MoDAR: Using Motion Forecasting for 3D Object Detection in Point Cloud Sequences

Tag: WED-AM-105


Yingwei Li*


Charles Qi*

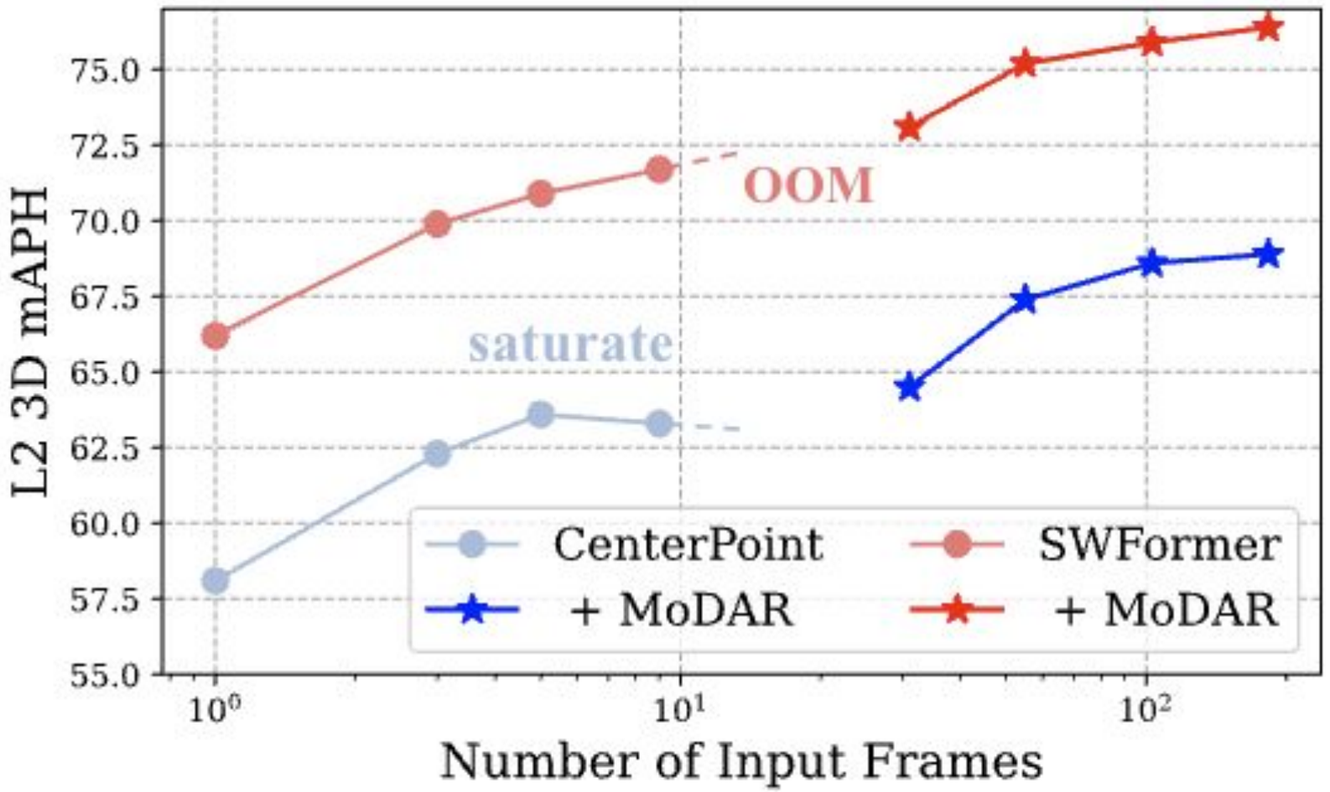

Yin Zhou


Chenxi Liu


Drago Anguelov

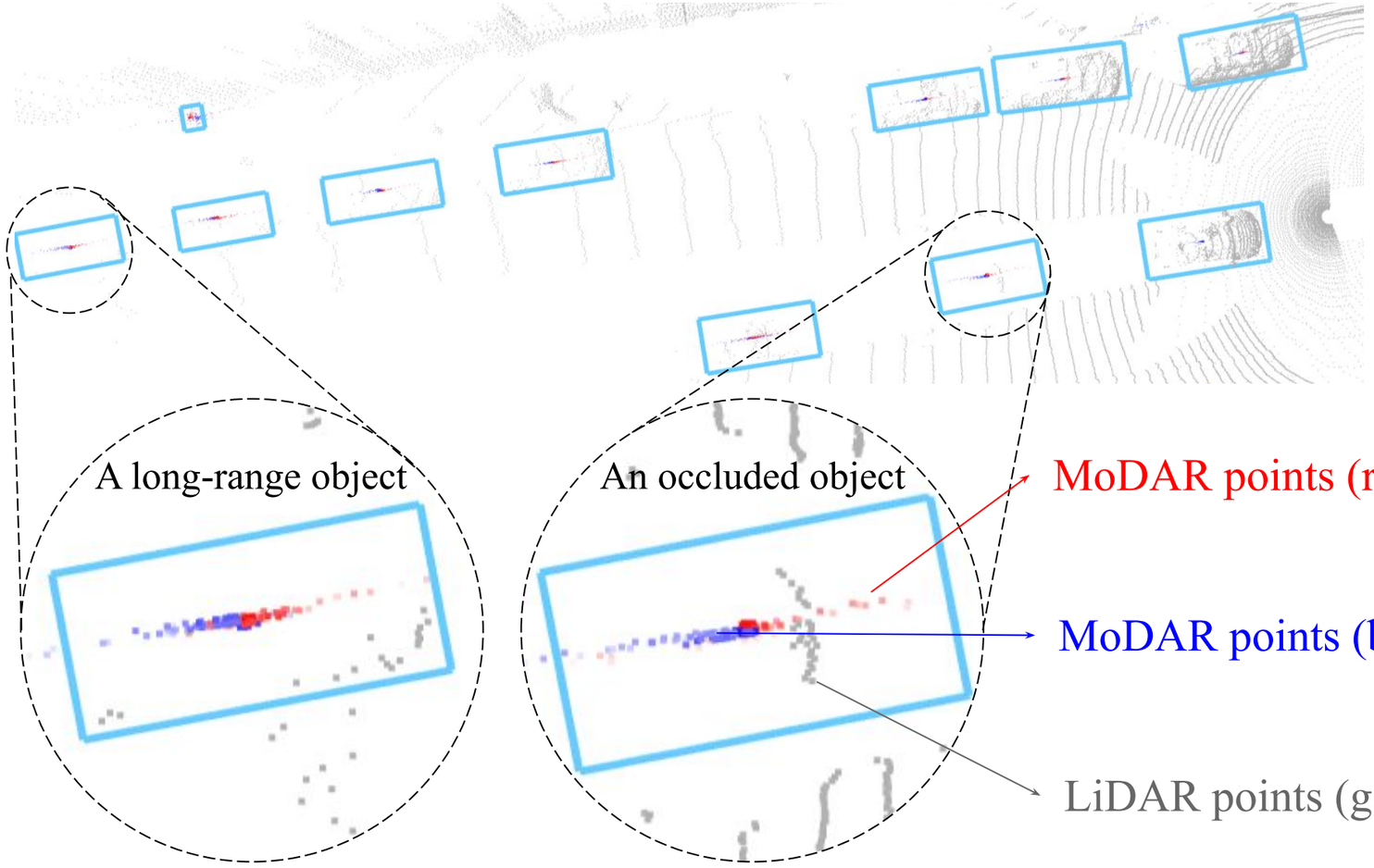
WAYMO

Scaling up Detection with Sequence Input



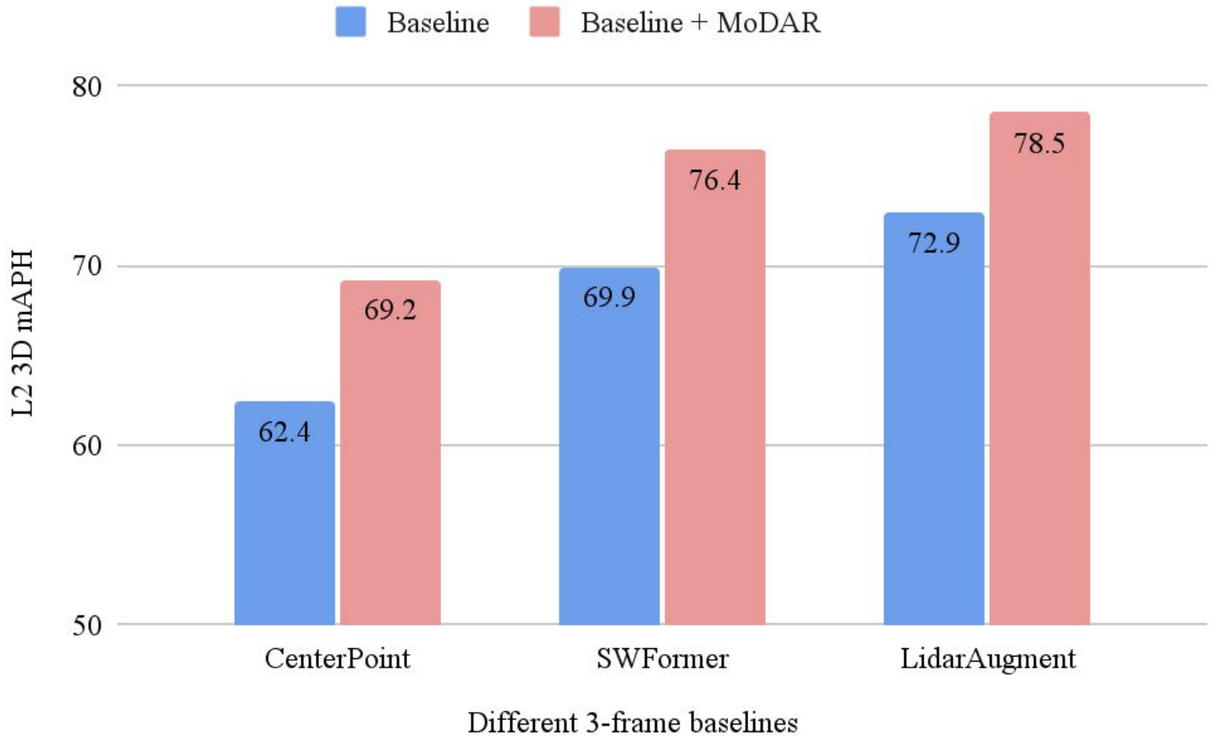
