ECE208 Project Presentation

SGNET: REAL-TIME MULTI-TASK AUTONOMOUS VEHICLE VISION

By Enis zuo, Haixi Zhang, Yufeng Yang

Acknowledgement & Appreciation

Suggestion Offered by Dr. Zhiyao Duan

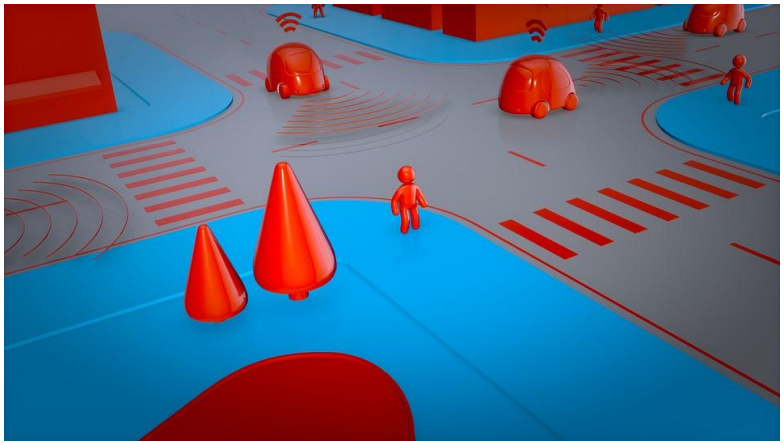
Traffic Light Labels Provided by Zirui Ling, RA, UR IntelliArch Lab

Computation Resource Sponsored by UR IntelliArch Lab, Director Dr. Tong Geng

Linux Setup by Zhuo Liu & Zhenyu Pan, Ph.D Candidates, UR ECE

Structures:

1. Current Problems
2. Related Works
3. Model Description
4. Results
5. Main Contributions



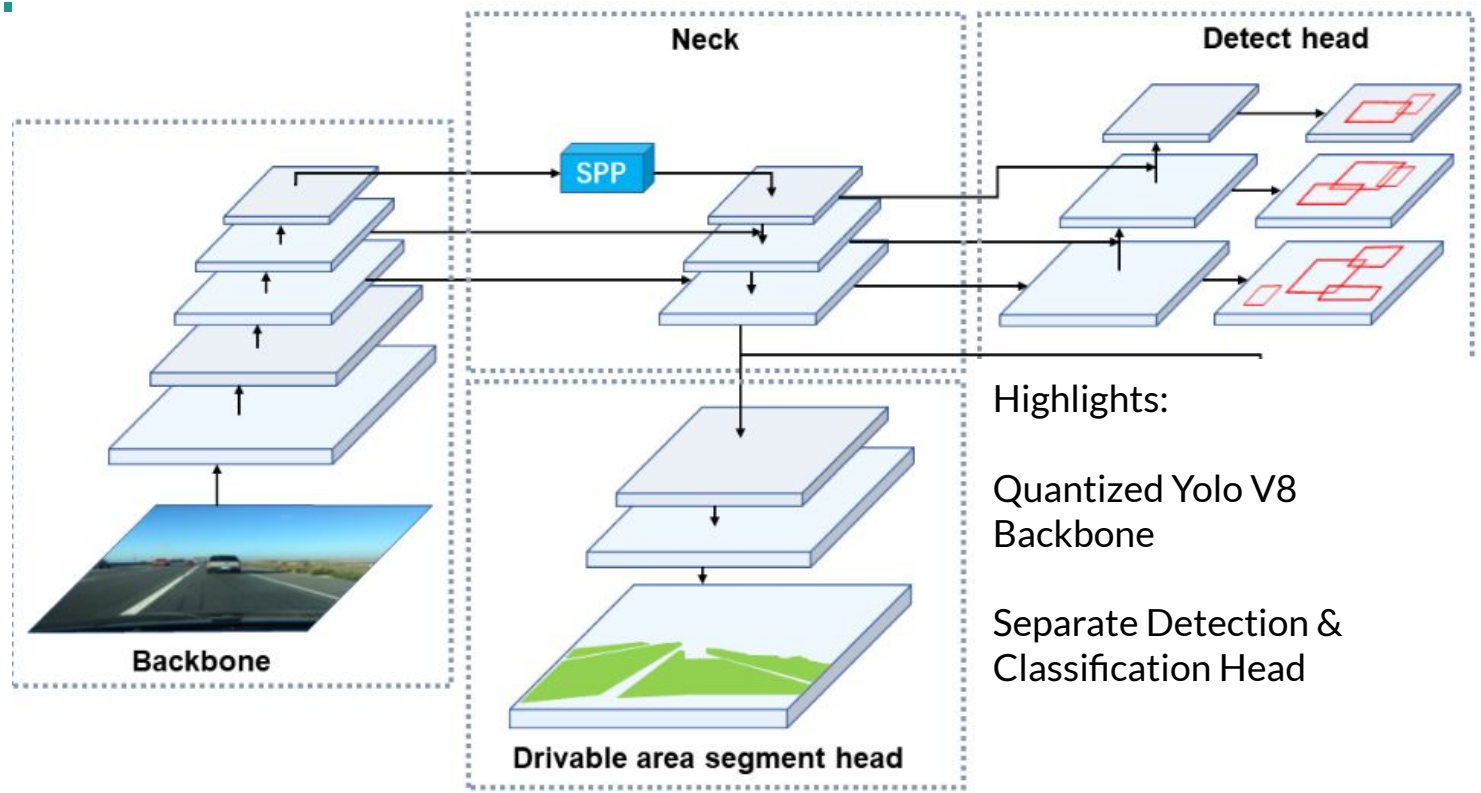
Current Solution & Problems

Multi-type sensors may not an ideal solution:

- **Fusion of different sensors- inconsistencies.**
- **Expensive sensors, intense computation needs.**

Fig 1 Autonomous Demonstration: https://vayyar.com/

Solution: Multi-task Model?



Highlights:

Quantized Yolo V8 Backbone

Separate Detection & Classification Head

Related Works : Interactive Object Detection

Two-stage algorithm- RCNN, Fast(er)-R(

Identify regions + Detect Objects

One-stage algorithm: SSD,YOLO

Fully convolution approach to dete objects

And the latest sota model is the Yo

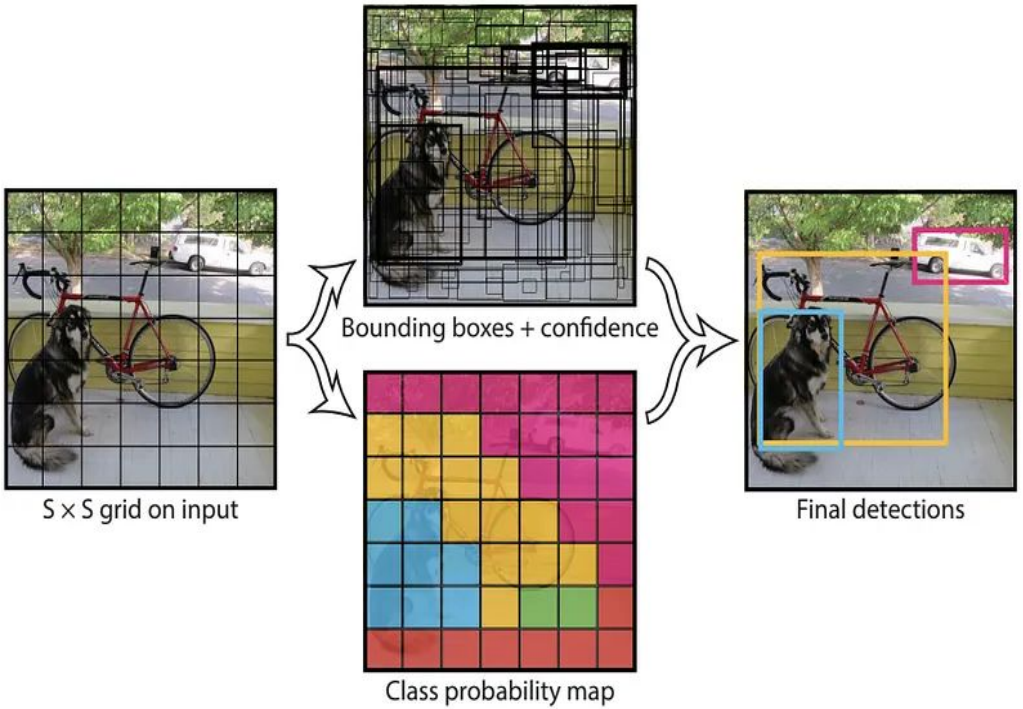


Figure 2 &3
https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

Related Work: SAM

Segment Anything Model:

Advantage: Generic

Problem: Size-Speed,
Non-task-specific

Fig 4 SAM example
credit:https://segment-anything.com/

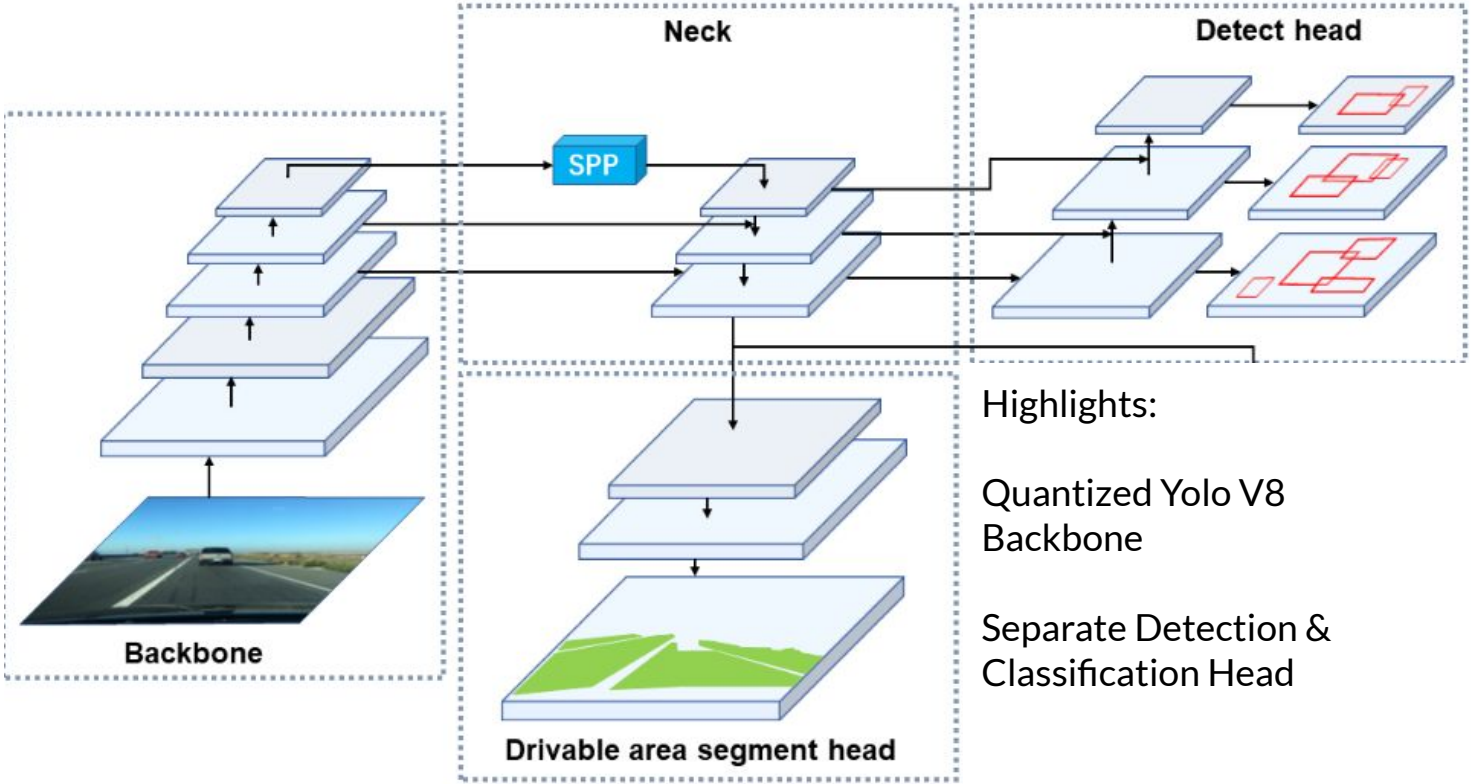
Models Description

SGnet: Detect and Segmentation Heads

Split the target into different networks (lossees) to prevent inter-limitations instead of conventional end-to-end training.

Combining real-time semantic segmentation, object detection in a single model, achieving similar accuracy and speed with single task neural network

Solution: Multi-task Model?



