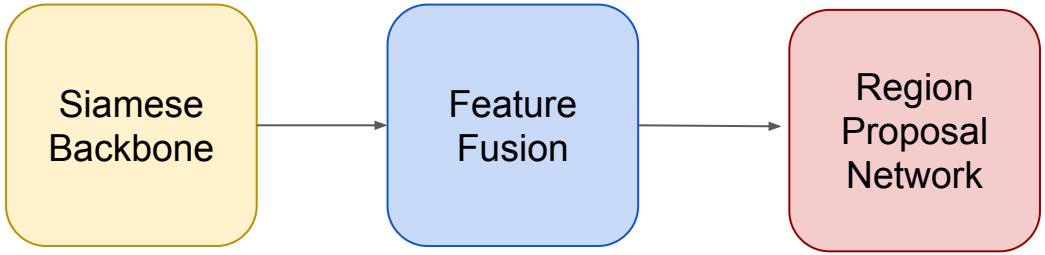
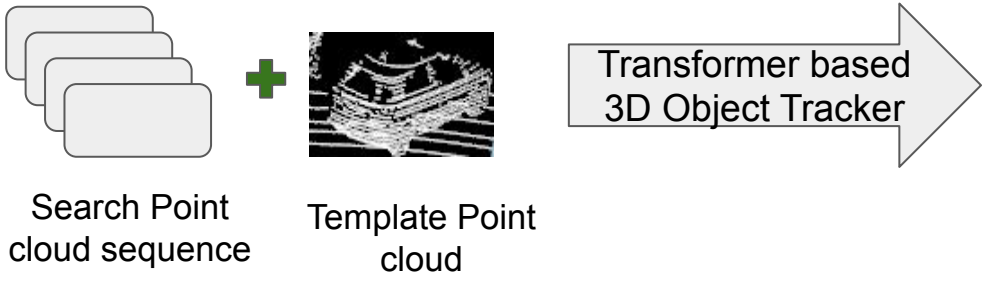
Transformer-based 3D Single Object Tracker

