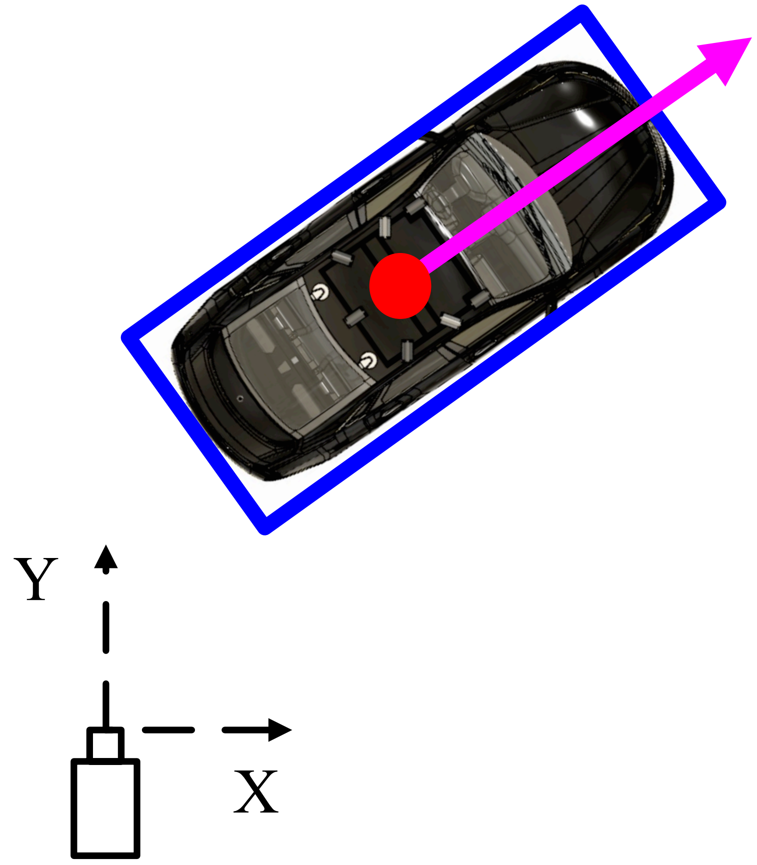
Using 2D Object Detectors for Self-Driving Perception

Course 3, Module 4, Lesson 4

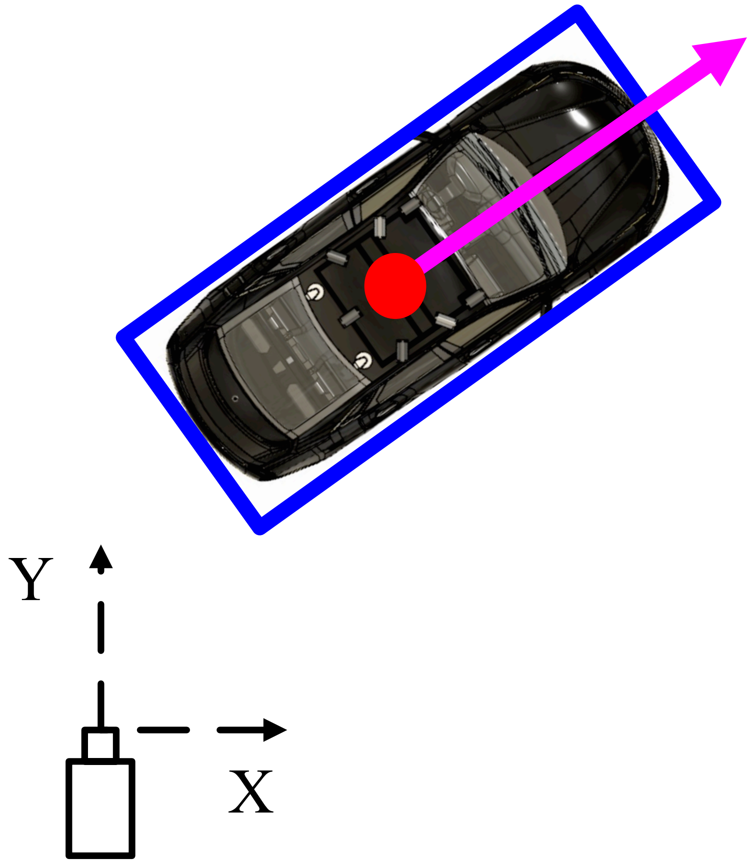
UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

