Deep Learning for Object Detection in Video Surveillance

Michelangelo Fiore & Florian Matusek

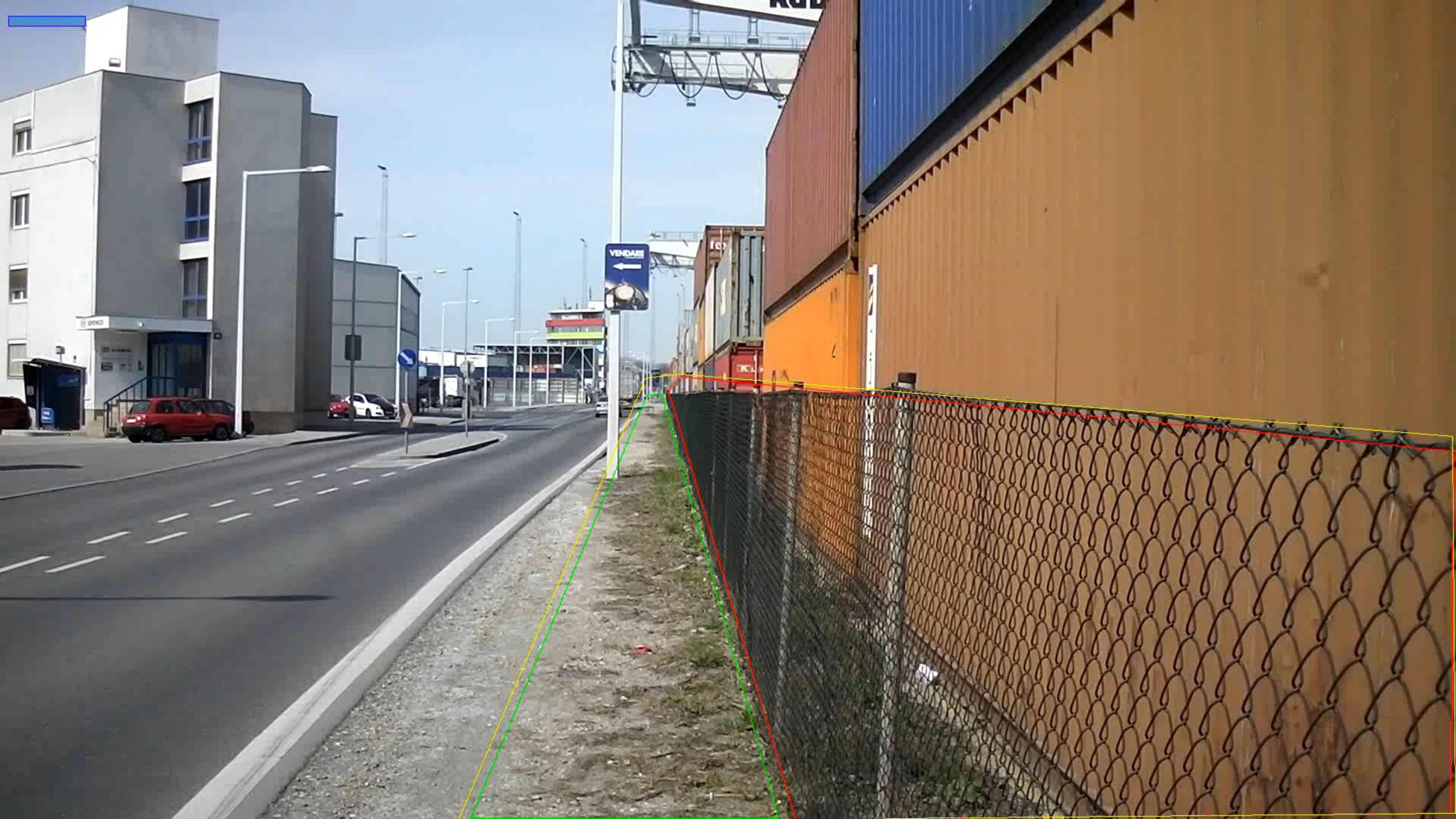
KiwiSecurity
AUTOMATING VIDEO SURVEILLANCE