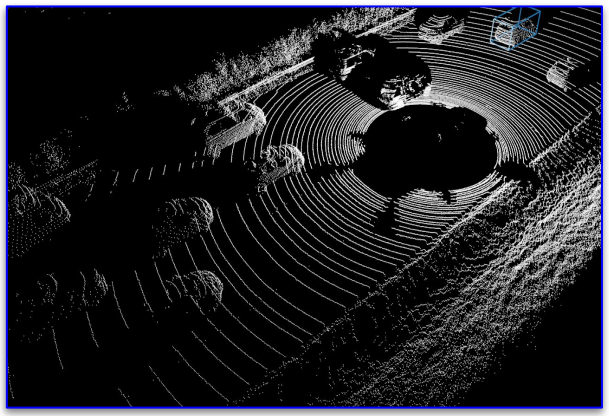
Akshay Kumar Sureddy

Roshini Pulishetty

Goal: Transformers as Feature Fusion Networks

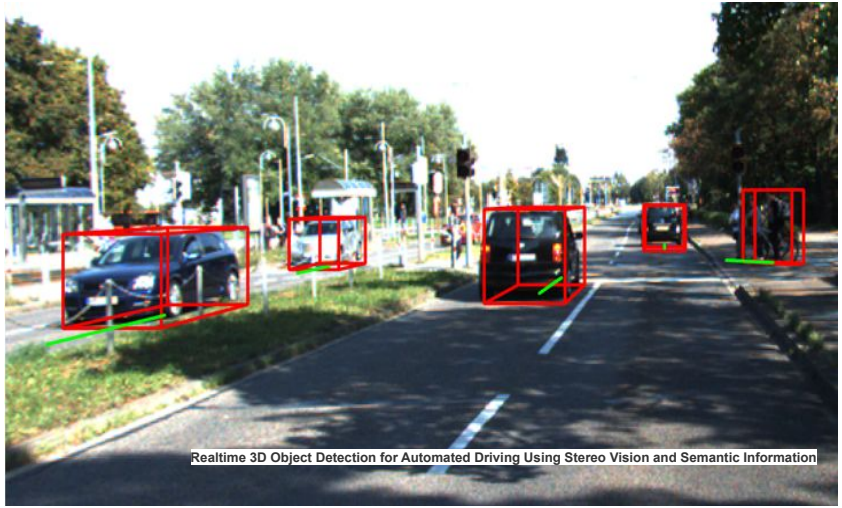






Motivation: 3D Object Tracking for Self-Driving Cars

Robust methods for Real-time Object Localization

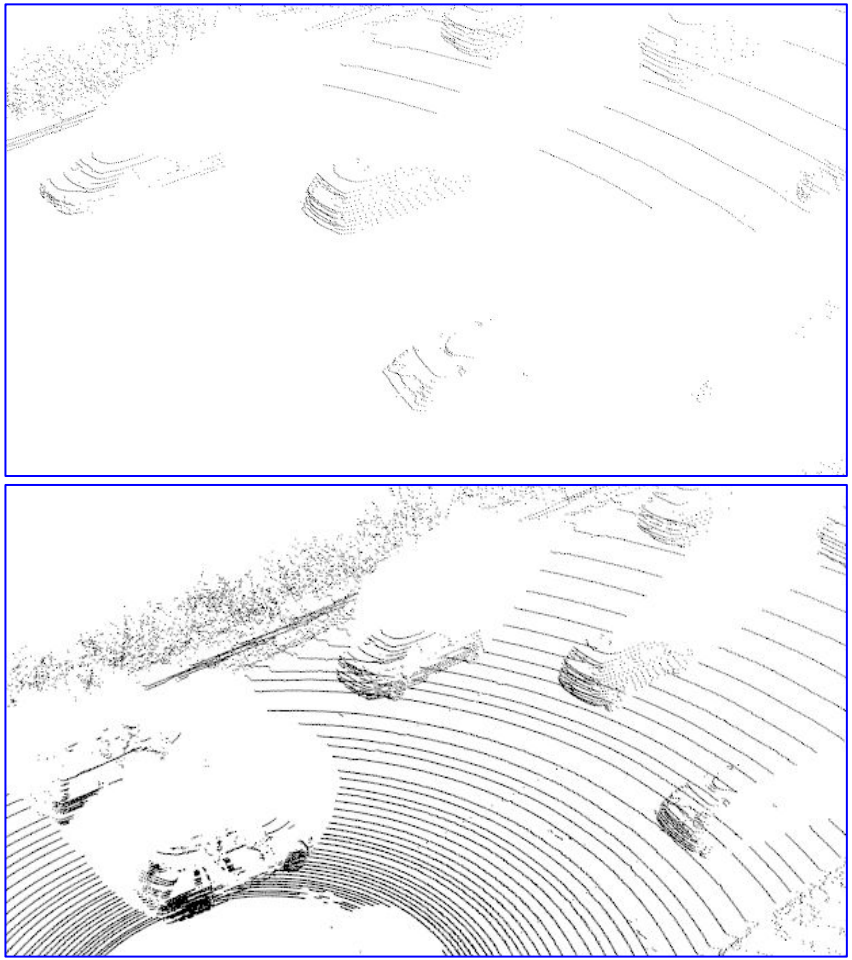


Limited Depth Information

Illumination Issues

Key Challenges:

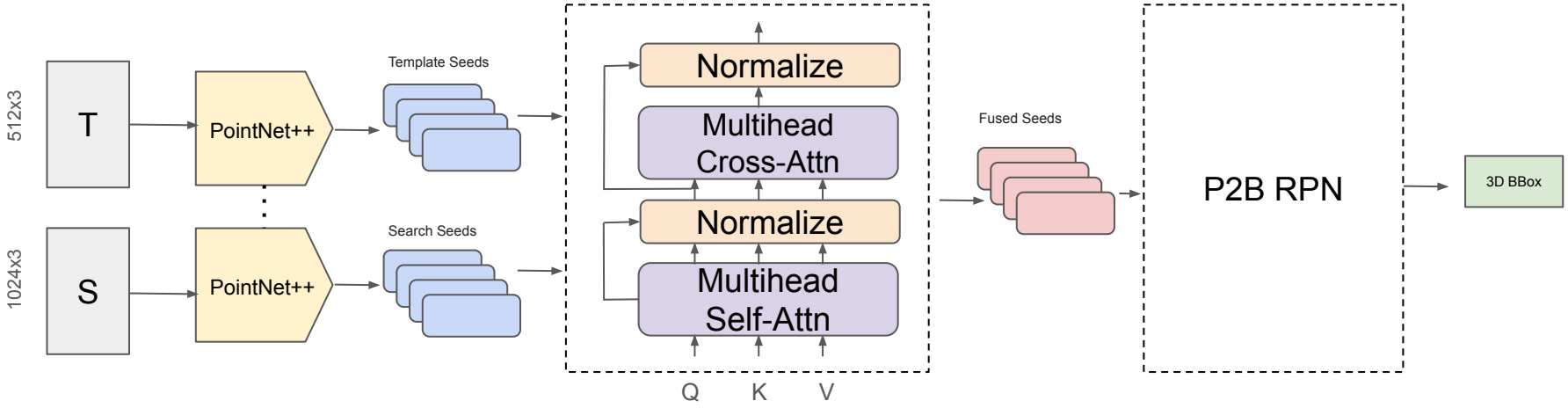
- Point Cloud sparsity
- Non-Rigid Objects
- Permutation Invariance
- Real-time tracking
- Occlusions in Point Clouds