3D Object Detection



- Estimating the:
  - Category Classification
  - Position of the centroid in 3D
  - Extent in 3D
  - Orientation in 3D

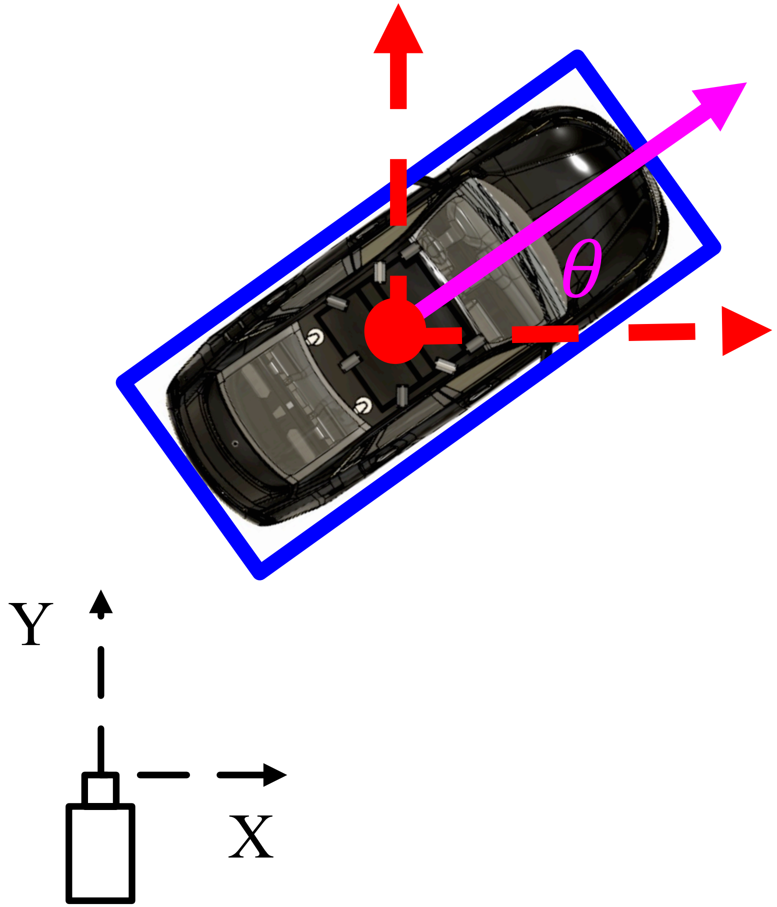
3D Object Detection



- Estimate the:
  - Car, pedestrian, cyclist
  - $[x, y, z]$
  - $[l, w, h]$
  - $[\phi, \psi, \theta]$

yaw

3D Object Detection



- Estimate the:
  - Car, pedestrian, cyclist
  - $[x, y, z]$
  - $[l, w, h]$
  - $[\phi, \psi, \theta]$

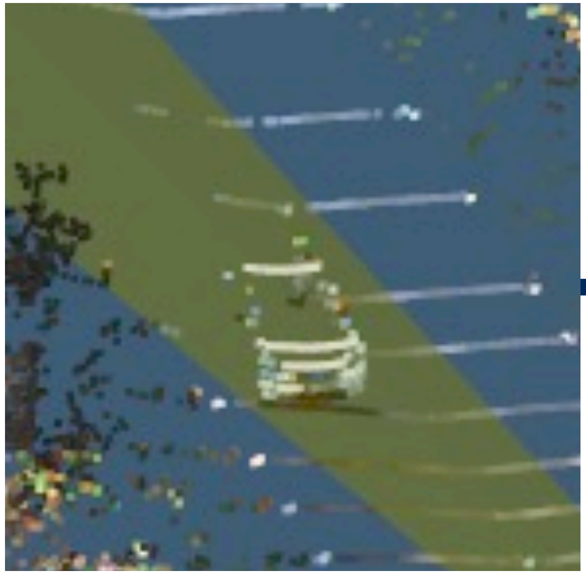
From 2D → 3D Object Detection

Data From: Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the KITTI vision benchmark suite." *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012.

