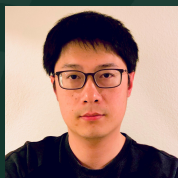


SeaBird: Segmentation in Bird's View Improves Mono3D of Large Objects (CVPR2024)



Abhinav Kumar¹



Yuliang Guo²



Xinyu Huang²



Liu Ren²



Xiaoming Liu¹



¹Michigan State University



BOSCH

²Bosch Research North America

Large Object Detection is Harder

A Tesla driver was killed after smashing into a **firetruck** on a California highway

February 20, 2023 - 4:42 PM ET

By Roshan Fernandez



Courtesy:

<https://www.npr.org/2023/02/20/1158367204/tesla-driver-killed-california-firetruck-nhtsa>

2 die when Tesla crashes into parked **tractor-trailer** in Florida



By Brittany Caldwell, WFTV.com

July 08, 2022 at 5:29 pm EDT

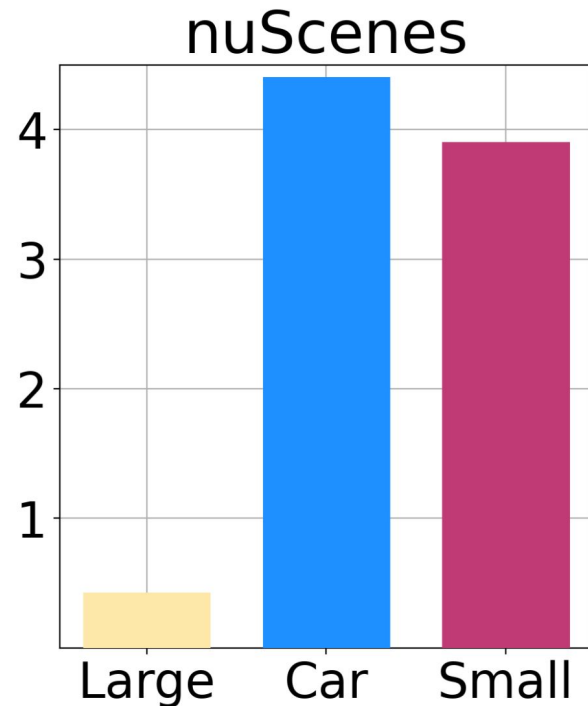
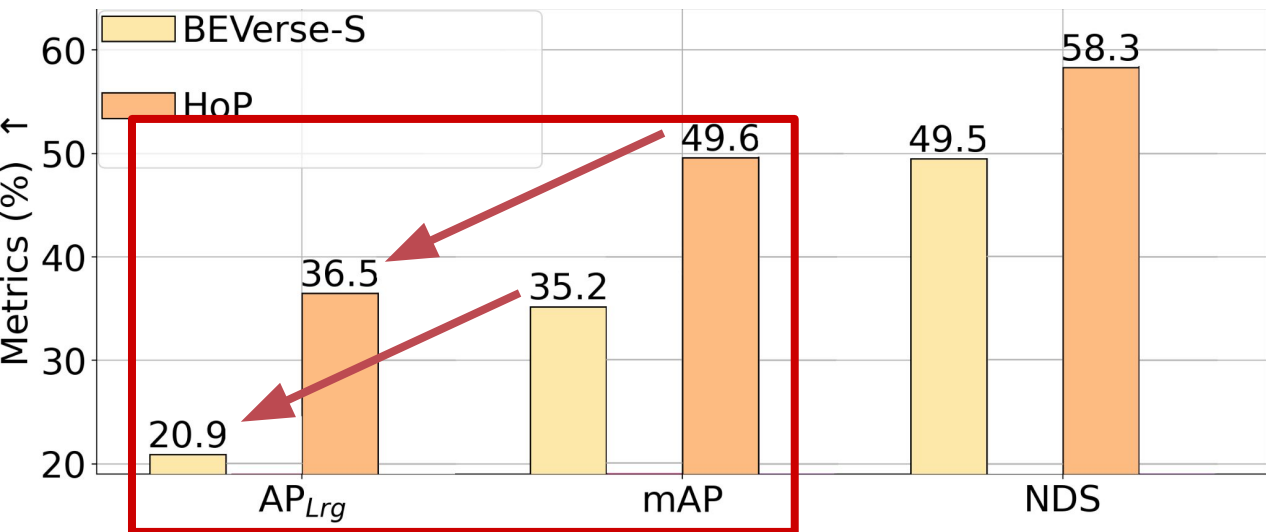