Detection from MoDAR and LiDAR Points

MoDAR models long-term sequence information with behavior prediction (BP) model.



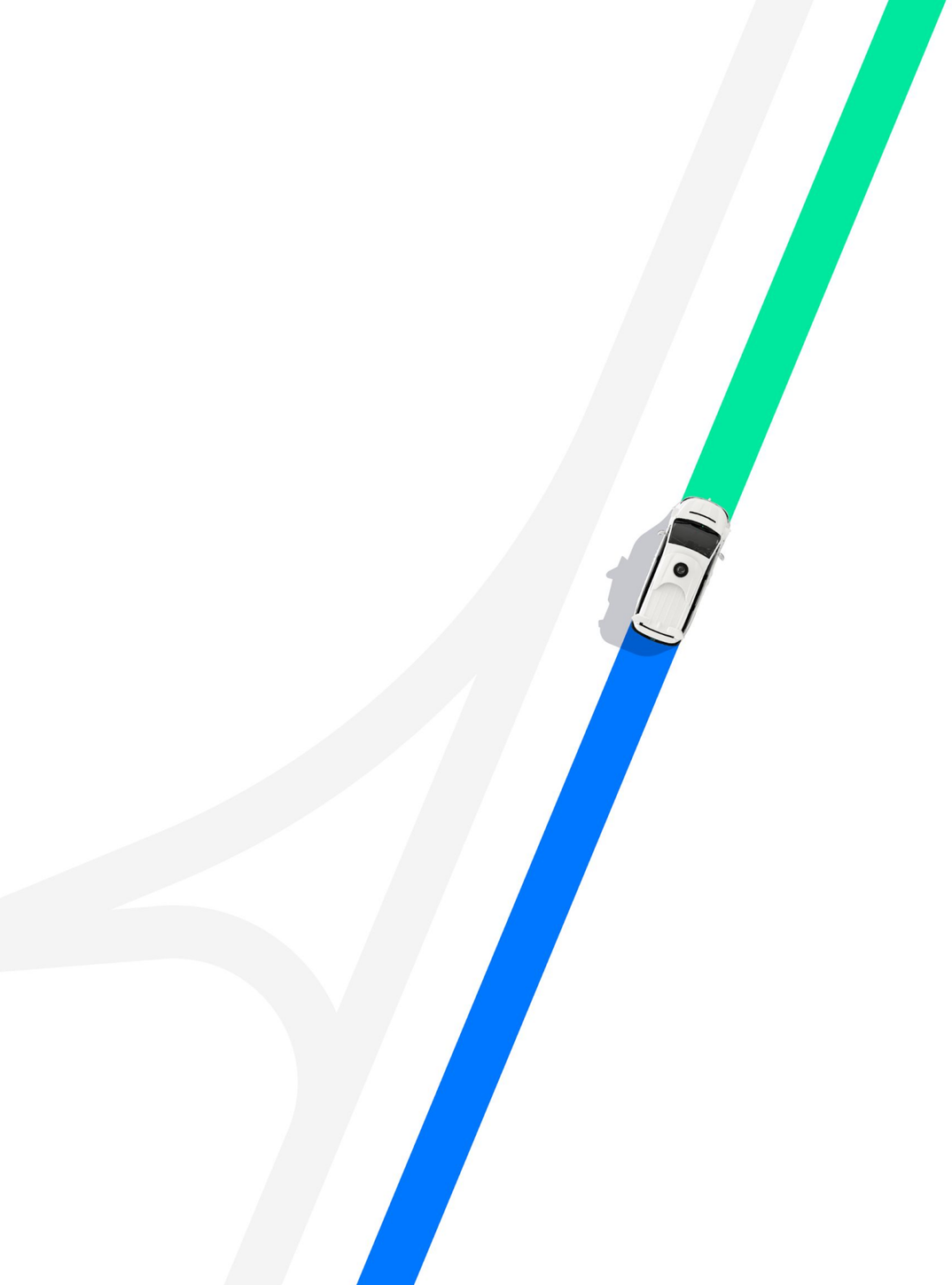
Improving on Various Detectors

MoDAR achieves significant improvements on top of popular and state-of-the-art 3D point-cloud based detectors.



