizon
4:20 PM
34%
Albums
puppy or bagel
Select
Reply to karen zack
Home
Notifications
Moments
Messages
Me


Verizon
4:20 PM
76%
Albums
chihuahua or muffin
Select
@teenybiscuit

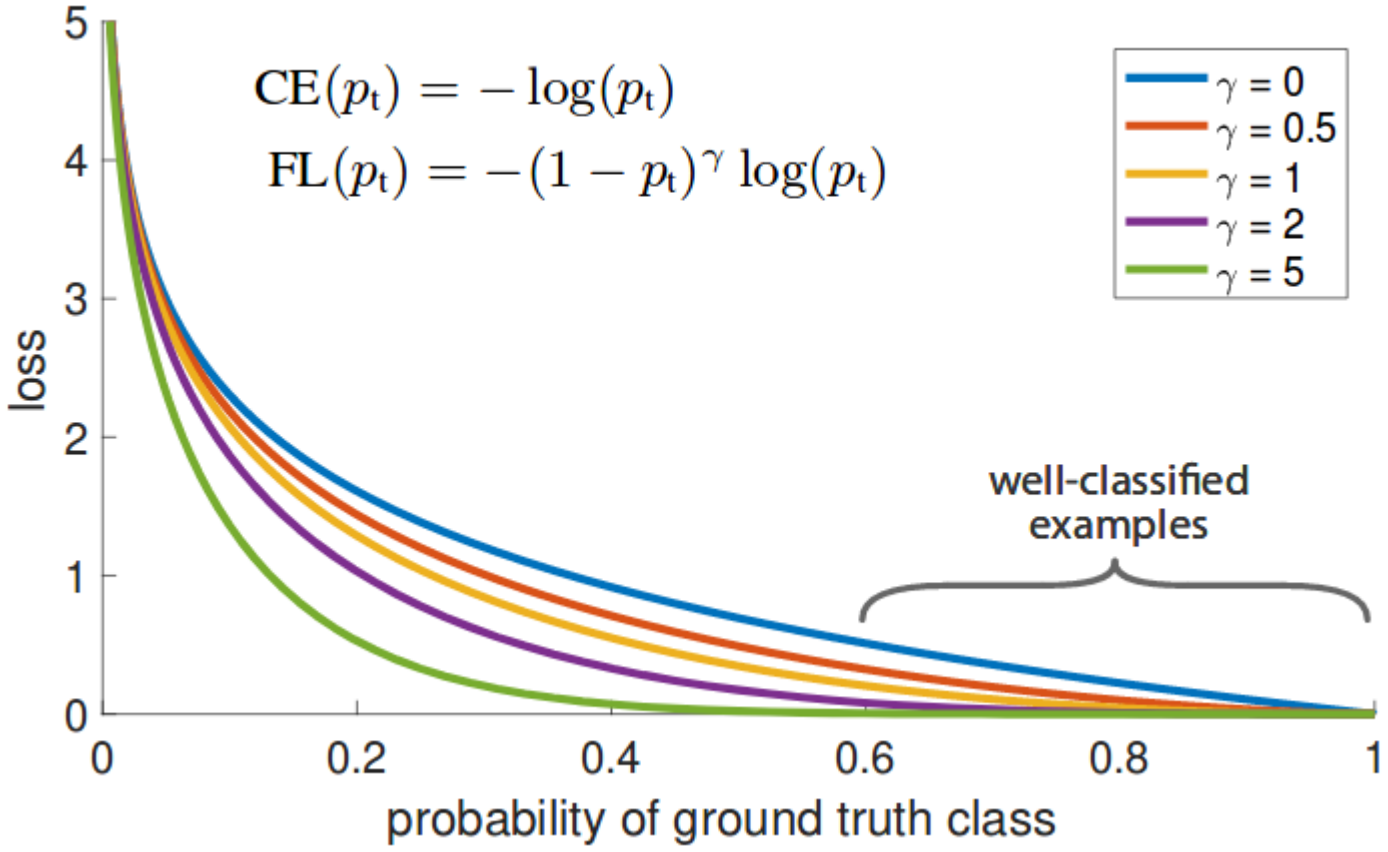
