Aerial Object Detection

PHD DEFENSE

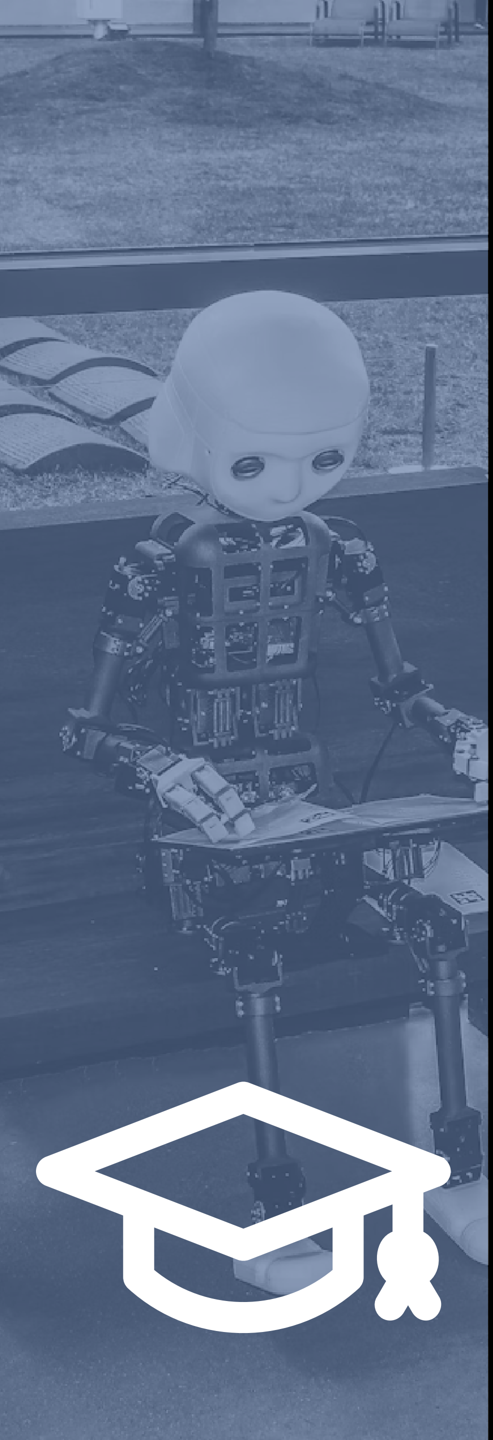
Tanguy Ophoff
26 June 2023

KU LEUVEN eavise

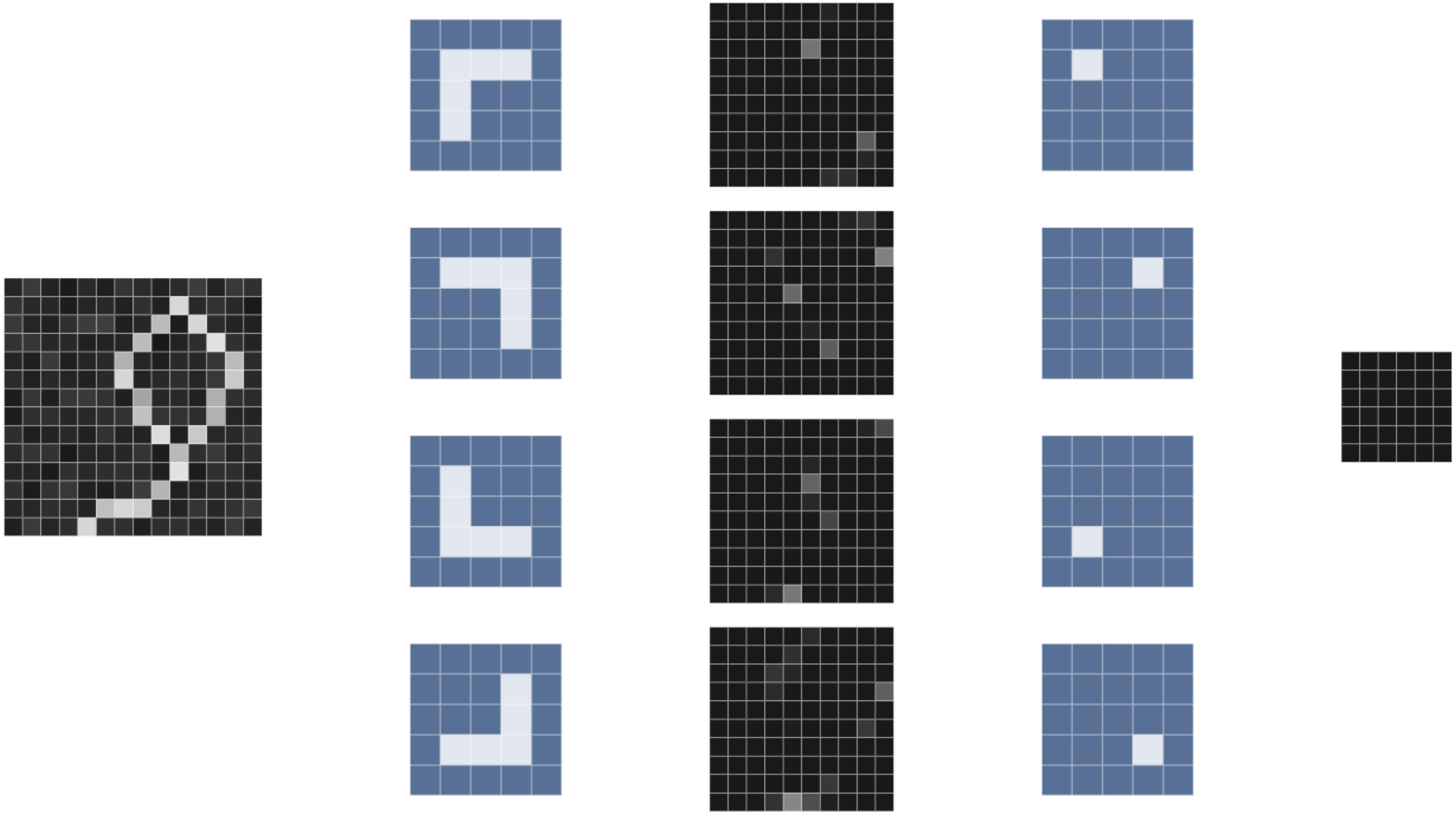
How can we adapt detection algorithms to work on remote sensing data?

How to combine color and depth data to improve detection models?

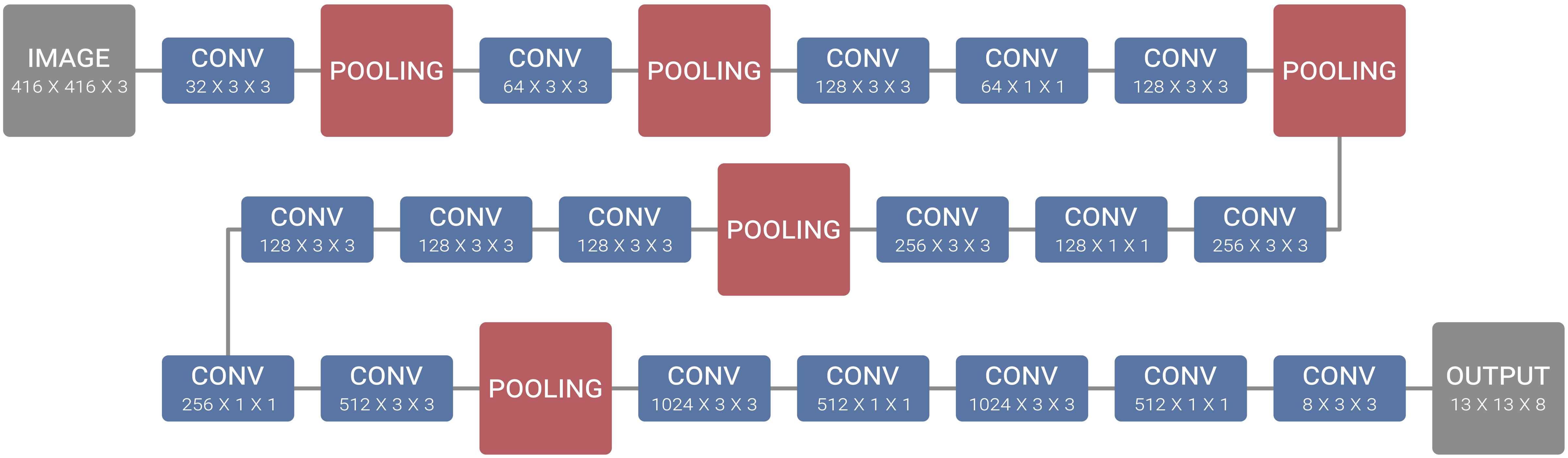
How much can we speed up our models whilst maintaining the accuracy?