Courtesy:

<https://www.wftv.com/news/local/2-die-when-tesla-crashes-into-parked-tractor-trailer-florida/KJGMHHYTQZA2HNAHWL2OFSVIPM/>

Most accidents involve **large objects**.

Large Object Detection is Harder



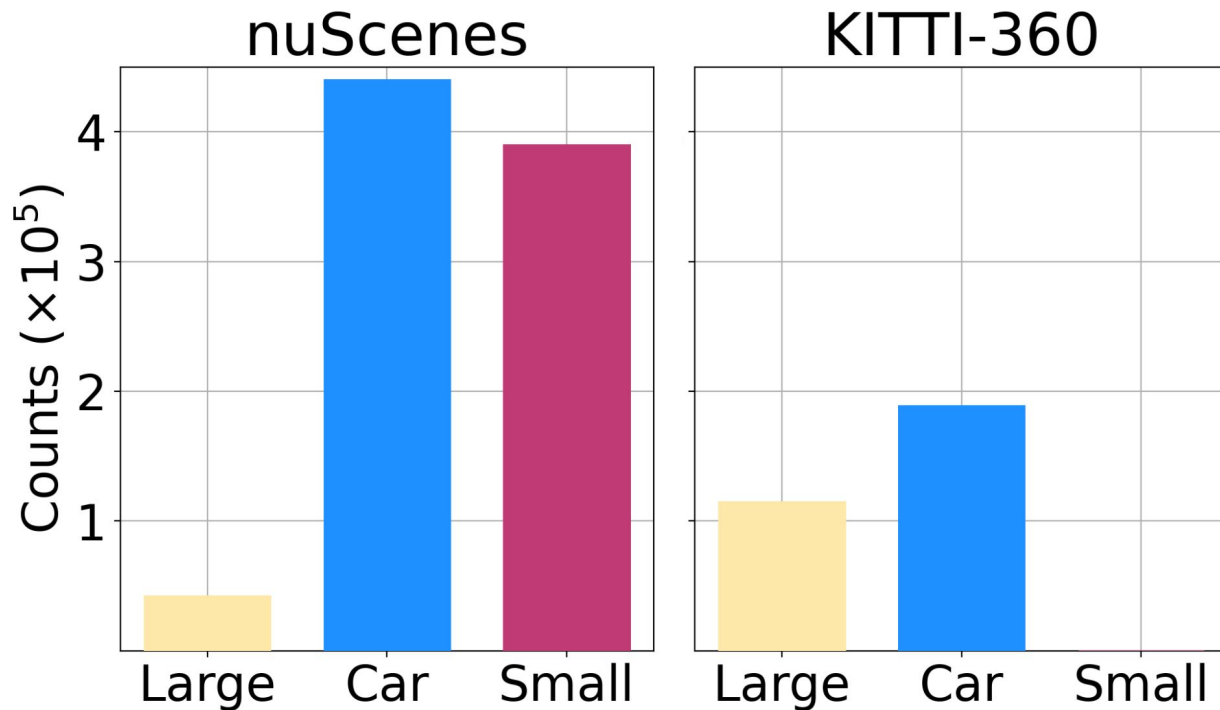
Possible Reasons:

- Training Data Scarcity [A]
- Larger Receptive Field Requirements [B]