Focal loss

$$\mathrm{FL}(p_t) = -(1-p_t)^{\gamma}\log(p_t)$$

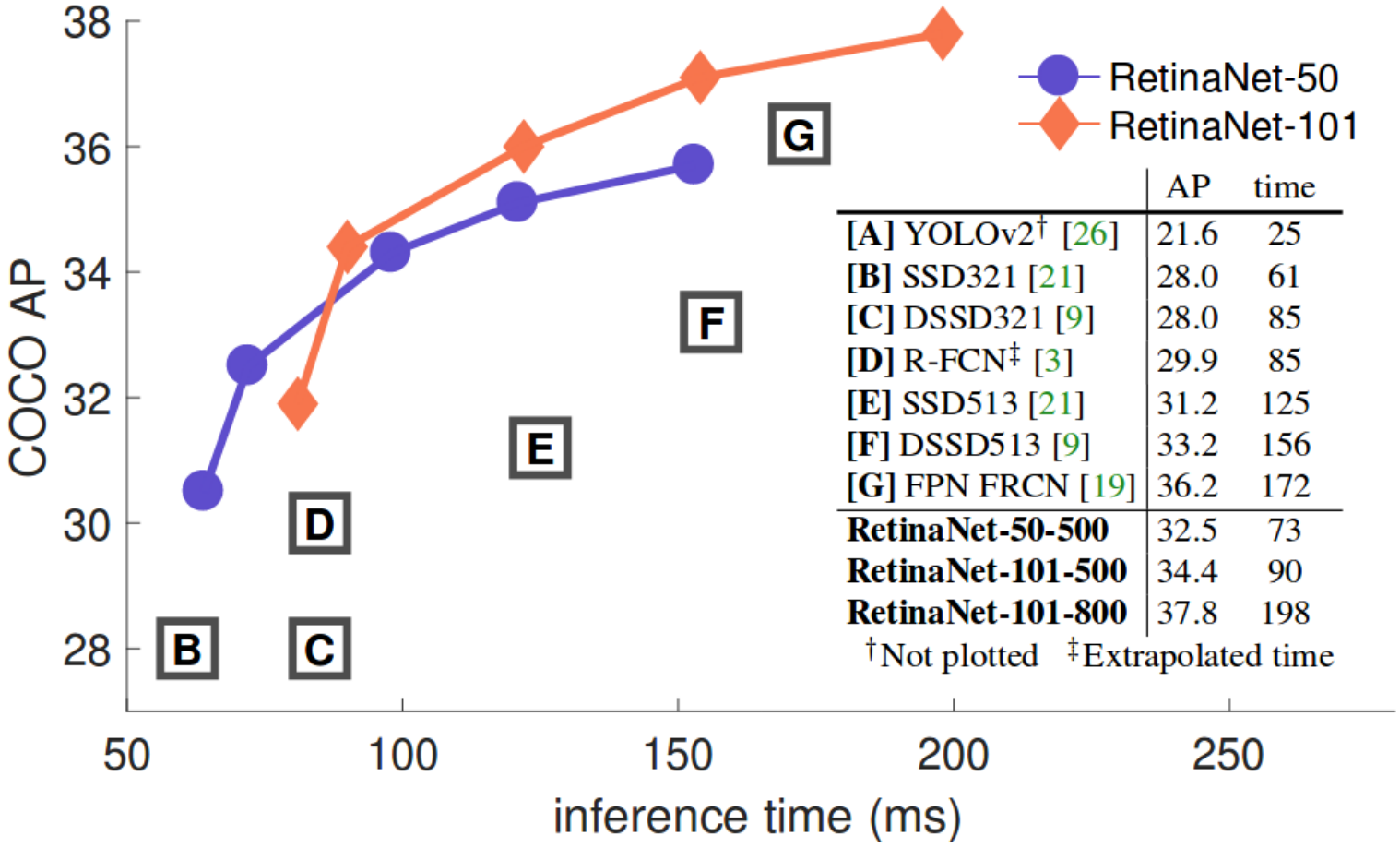
$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise,} \end{cases}$$