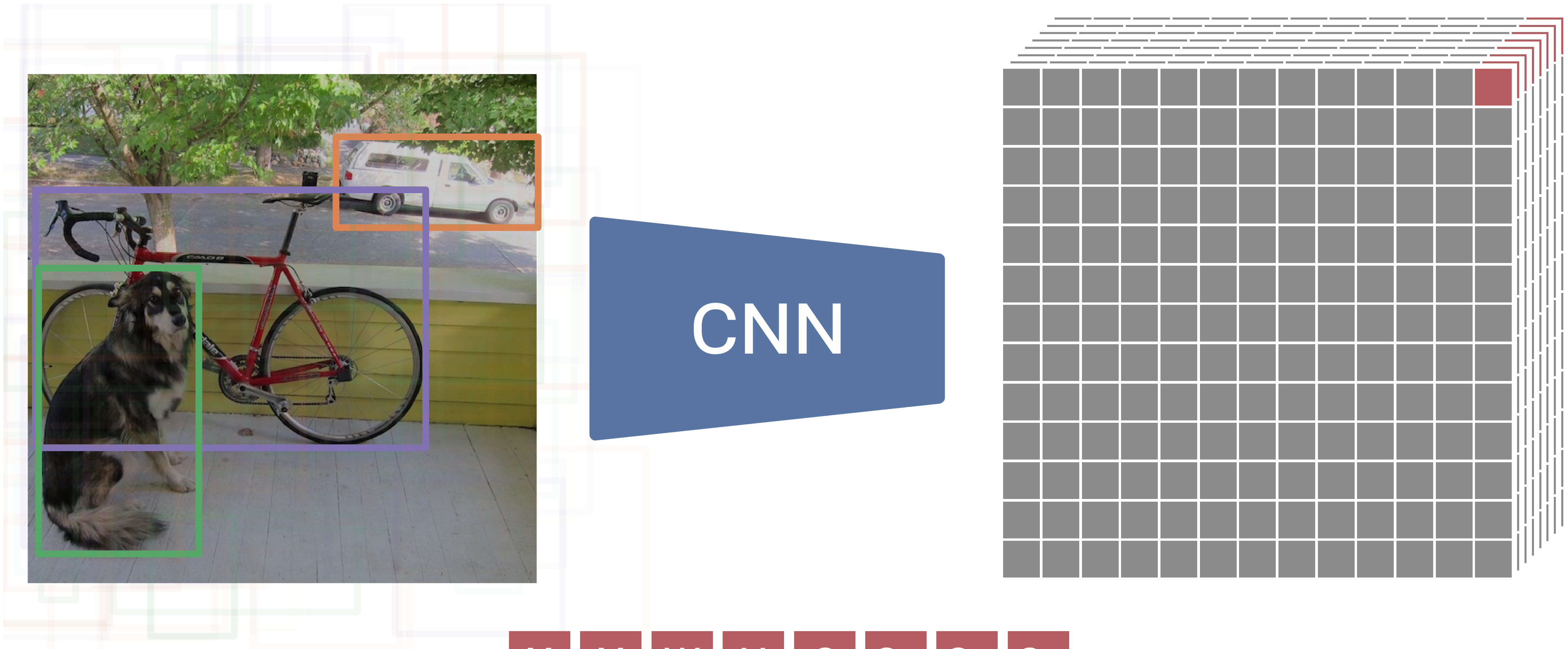
CONVOLUTION



NEURAL NETWORK

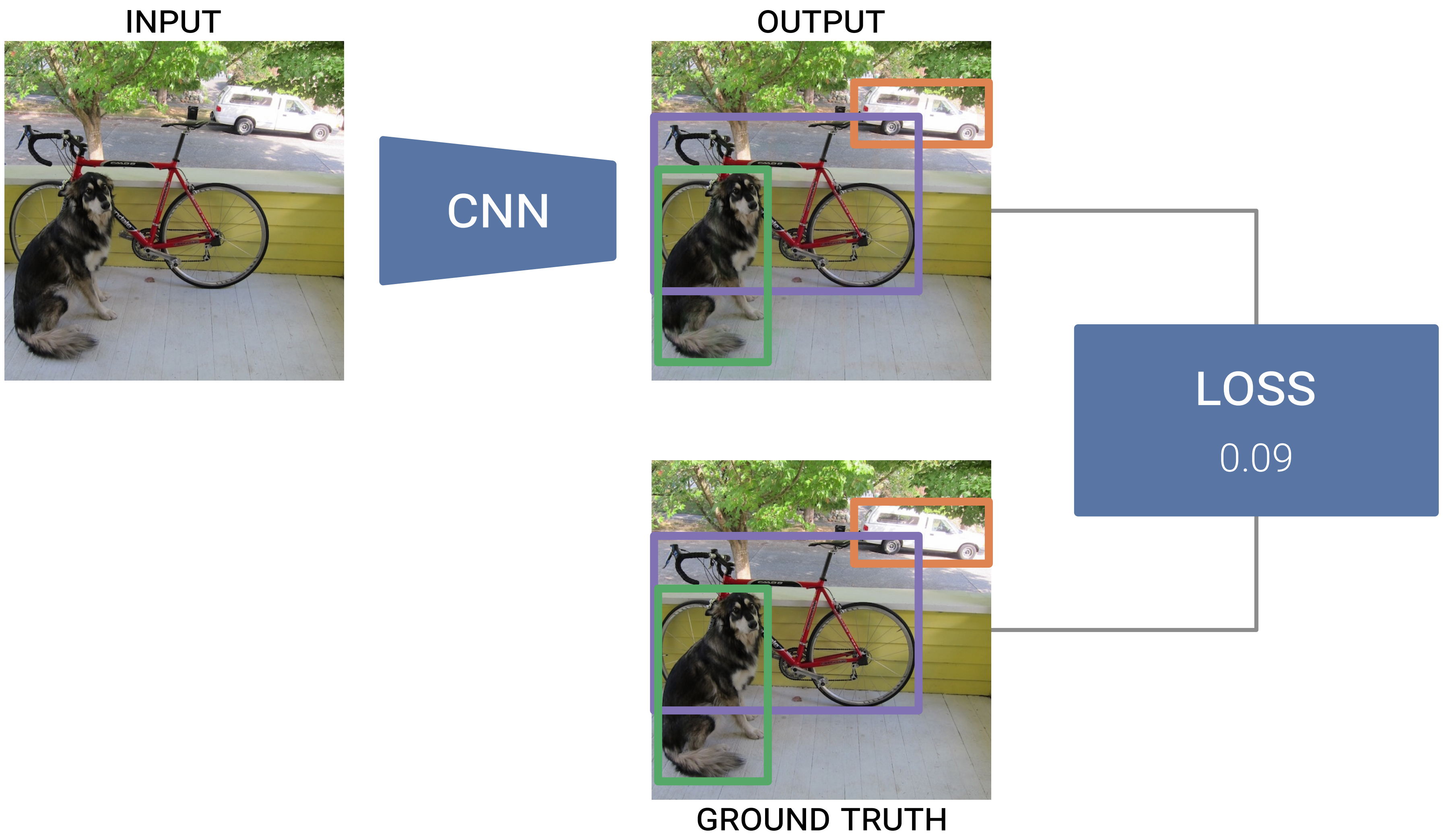


OBJECT DETECTION

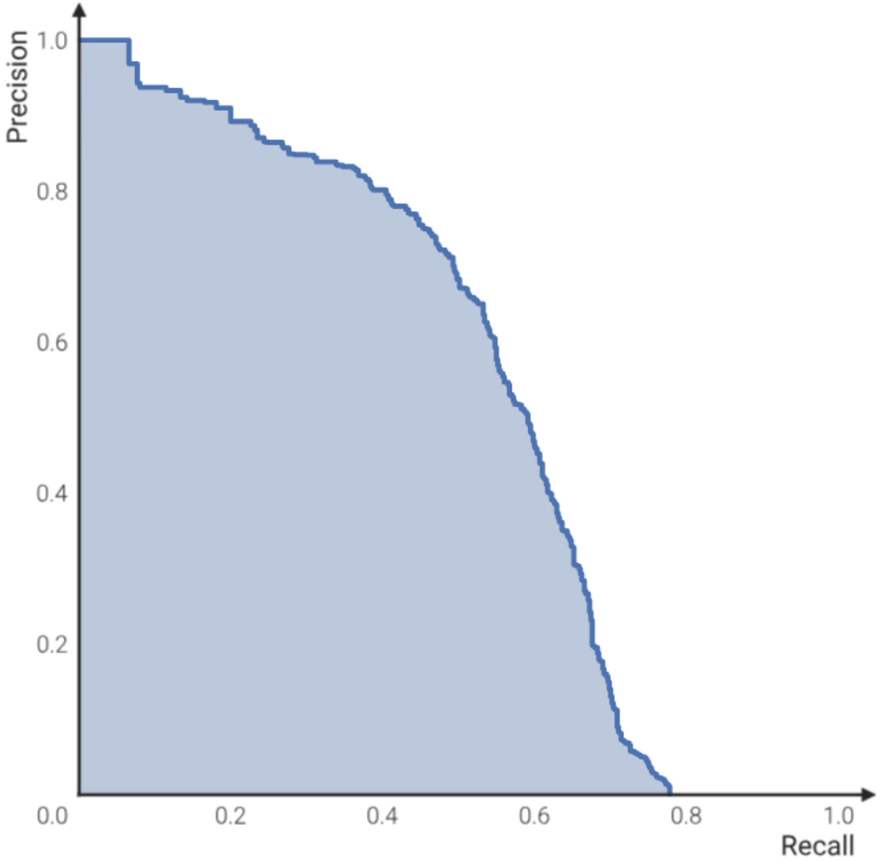


X	Y	W	H	C	C_0	C_1	C_2

TRAINING



STATISTICS



Precision

How many of the detected objects are correct?

Recall

How many correct objects are detected?

Average Precision

Area under the curve

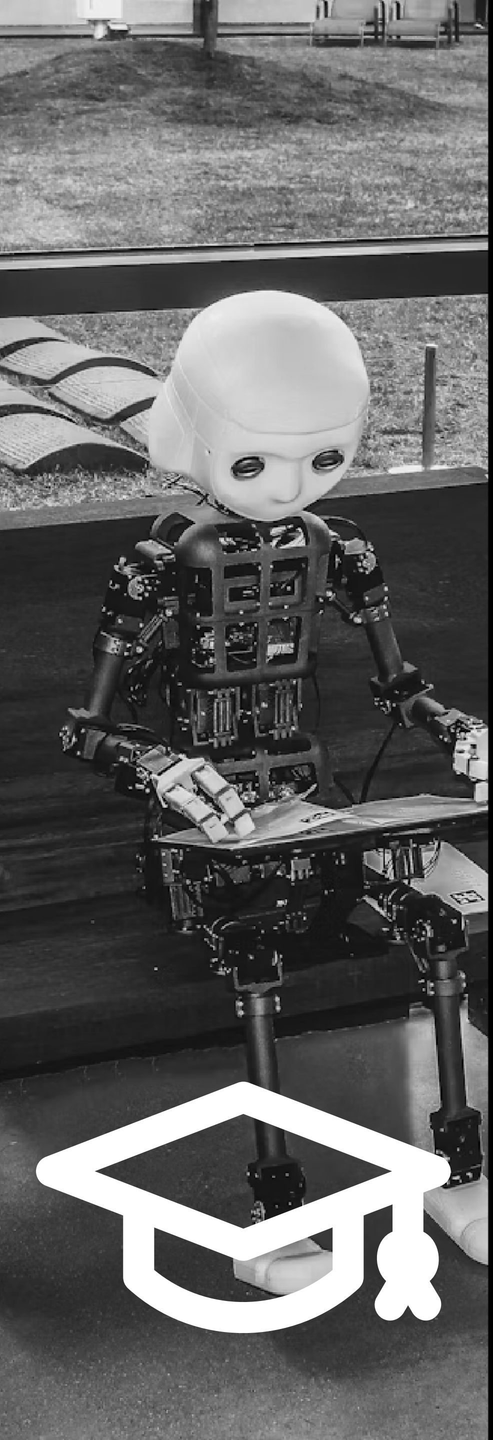
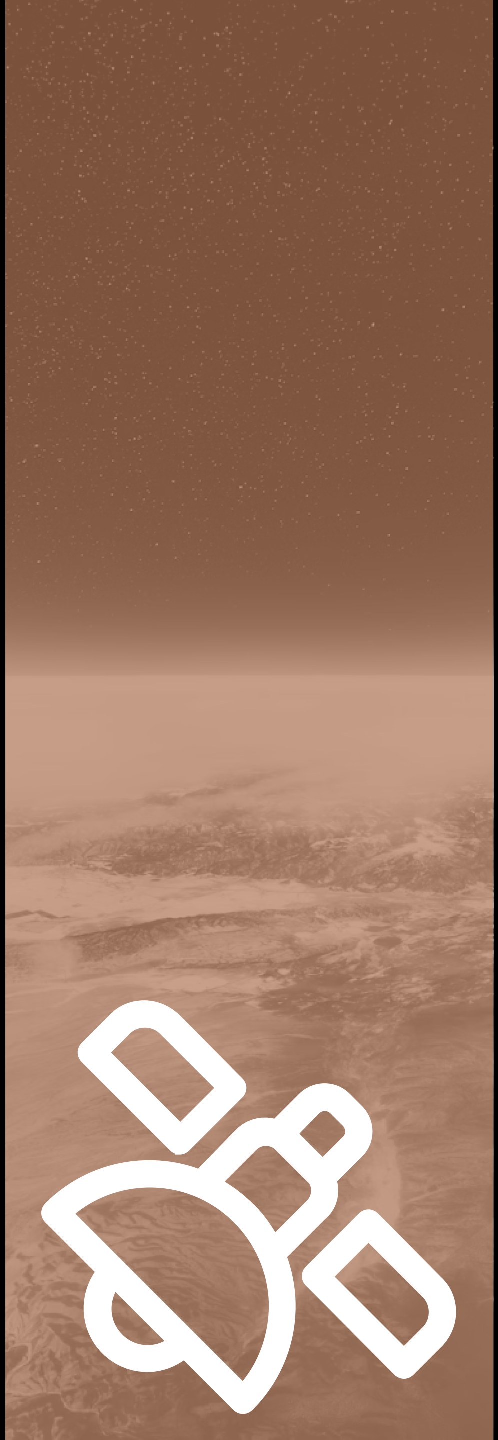
SUMMARY

How do neural networks detect objects?

- Convolution filters find features
- Stack many convolutions to create a network
- Regress detection coordinates and confidences
- Train the model with many examples

How can we evaluate the detection performance?

- Precision tells how many of the detections are correct
- Recall tells how many objects have been successfully detected
- Sweep the confidence to find an optimal precision-recall trade-off
- AP provides a single value to easily compare models