From 2D → 3D Object Detection

(raw data)
X Y Z
or normalized X Y Z

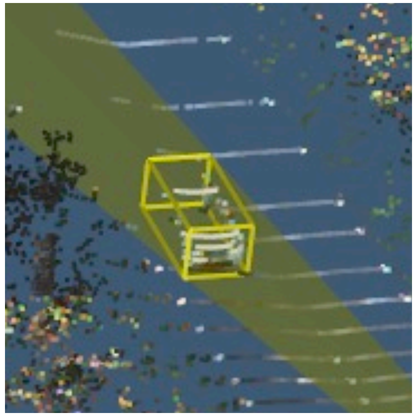
or



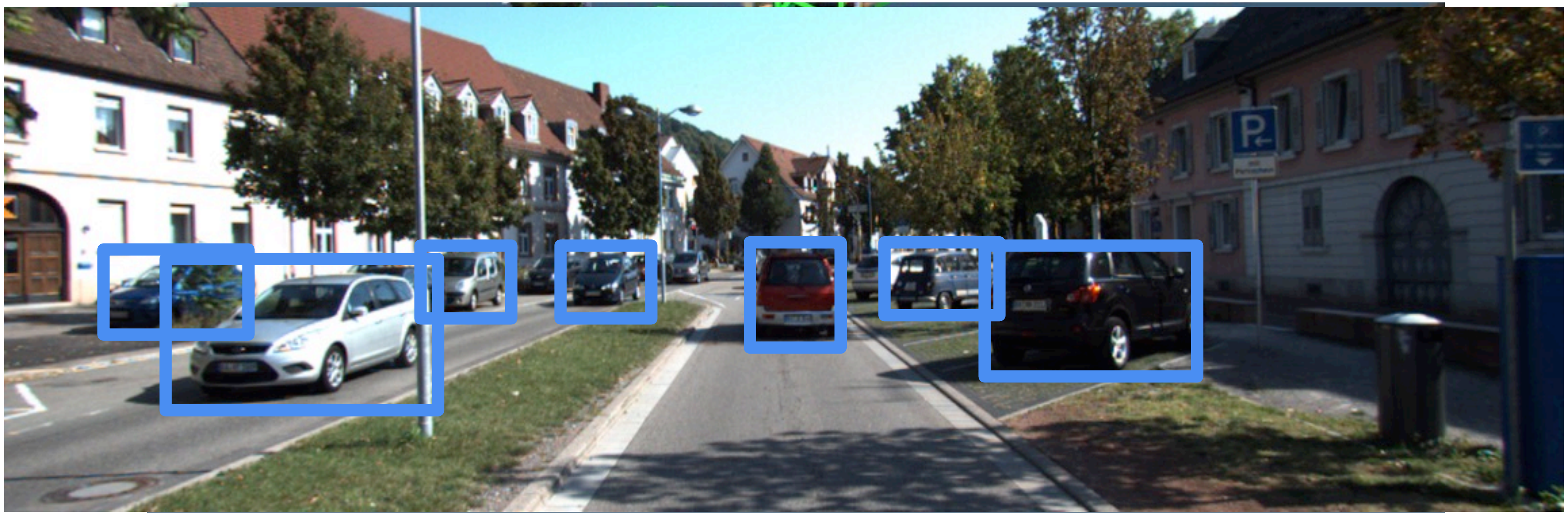
Histogram of
X Y Z

Neural
Network

$[x, y, z], [l, w, h], \theta$



Typical 3D object detection results



Code at: https://github.com/kujason/avod

From 2D → 3D Object Detection

- Advantages:
  - Allows exploitation of mature 2D object detectors, with high precision and recall
  - Class already determined from 2D detection
  - Does not require prior scene knowledge, such as ground plane location