Number of video surveillance cameras world-wide

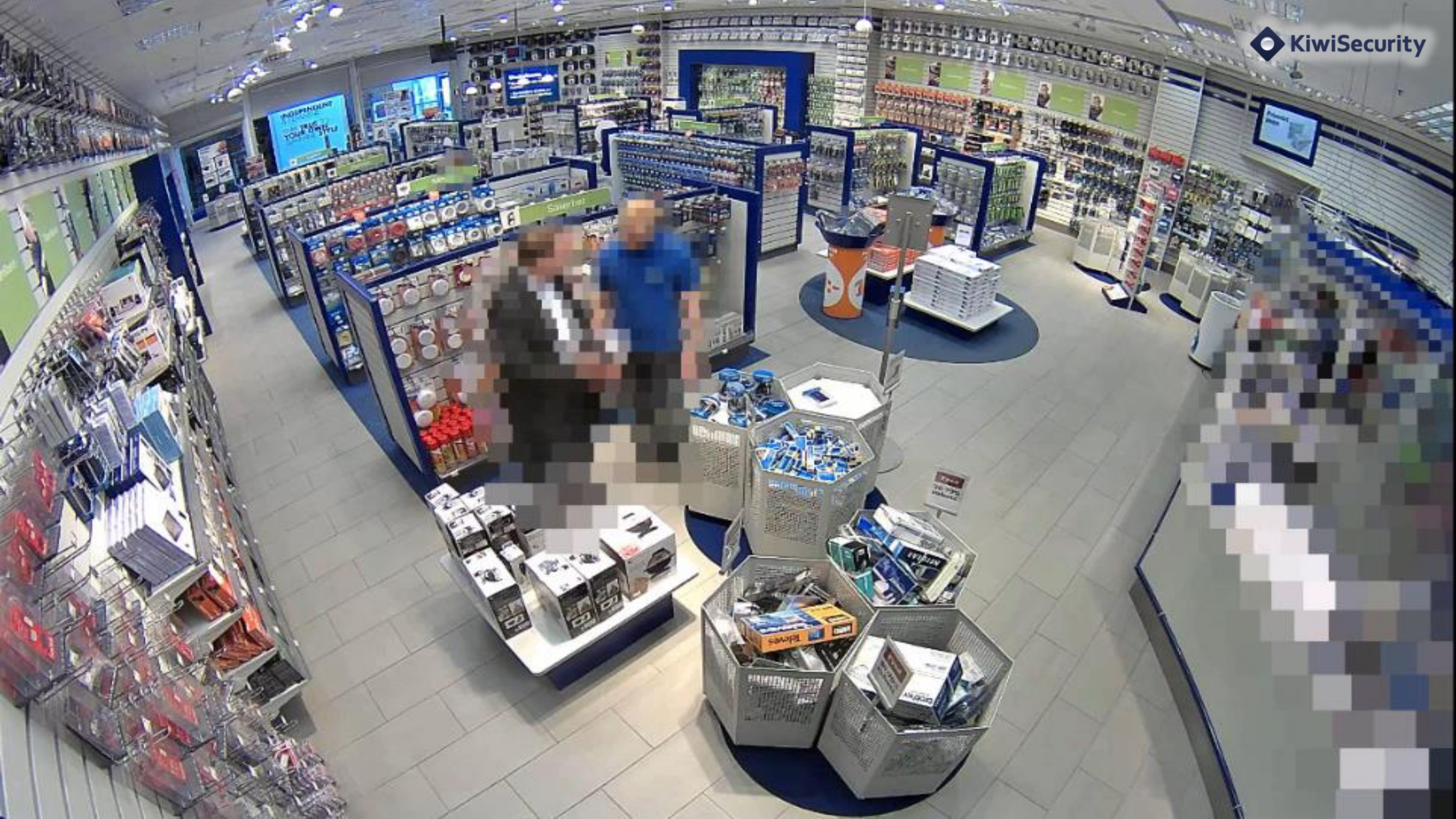
300m+

Incidents missed by security operators after 20 minutes

95%



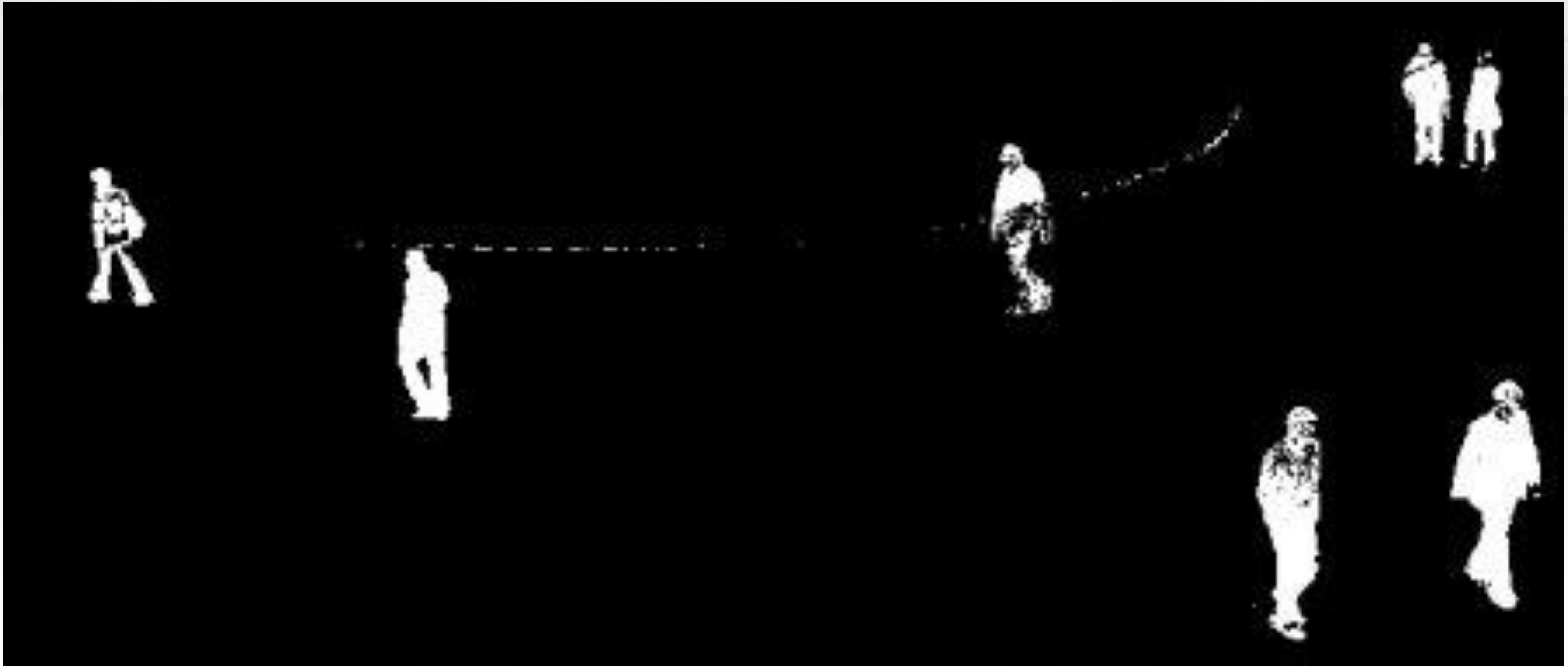
VENDARE


KiwiSecurity

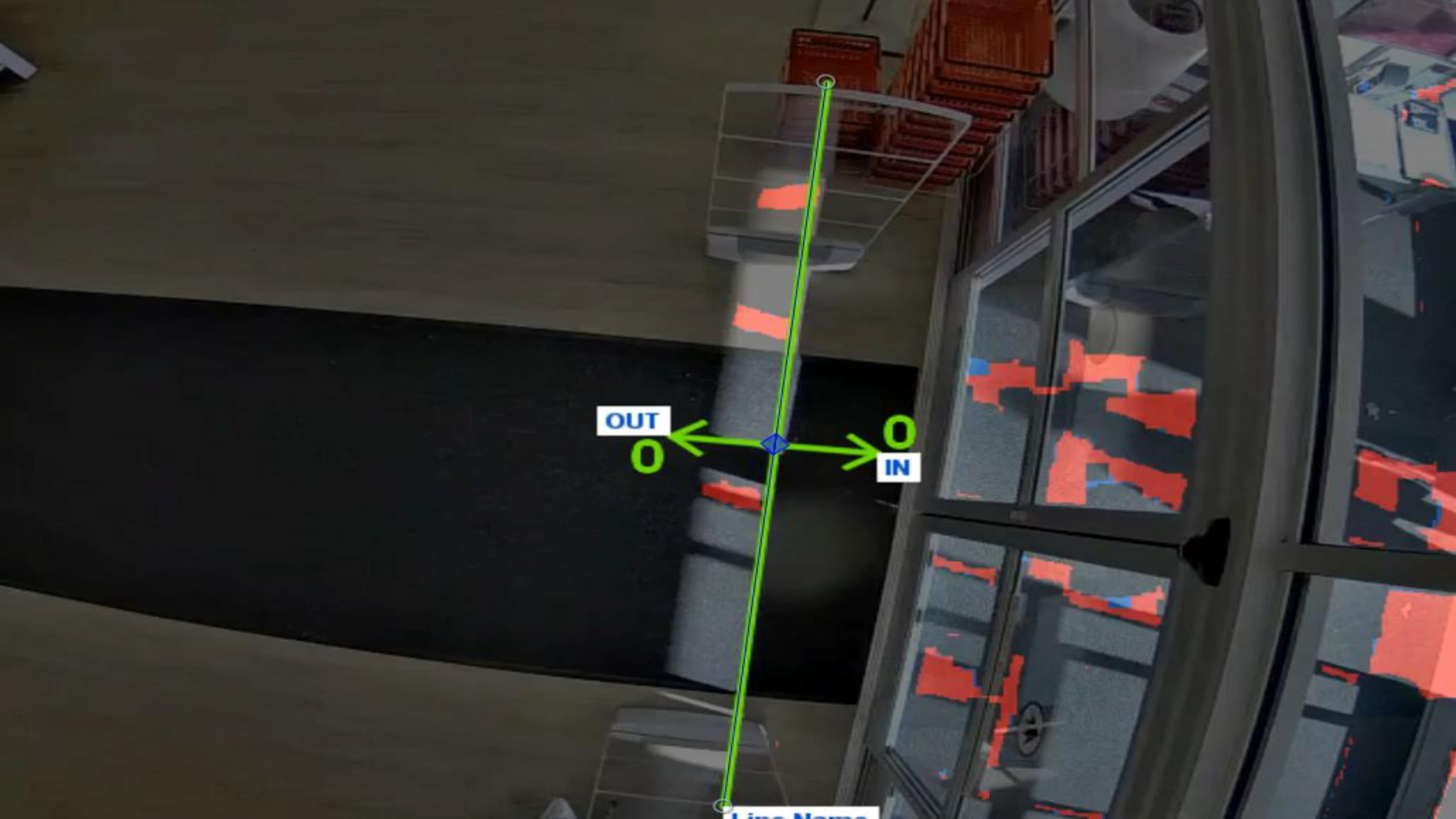
So where do we go from here?