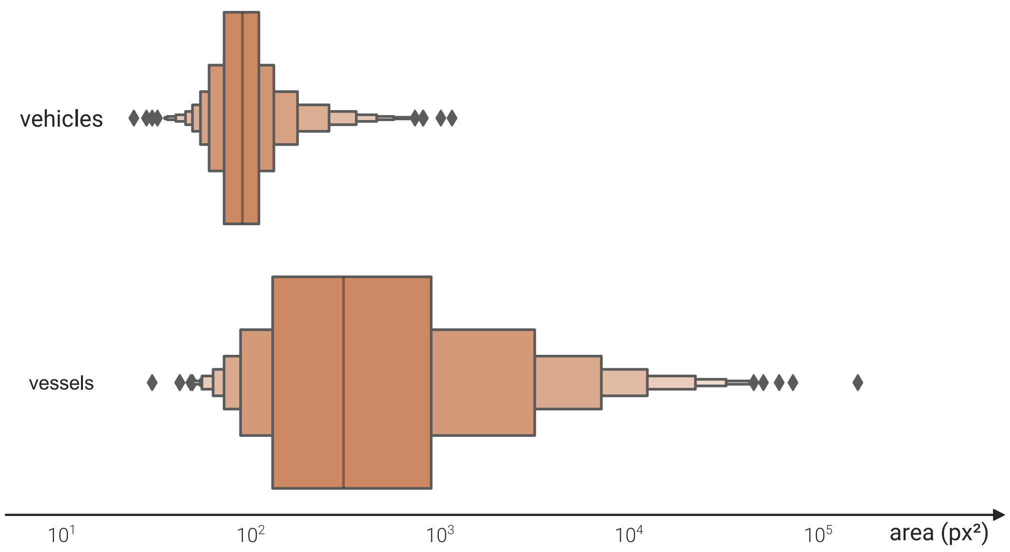
PROJECT

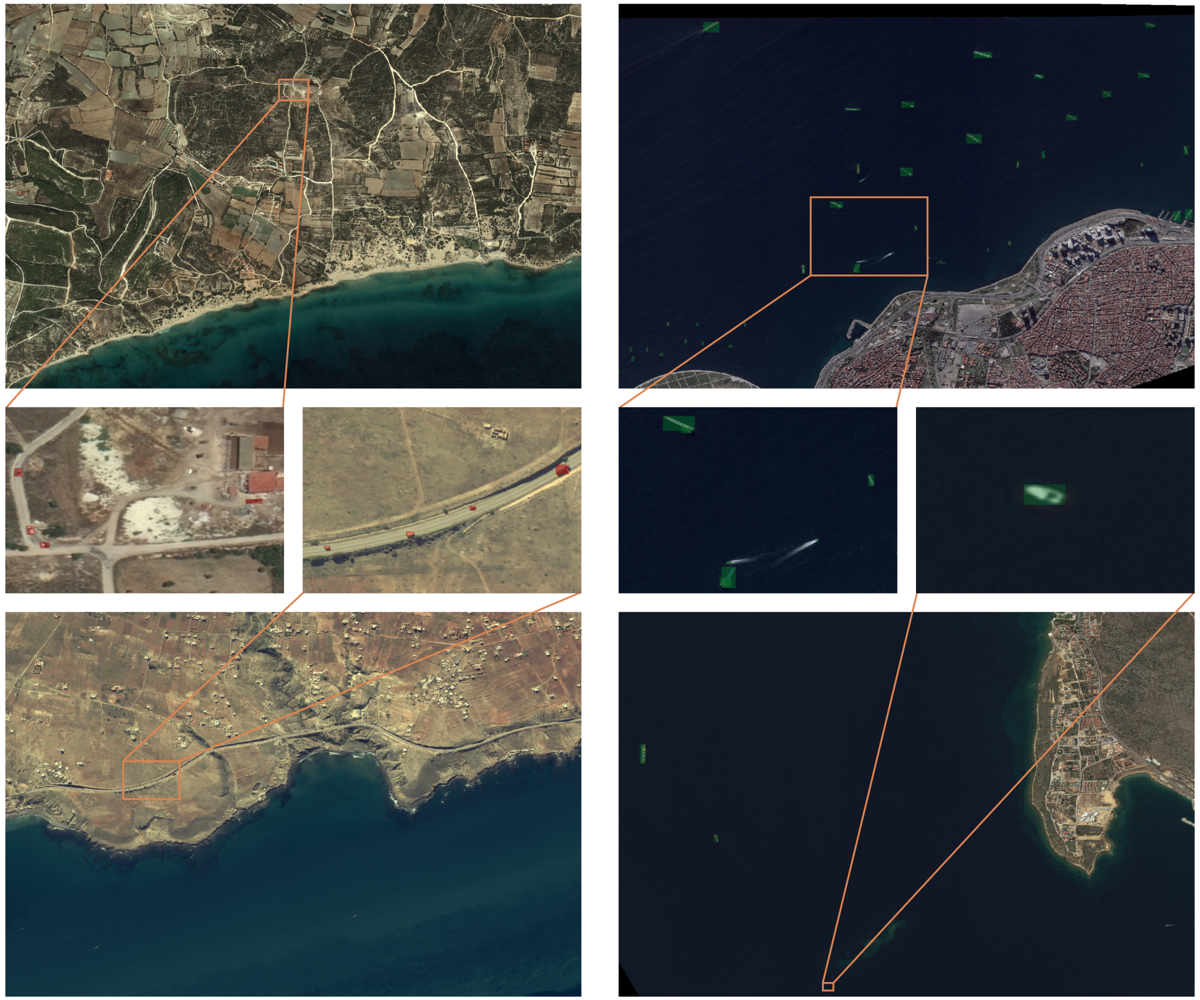
OBJECTIVE

Provide a tool to automatically detect and classify objects in satellite imagery

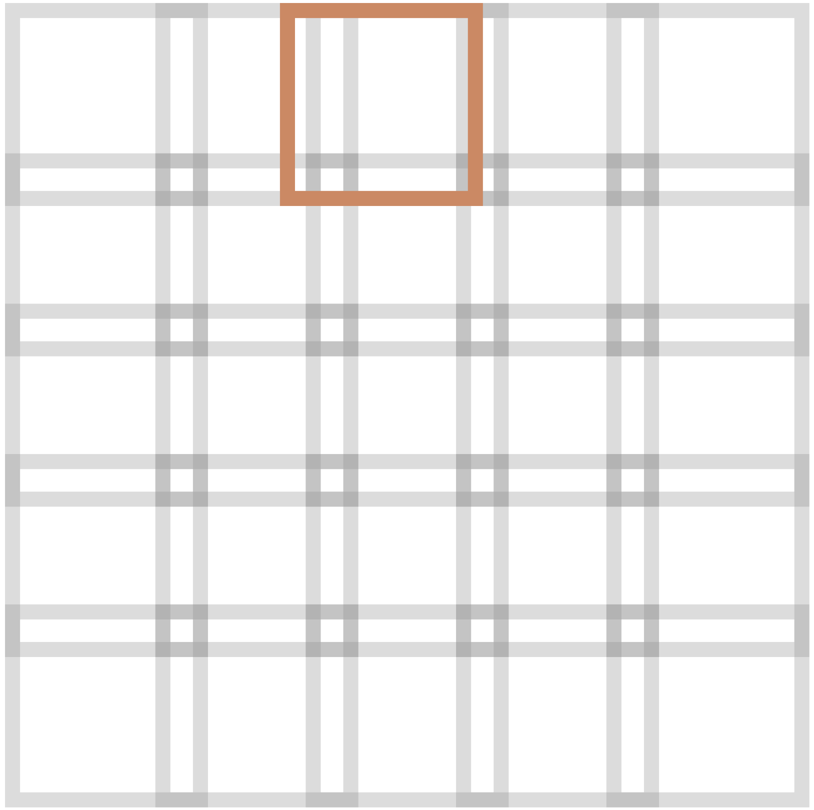
SATELLITE DETECTION

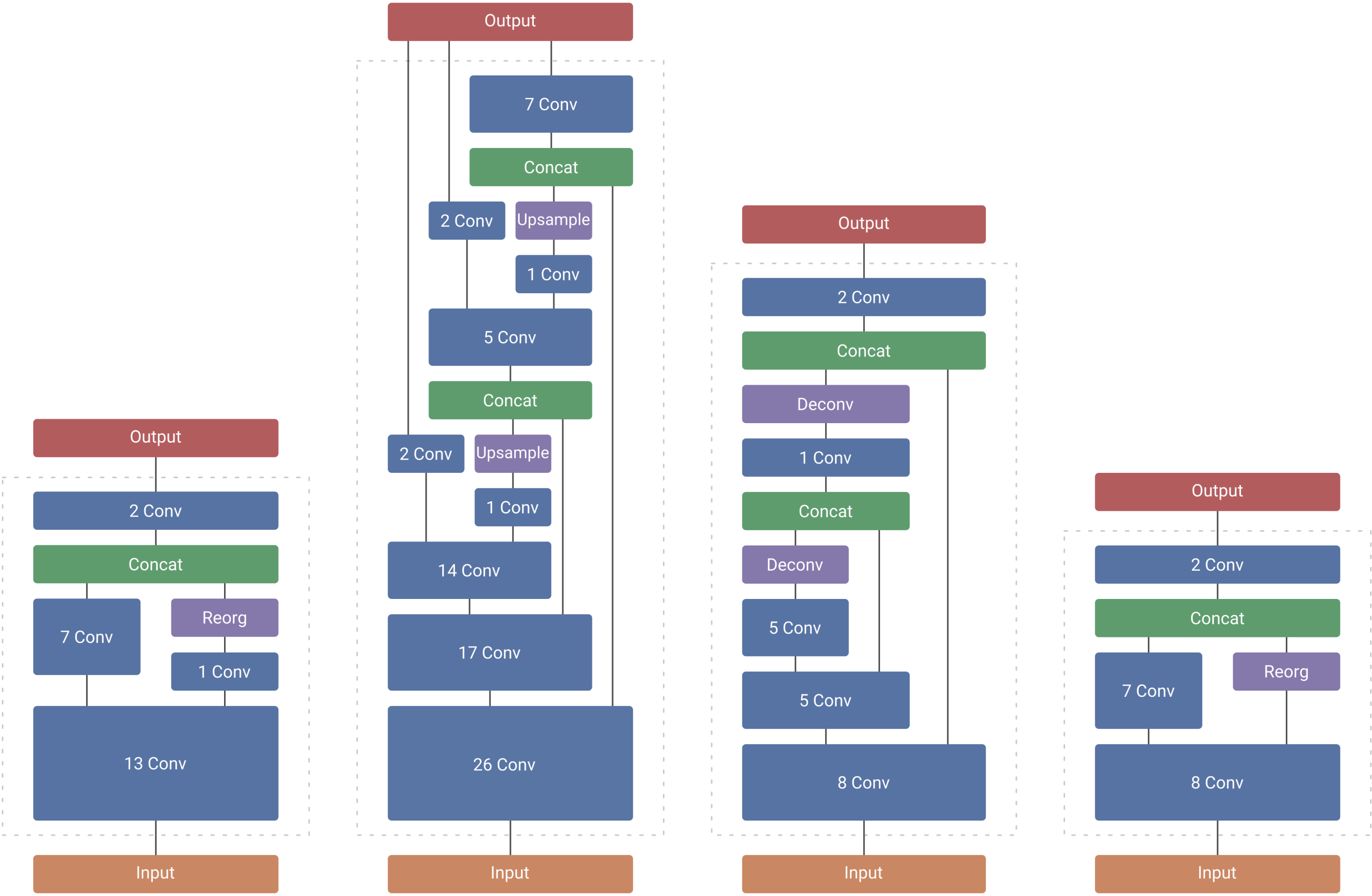
	Vehicles	Vessels
Region	641 km²	676 km²
Resolution	0.3m - 0.5m	0.3m - 0.5m
Objects	4075	1096





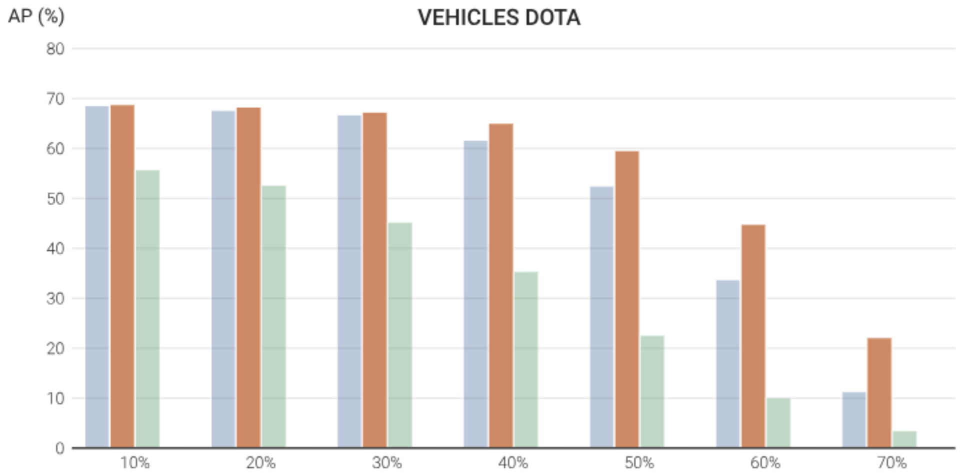
METHODOLOGY

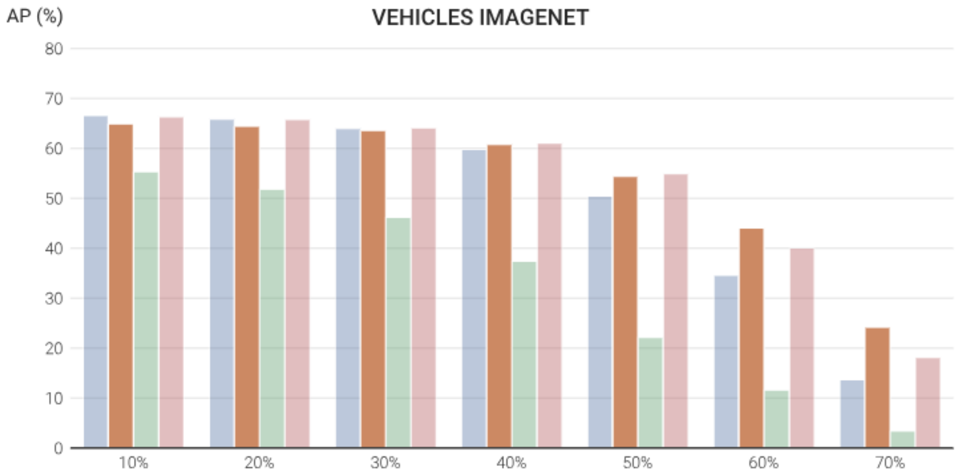


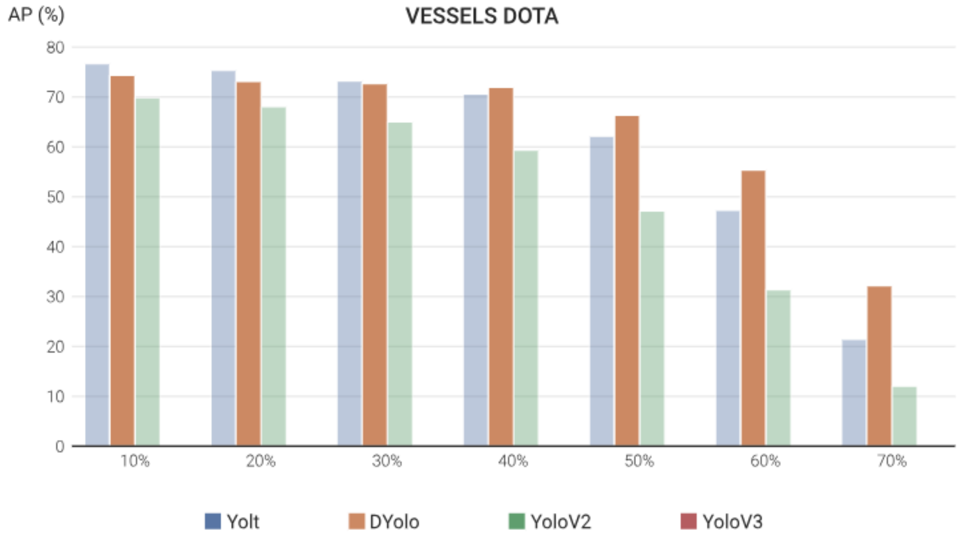


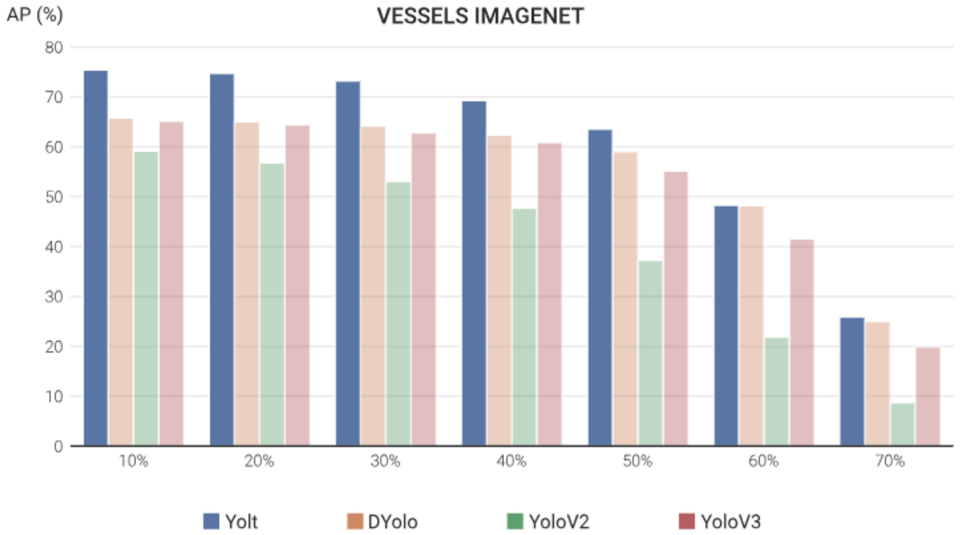
YOLO II YOLO III DYOLO YOLT

RESULTS


AP (%)
VEHICLES DOTA
80
70
60
50
40
30
20
10
0
10%
20%
30%
40%
50%
60%
70%


AP (%)
VEHICLES IMAGENET
80
70
60
50
40
30
20
10
0
10%
20%
30%
40%
50%
60%
70%


AP (%)
VESSELS DOTA
80
70
60
50
40
30
20
10
0
10%
20%
30%
40%
50%
60%
70%
Yolt
DYolo
YoloV2
YoloV3

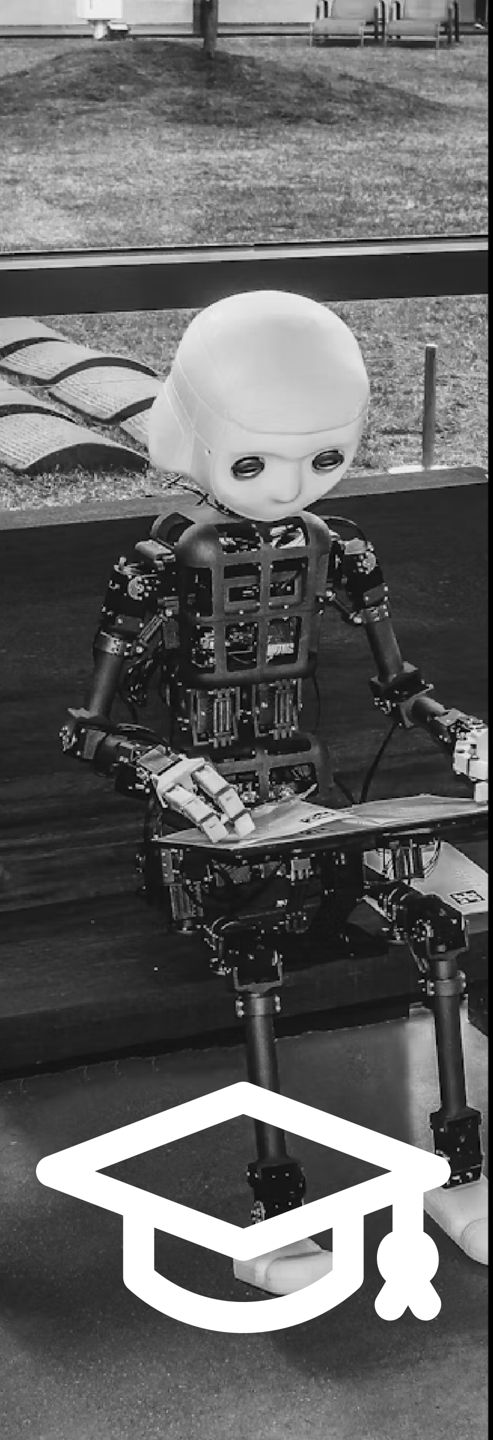

AP (%)
VESSELS IMAGENET
80
70
60
50
40
30
20
10
0
10%
20%
30%
40%
50%
60%
70%
Yolt
DYolo
YoloV2
YoloV3

CONTRIBUTIONS

How can we adapt detection algorithms to work on remote sensing data?

- We developed a sliding window technique
- Pretrained weights from similar data improves the results
- D-Yolo works the best on this data

1. T. Ophoff, S. Puttemans, V. Kalogirou, J.-P. Robin, and T. Goedemé. "Vehicle and Vessel Detection on Satellite Imagery: A Comparative Study on Single-Shot Detectors". In: *Remote Sensing* 12.7 (2020).