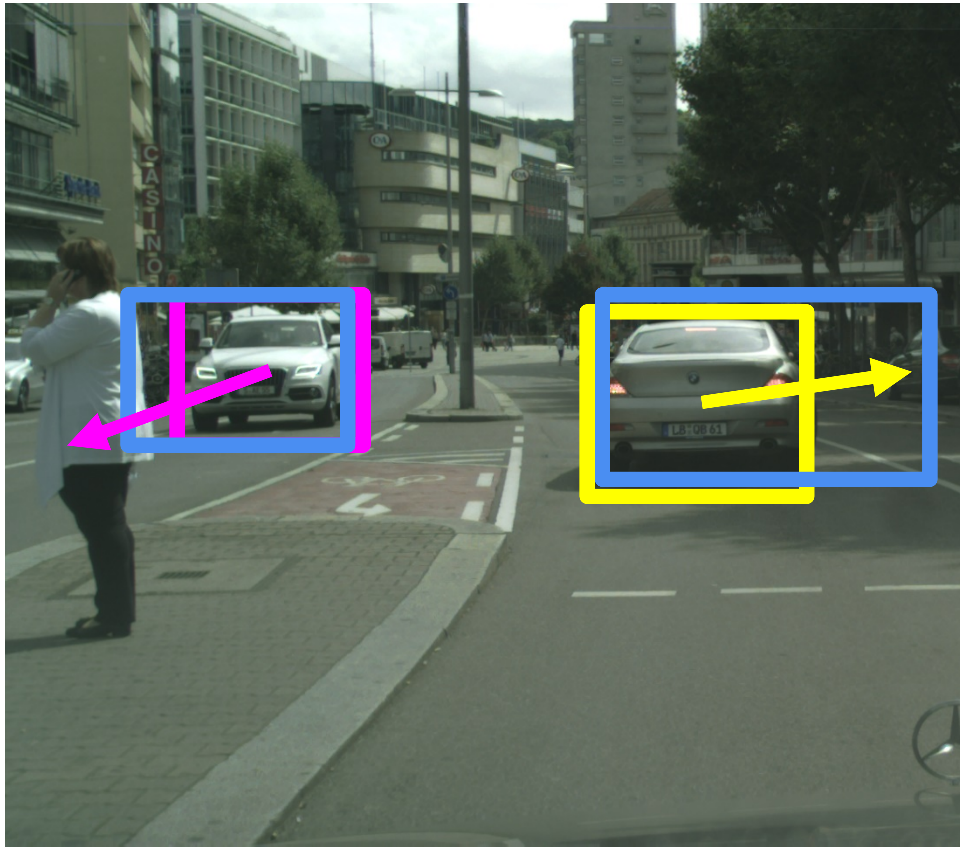
- Disadvantages:
  - The performance of the 3D estimator is bounded by the performance of the 2D detector
  - Occlusion and truncation are hard to handle from 2D only
  - 3D estimator needs to wait for 2D detector, inducing **latency** in our system

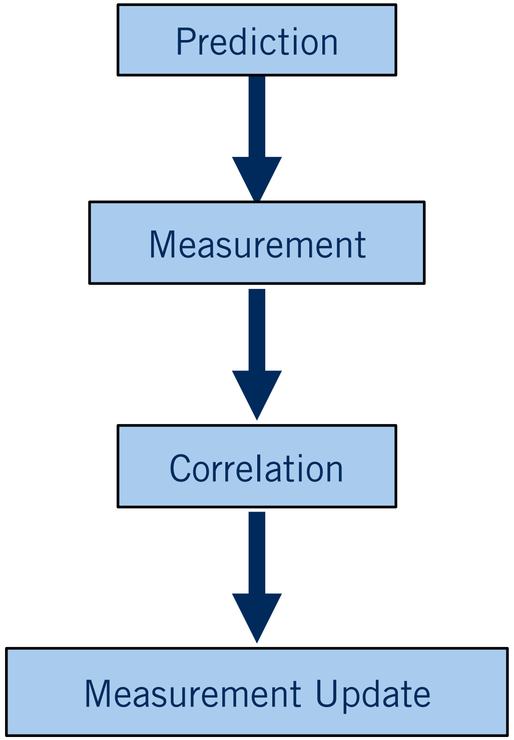
delayed perception

2D Object Tracking

- **Detection**: We detect the object independently in each frame and can record its position over time
- **Tracking**: We use image measurements to estimate position of object, but also incorporate position predicted by dynamics, i.e., our expectation of object's motion pattern
- **Tracking Assumptions:**
    - Camera is not moving instantly to new viewpoint
    - Objects do not disappear and reappear in different places in the scene
    - If the camera is moving, there is a gradual change in pose between camera and scene