[CenterPoint] Yin, Tianwei, Xingyi Zhou, and Philipp Krahenbuhl. "Center-based 3d object detection and tracking." CVPR. 2021.
[SWFormer] Sun, Pei, et al. "Swformer: Sparse window transformer for 3d object detection in point clouds." ECCV. 2022.
[LidarAugment] Leng, Zhaoqi, et al. "LidarAugment: Searching for Scalable 3D LiDAR Data Augmentations." ICRA. 2023.

WAYMO



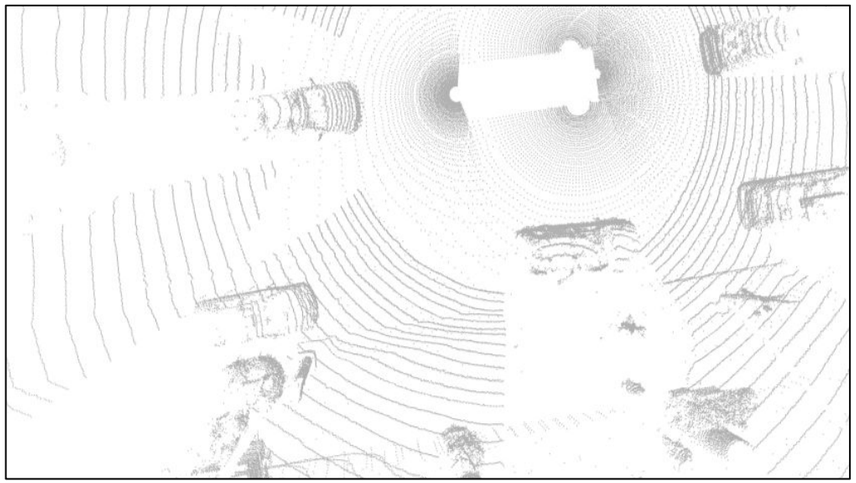
Agenda

- Related Works
- Method
- Experimental Results
- Conclusion and Future Work

WAYMO

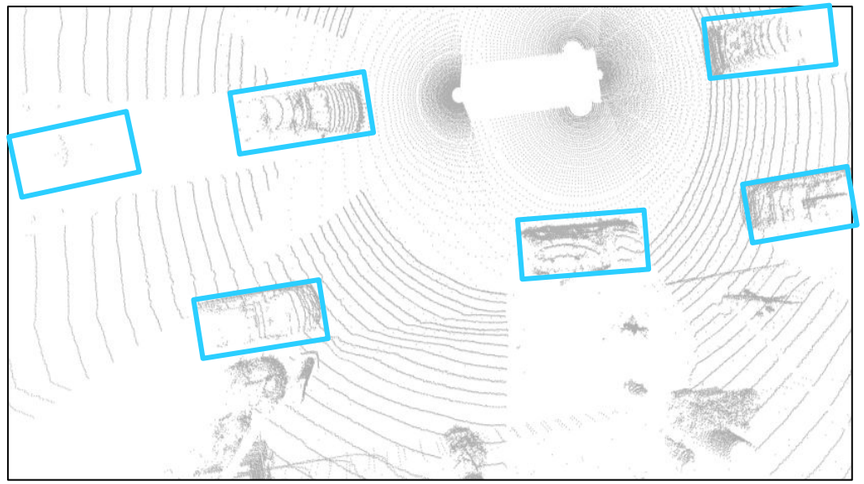
Related Works

3D Object Detection on Point Clouds



Point Clouds

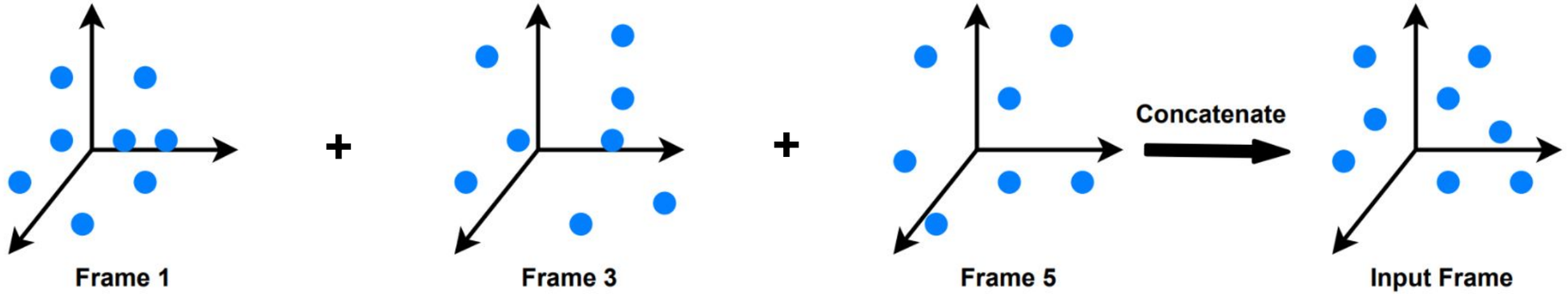
Detection



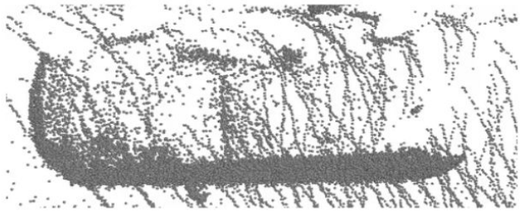
Point Clouds
with 3D Bounding Boxes

Multi-frame 3D object detection

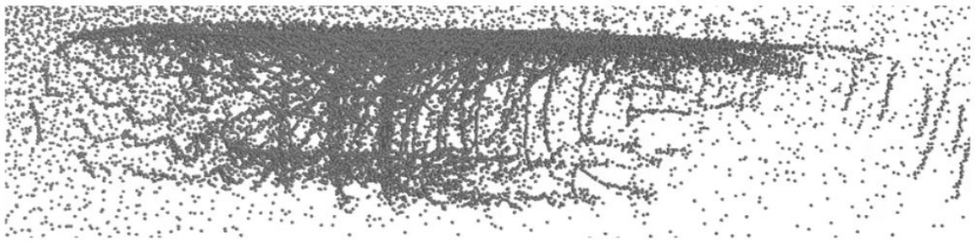
How to leverage long-term sequence information for 3D object detection?
Previous method – Point Concatenation [1, 2]