PROJECT

OBJECTIVE

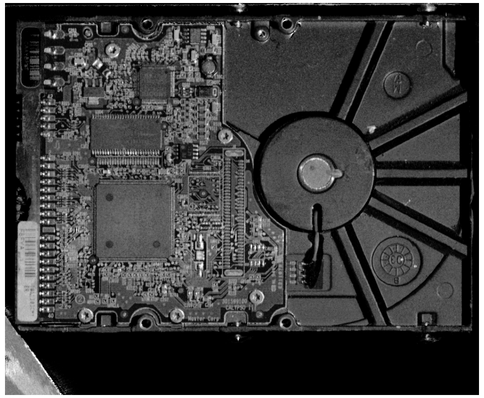
Improve the accuracy of object detection networks
by combining color and depth images

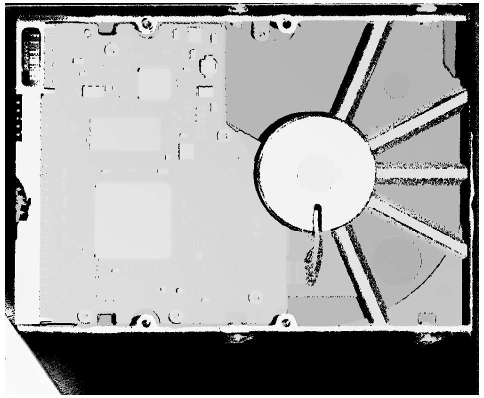
RGBD FUSION

EPFL RELABELED

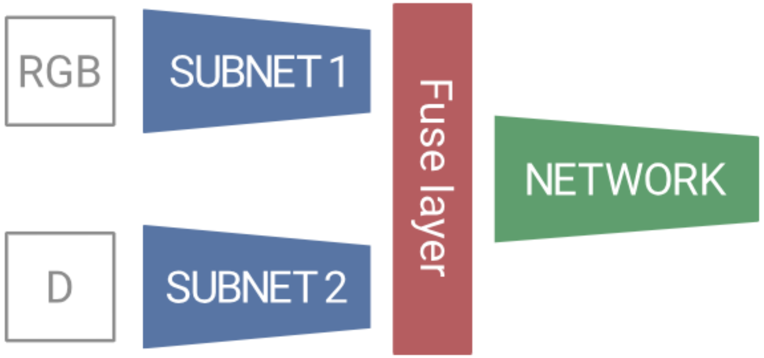
KITTI

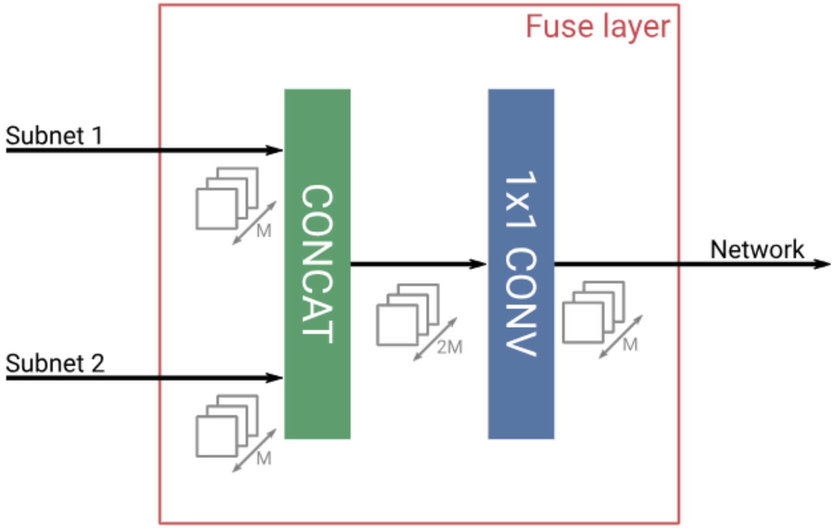
PCB SCREWS

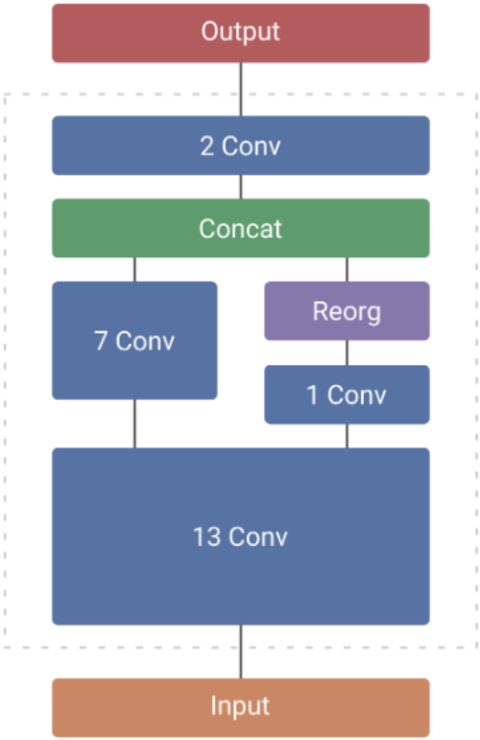




METHODOLOGY

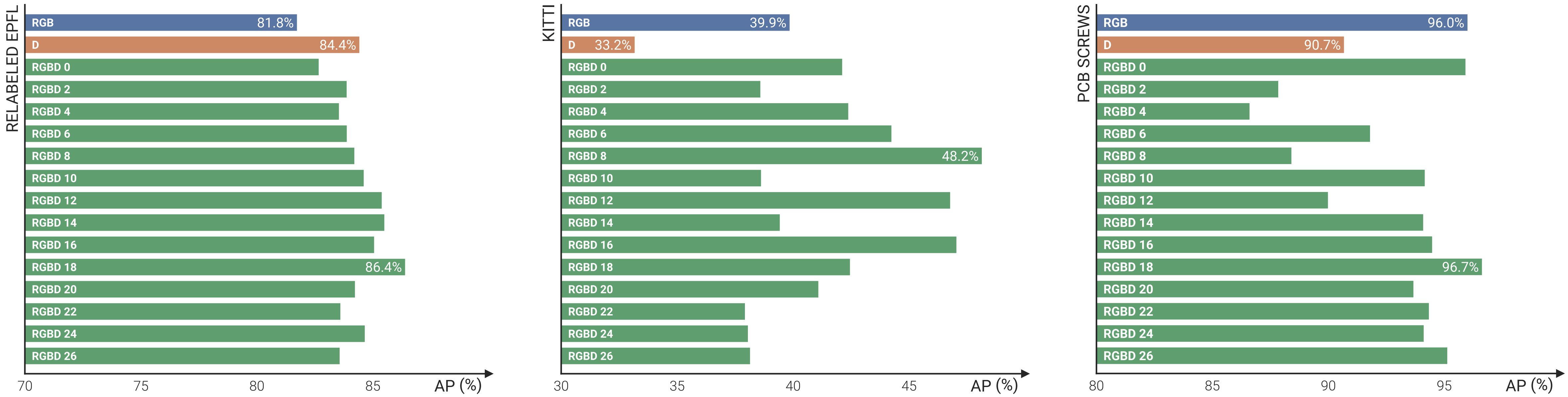






YOLO II

RESULTS

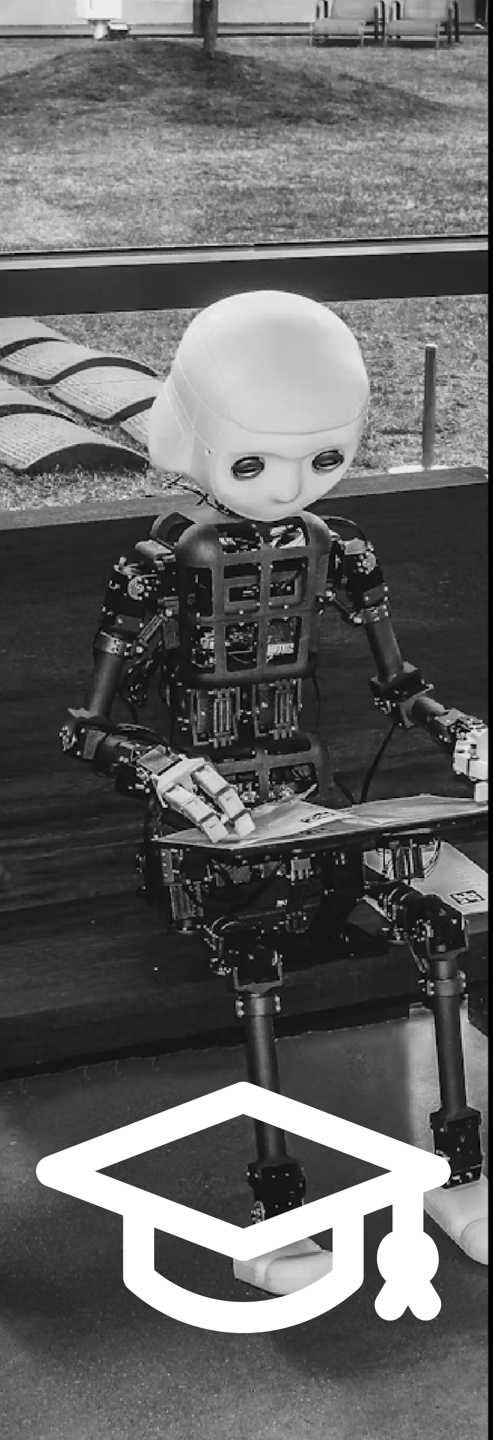

RELABELED EPFL
RGB 81.8%
D 84.4%
RGBD 0
RGBD 2
RGBD 4
RGBD 6
RGBD 8
RGBD 10
RGBD 12
RGBD 14
RGBD 16
RGBD 18 86.4%
RGBD 20
RGBD 22
RGBD 24
RGBD 26
70 75 80 85 AP (%)
KITTI
RGB 39.9%
D 33.2%
RGBD 8 48.2%
30 35 40 45 AP (%)
PCB SCREWS
RGB 96.0%
D 90.7%
RGBD 18 96.7%
80 85 90 95 AP (%)

CONTRIBUTIONS

How to combine color and depth data to improve detection models?

- We developed a transparent fuse layer
- RGBD improved the results on 3 different datasets
- Midway to late fusion is optimal

2. T. Ophoff, K. Van Beeck, and T. Goedemé. “Improving Real-Time Pedestrian Detectors with RGB+Depth Fusion”. In: 15th *AVSS* (2018).
3. T. Ophoff, K. Van Beeck, and T. Goedemé. “Exploring RGB+Depth Fusion for Real-Time Object Detection”. In: *Sensors* 19.4 (2019).



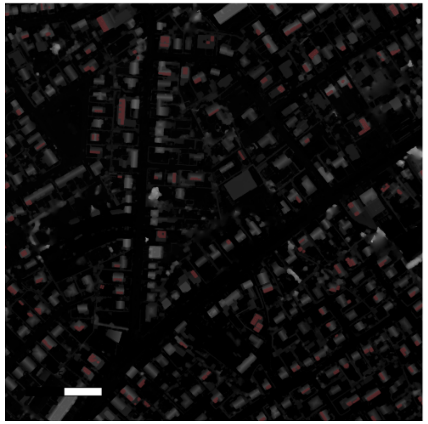

PROJECT

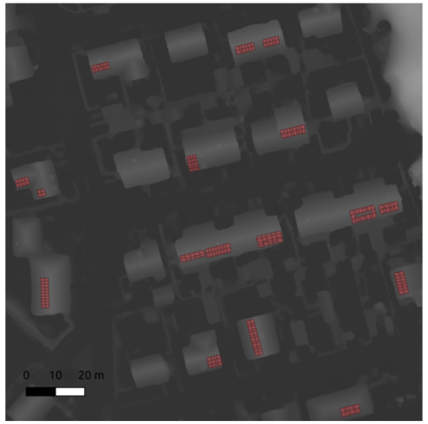
OBJECTIVE

Automatically detect objects in aerial imagery,
whilst combining data from multiple sources and sensors