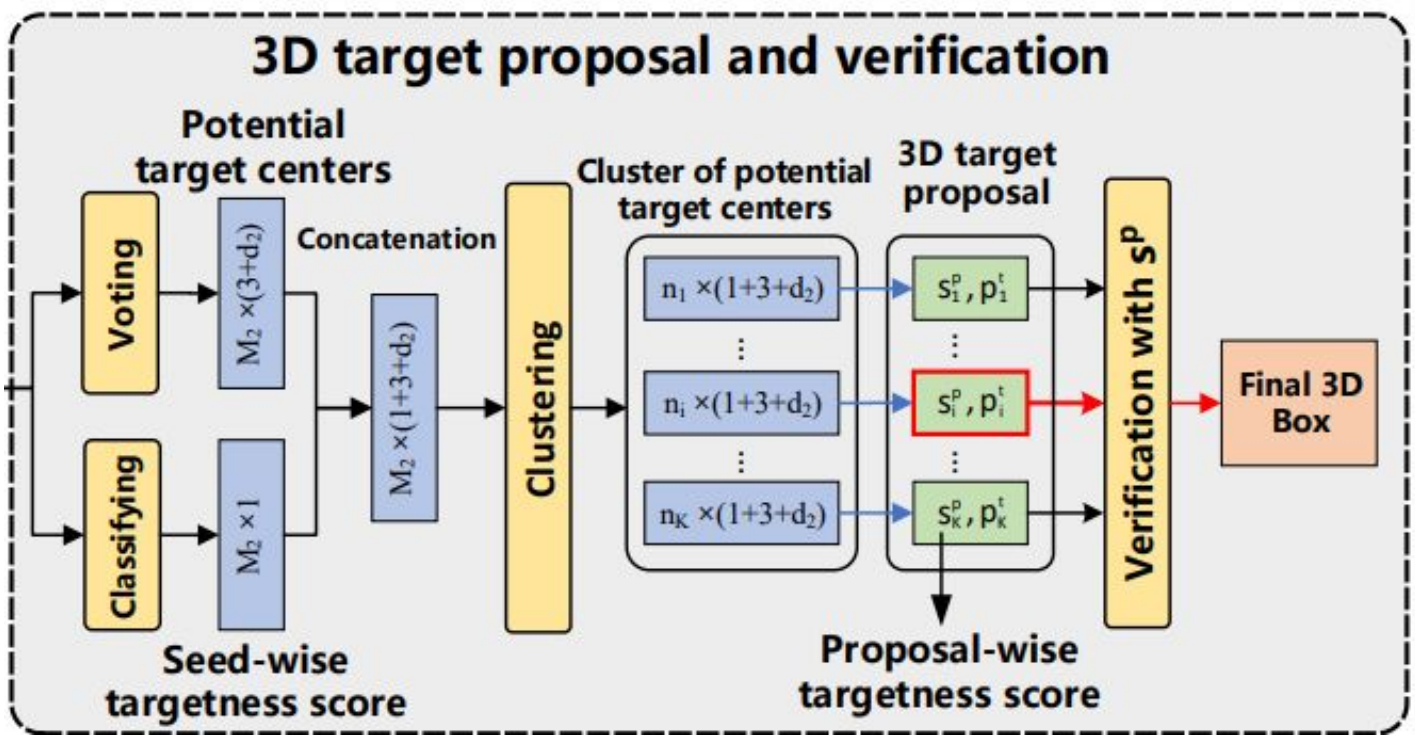
Related Work:

- SC3D
  - pioneer work using Siamese paradigm for object tracking for the first time.
  - Deterministic
- P2B
  - Point to Box relation within Siamese architecture.
  - End-to-end trainable
- OSP2B
  - Efficient network for parallelizing voting the regions and determining target-ness scores.