Highlights:

Quantized Yolo V8 Backbone

Separate Detection & Classification Head

Credit: Wu, D., Liao, MW., Zhang, WT. *et al.* YOLOP: You Only Look Once for Panoptic Driving Perception. *Mach. Intell. Res.* 19, 550–562 (2022). https://doi.org/10.1007/s11633-022-1339-y

Convolution Quantization- A Modified DSConv

Weight Quantization: similar weights grouped

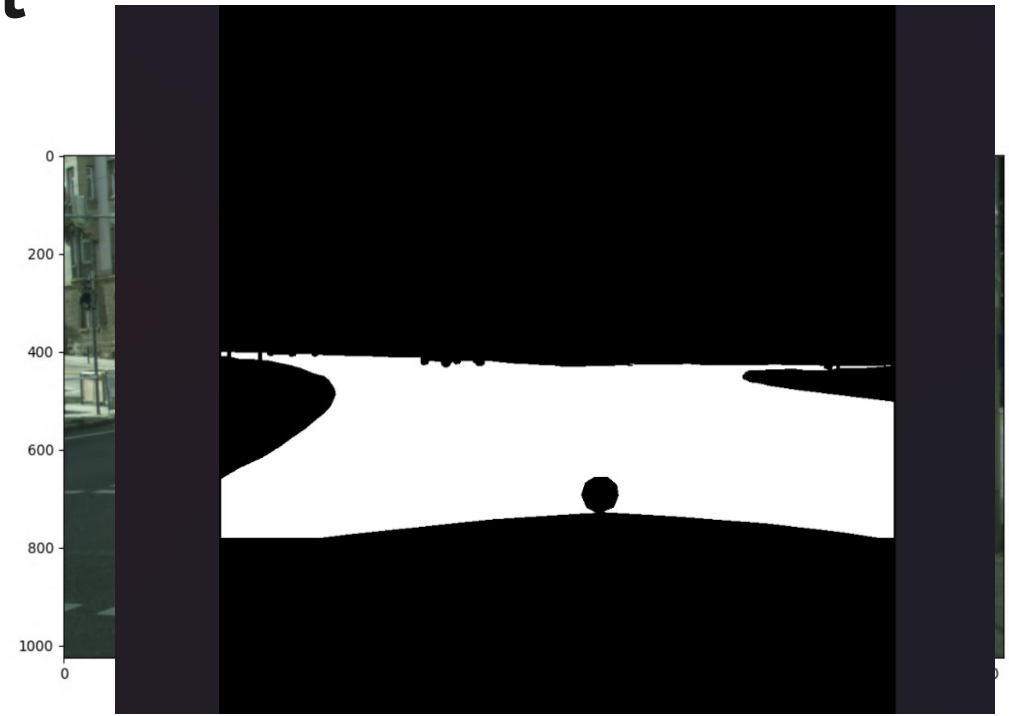
Dynamic Sparsity: non-zero weights change pattern during training

Competitive Accuracy: Dropout Effects in Neural Network

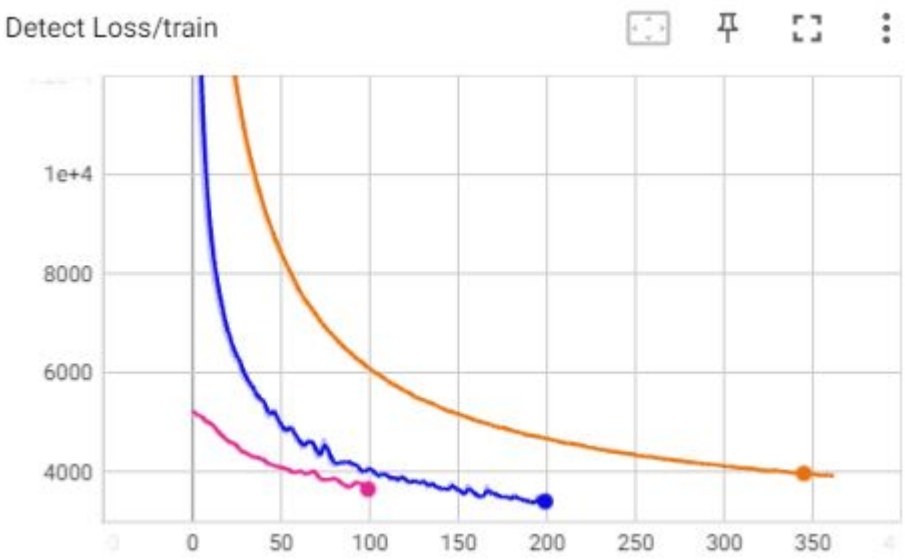
*Gennari, M., Fawcett, R., & Prisacariu, V. A. (2019). DSConv: Efficient Convolution Operator. arXiv preprint arXiv:1901.01928.