PLANE DETECTION

SOLAR PANELS





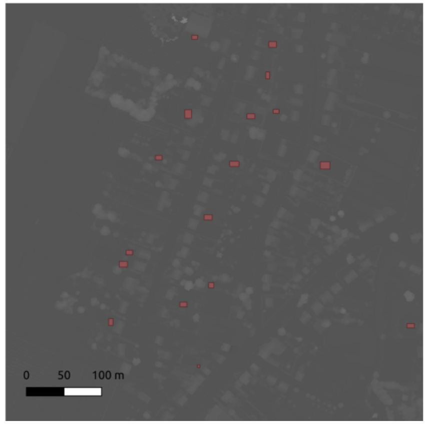
RGB

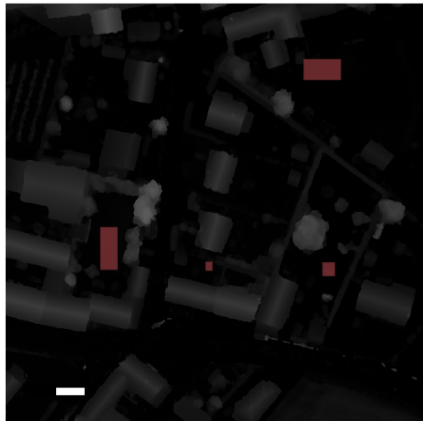
3cm GSD

DEPTH

25cm GSD

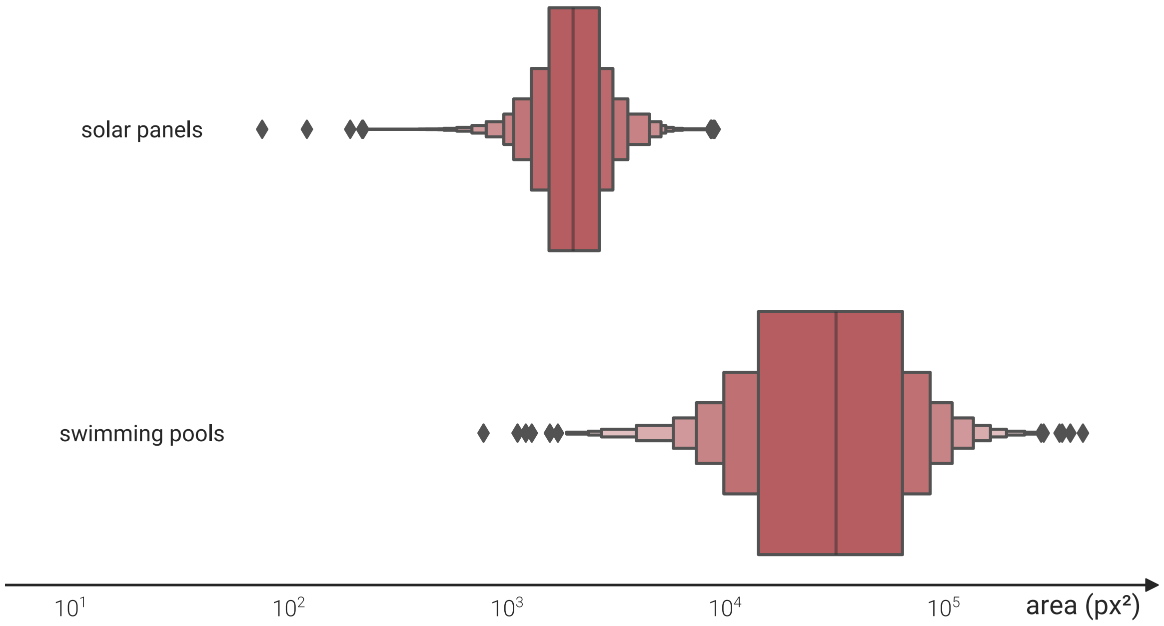
SWIMMING POOLS



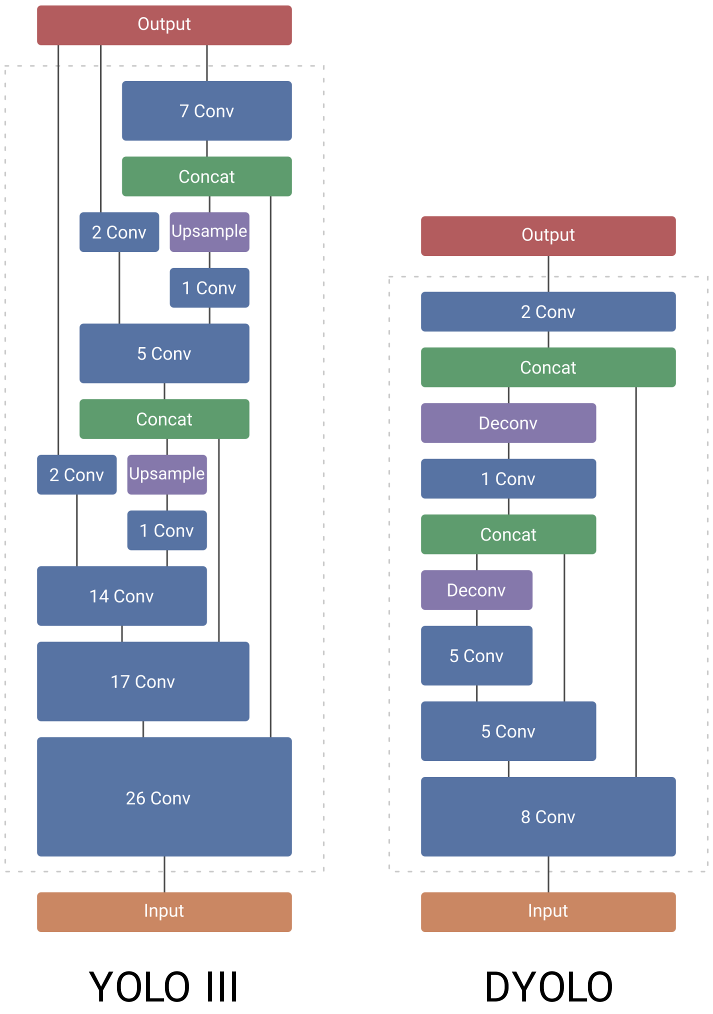


DATASET

	Solar Panels	Swimming Pools
Region	10.1 km²	17.3 km²
Objects	32970	3000

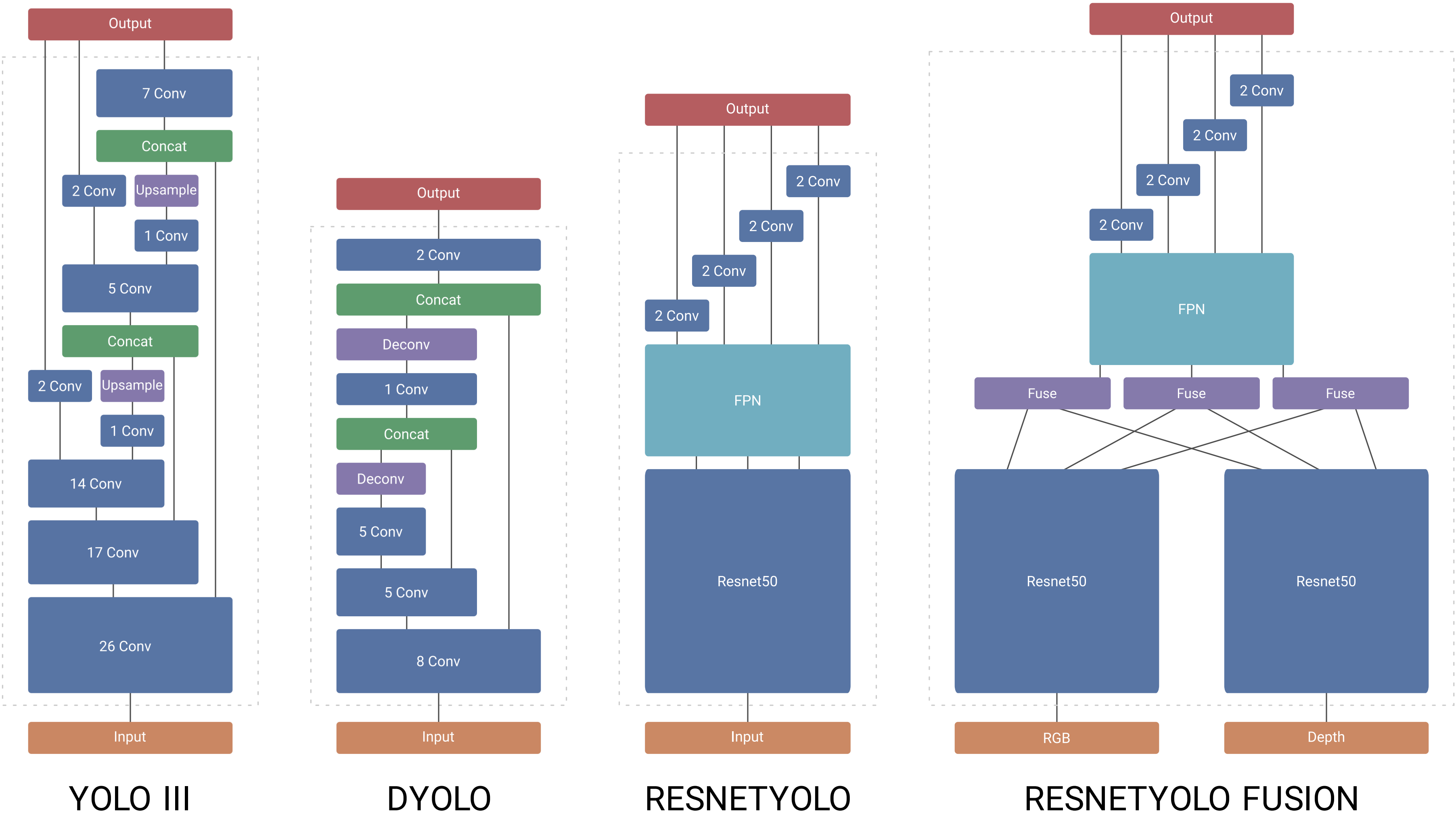


BASELINE



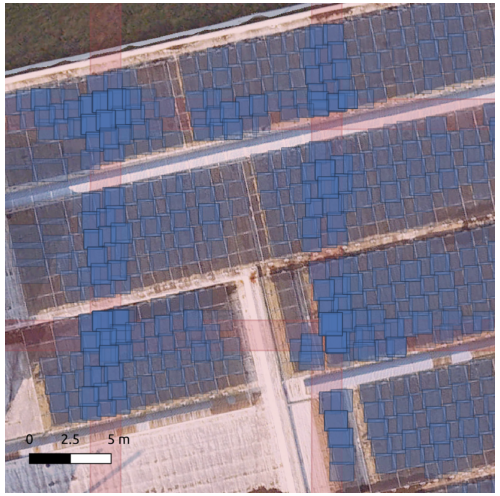
	Solar Panels	Swimming Pools
DYOLO	59.67%	**25.08%**
YOLO III	**62.96%**	23.73%

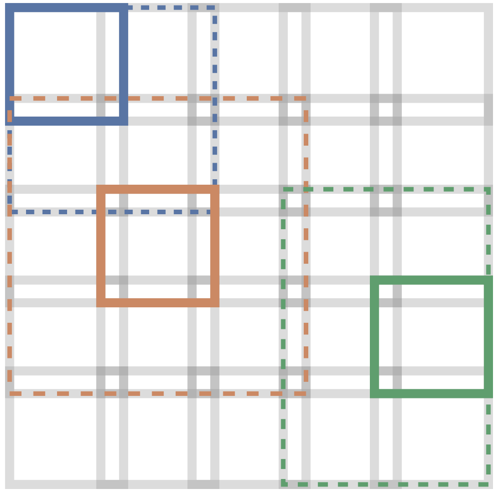
MODELS

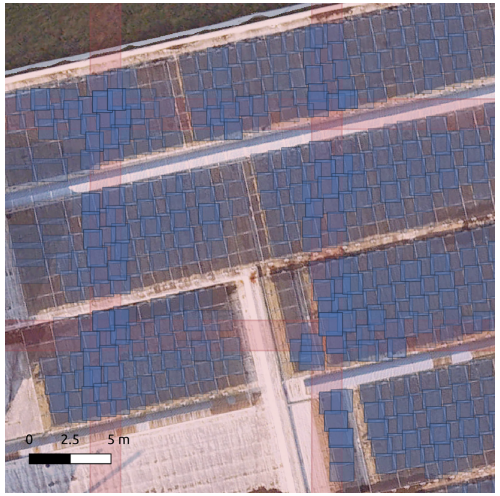


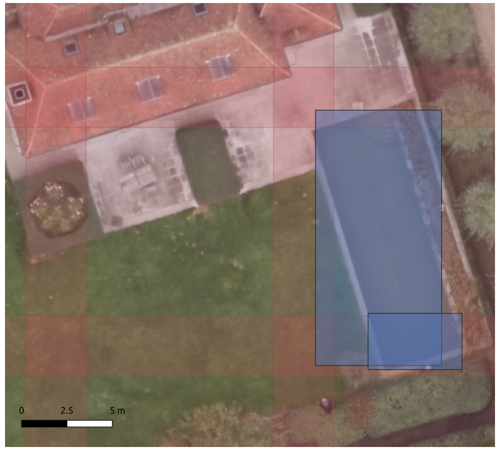
Output
7 Conv
Concat
2 Conv
Upsample
1 Conv
5 Conv
Concat
2 Conv
Upsample
1 Conv
14 Conv
17 Conv
26 Conv
Input
YOLO III
Output
2 Conv
Concat
Deconv
1 Conv
Concat
Deconv
5 Conv
5 Conv
8 Conv
Input
DYOLO
Output
2 Conv
2 Conv
2 Conv
2 Conv
FPN
Resnet50
Input
RESNETYOLO
Output
2 Conv
2 Conv
2 Conv
2 Conv
FPN
Fuse
Fuse
Fuse
Resnet50
Resnet50
RGB
Depth
RESNETYOLO FUSION

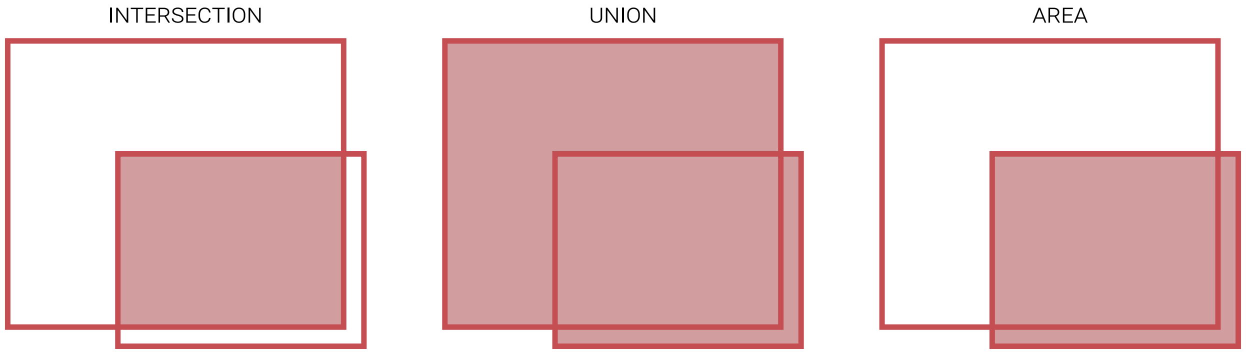
POST PROCESSING



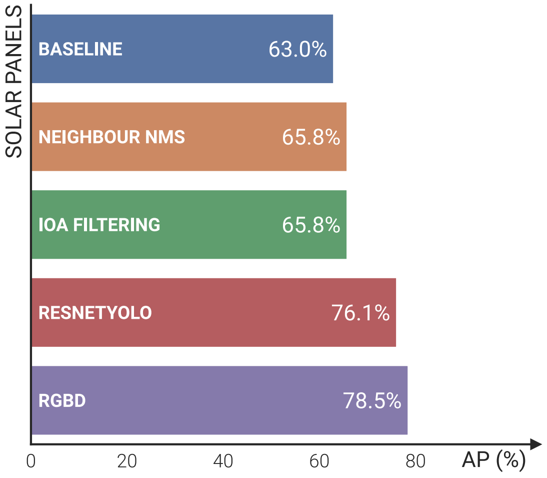




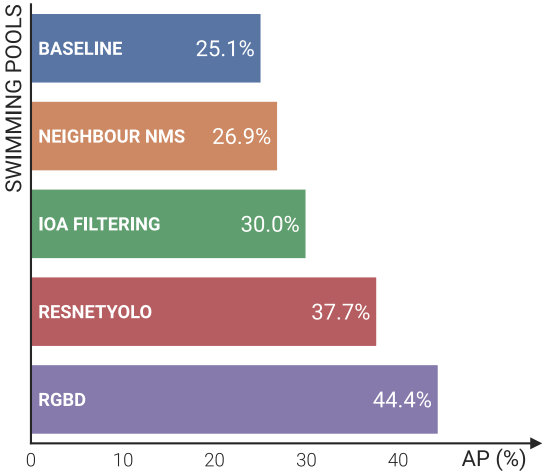




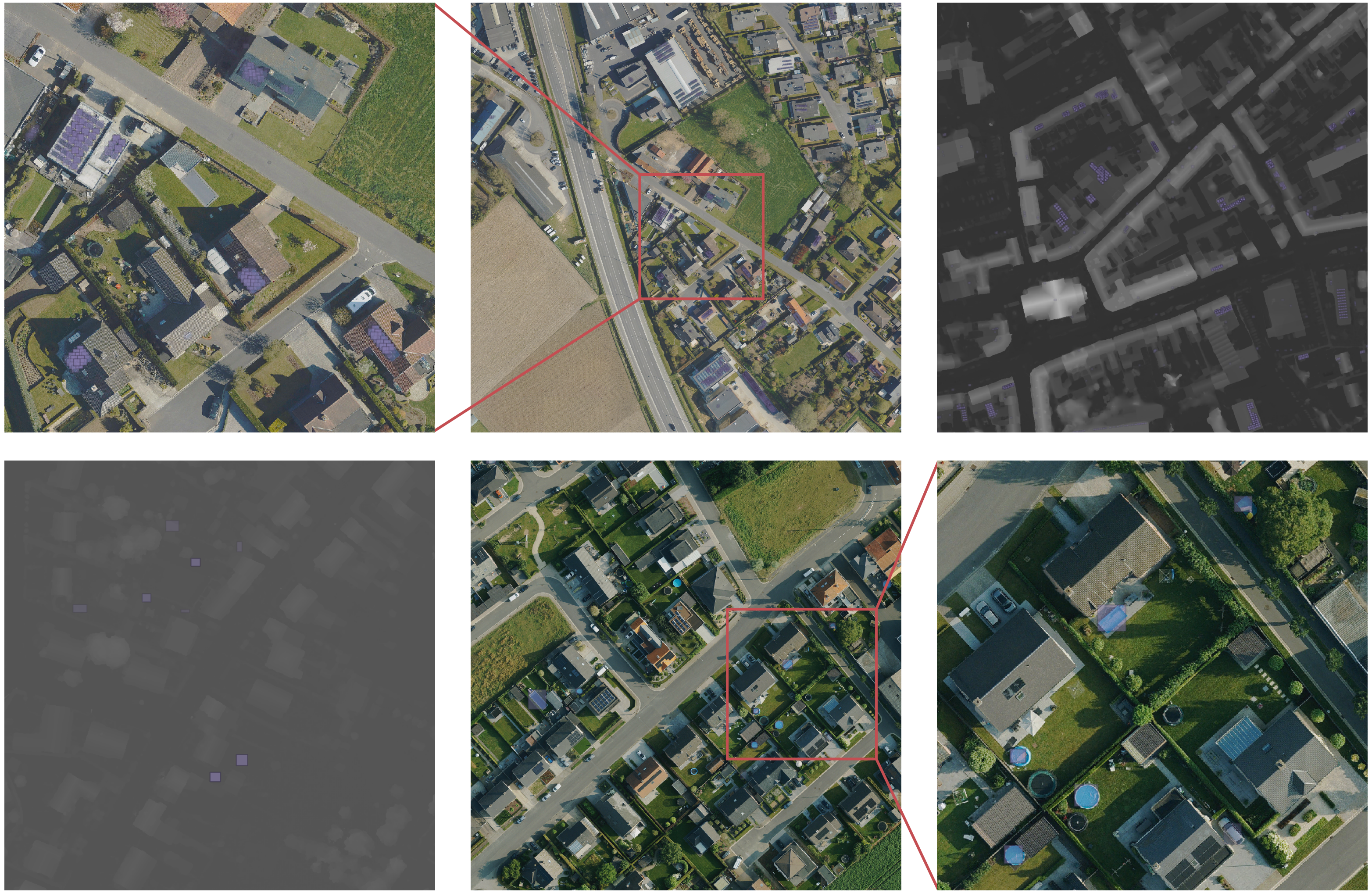
RESULTS



SOLAR PANELS
BASELINE 63.0%
NEIGHBOUR NMS 65.8%
IOA FILTERING 65.8%
RESNETYOLO 76.1%
RGBD 78.5%
0 20 40 60 80
AP (%)



SWIMMING POOLS
BASELINE 25.1%
NEIGHBOUR NMS 26.9%
IOA FILTERING 30.0%
RESNETYOLO 37.7%
RGBD 44.4%
0 10 20 30 40
AP (%)



CONTRIBUTIONS

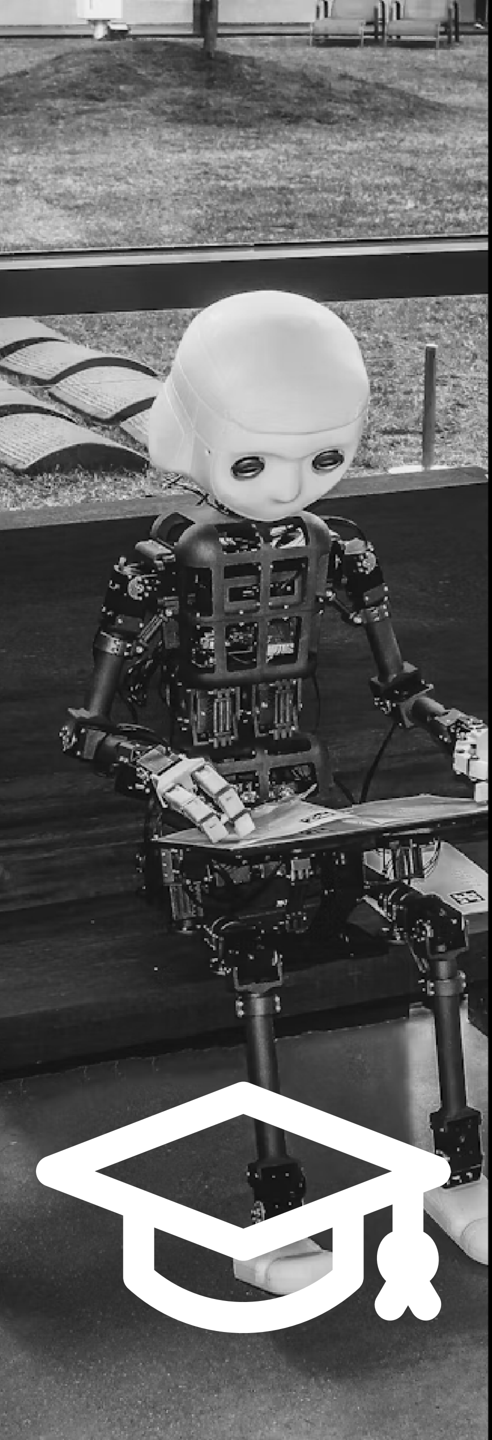
How can we adapt detection algorithms to work on remote sensing data?