Background subtraction

Background subtraction

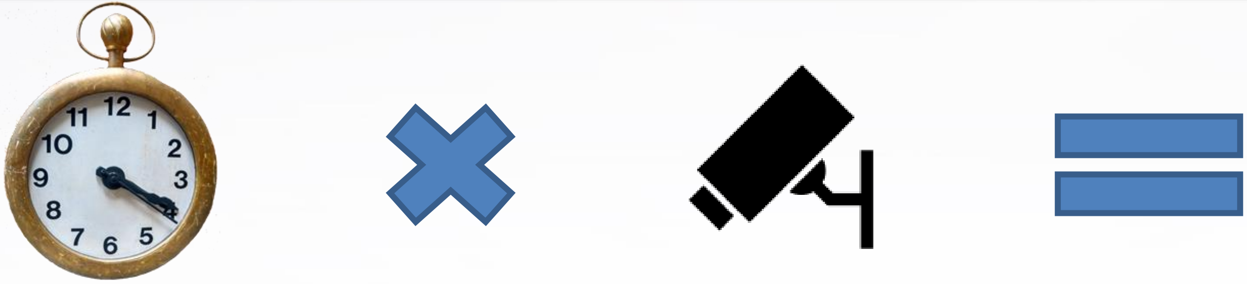


Some Pitfalls


OUT
0
0
IN
Line Name

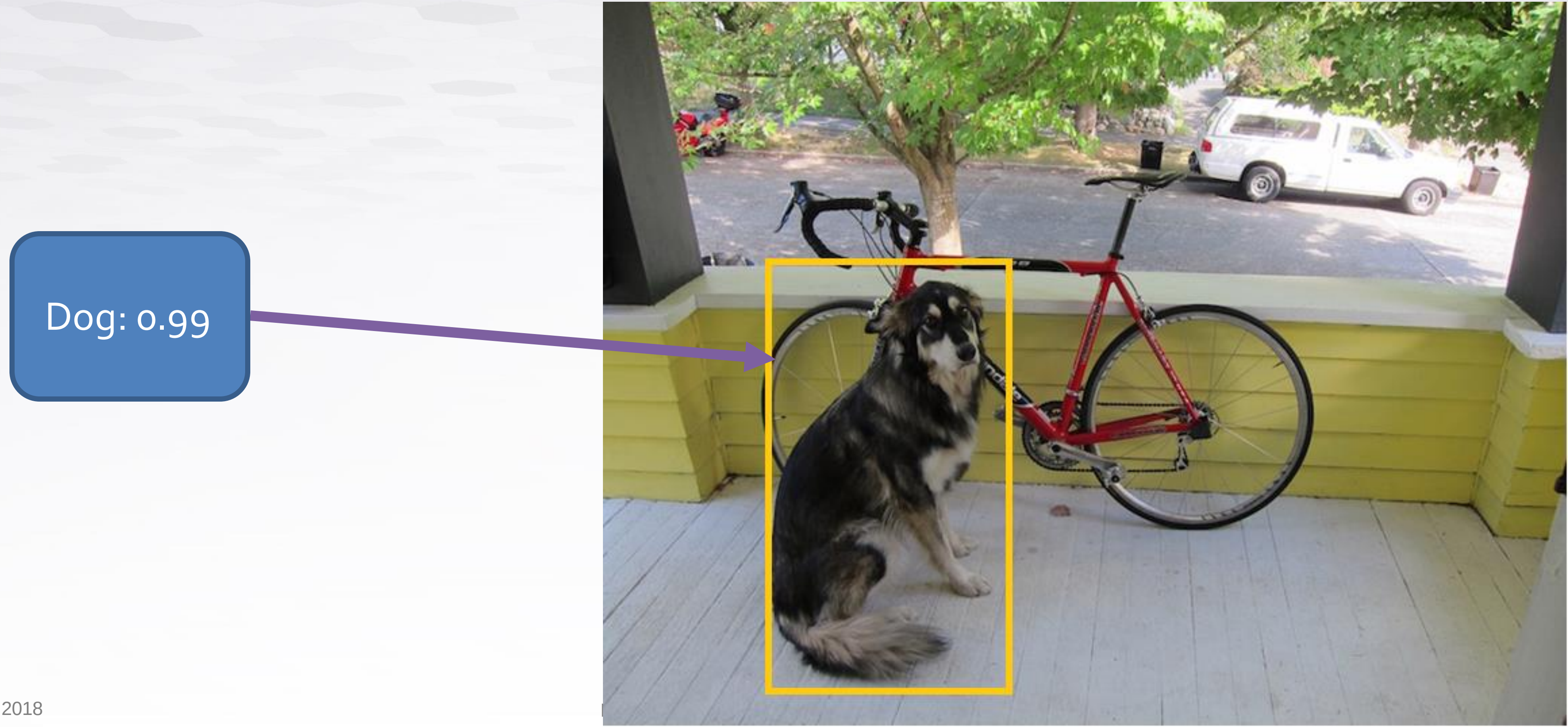
Many Parameters

Hard to Tune



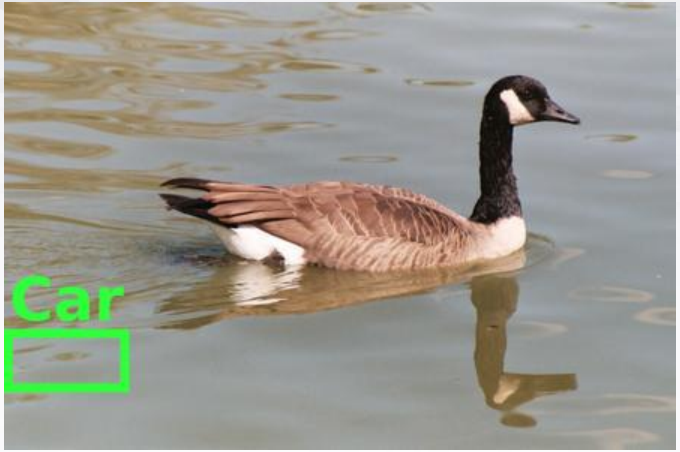
Too Much

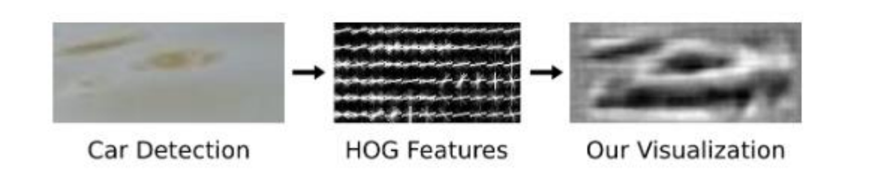
Machine Learning Approaches



Dog: 0.99

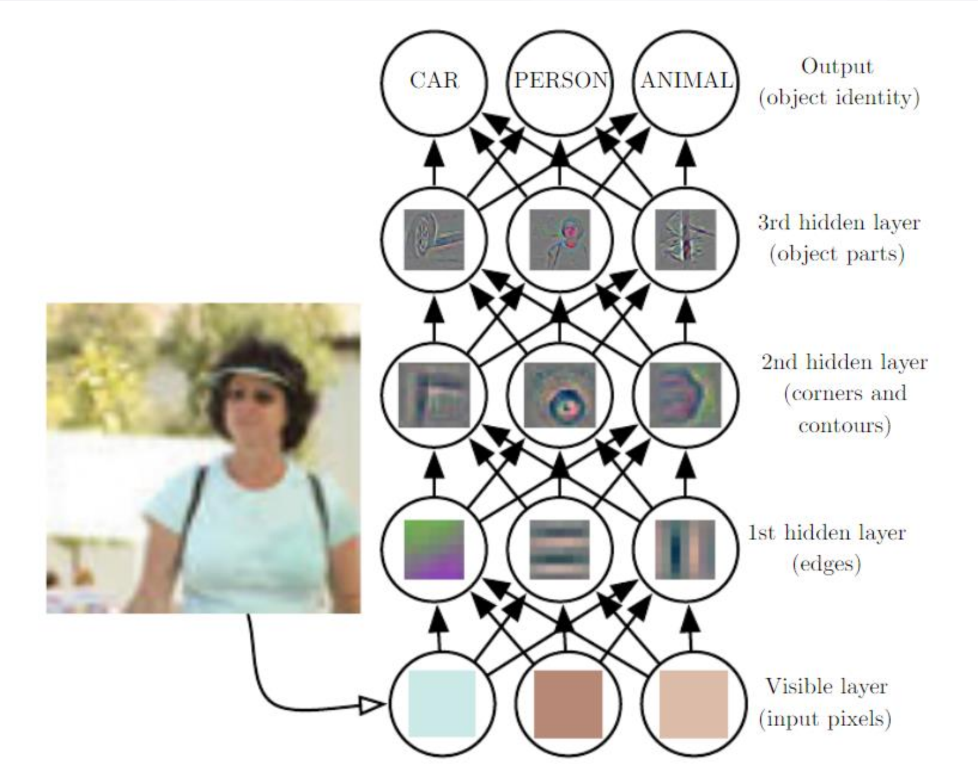
Choosing the Right Features



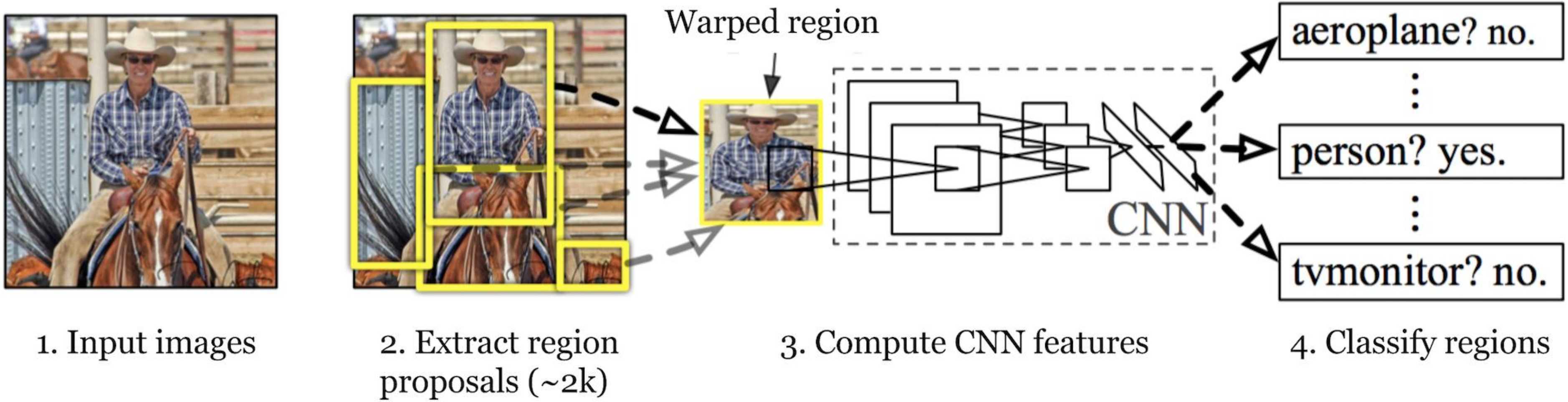


C. Vondrick, A. Khosla, T. Malisiewicz, A. Torralba. "HOGgles: Visualizing Object Detection Features" *International Conference on Computer Vision* (ICCV), Sydney, Australia, December 2013.