$$\mathrm{CE}(p_t) = -\log(p_t)$$

$$\mathrm{FL}(p_t) = -(1-p_t)^{\gamma}\log(p_t)$$



RetinaNet
Feature Pyramid Network + Focal Loss



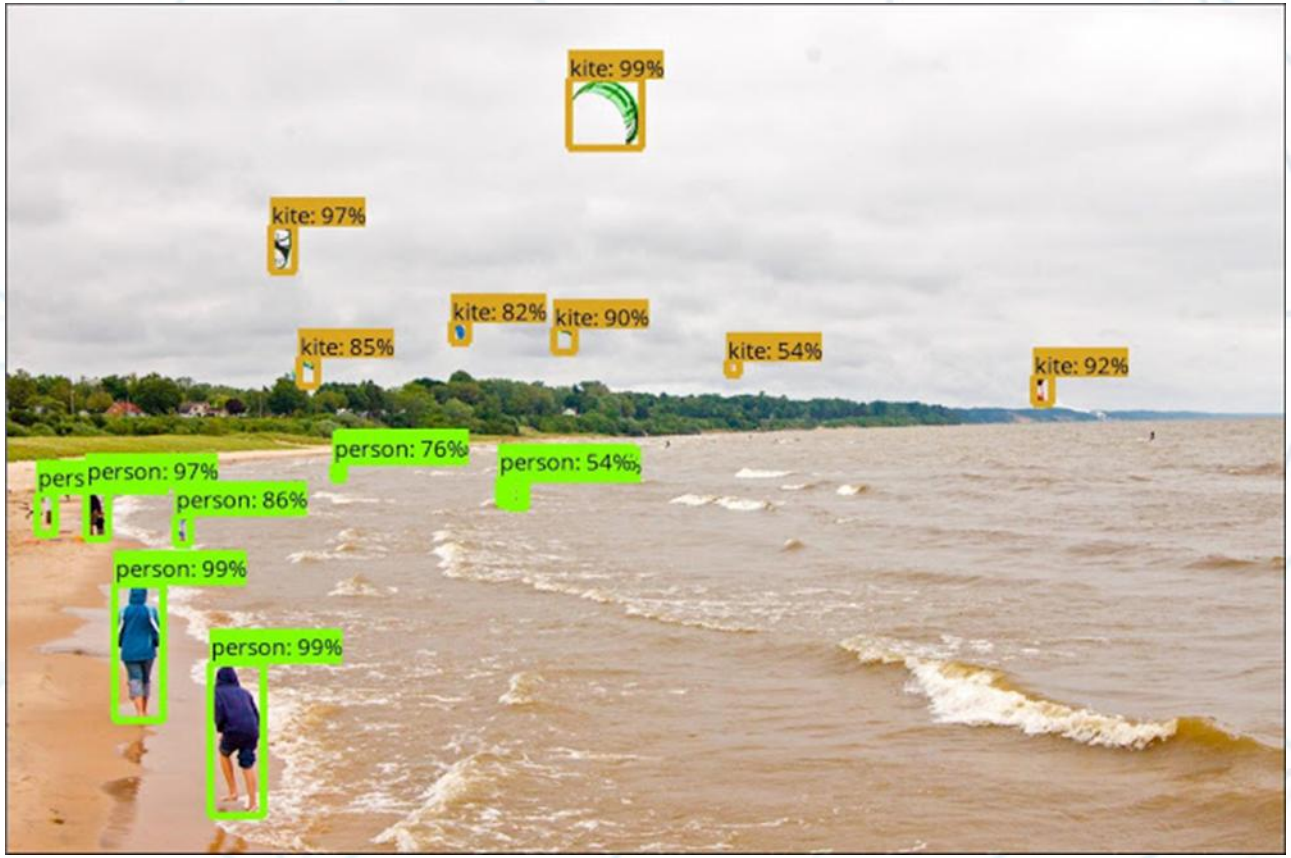
	AP	time
[A] YOLOv2† [26]	21.6	25
[B] SSD321 [21]	28.0	61
[C] DSSD321 [9]	28.0	85
[D] R-FCN‡ [3]	29.9	85
[E] SSD513 [21]	31.2	125
[F] DSSD513 [9]	33.2	156
[G] FPN FRCN [19]	36.2	172
RetinaNet-50-500	32.5	73
RetinaNet-101-500	34.4	90
RetinaNet-101-800	37.8	198