Pipeline


512x3
T
PointNet++
Template Seeds
1024x3
S
PointNet++
Search Seeds
Normalize
Multihead
Cross-Attn
Normalize
Multihead
Self-Attn
Q
K
V
Fused Seeds
P2B RPN
3D BBox

Pipeline - RPN (P2B)



Ref: P2B

Dataset

- KITTI Dataset
  - Velodyne - 21 Training Sequences
  - 0-16 used for training
  - 17-18 for validation
  - 19-20 for testing

https://www.cvlibs.net/datasets/kitti/index.php

Results

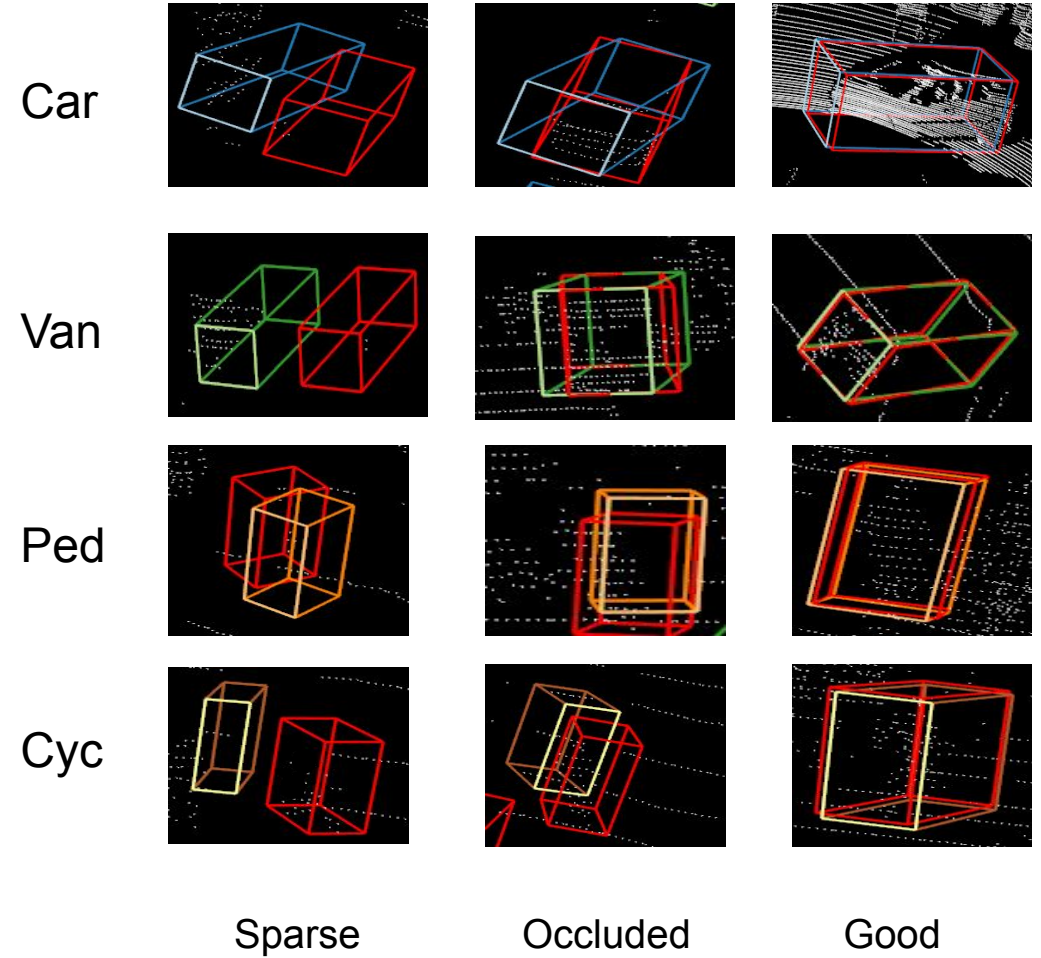
Model	Precision/Success				
	Car	Pedestrian	Van	Cyclist	Mean
P2B	72.8 / 56.2	49.6 / 28.7	48.4 / 40.8	44.7 / 32.1	60.0 / 42.4
OSP2B	82.3 / 67.5	85.1 / 53.6	66.2 / 56.3	90.5 / 65.6	82.3 / 60.5
LTTR	77.1 / 65.0	56.8 / 33.2	48.4 / 35.8	89.9 / 66.2	65.8 / 48.7
Ours	**72.0 / 56.8**	**63.9 / 35.7**	**52.8 / 45.6**	**46.5 / 34.1**	**66.0 / 45.5**

Success: Area Under the curve (AUC) determined by comparing the bounding box overlap threshold with the percentage of correctly tracked frames.