[A] Zhu et al., CBGS for point cloud 3D detection, CVPR Workshop 2019

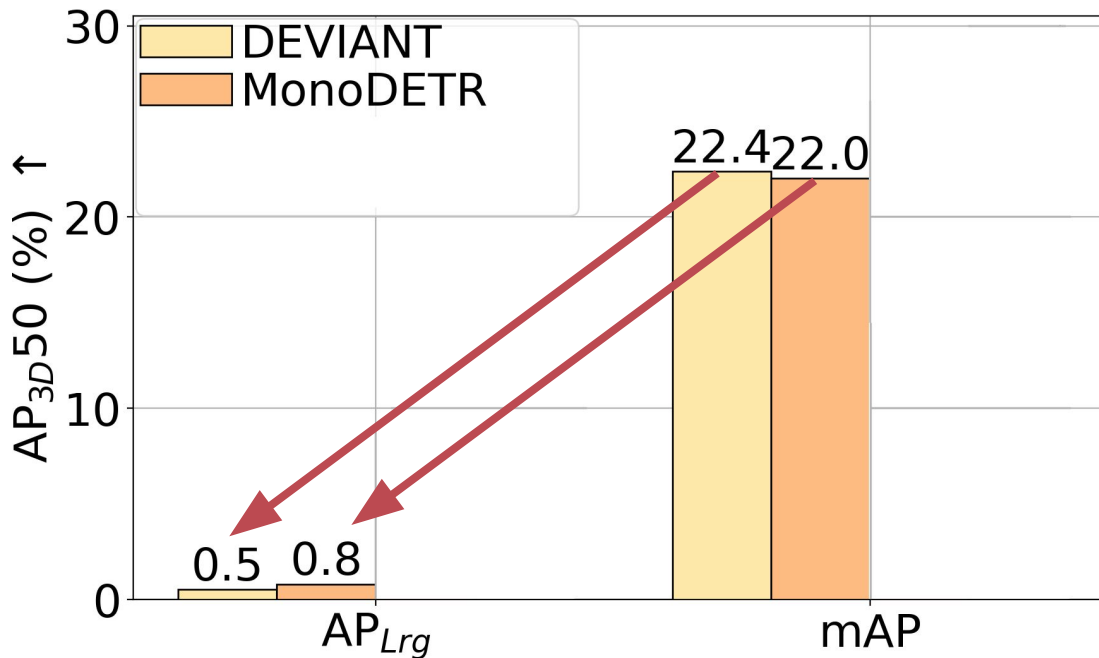
[B] Chen Wu, Waymo Keynote Talk, CVPR Workshop on Autonomous Driving 2023

Is Data Scarcity the Real Reason?



- KITTI-360 nearly balanced ratio of Large : Cars = 1:2

Is Data Scarcity the Real Reason?