✅ Stationary Object



❌ Moving Object

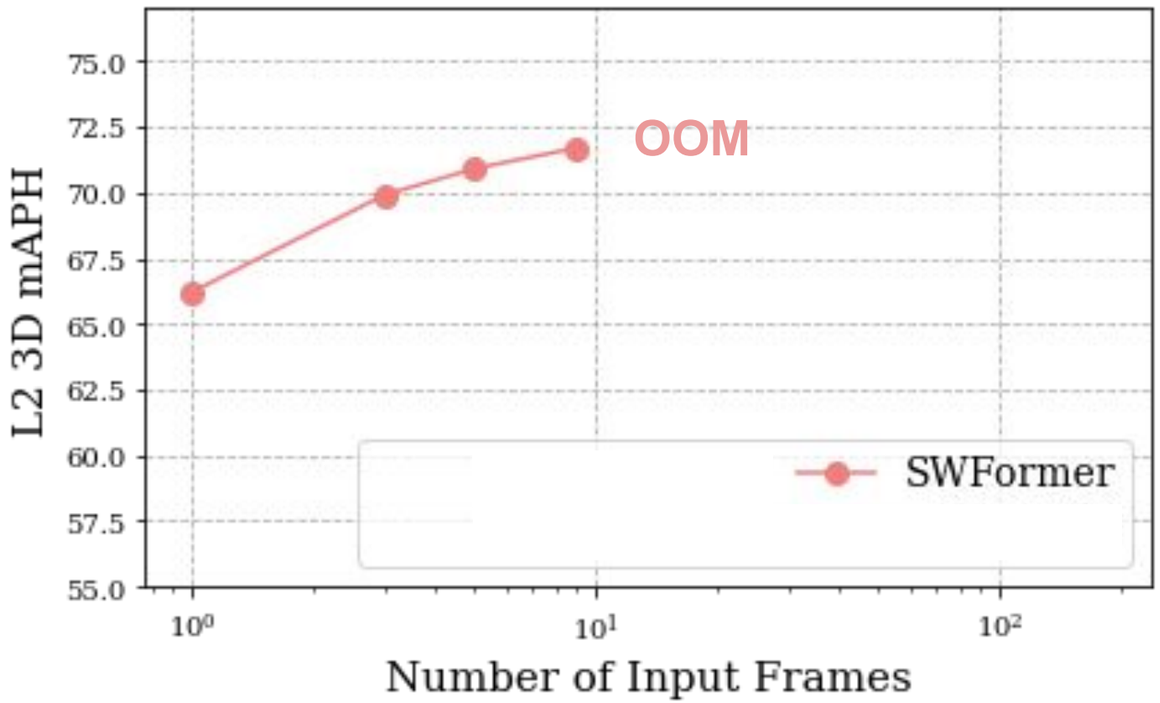


[1] Li, Yingwei, et al. "Deepfusion: Lidar-camera deep fusion for multi-modal 3d object detection." CVPR. 2022.
[2] Chen, Xuesong, et al. "MPPNet: Multi-Frame Feature Intertwining with Proxy Points for 3D Temporal Object Detection." ECCV. 2022.

WAYMO

Point Concatenation

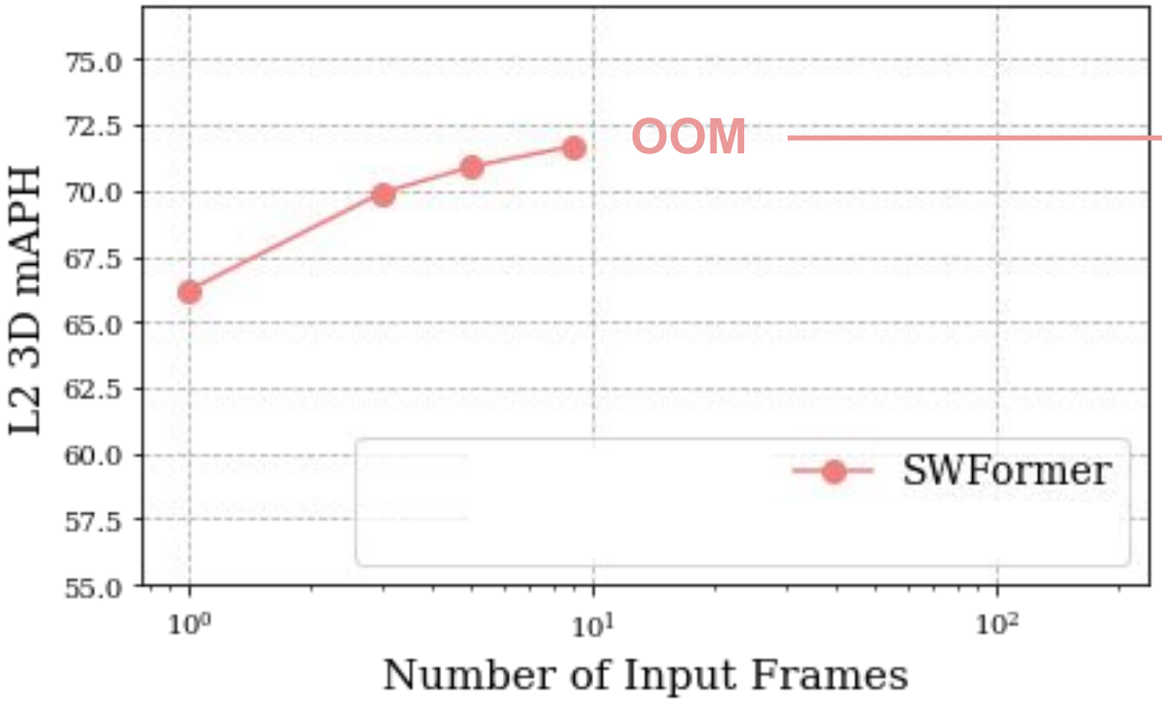
How to leverage long-term sequence information for 3D object detection?



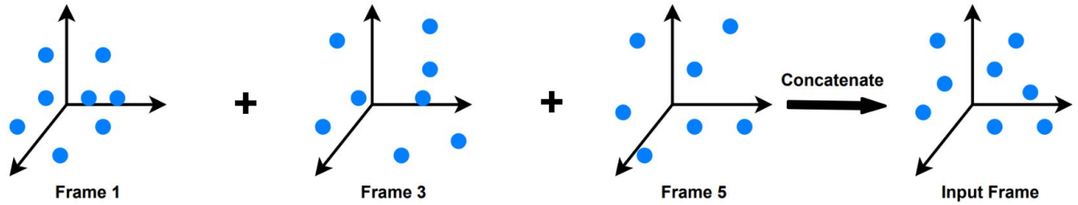
WAYMO

Point Concatenation

How to leverage long-term sequence information for 3D object detection?