†Not plotted ‡Extrapolated time

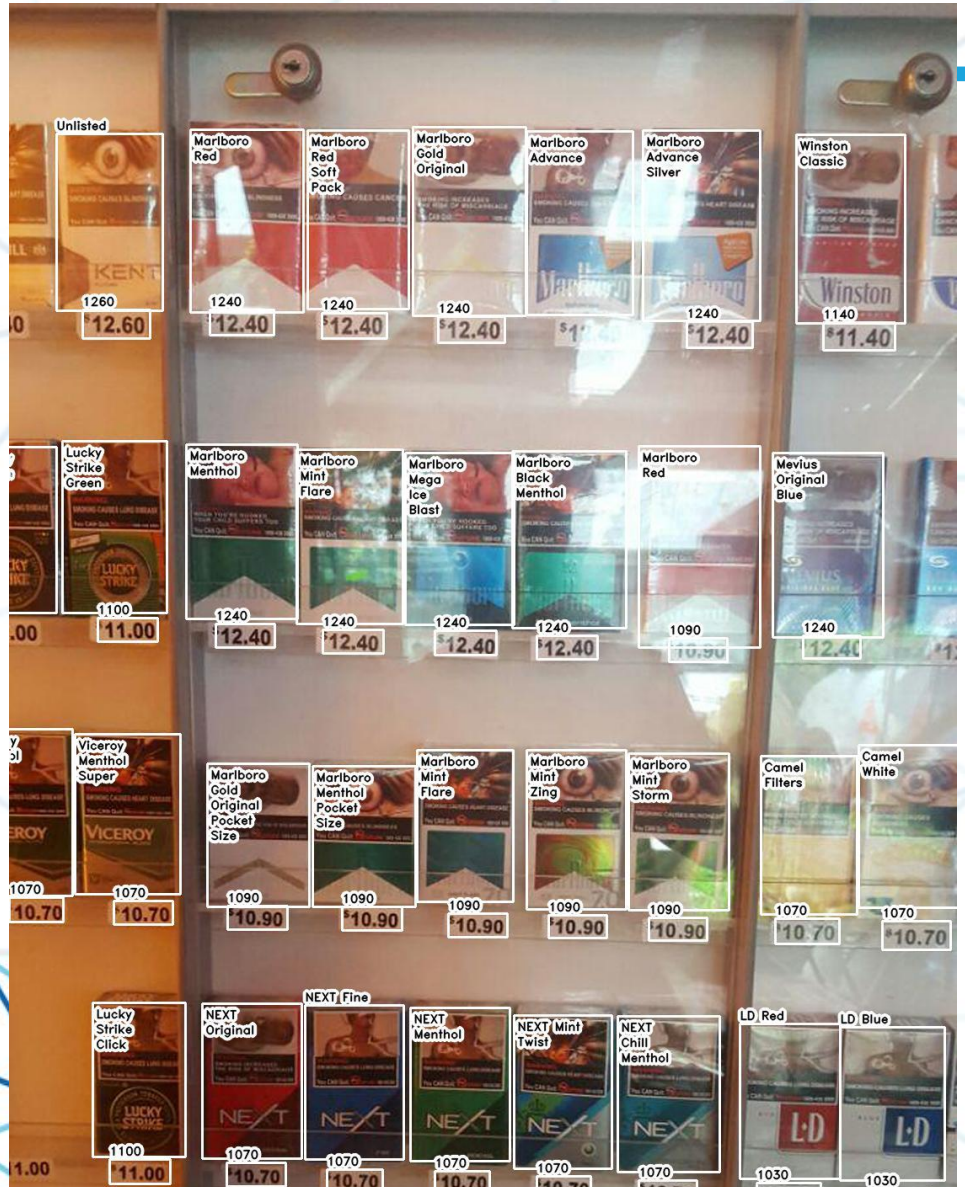
Tensorflow APIs

https://research.googleblog.com/2017/06/supercharge-your-computer-vision-models.html

- Several algorithms (e.g. Faster RCNN)
- Trained weights (i.e. ready models)
- Jupyter notebooks