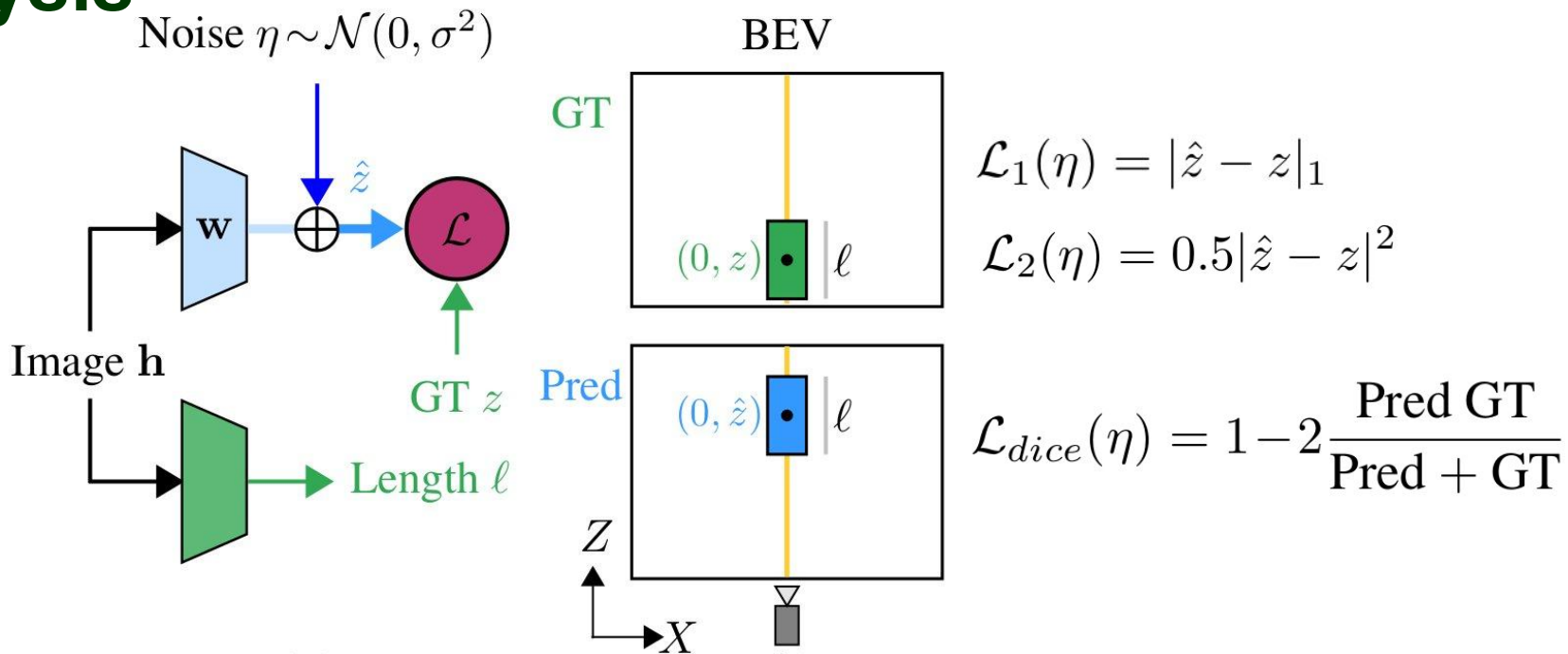


- SoTA Frontal KITTI detectors bad even on balanced KITTI-360 dataset.
- Data scarcity is NOT the only reason.

Zhang et al., MonoDETR: Depth guided transformer for Mono3D, ICCV 2023

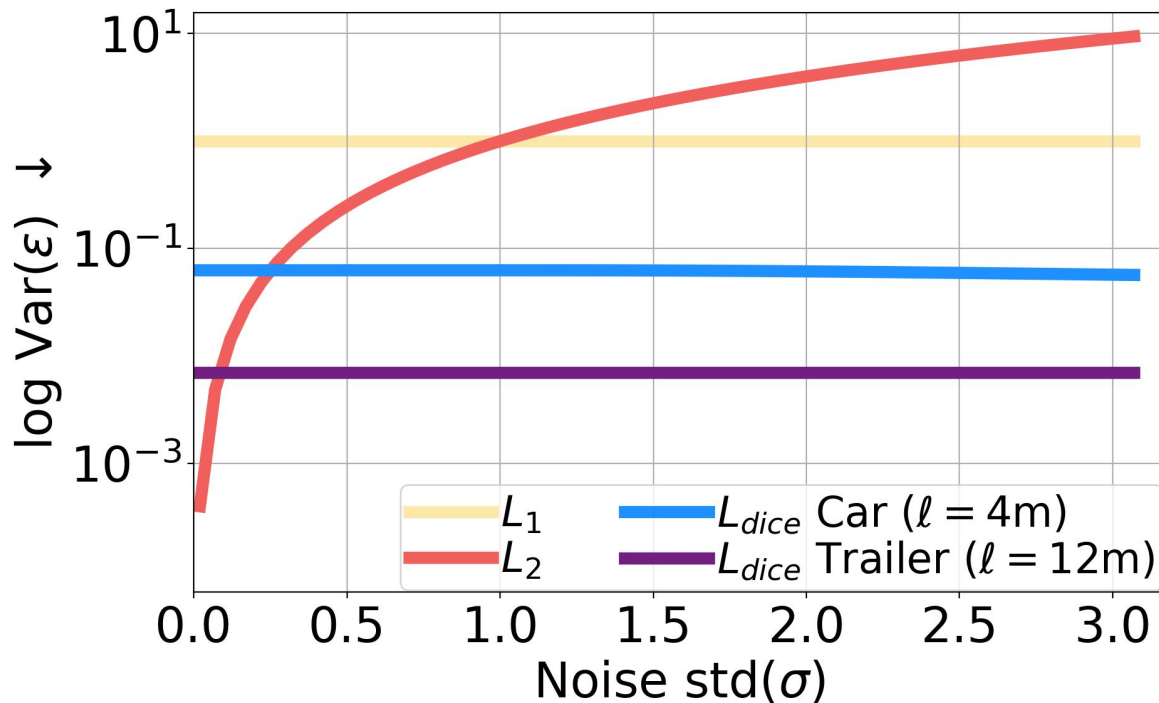
Kumar et al., DEVIANT: Depth Equivariant Network for Mono3D, ECCV 2022