Feature Extraction With Deep Learning

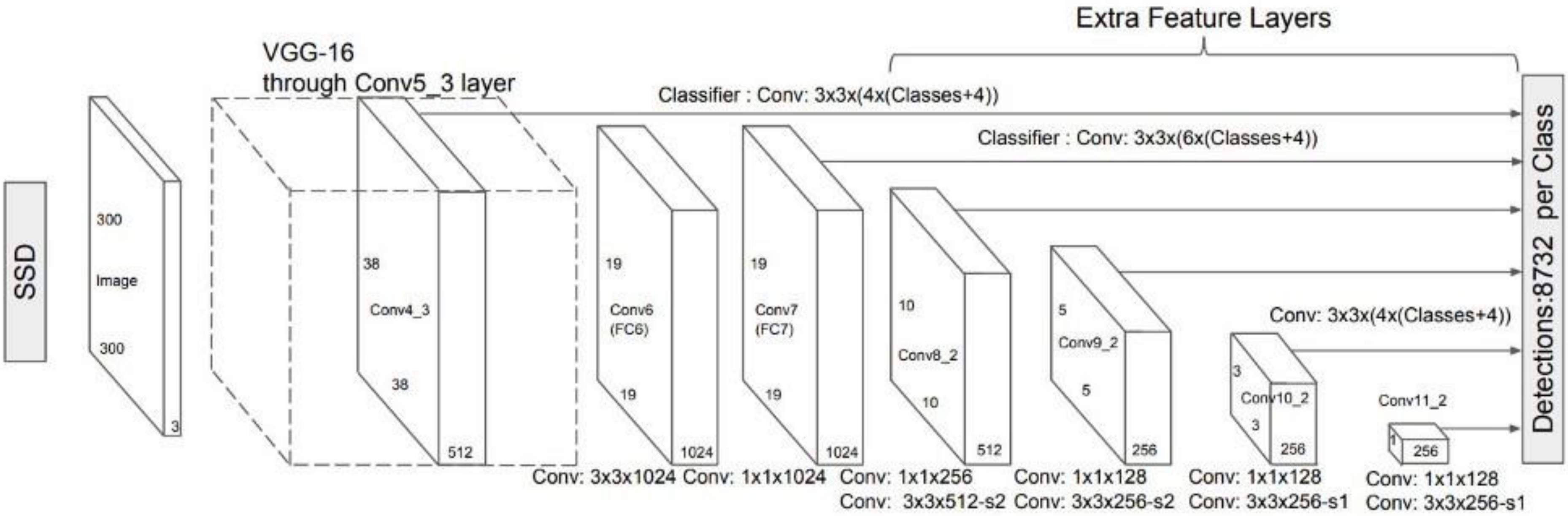


Object Detection with Deep Learning: Region Based Approach



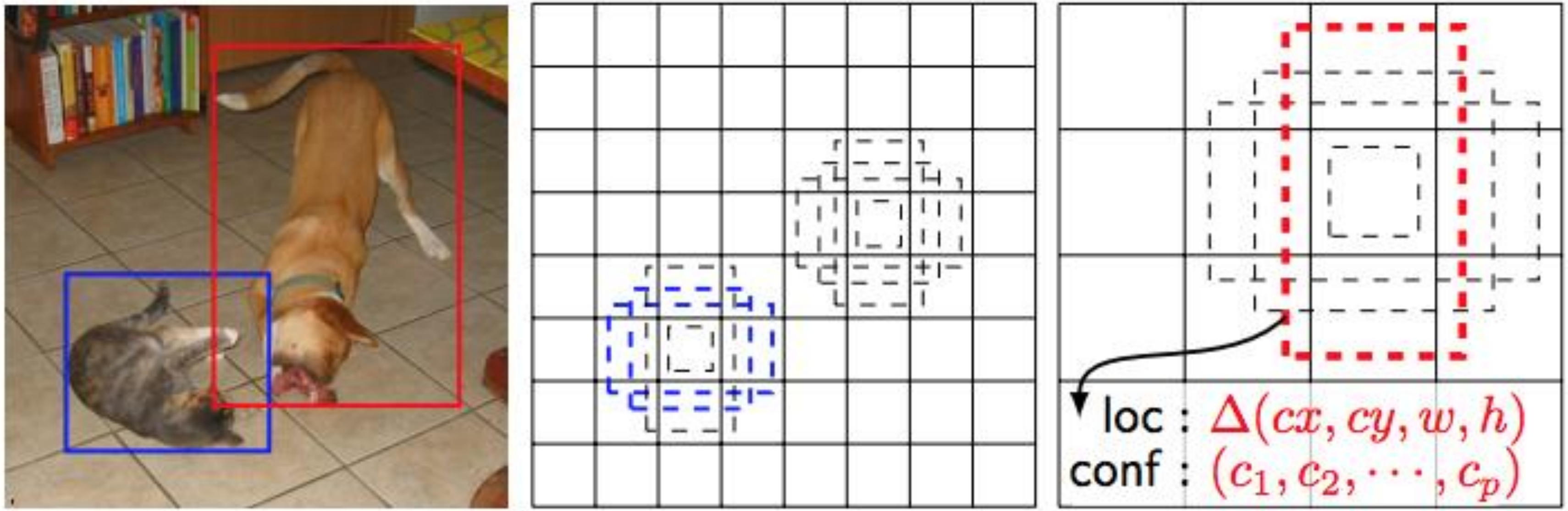
Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2014.

Object Detection with Deep Learning: Single Shot Approach



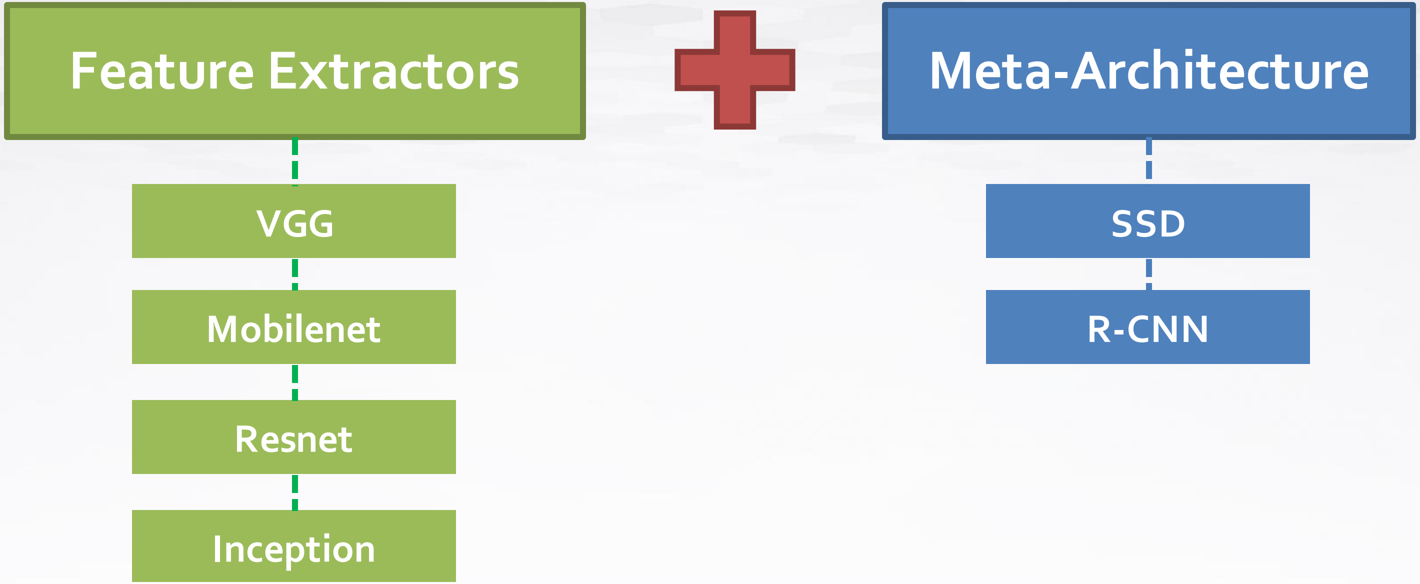
Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, 2016