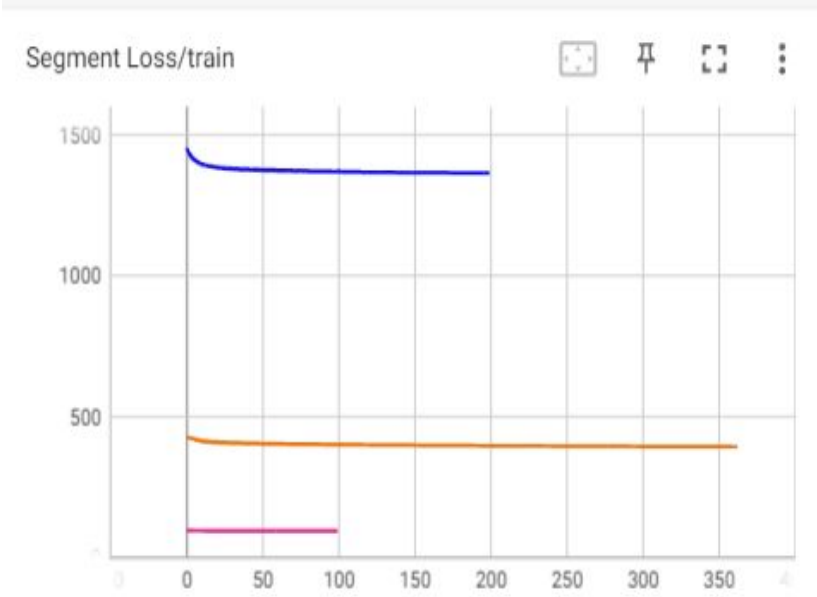
Experiments: Dataset

- Fine segmentation from Cityscapes, bbox generated using conversion tool
- Self-labeled traffic lights by Anylabeling tool.
- Classes: Car, Pedestrian, Truck/Bus, Rider, Traffic Light, Bicycle
- Train: 2975, Test: 500



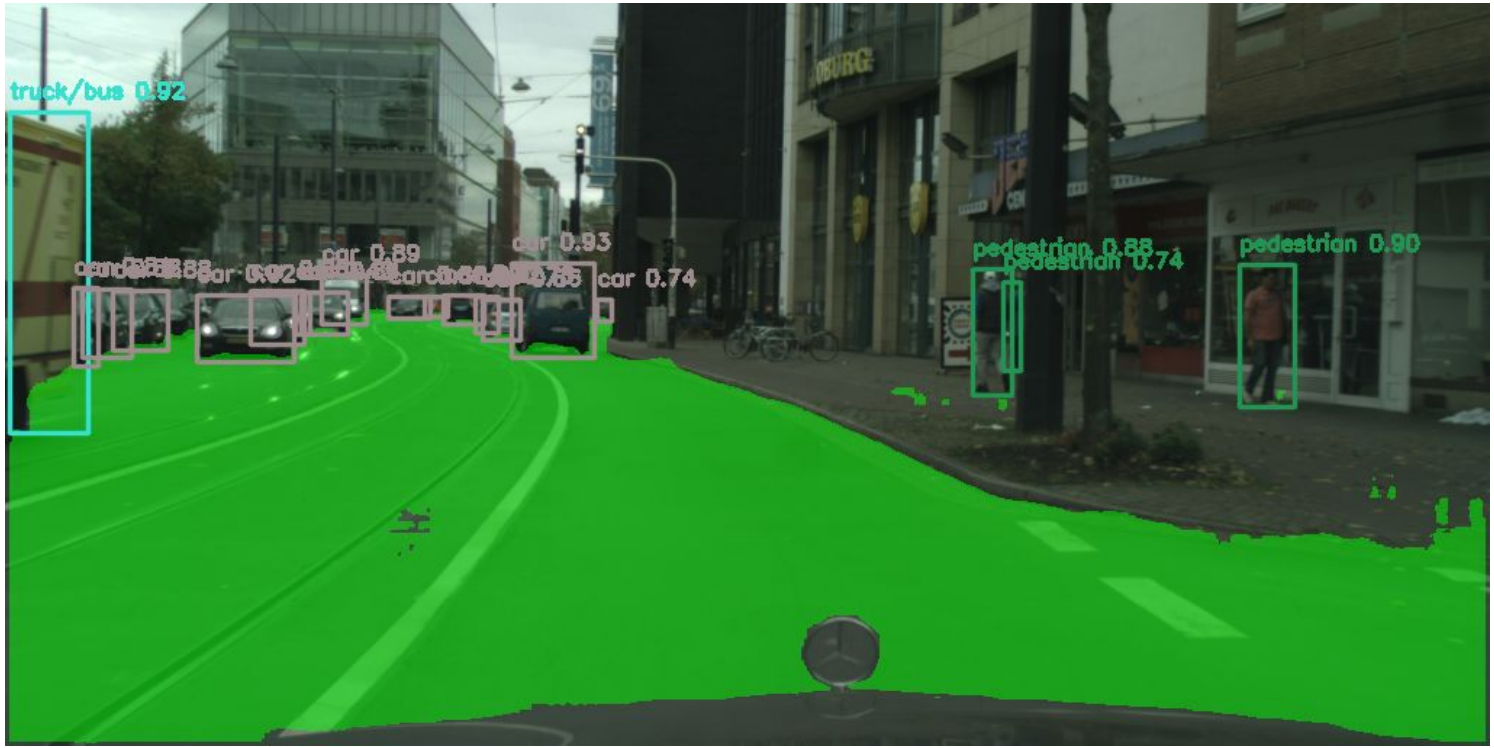
Results of Loss:





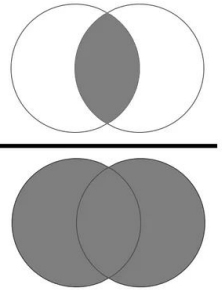
* Trained & Evaluated on Single Nvidia RTX 3090 GPU

Qualitative Result



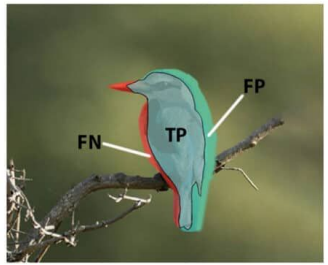
Quantitative Results: Evaluation on Segmentation

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



Ground Truth Mask

Predicted Mask



Network	mIoU(%) [train]	mIoU(%) [test]	Speed(fps)
DSNet	94.2	88.5	96.4
Yolo V8	93.8	88.8	99.4

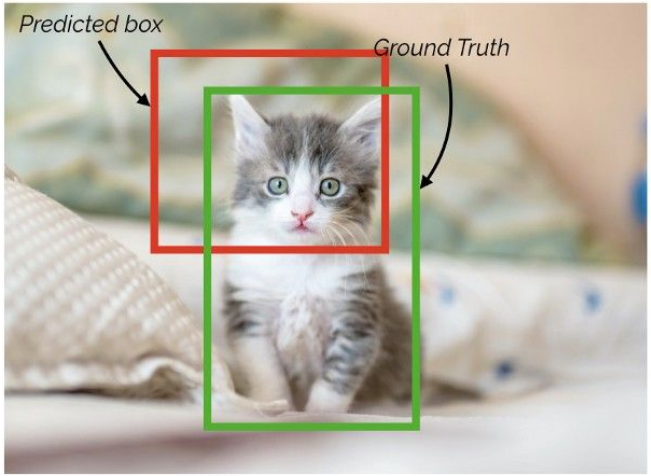
Quantitative Results: Precision & Recall

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

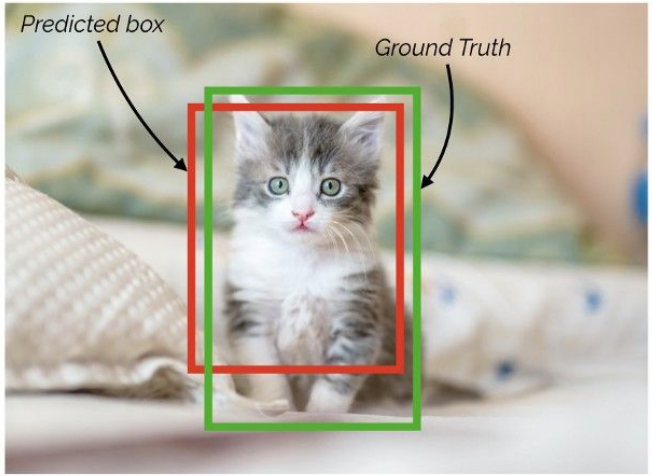
If IoU threshold = 0.5

False Positive (FP)