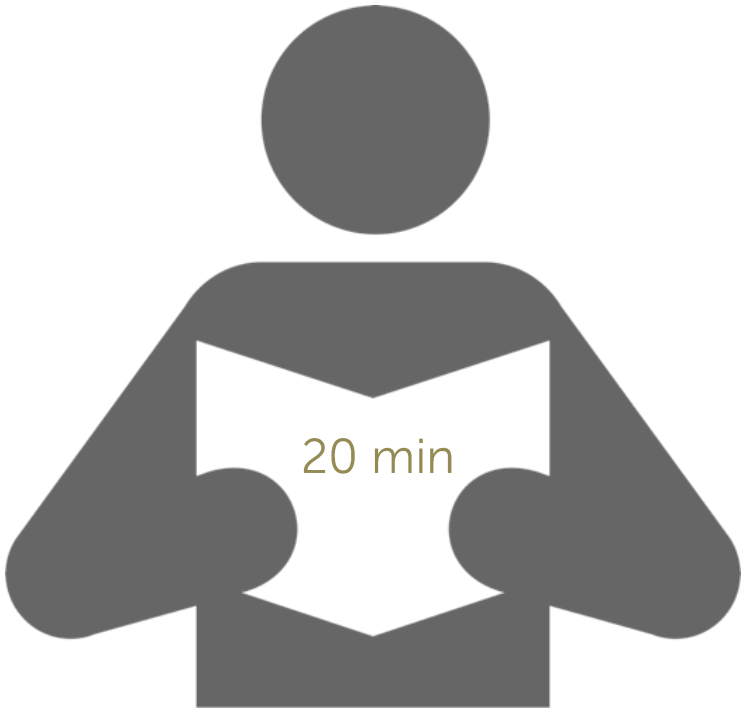


Detection @ Fornax



ShelfWise

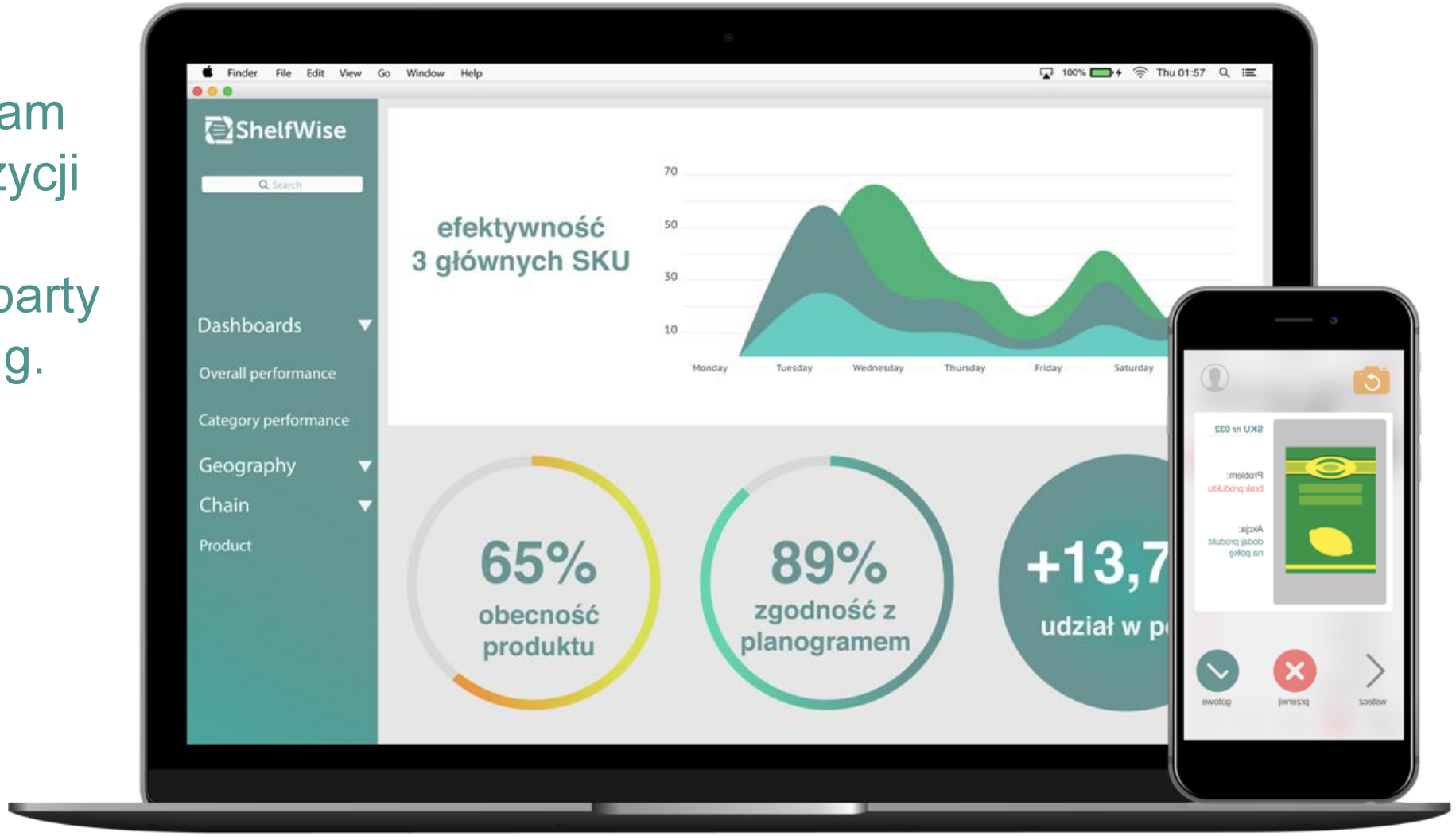
Zupełnie nowy standard pracy osób odpowiedzialnych za ułożenie produktów na półce



Business Case

ShelfWise

Inteligentny program do analizy ekspozycji półki sklepowej wykorzystujący oparty o machine learning.



FORNAX

Rafał Cycoń
CTO / co-founder

rafal.cycon@fornax.ai

WWW.FORNAX.AI