- We further increased our results with scene-specific post-processing
- Deeper networks work well with enough data
- ResnetYolo with selectable heads is a prime candidate for remote sensing detection

How to combine color and depth data to improve detection models?

- Our RGBD fusion technique transfers perfectly to remote sensing
- The technique works with deeper networks as well

4. T. Ophoff, K. Van Beeck, and T. Goedemé. “Improving Object Detection in VHR Aerial Orthomosaics”. In: *ECCV Workshops* (2022).



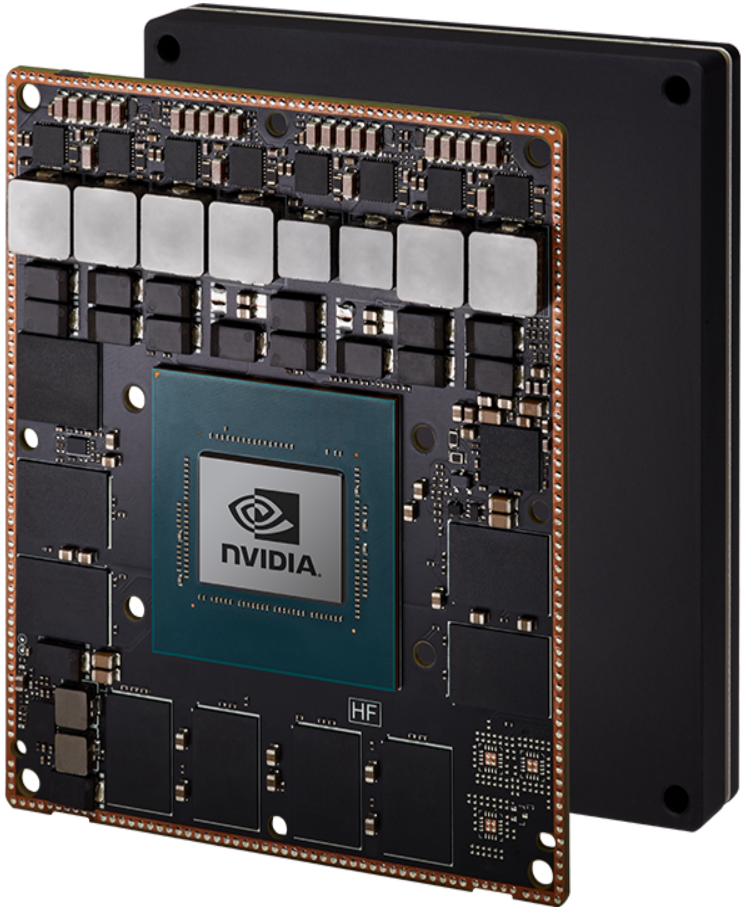


PROJECT

OBJECTIVE

Design faster and smaller object detection networks without deteriorating the accuracy