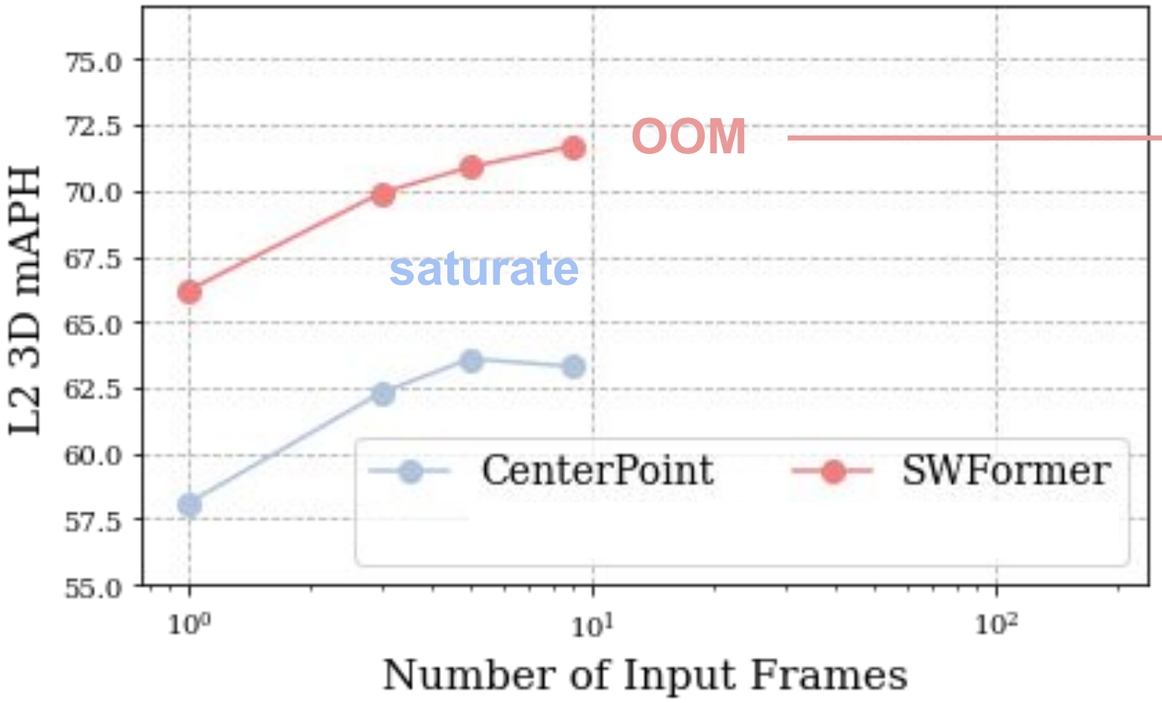
OOM → OOM is caused by concatenating too many points.



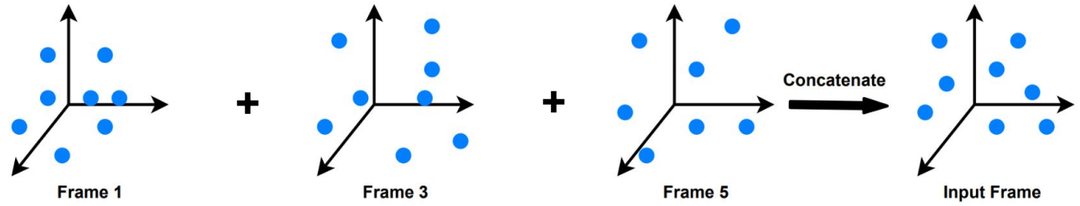
WAYMO

Point Concatenation

How to leverage long-term sequence information for 3D object detection?



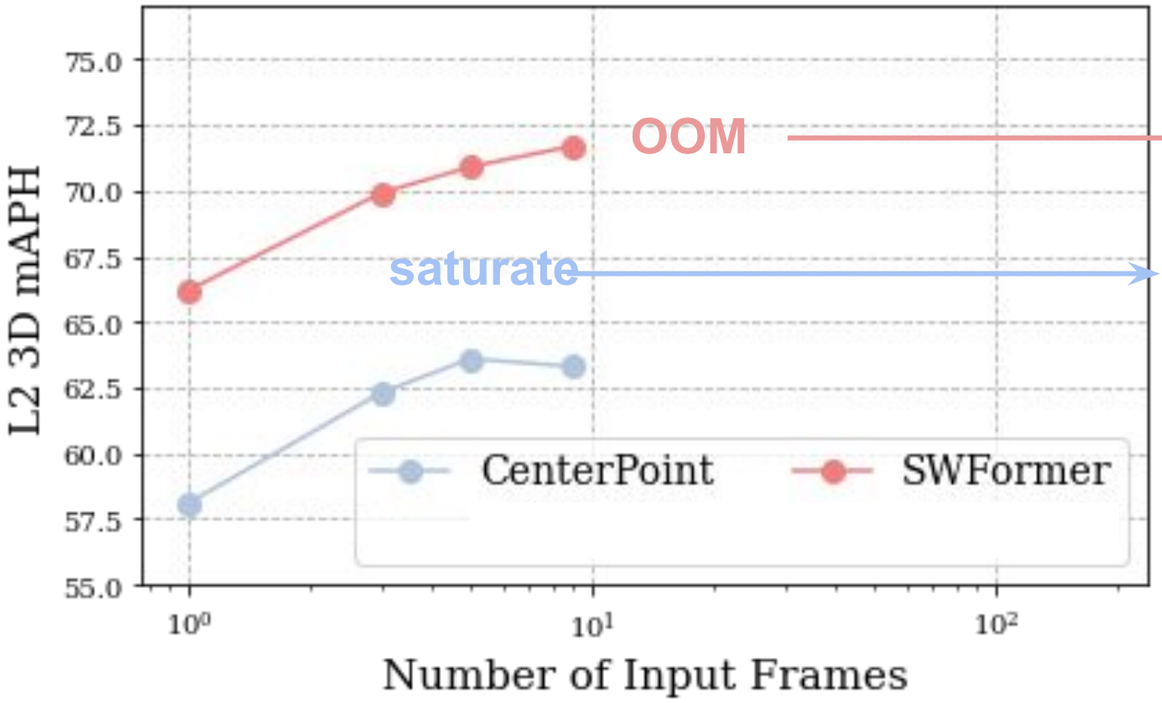
OOM → OOM is caused by concatenating too many points.



WAYMO

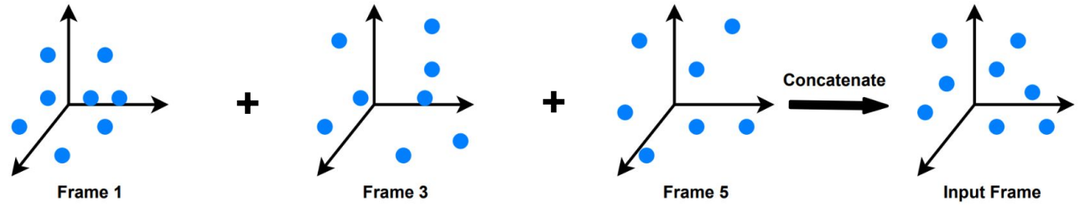
Point Concatenation

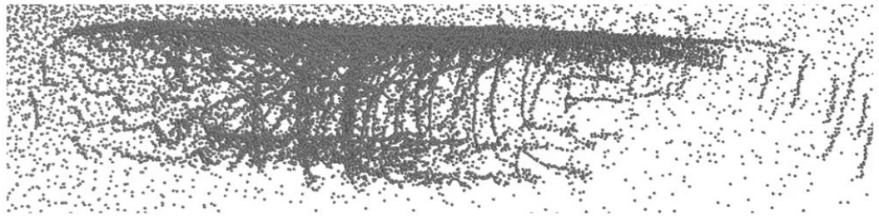
How to leverage long-term sequence information for 3D object detection?



OOM → OOM is caused by concatenating too many points.

saturate → Saturation is due to misalignment of moving object[1].