Detection at Different Scales


loc : $\Delta(cx, cy, w, h)$
conf : $(c_1, c_2, \cdots, c_p)$

(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

Choosing a Detector

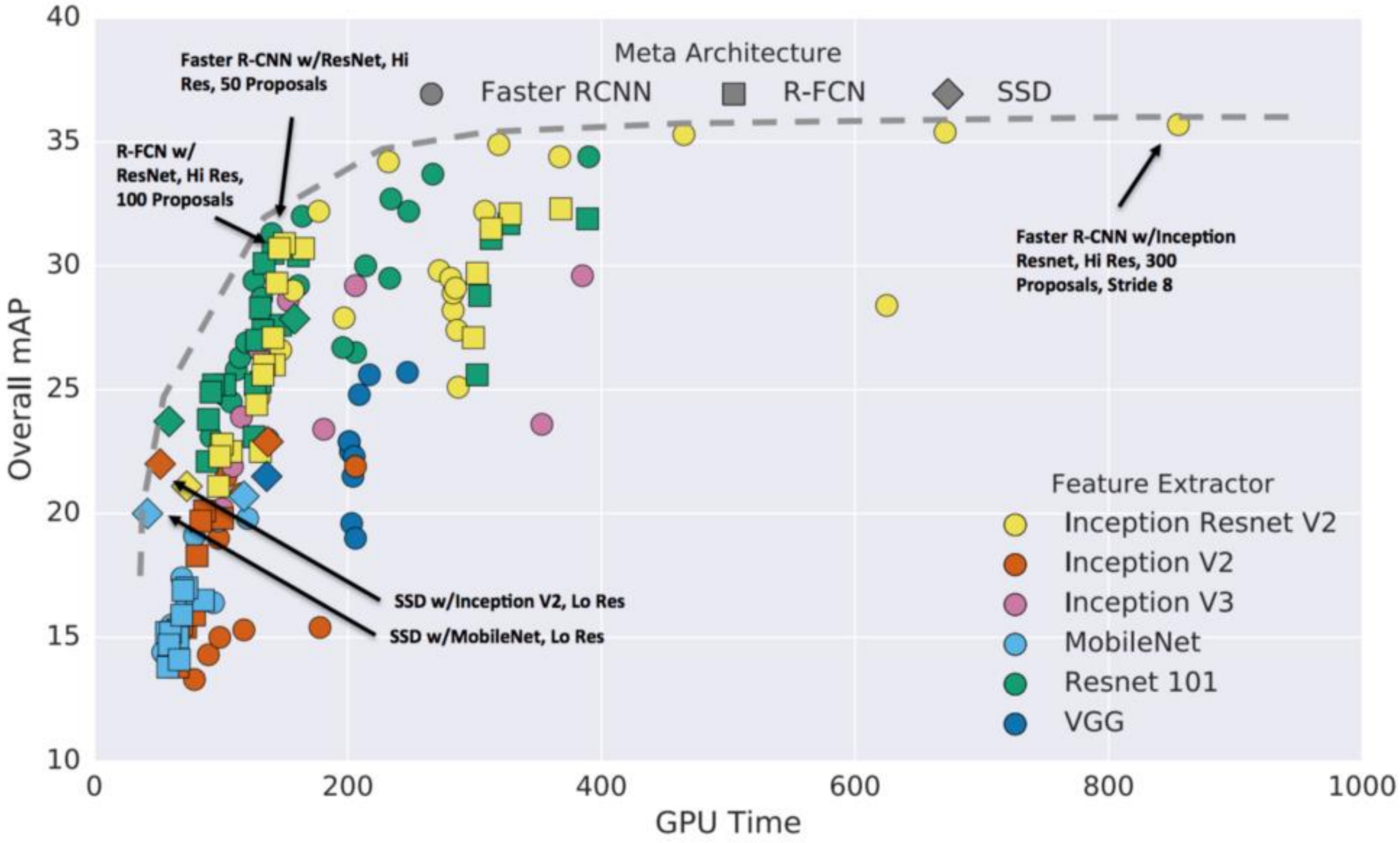


Feature Extractors

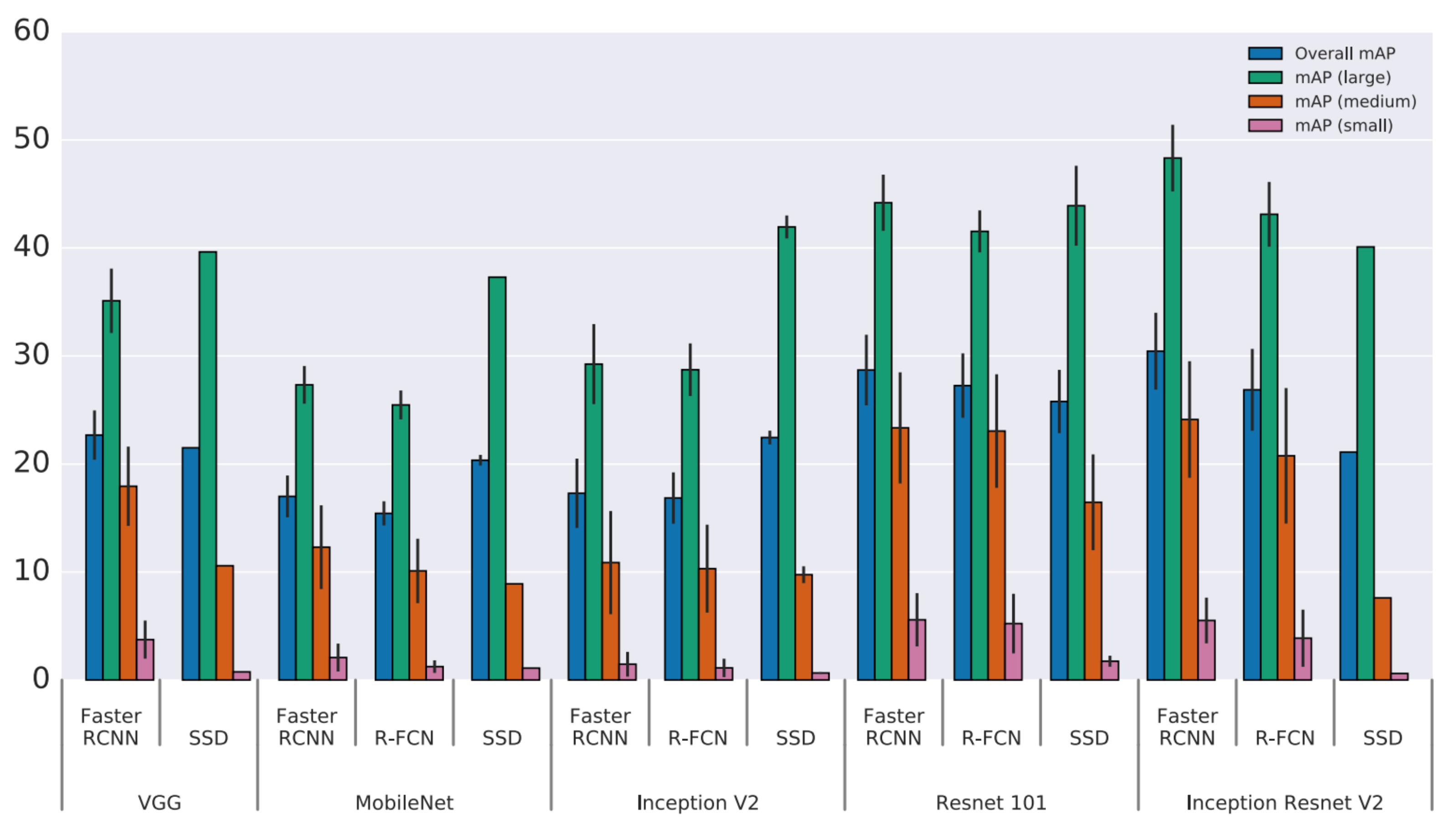
- VGG
- Mobilenet
- Resnet
- Inception

Meta-Architecture

- SSD
- R-CNN

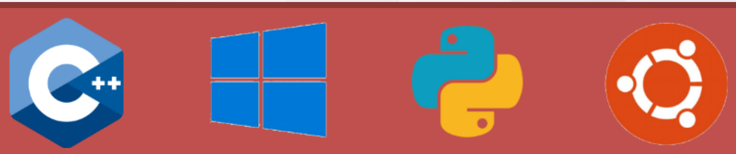


Huang, Jonathan, et al. "Speed/accuracy trade-offs for modern convolutional object detectors." *IEEE CVPR*. 2017.