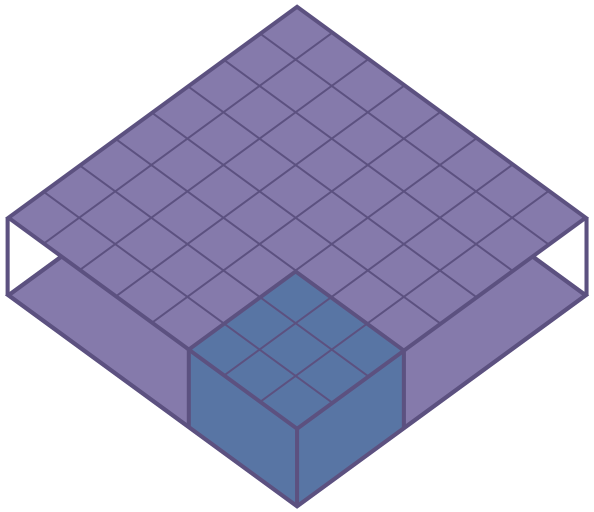
SPEED OPTIMIZATIONS

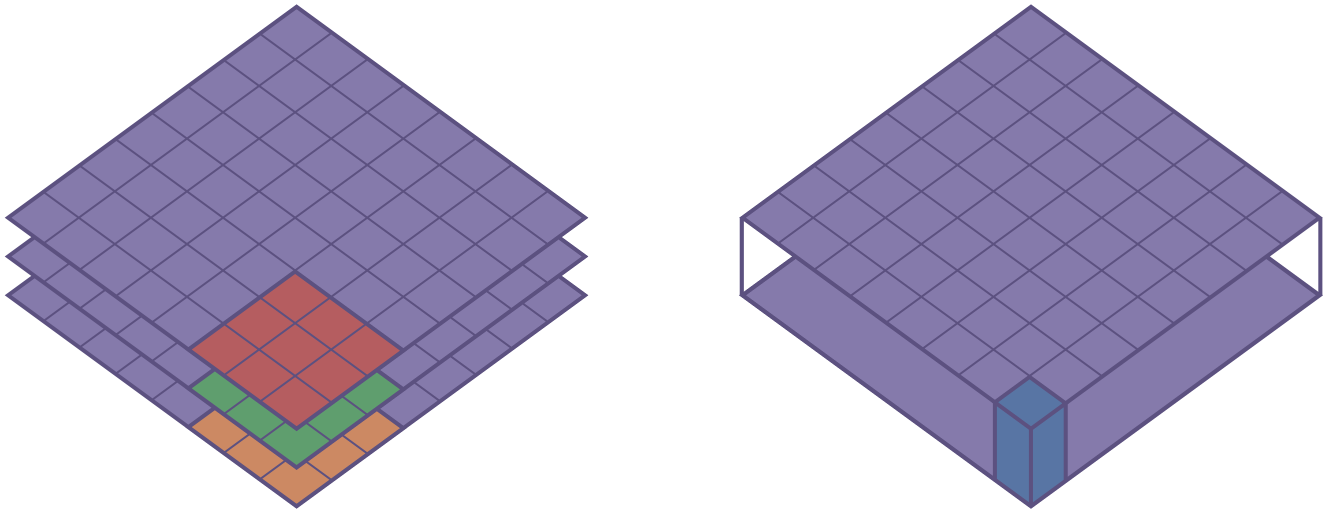


PASCAL VOC

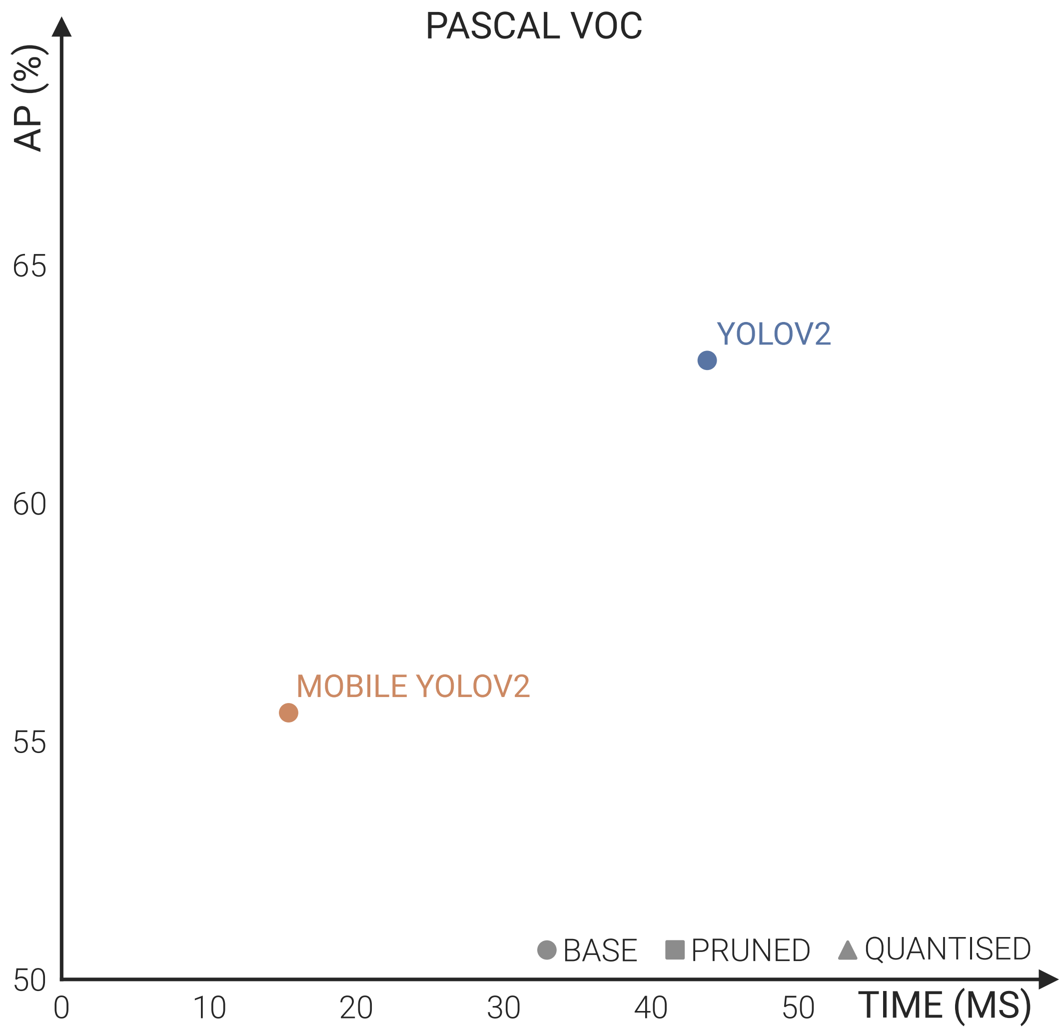
IR SURVEILLANCE

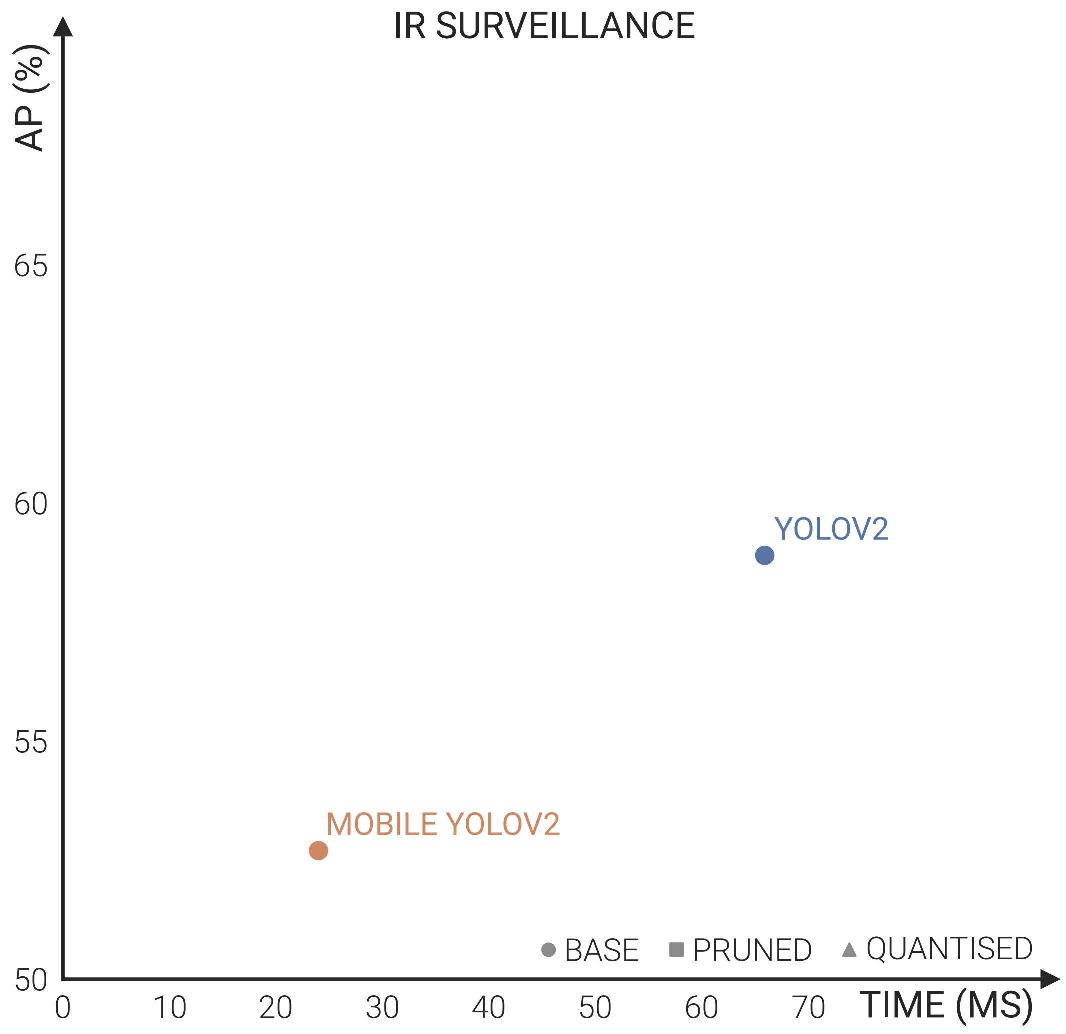
MOBILE CONVOLUTIONS



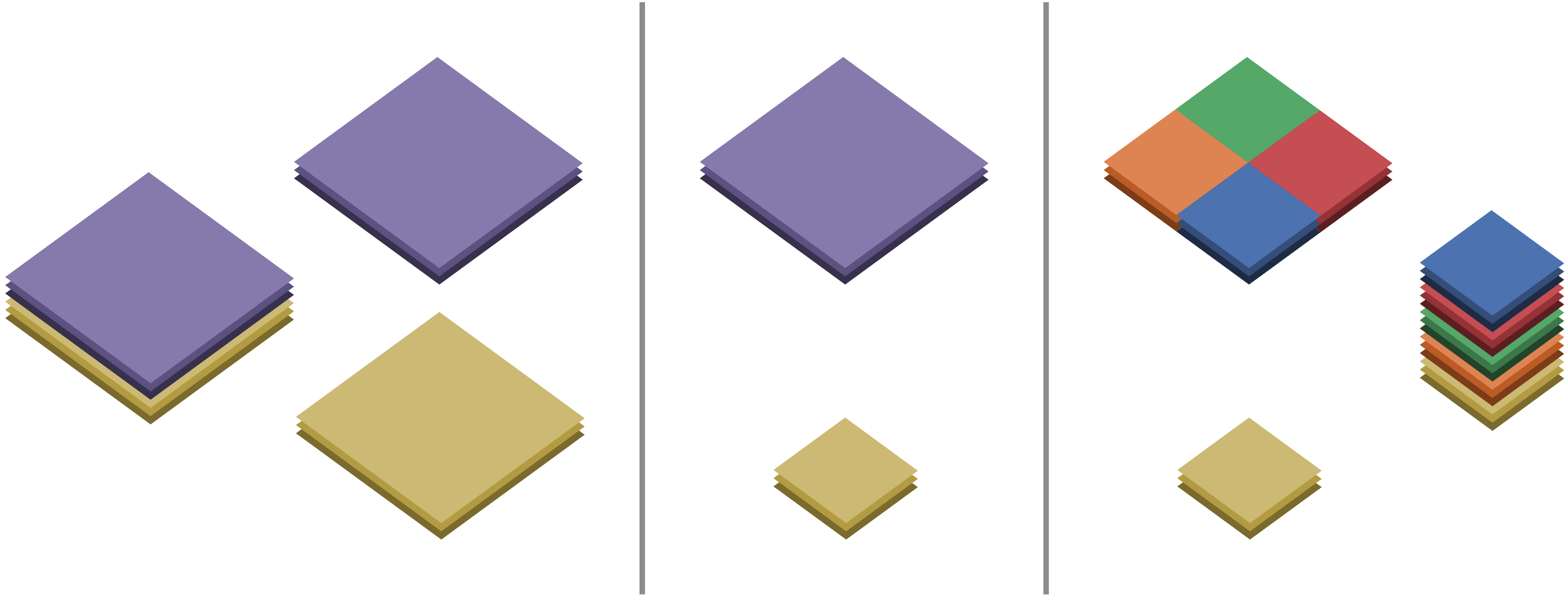


RESULTS

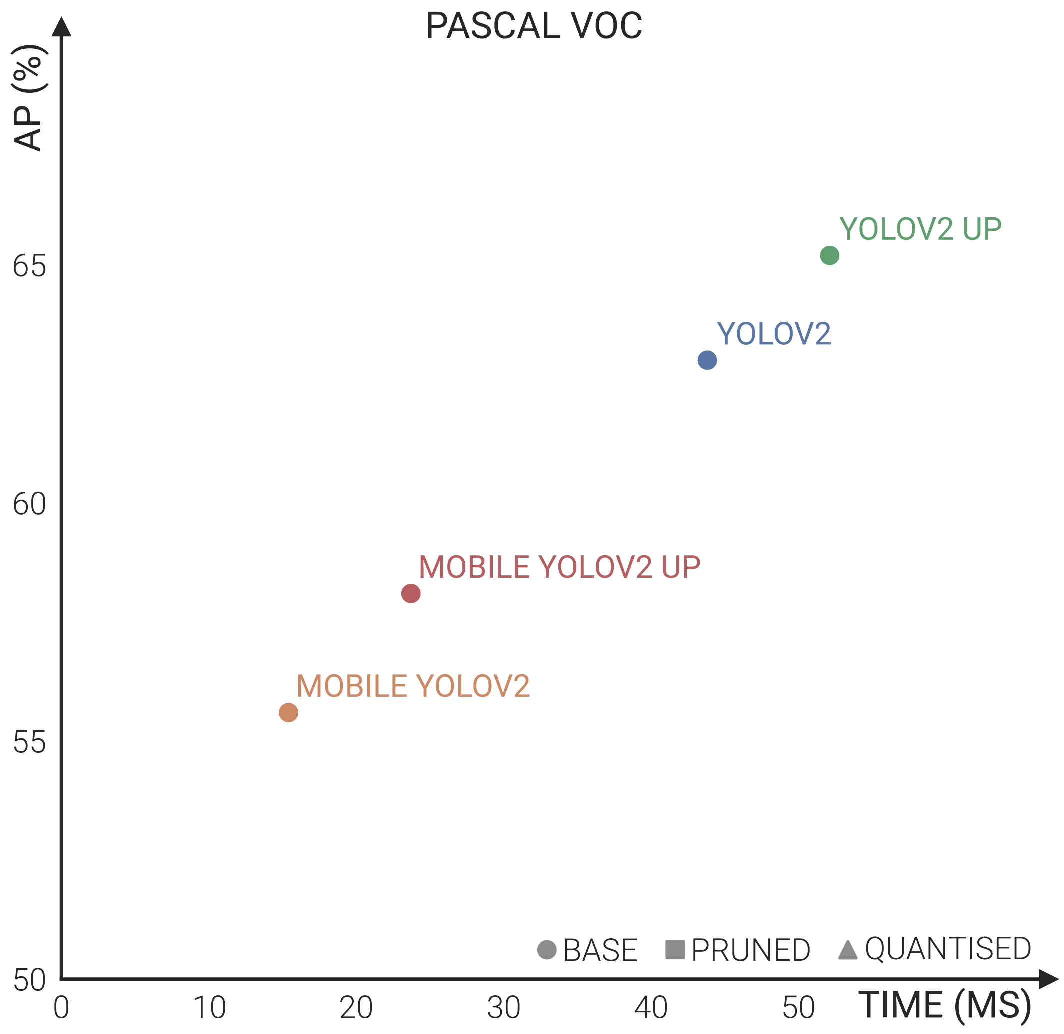

PASCAL VOC
AP (%)
65
60
55
50
YOLOV2
MOBILE YOLOV2
BASE
PRUNED
QUANTISED
0
10
20
30
40
50
TIME (MS)


IR SURVEILLANCE
AP (%)
65
60
55
50
YOLOV2
MOBILE YOLOV2
BASE
PRUNED
QUANTISED
0
10
20
30
40
50
60
70
TIME (MS)

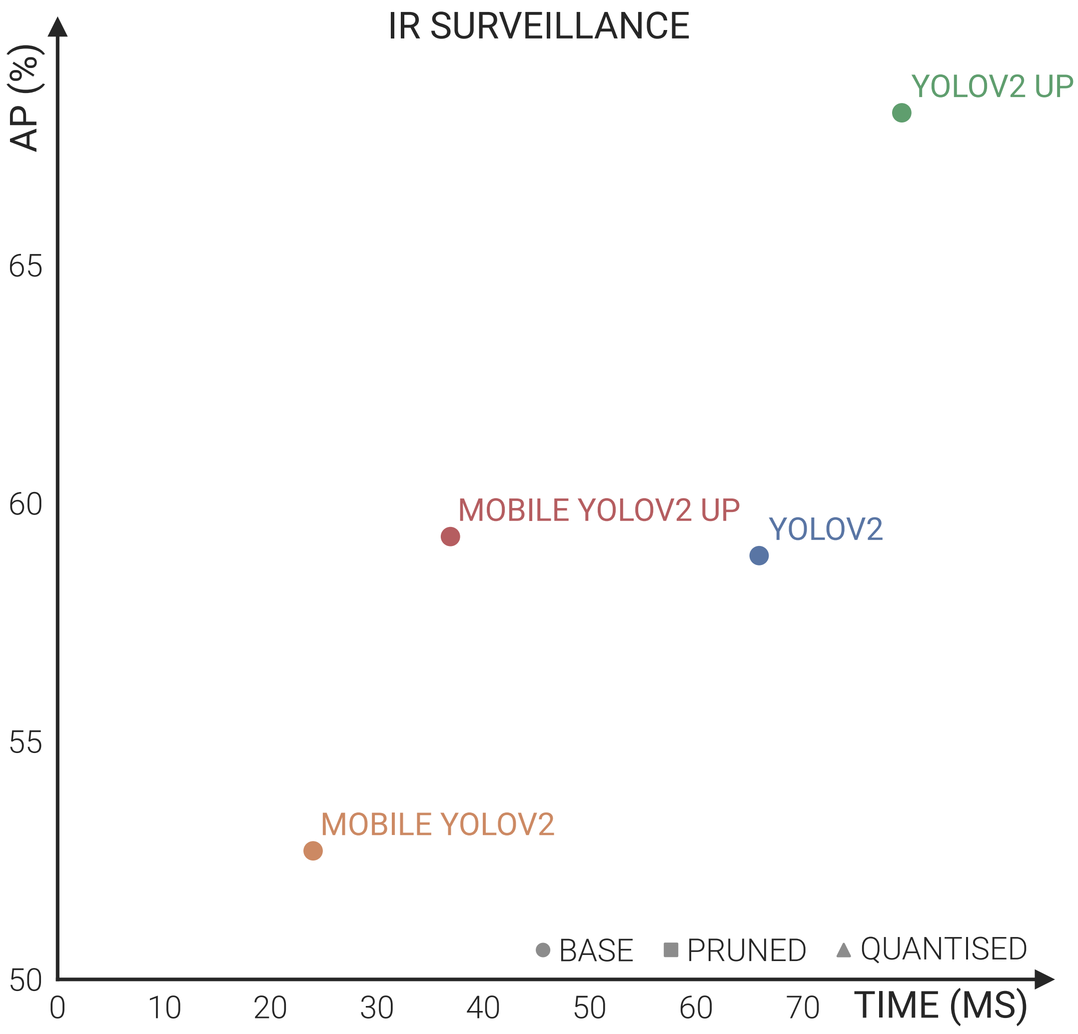
UPSAMPLE



RESULTS

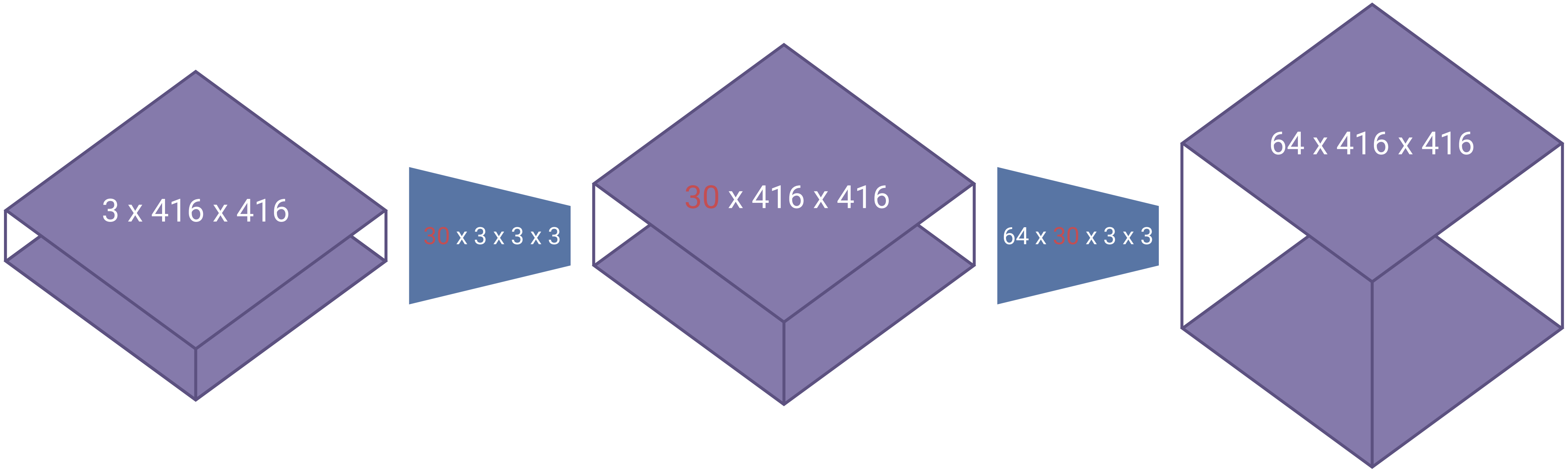


PASCAL VOC
AP (%)
65
60
55
50
YOLOV2 UP
YOLOV2
MOBILE YOLOV2 UP
MOBILE YOLOV2
BASE
PRUNED
QUANTISED
0
10
20
30
40
50
TIME (MS)

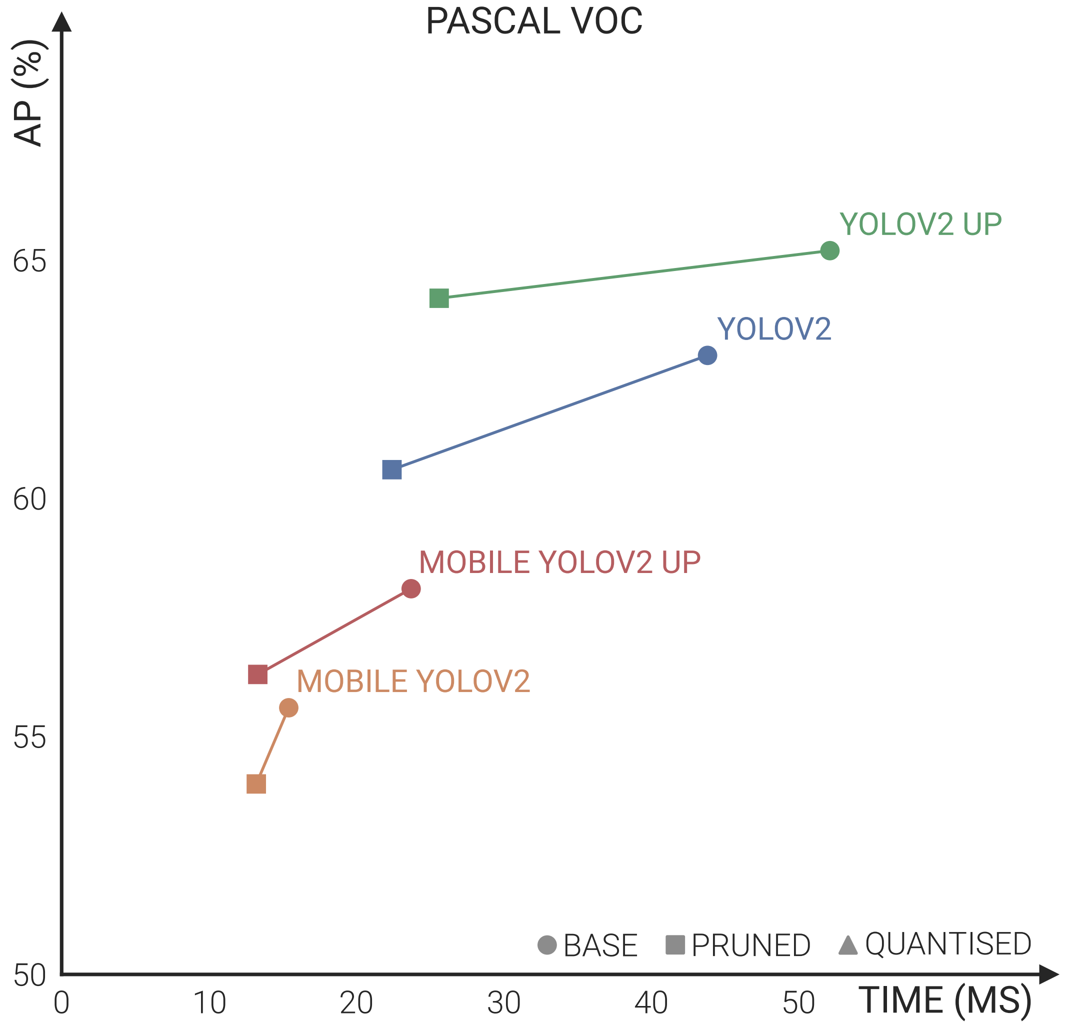


IR SURVEILLANCE
AP (%)
65
60
55
50
YOLOV2 UP
MOBILE YOLOV2 UP
YOLOV2
MOBILE YOLOV2
BASE
PRUNED
QUANTISED
0
10
20
30
40
50
60
70
TIME (MS)