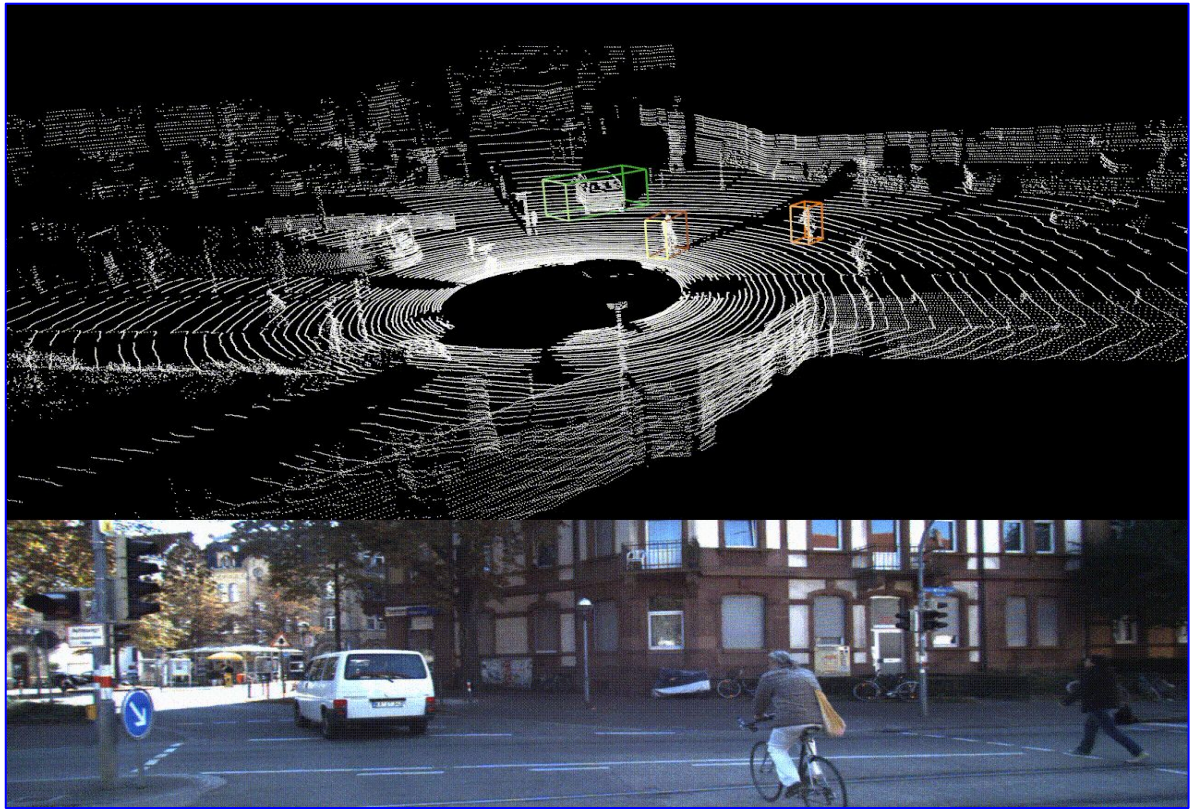
Precision: Percentage of frames where the tracked object's centers of the are within 2 meter from the groundtruth

Shortcomings

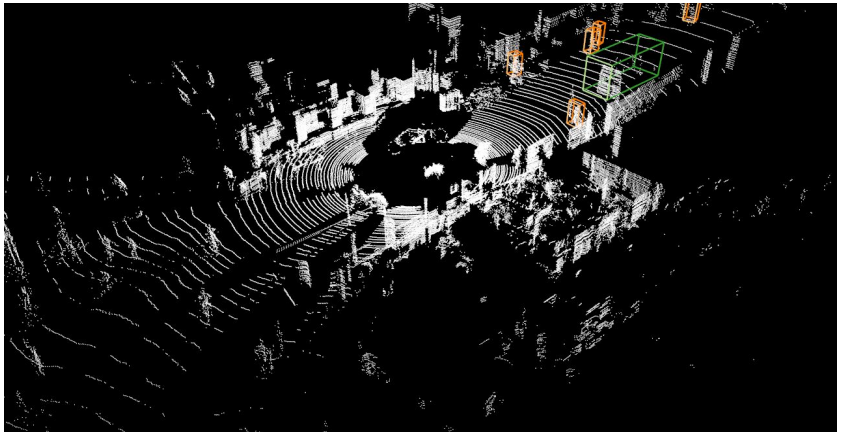
- Sparse Point Clouds
- Occluded Point Clouds