60
50
40
30
20
10
0
Overall mAP
mAP (large)
mAP (medium)
mAP (small)
Faster RCNN
SSD
Faster RCNN
R-FCN
SSD
Faster RCNN
R-FCN
SSD
Faster RCNN
R-FCN
SSD
Faster RCNN
R-FCN
SSD
VGG
MobileNet
Inception V2
Resnet 101
Inception Resnet V2

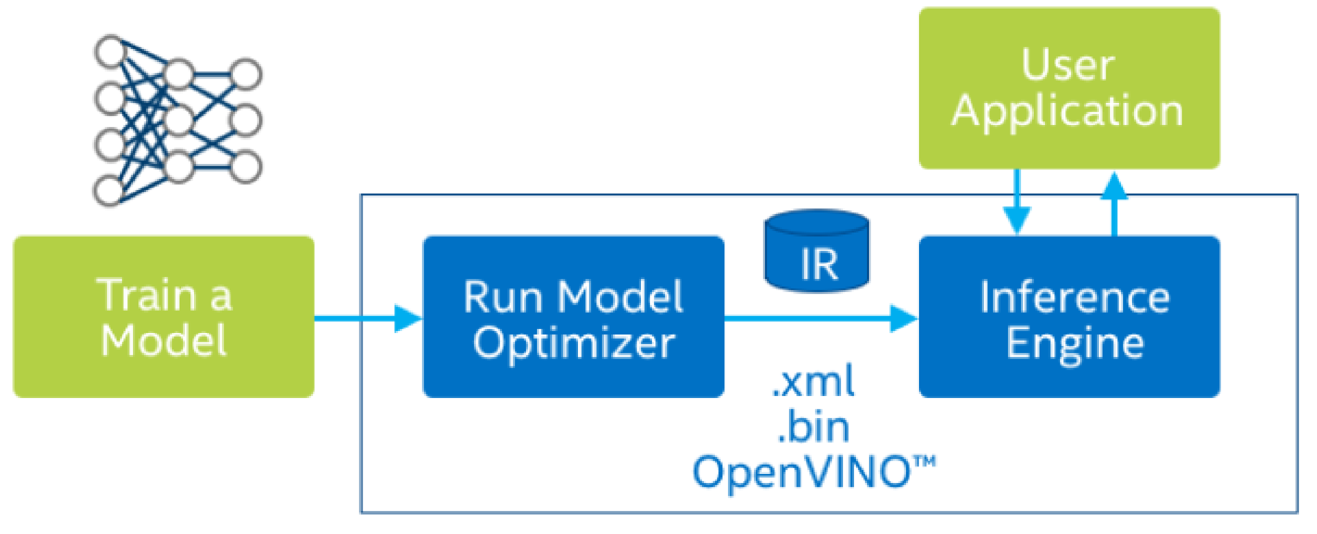
Understanding Your Requirements

Our Frameworks: Apache MXNet

OS and Language support	
Documentation	Good API and GLUON
Models Support	Many Object Detection Models
Performances	Fast Performances on Training and Inference