Analysis



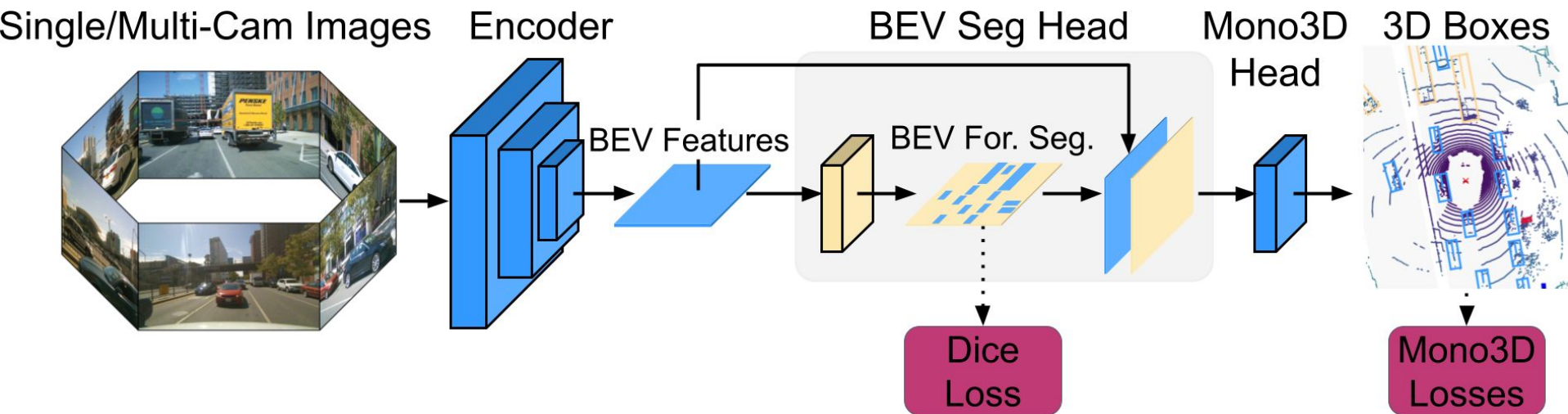
- Performance = function(Representation, **Loss**, **Noise**)
- Mono3D networks **sensitive to large noise** from larger objects.