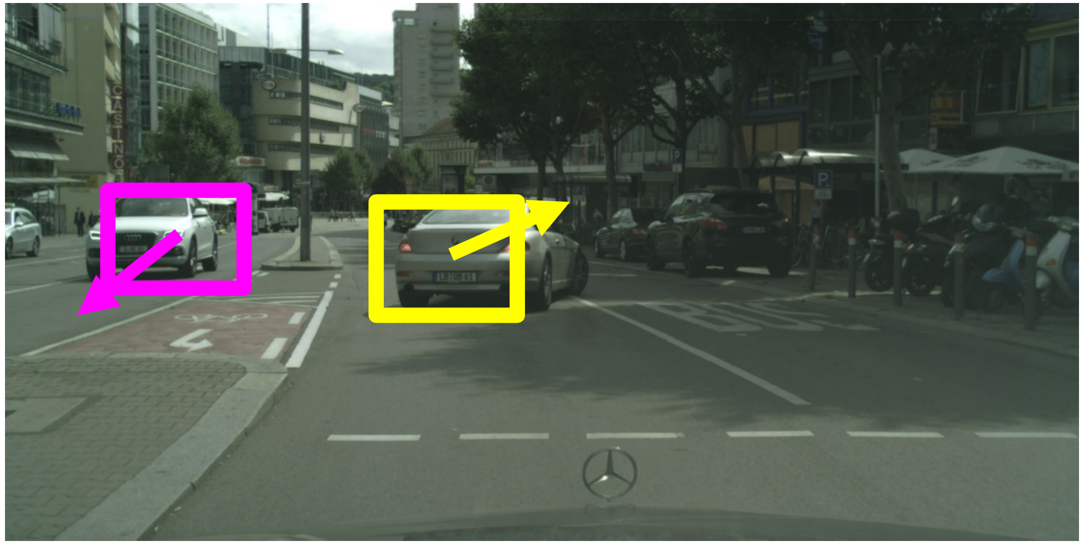
2D Object Tracking





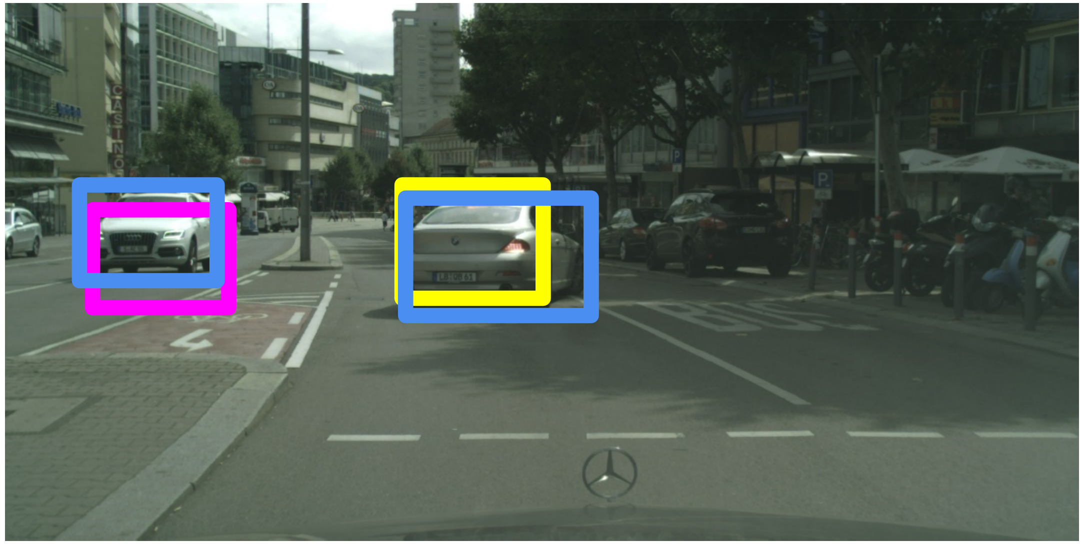
Object Tracking: Prediction

- Each object will have a predefined **motion model** in image space
- **Example:** $p_k = p_{k-1} + v_k \Delta t + \mathcal{N}(0, \Sigma)$ Noise