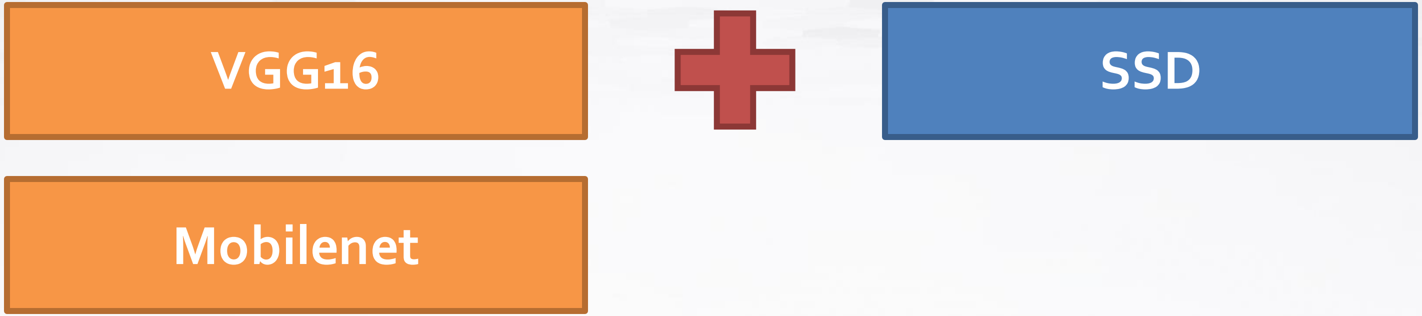
https://mxnet.apache.org

Our Frameworks: Intel OpenVINO



https://software.intel.com/en-us/openvino-toolkit

Our Detector



VGG16

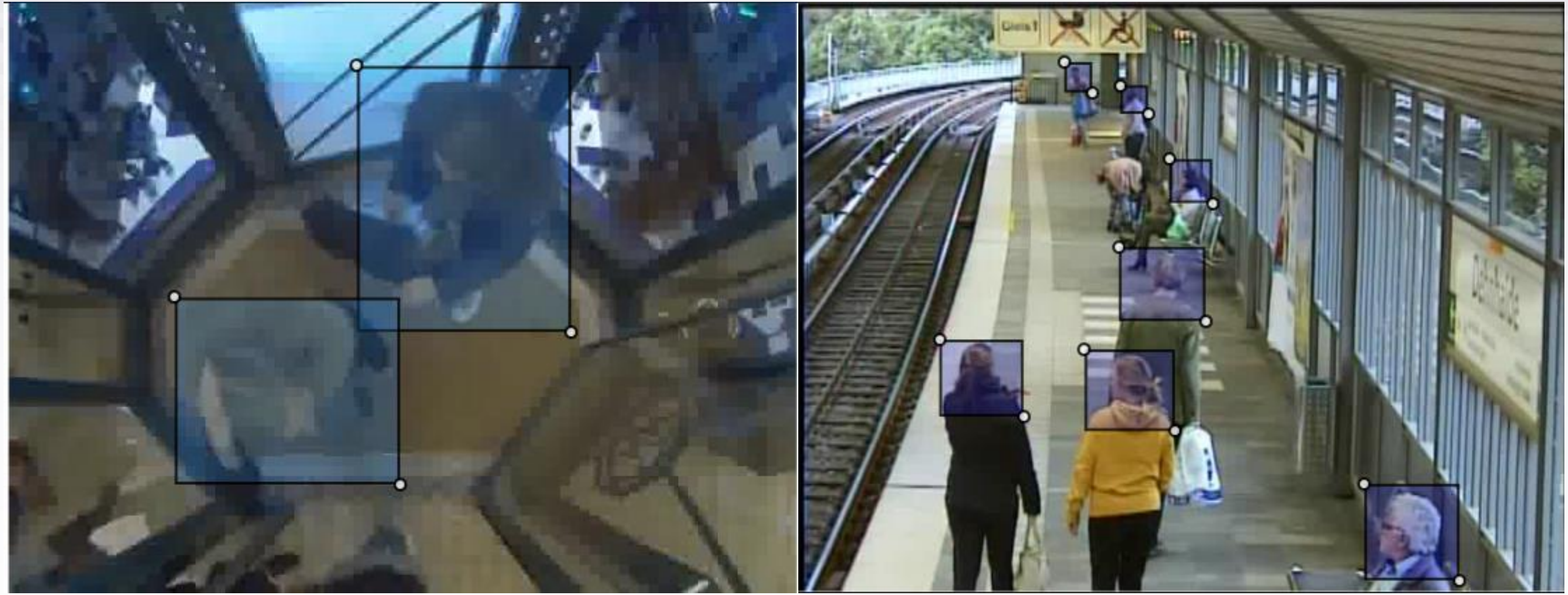
+

SSD

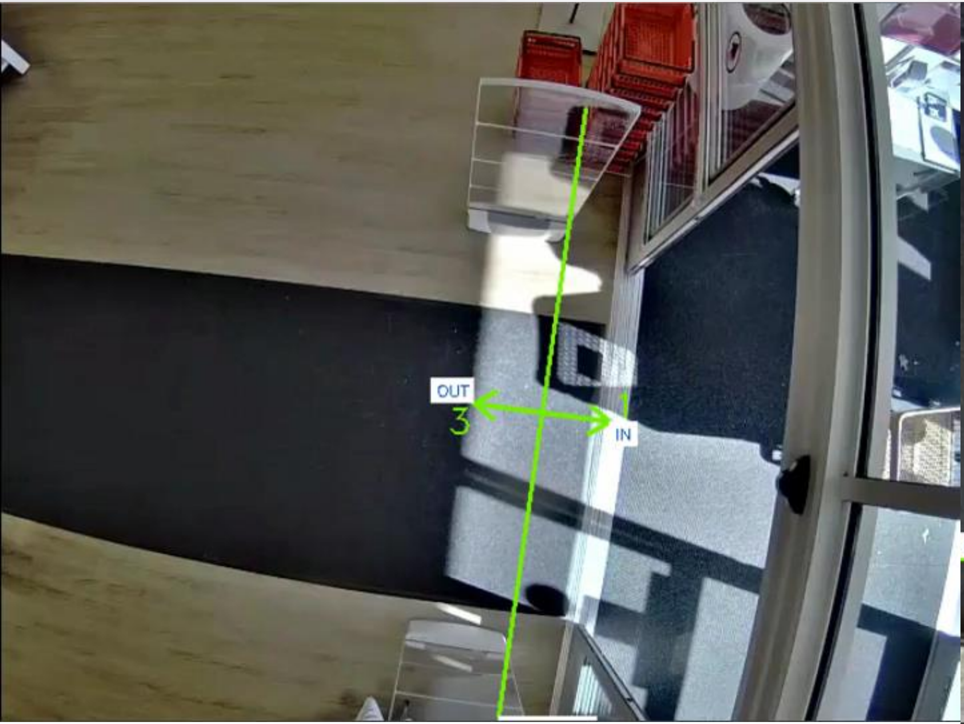
Mobilenet

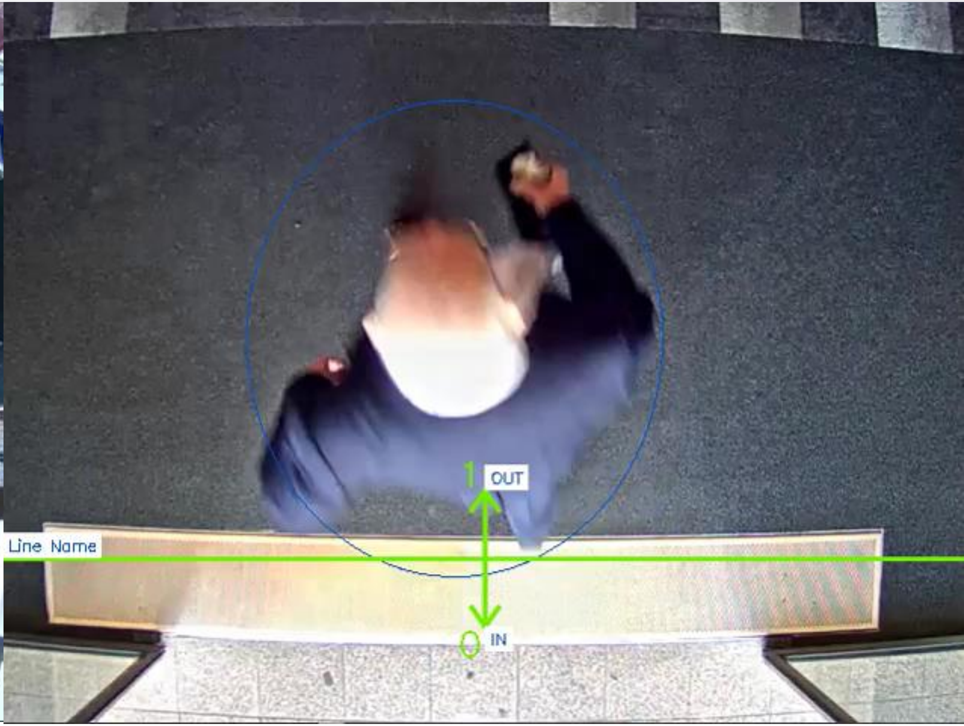
Building a Dataset

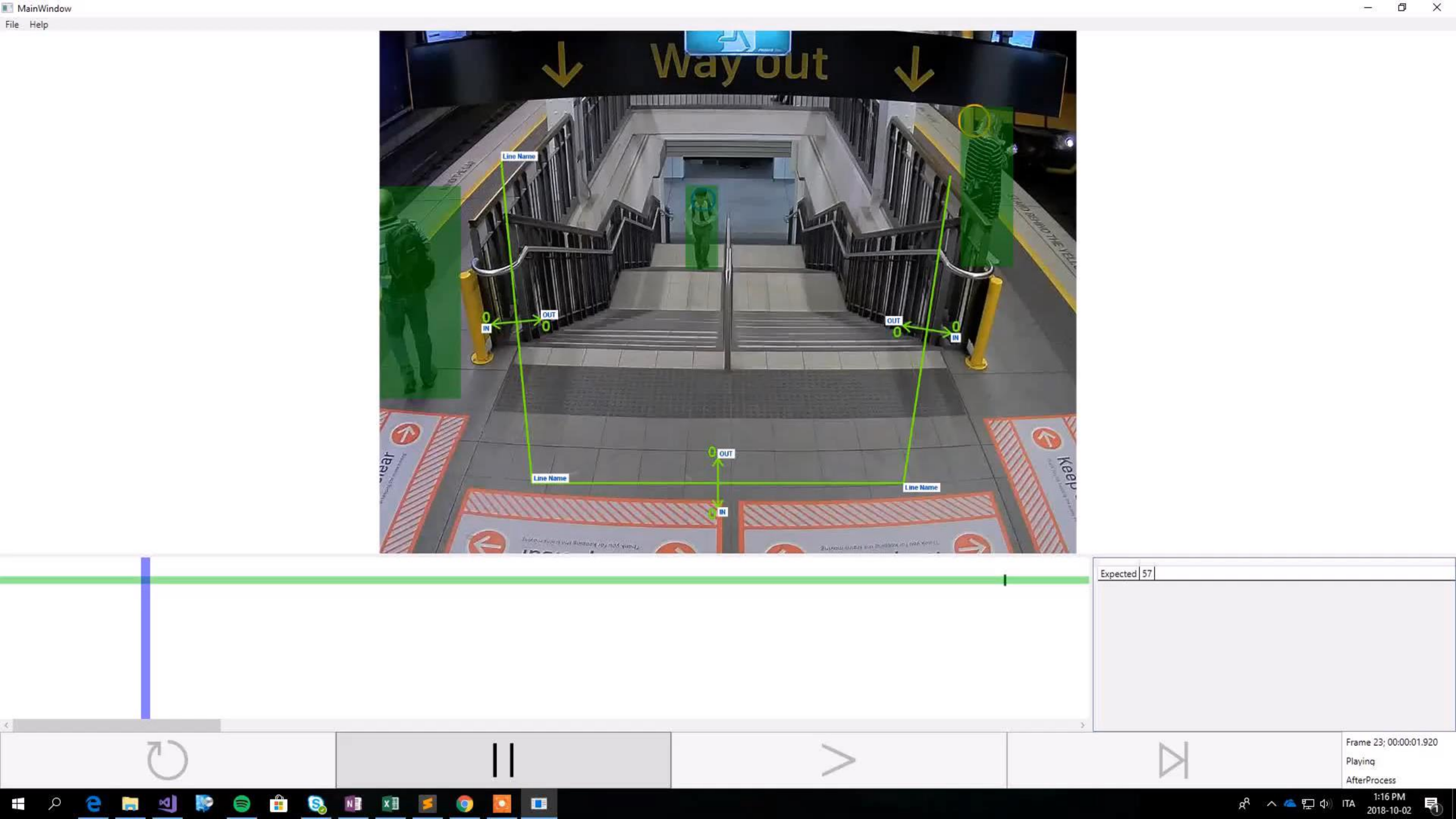
People Detection Dataset



Some Results







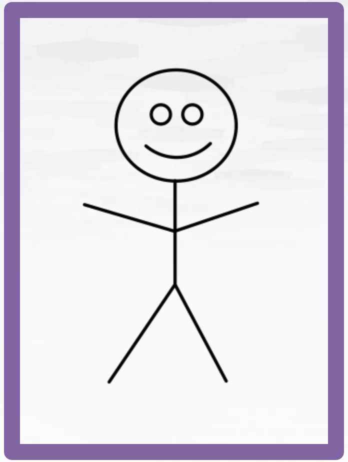
Expected 57

Frame 23; 00:00:01.920

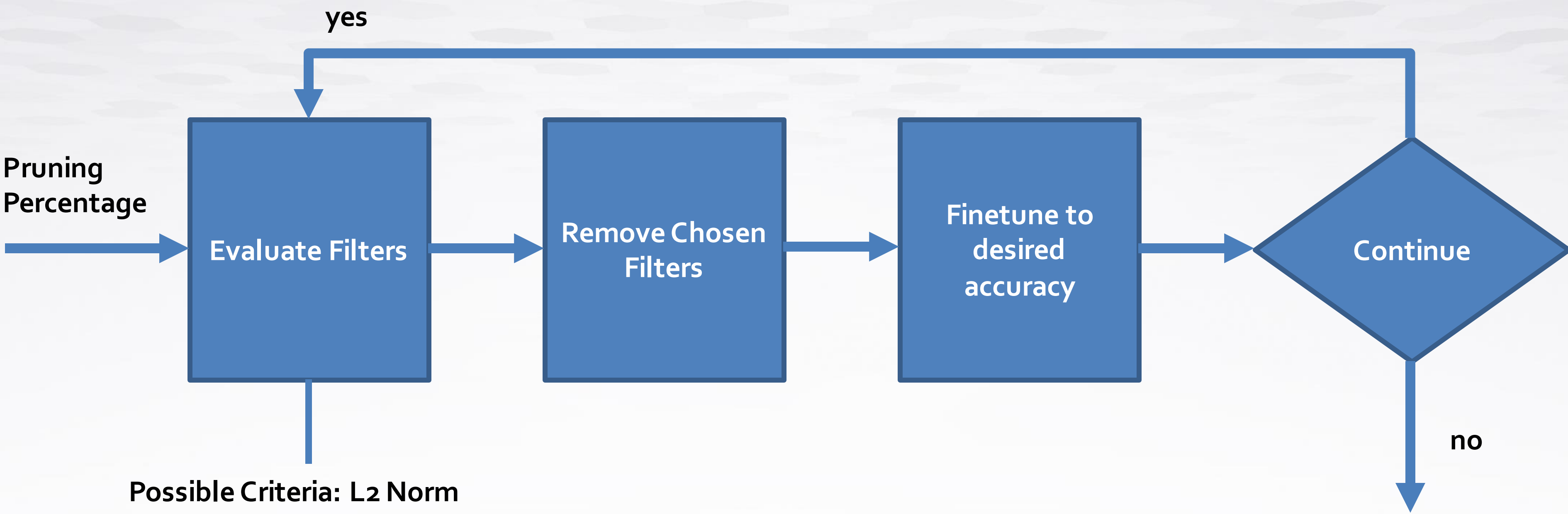
Playing

AfterProcess

Improving the Performances
Reducing the Model Complexity



Model Pruning



Molchanov, Pavlo, et al. "Pruning convolutional neural networks for resource efficient inference." (2016).

Pruning Results: VGG16

Greatly Reduced Model Size

Some layers cut by 75%