Analysis



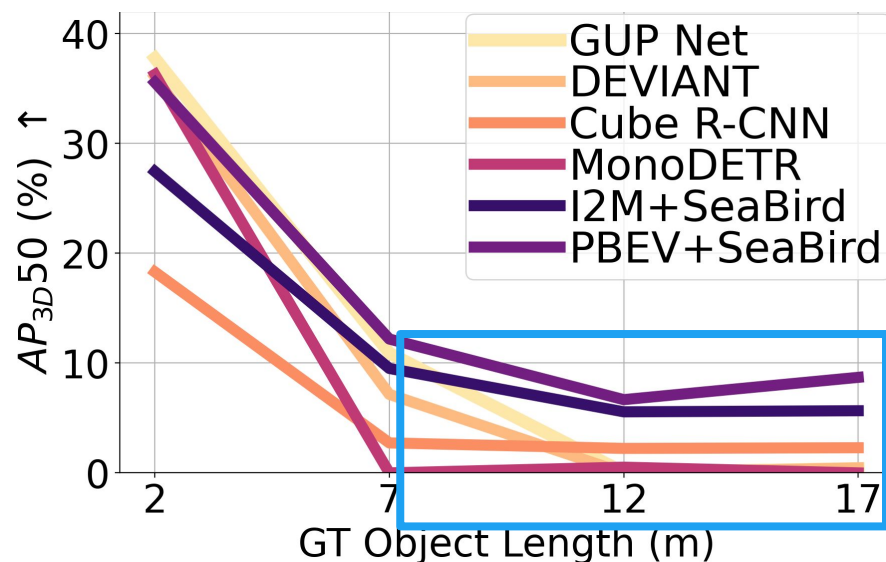
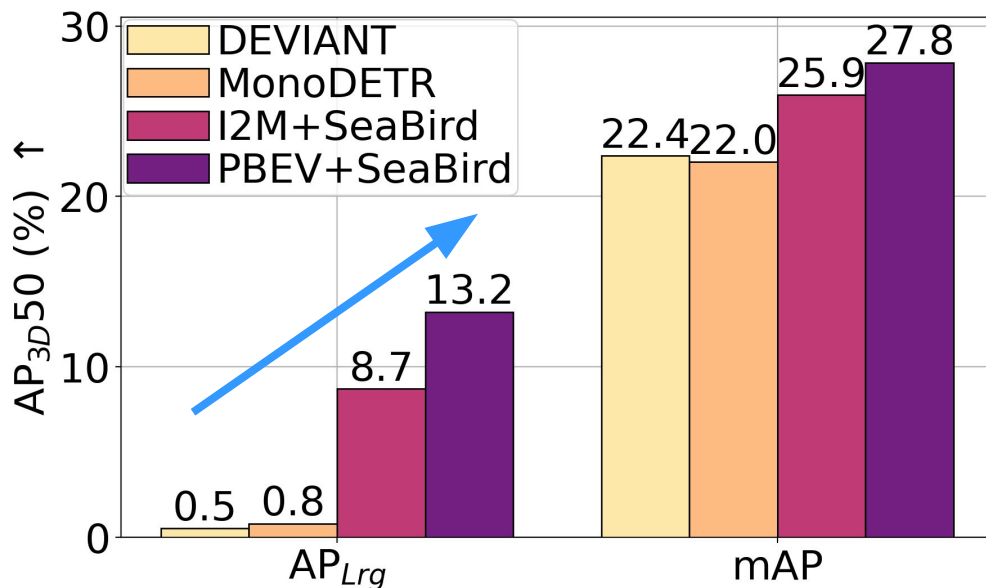
- Dice loss > Regression losses under large noise (large objects).