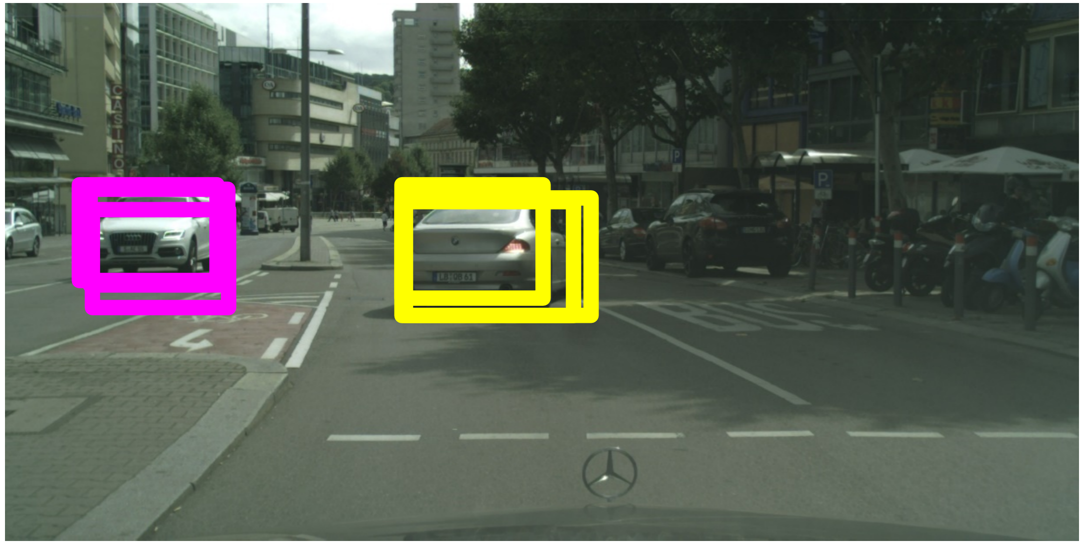
Object Tracking: Correlation

- Get **Measurement Bounding Boxes** from 2D detector.
- Correlate prediction with the **highest** IOU measurement



Object Tracking: Update

- The prediction and measurement are fused as part of the **Kalman Filter Framework**



Object Tracking

- For each frame, we start new track if a measurement has no correlated prediction
- We also terminate inconsistent tracks, if a predicted object does not correlate with a measurement for a **preset** number of frames
- The same methodology can be used to track objects in 3D!

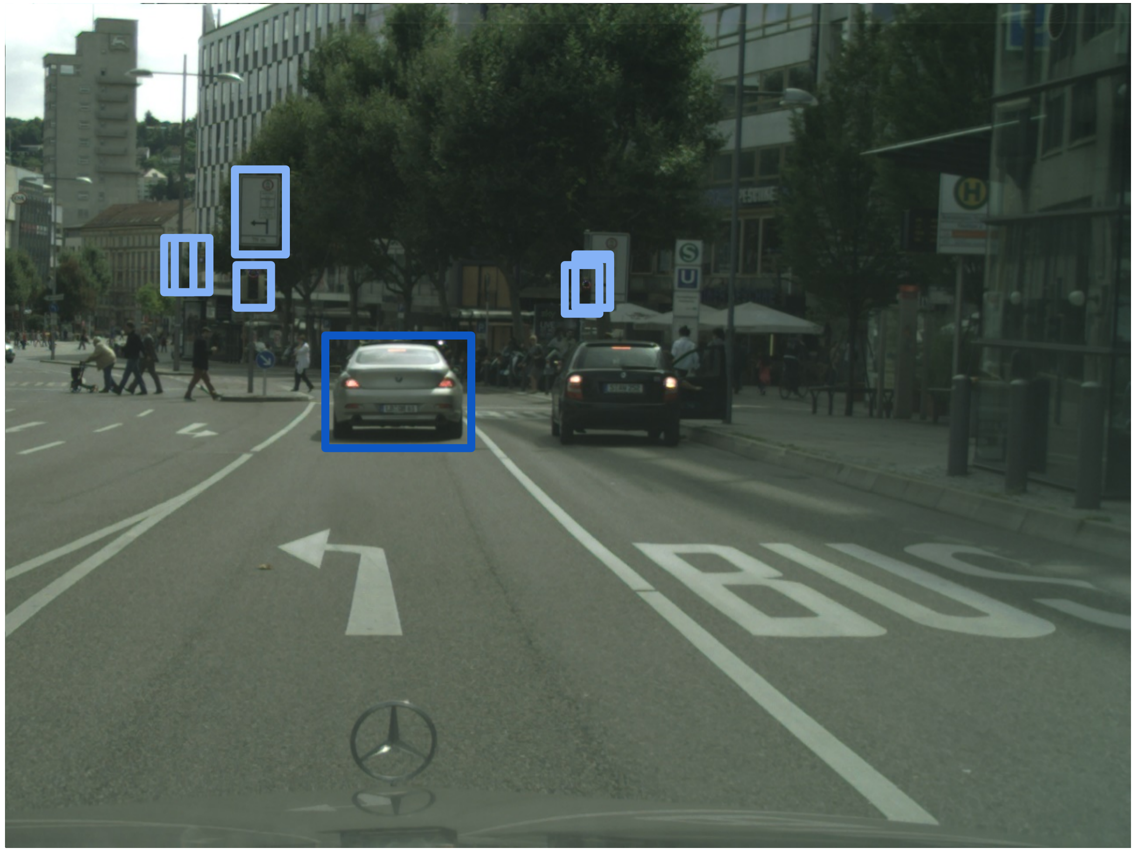
Traffic Sign and Traffic Signal Detection