Accuracy Decreased by 0.005

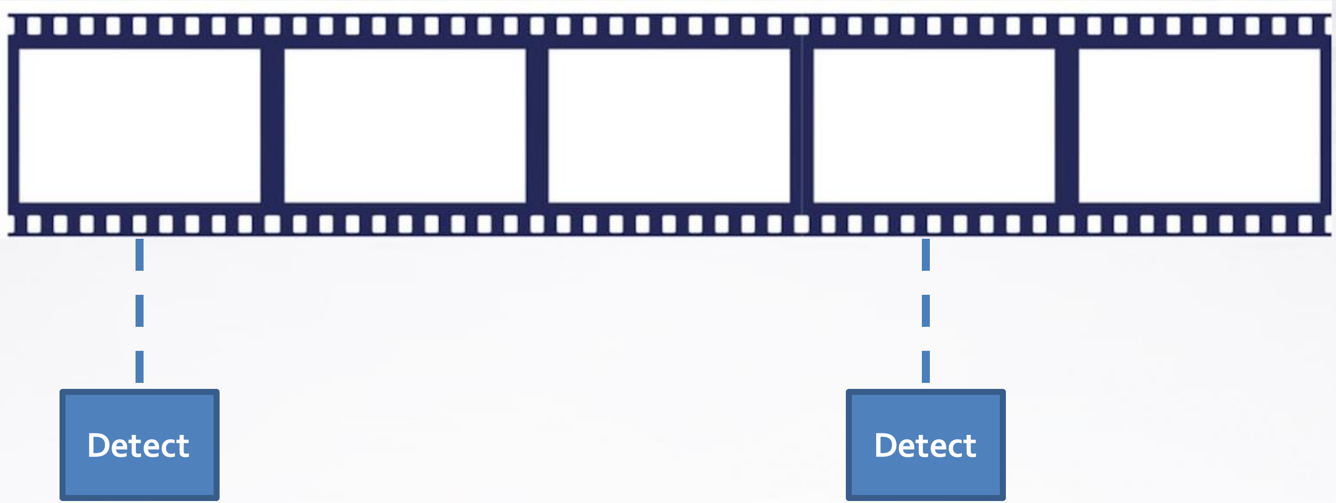
More than Doubled FPS

30 VS 70

All experiments where done on an Intel Core i7-7800X CPU with a Nvidia Quadro P2000 GPU

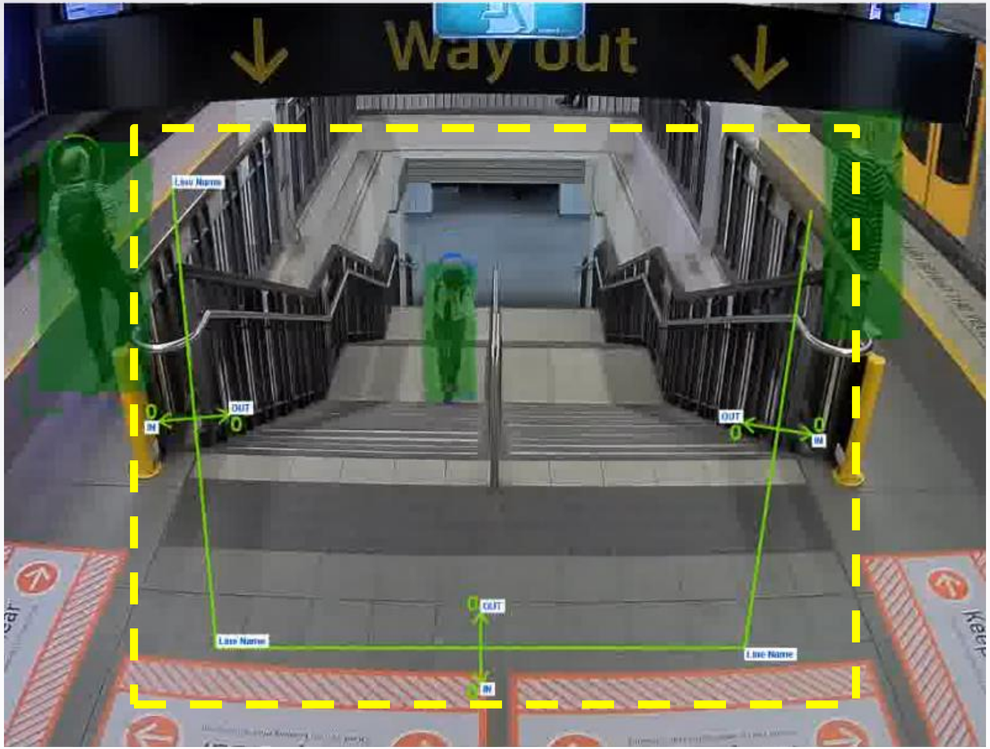
Improving the Performances

Do We Need to Detect at Every Frame?



Memory Consumption Can Still Be An Issue

Improving the Performances
Reducing the Detection Area



1 Reduce Distortion

2 Increase Object Size

Significant Performance Increase (6%)

If you don't need to beat ImageNet, **don't try to**

Use **every trick** you can

Future Plans

- Explore other domains
- Create our own feature extractor
- Introduce temporal information in the model

Interested?

Michelangelo Fiore
m.fiore@kiwisecurity.com

Florian Matusek
f.matusek@kiwisecurity.com

www.kiwisecurity.com

PS: We’re hiring!