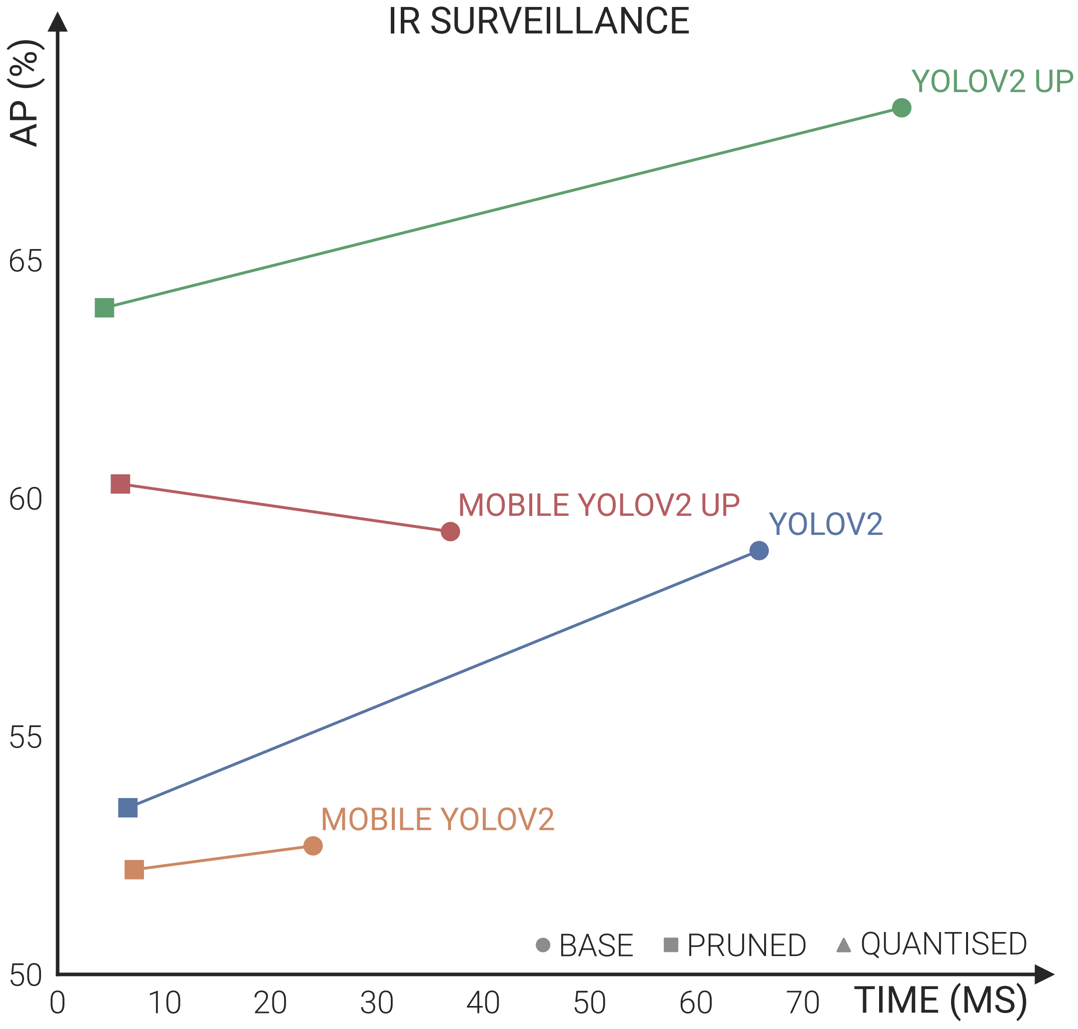
PRUNING



RESULTS



PASCAL VOC
AP (%)
65
60
55
50
0
10
20
30
40
50
TIME (MS)
YOLOV2 UP
YOLOV2
MOBILE YOLOV2 UP
MOBILE YOLOV2
BASE
PRUNED
QUANTISED



IR SURVEILLANCE
AP (%)
65
60
55
50
0
10
20
30
40
50
60
70
TIME (MS)
YOLOV2 UP
MOBILE YOLOV2 UP
YOLOV2
MOBILE YOLOV2
BASE
PRUNED
QUANTISED

QUANTISATION

3.1415

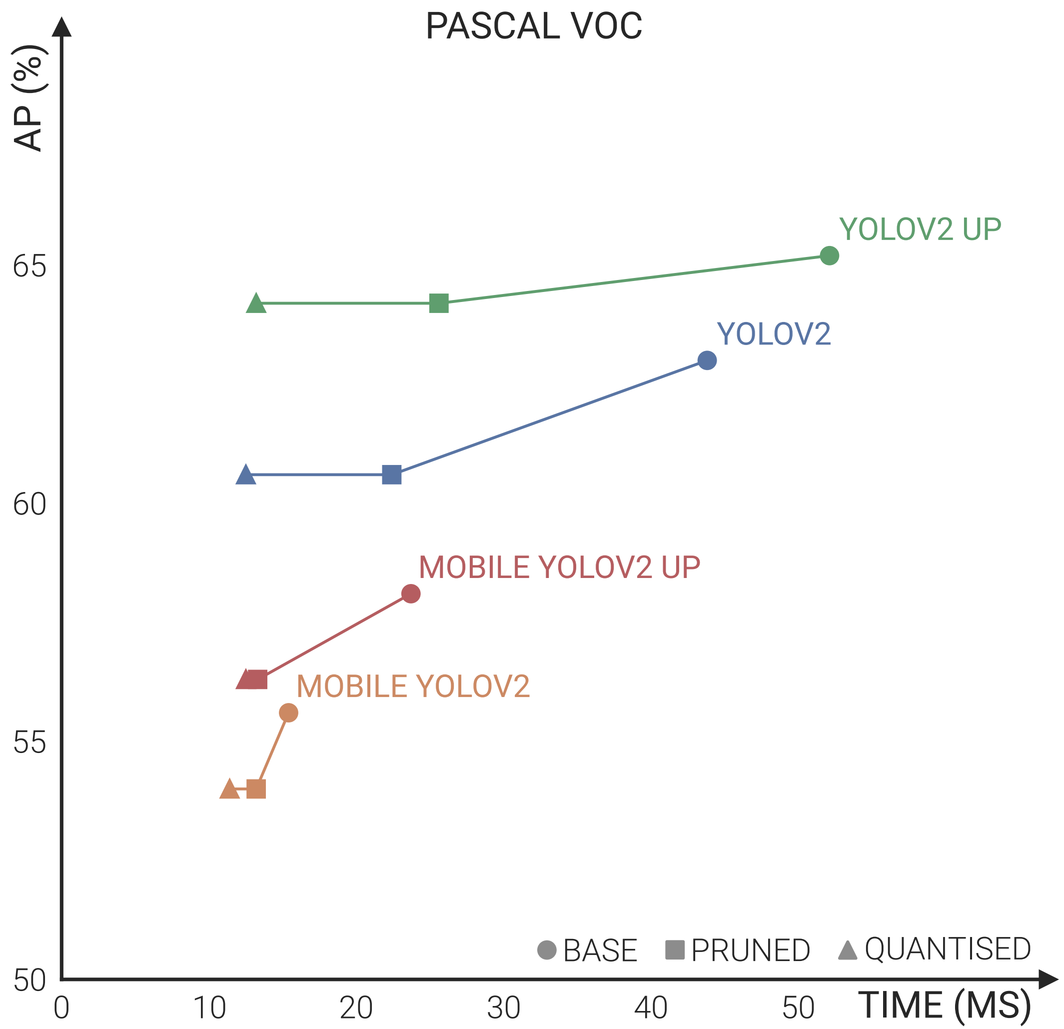
S E E E E E E E E M M M M M M M M M M M M M M M M M M M M M M M

FP32

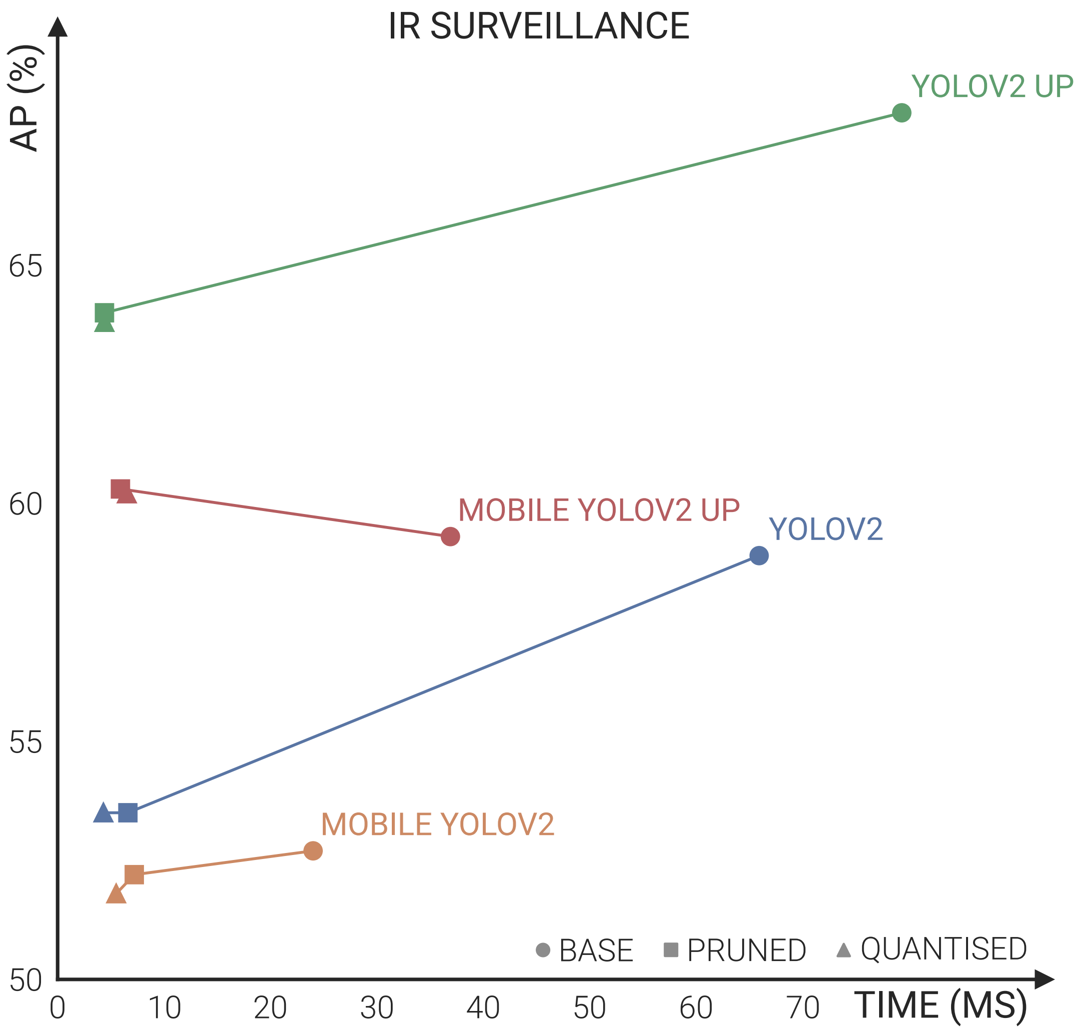
FP16

S E E E E E M M M M M M M M M M

RESULTS



PASCAL VOC
AP (%)
65
60
55
50
0
10
20
30
40
50
TIME (MS)
YOLOV2 UP
YOLOV2
MOBILE YOLOV2 UP
MOBILE YOLOV2
BASE
PRUNED
QUANTISED



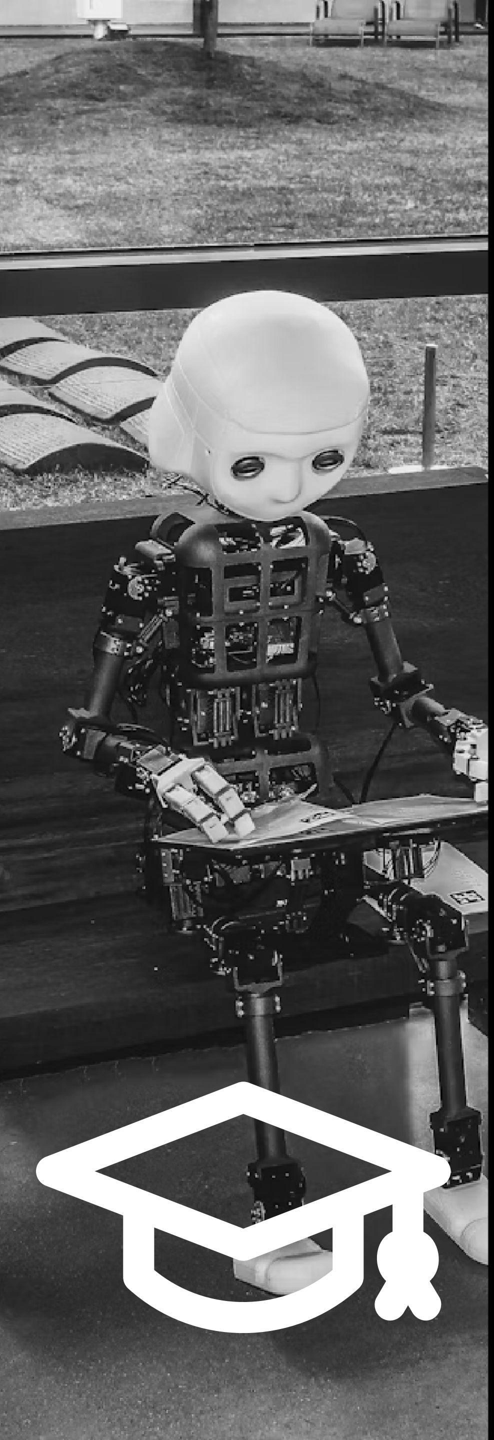
IR SURVEILLANCE
AP (%)
65
60
55
50
0
10
20
30
40
50
60
70
TIME (MS)
YOLOV2 UP
MOBILE YOLOV2 UP
YOLOV2
MOBILE YOLOV2
BASE
PRUNED
QUANTISED

CONTRIBUTIONS

How much can we speed up our models whilst maintaining the accuracy?

- Blindly applying all optimizations does not yield the best results
- On Pascal VOC, we managed to make our model 4x faster
- On LWIR, we made our model 15x faster
- More constrained problems allow for more reduction in complexity

5. T. Ophoff, C. Gullentops, K. Van Beeck, and T. Goedemé. “Investigating the Potential of Network Optimization for a Constrained Object Detection Problem”. In: *Journal of Imaging* 7.4 (2021).





SLIDING WINDOW

SCENE-SPECIFIC PROCESSING

RESNETYOLO

TRANSPARENT FUSE LAYER

MID-LATE FUSION

DIFFERENT USE CASES

REMOTE SENSING

CAREFUL SELECTION

CONSTRAINEDNESS

ACADEMIC 4x

INDUSTRIAL 15x