

Camera Input Image

YOLOv7 Instance Segmentation

Frame at Time t

HD Map Heading Lookup Grids

Pixel Mask

2D Bounding Box

Category

Yaw Option Histogram

Bottom Contour Extraction and 3D Projection

Outlier Filtering

Map Grid Lookup

Detected Instance n

3D Bottom Contour

VRU
Detection

Map- and Tracking-Guided L-Shape Fitting (LSF)

L-Shape-Fitting and BEV Dimension Filter

Yaw Proposal θ_k

Width

Length

Height and Position Regression

Position

Height

Historical Plausibility Scoring

Match to tracked objects

 $\text{argmax}(\text{Score}(\theta_k))$ Detected 3D
Object n

$$\text{Best} \begin{pmatrix} \vec{x}\vec{y}\vec{z} \\ \vec{l}\vec{w}\vec{h} \\ \theta \end{pmatrix} \text{Estimate}$$

Remember detected object

Tracked Objects

$$\begin{pmatrix} \vec{x}\vec{y}\vec{z}_{t-1}, \dots, \vec{x}\vec{y}\vec{z}_{t-T} \\ \text{length} \\ \text{width} \\ \text{height} \\ \vec{\text{bbox}}_{2d} \\ \text{yaw} \\ \text{category} \end{pmatrix}_{0 \leq i \leq N}$$