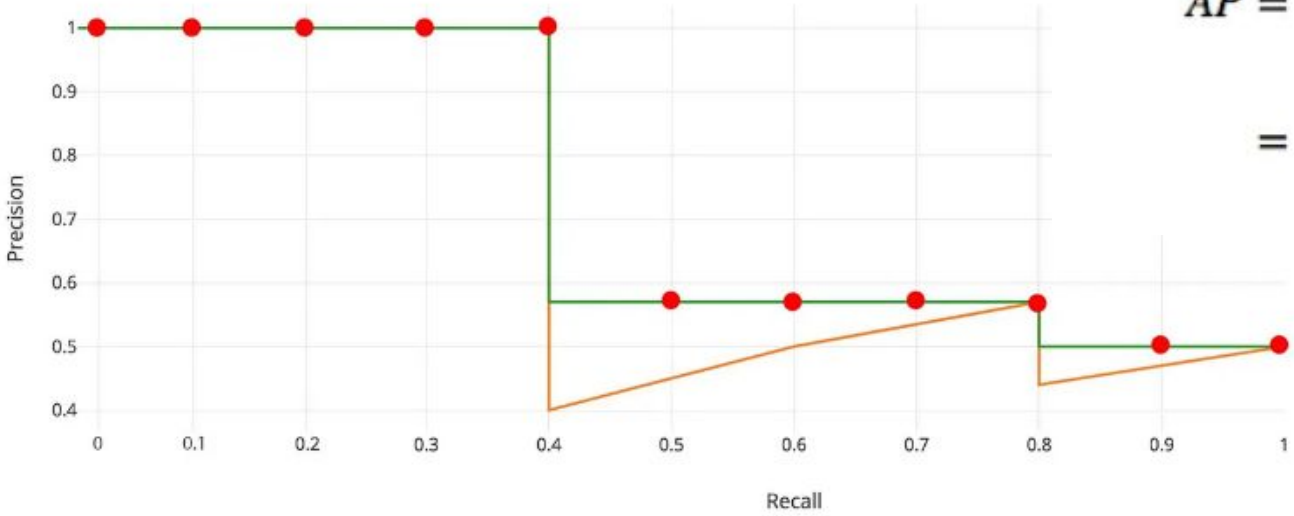
IoU = ~0.3

True Positive (TP)