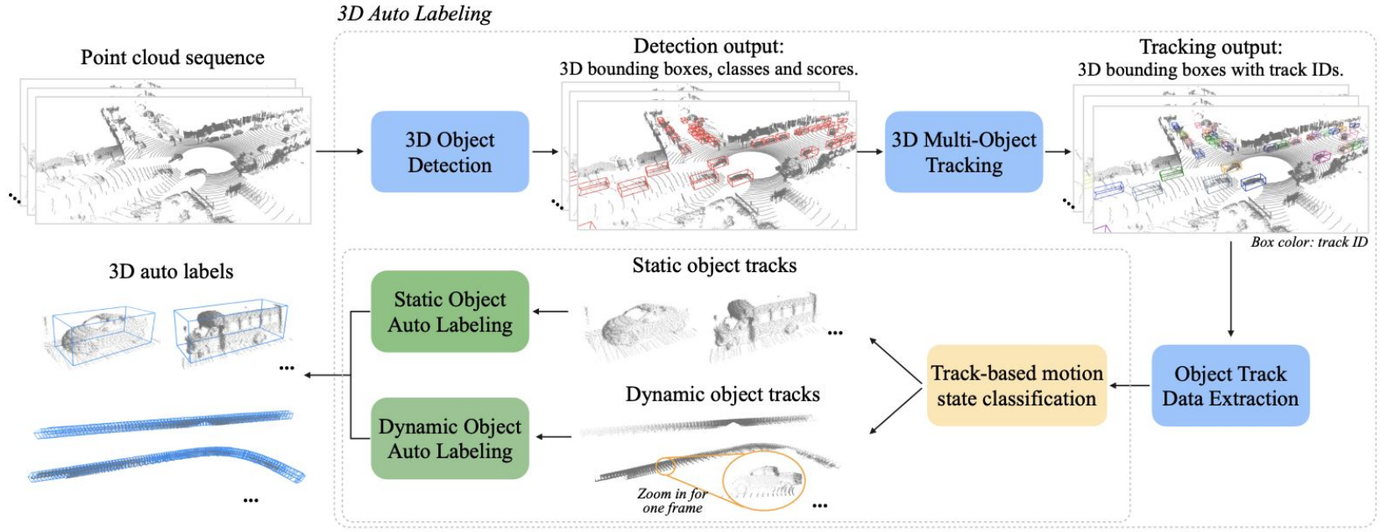
❌ Moving Object

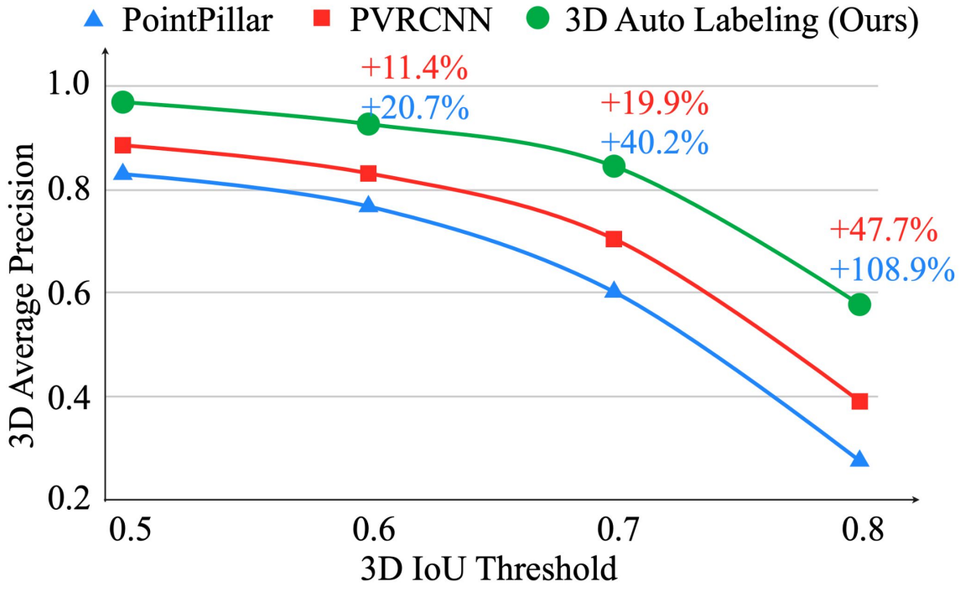
[1] Yang, Zetong, et al. "3d-man: 3d multi-frame attention network for object detection." CVPR. 2021.

WAYMO

Offboard Detection

Offboard detection can have access to more temporal information from the future and is less constraint in computation budget or problem formulation compared to onboard use cases.



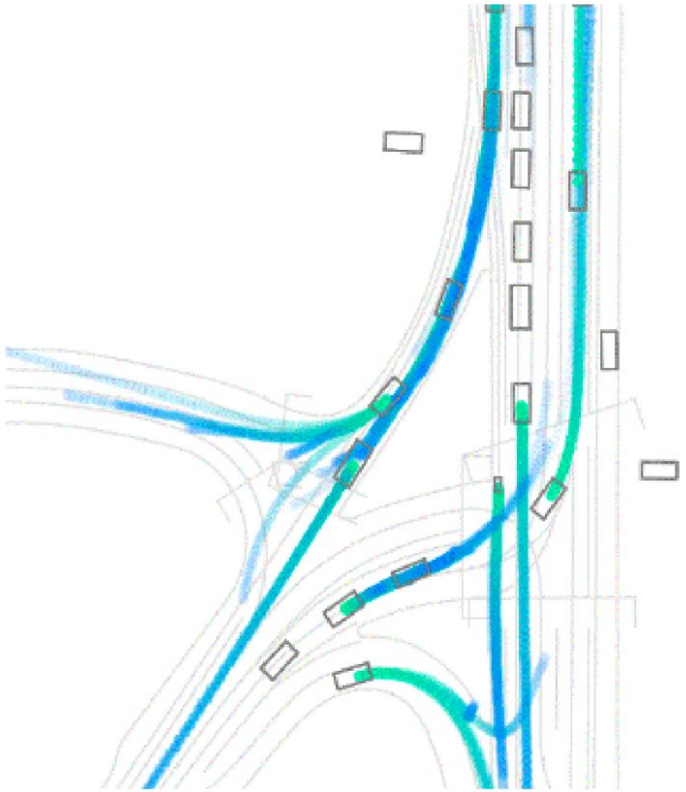


Qi, Charles R., et al. "Offboard 3d object detection from point cloud sequences." CVPR. 2021.

WAYMO

Motion Forecasting

Predicting the future behavior of road user. Also known as behavior prediction or trajectory prediction.