SeaBird Pipeline



- Train BEV Segmentation first with Dice Loss
- Finetune BEV Segmentation + Mono3D

KITTI-360 (Single-Camera) Results

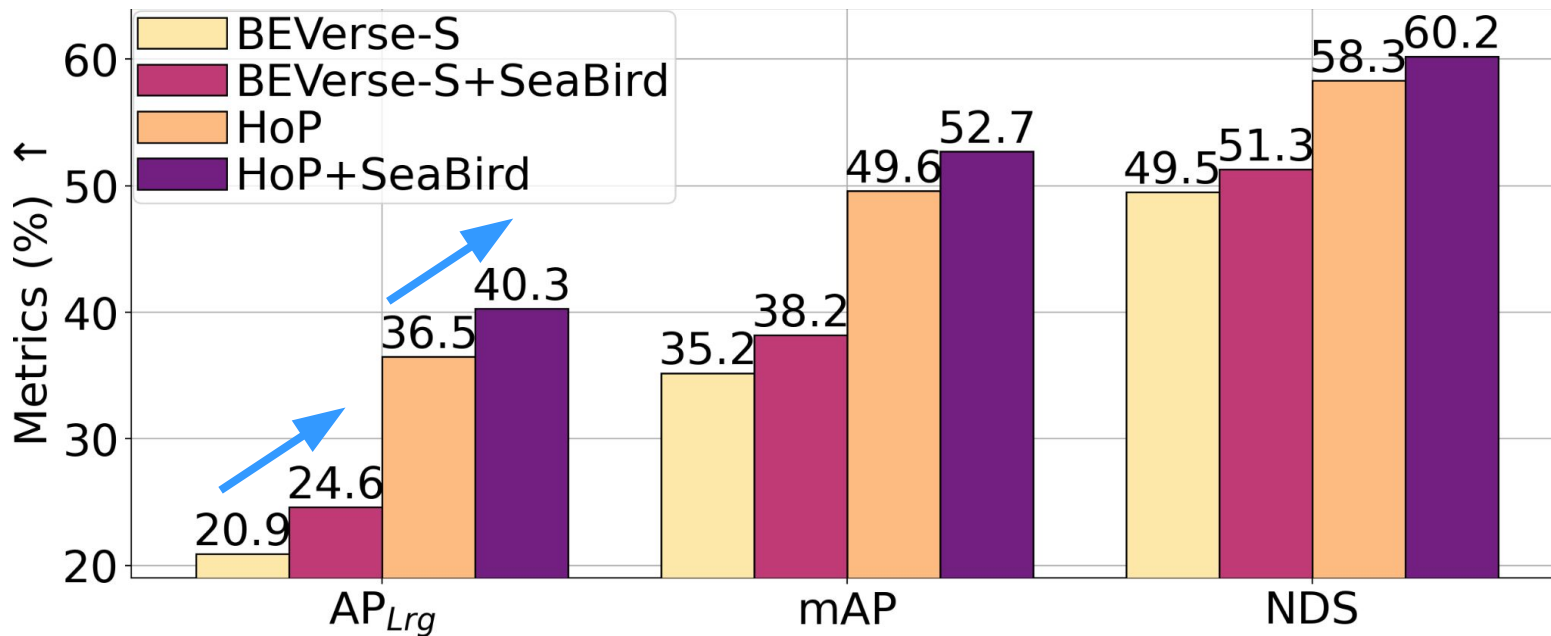


- SoTA on KITTI-360
- Outperforms frontal detectors and also old LiDAR detectors

Zhang et al., MonoDETR: Depth guided transformer for Mono3D, ICCV 2023

Kumar et al., DEVIANT: Depth Equivariant Network for Mono3D, ECCV 2022

nuScenes (Multi-Camera) Results

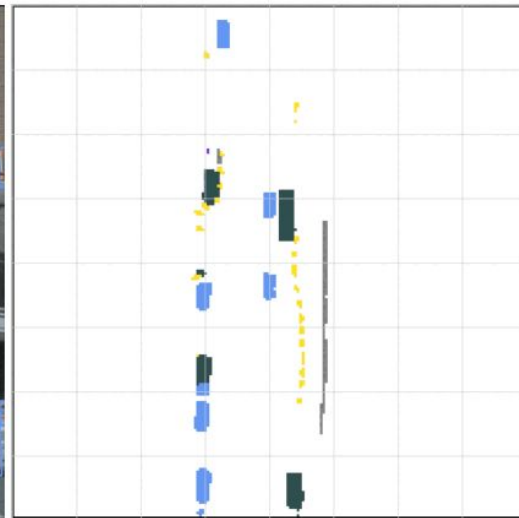
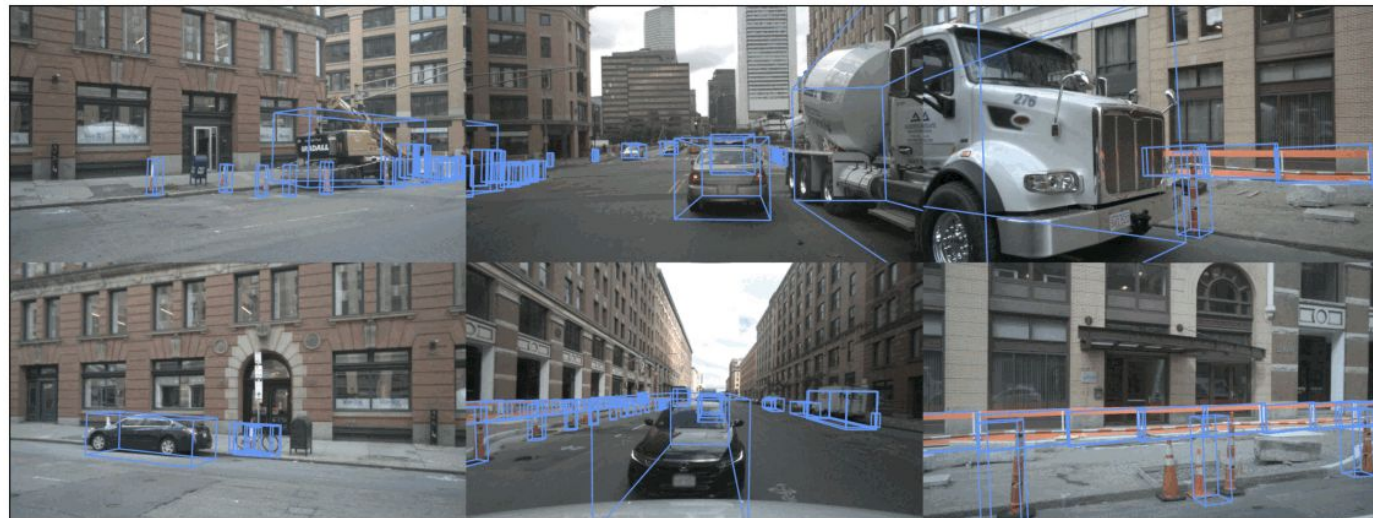


- SoTA on nuScenes
- Outperforms BEV-based detectors

Zong et al., Temporal enhanced training of multi-view 3D object detector via historical object prediction, ICCV 2023

Zhang et al., BEVerse: Unified perception and prediction in birds-eye-view for vision-centric autonomous driving, arXiv 2022

SeaBird Demo





Conclusions

- Large Object Detection = Representation (Front / BEV) + Loss
- Frontal detectors even with transformers do not work
- BEV detectors sub-optimal, improved by noise-robust Dice loss
- Dice loss > Regression losses under large noise (large objects)






Summary



 **SeaBird** Public 

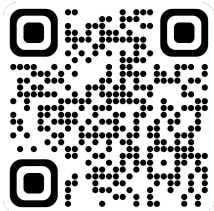
[CVPR 2024] Official PyTorch Code of SeaBird: Segmentation in Bird's View with Dice Loss Improves Monocular 3D Detection of Large Objects

 Python  50  4

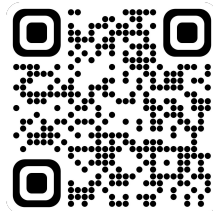
Support



Project Website



Demo



Code