IoU = ~0.7

Quantitative Results: Evaluation on Object detection

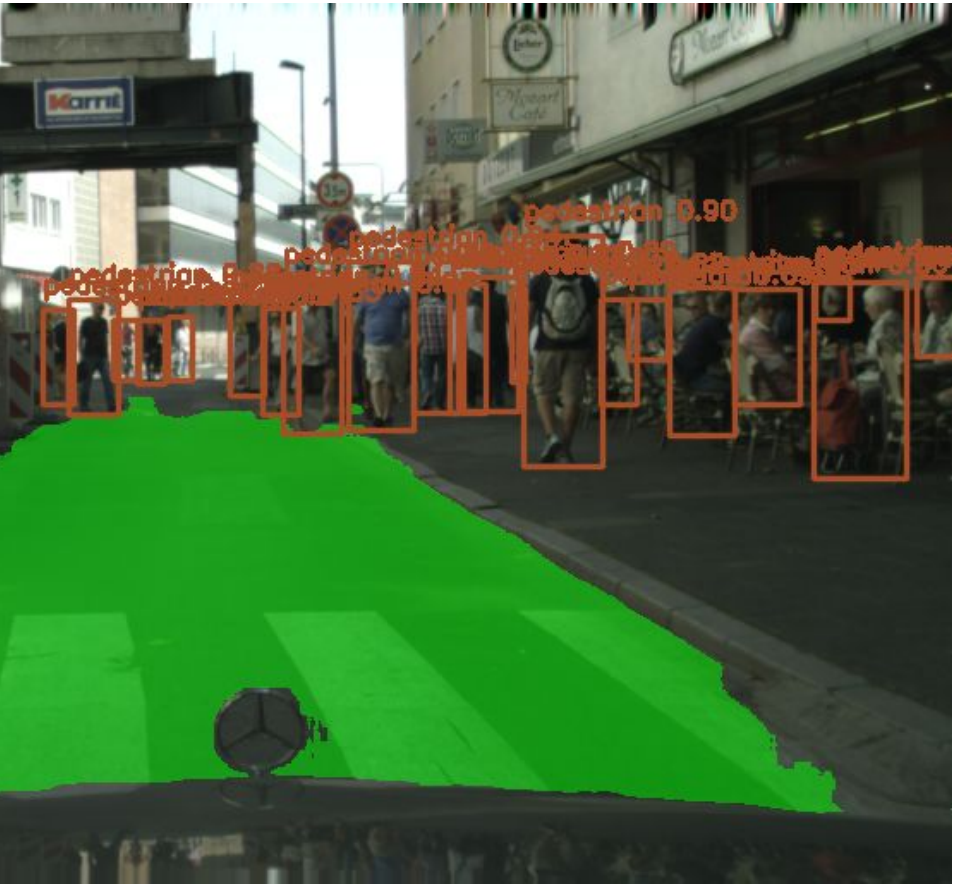


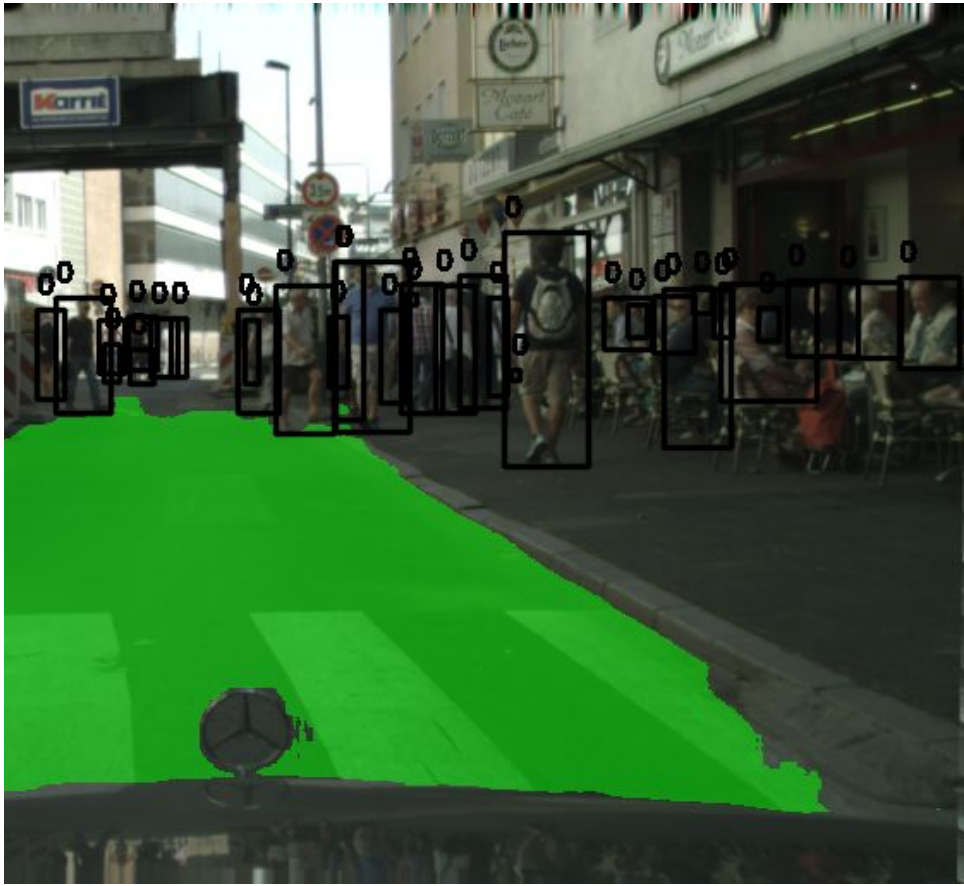
$$AP = \frac{1}{11} \sum_{r \in \{0.0,\ldots,1.0\}} AP_r$$

$$= \frac{1}{11} \sum_{r \in \{0.0,\ldots,1.0\}} p_{interp}(r)$$

Network	mAP50(%) [train]	mAP5095(%) [train]	mAP50(%) [test]	mAP5095(%) [test]	Speed(fps)
DSNet	90.7	75.3	44.0	22.6	96.4
Yolo V8	90.4	71.7	42.0	22.5	99.2

Low mAP, Data or Code?





Conclusion & Contribution

- **Offer Concept for Vehicle Perceptron with Monocular Images**
- **Enable Human-alike Pipeline for Autonomous Driving**
- **Base Model for Further Research on Trajectory Prediction**