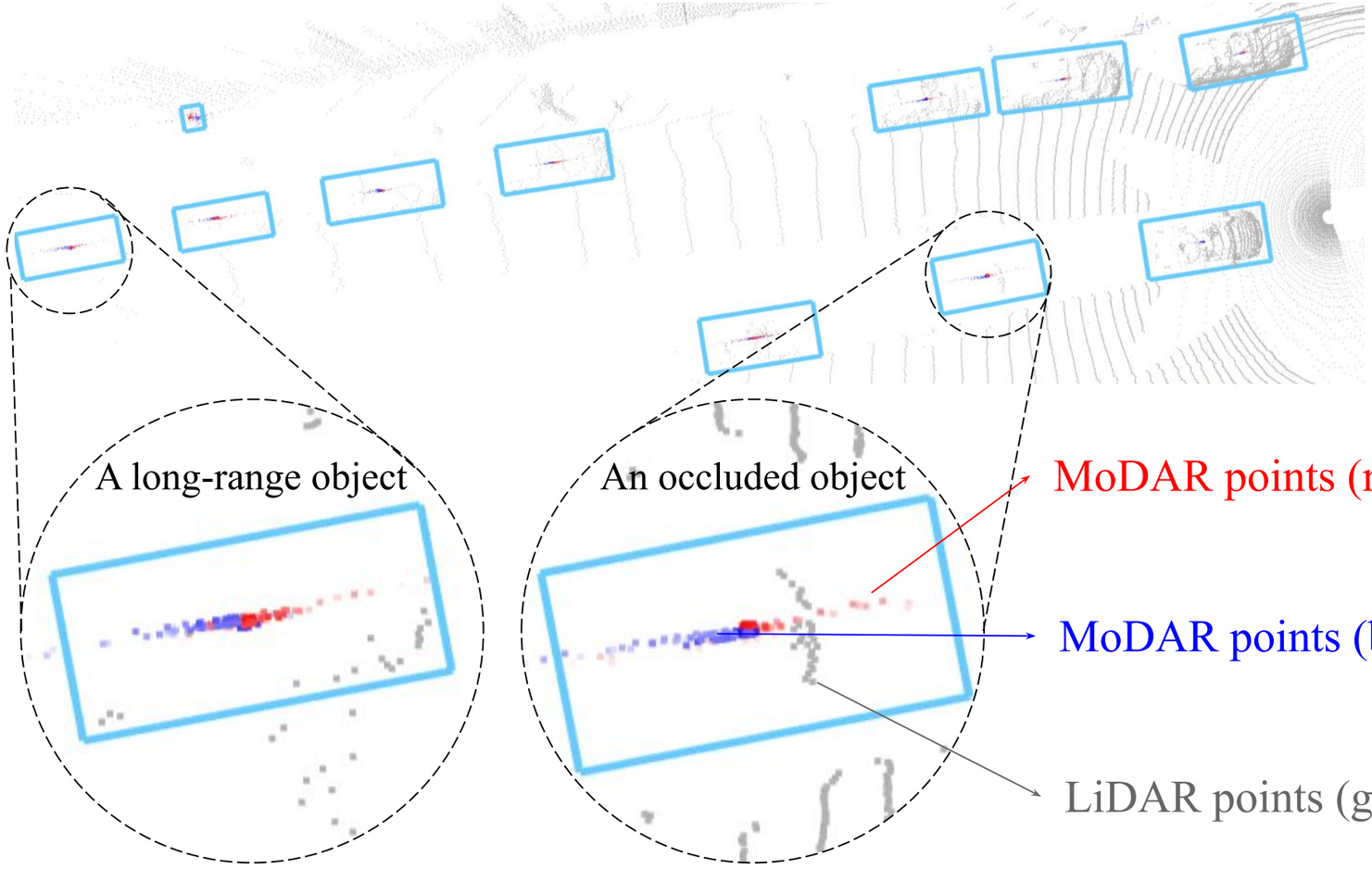
WAYMO

Varadarajan, Balakrishnan, et al. "Multipath++: Efficient information fusion and trajectory aggregation for behavior prediction." ICRA. 2022.

Using MoDAR for 3D object detection

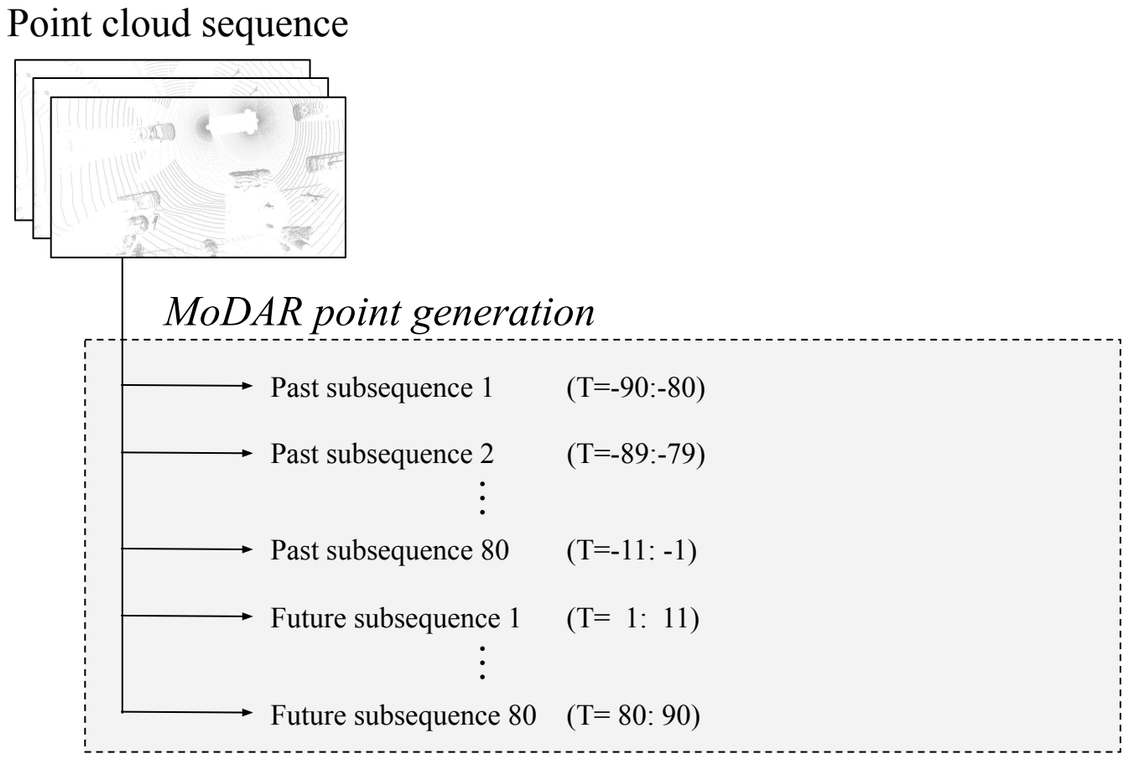
Using MoDAR for 3D object detection

MoDAR models long-term sequence information with behavior prediction (BP) model.