- Traffic signs and signals appear smaller in size compared to cars, two-wheelers, and pedestrians.

- **Traffic signs** are highly variable with many classes to be trained on.

- **Traffic signals** have different states that are required to be detected.

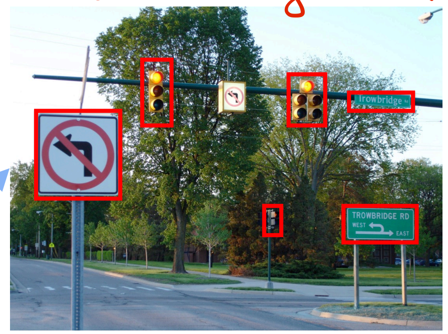
- In addition, traffic signals change state as the car drives!

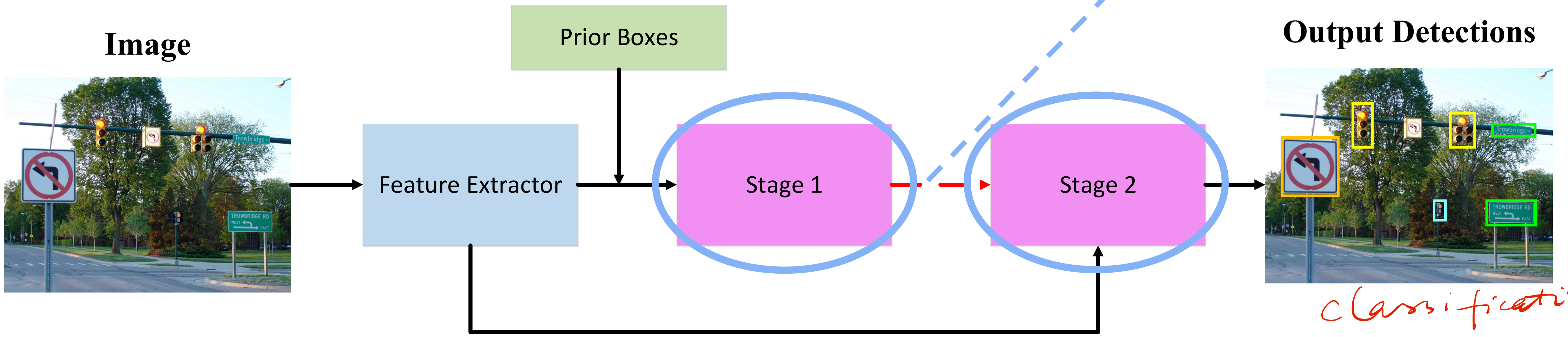


Traffic sign and signal detection

- 2D object detectors can be used to perform traffic sign and traffic signal detection without any modifications
- However, **multi-stage hierarchical models** have been shown to outperform the standard single stage object detectors

bounding box





classification of the bounding box

Summary

- The output of 2D object detectors can be extended to produce 3D object location and dimensions
- The output of 2D object detectors in consecutive frames can be used to track objects in 2D and in 3D
- The output of 2D object detectors can be used to detect traffic signs and signals, and determine the state of traffic signals