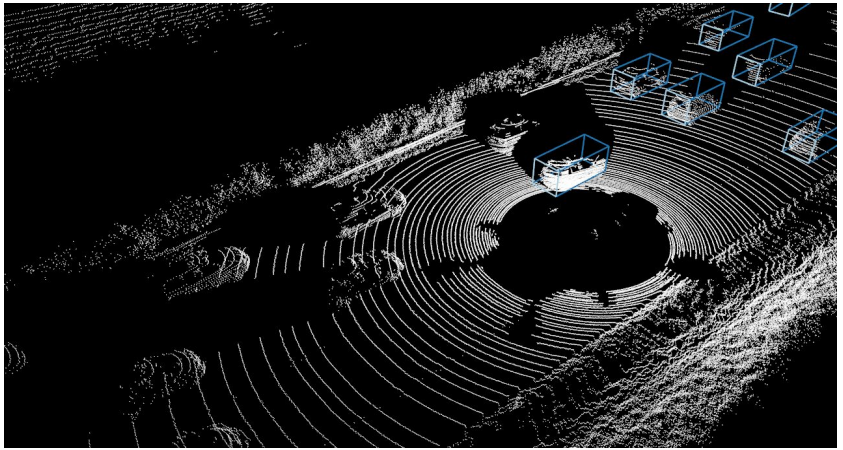


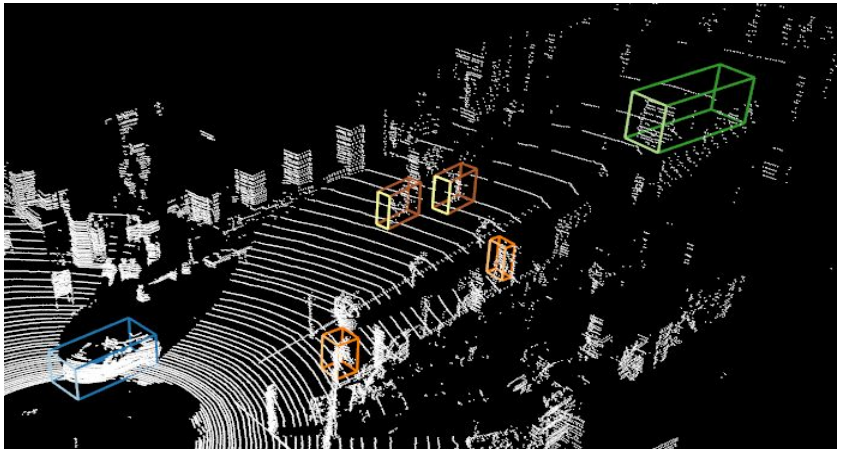
Sample Predictions

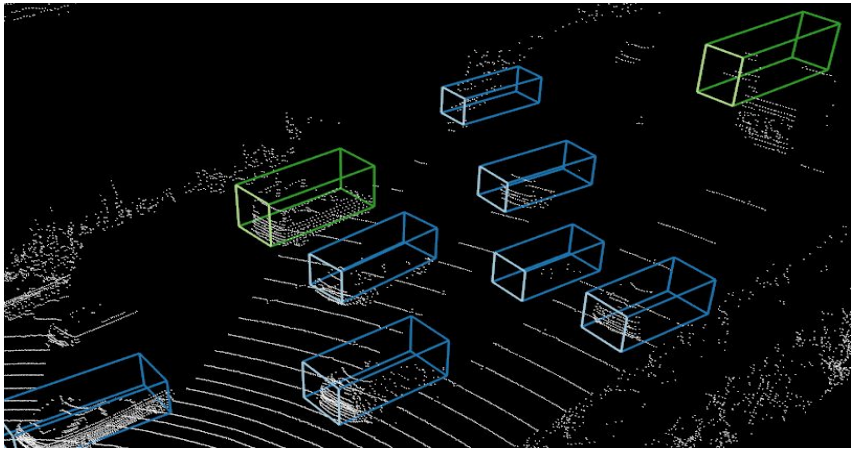


Sample Predictions









Future Improvements

- Point Cloud Completion for Occlusions and Sparse Point Clouds

- Using Image information (2D and 3D)