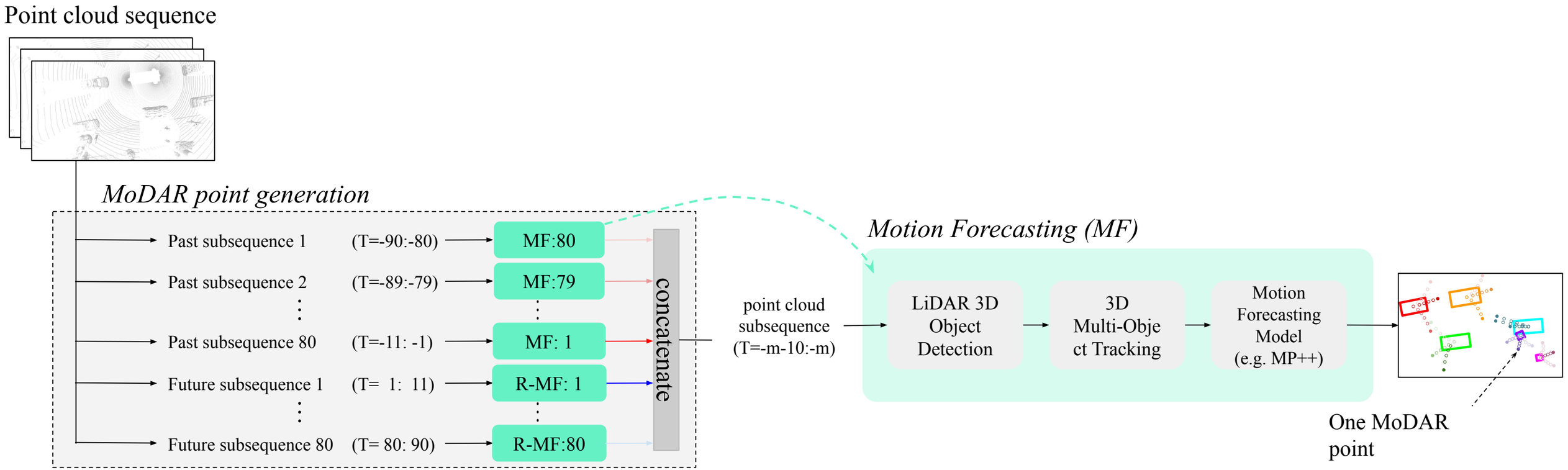


WAYMO

Using MoDAR for 3D object detection

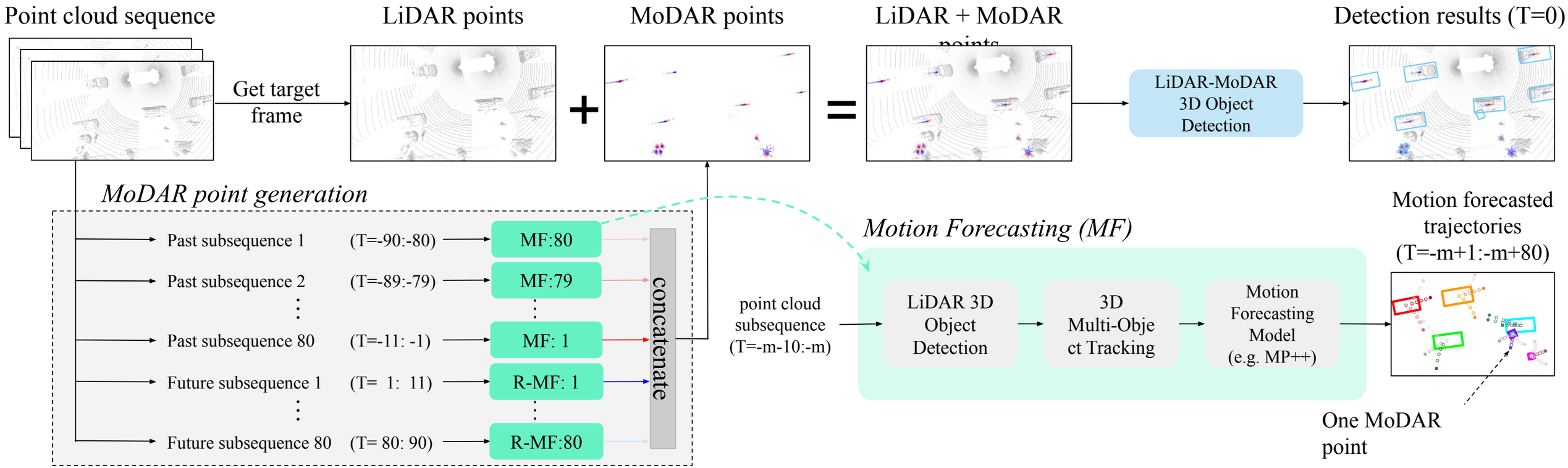


Using MoDAR for 3D object detection



WAYMO

Using MoDAR for 3D object detection

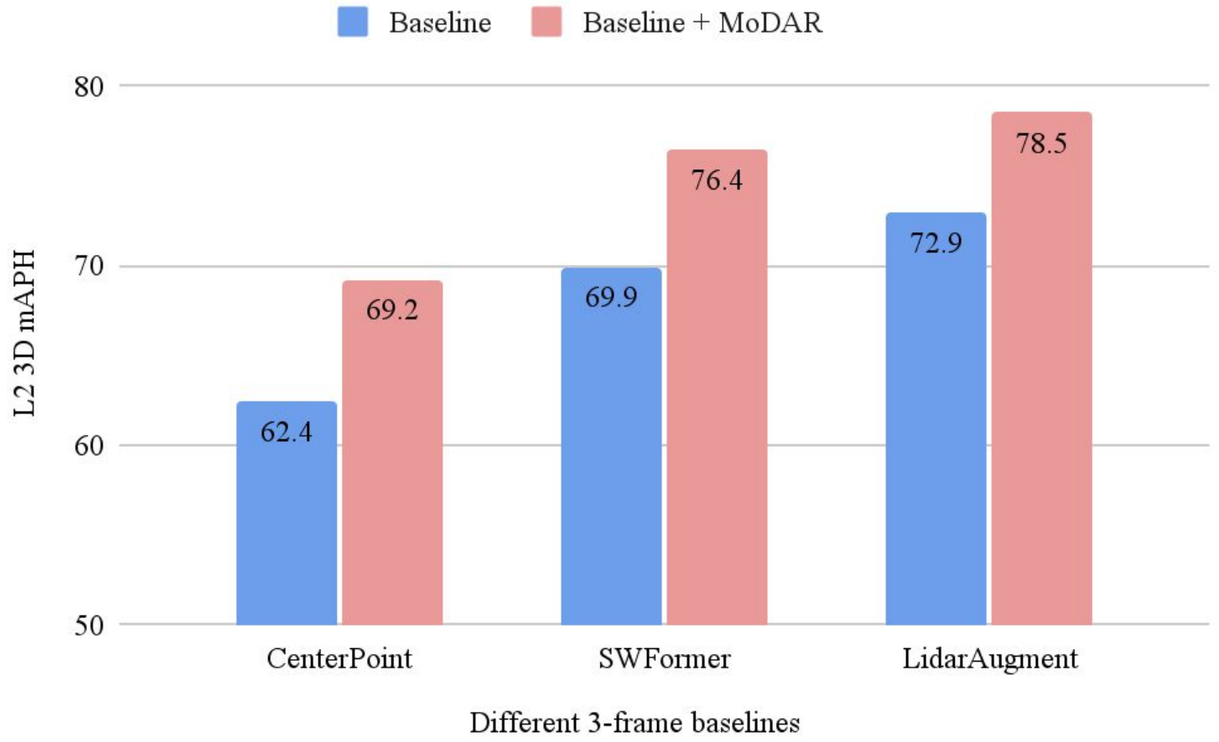


WAYMO

Experimental Results

MoDAR achieves SOTA results

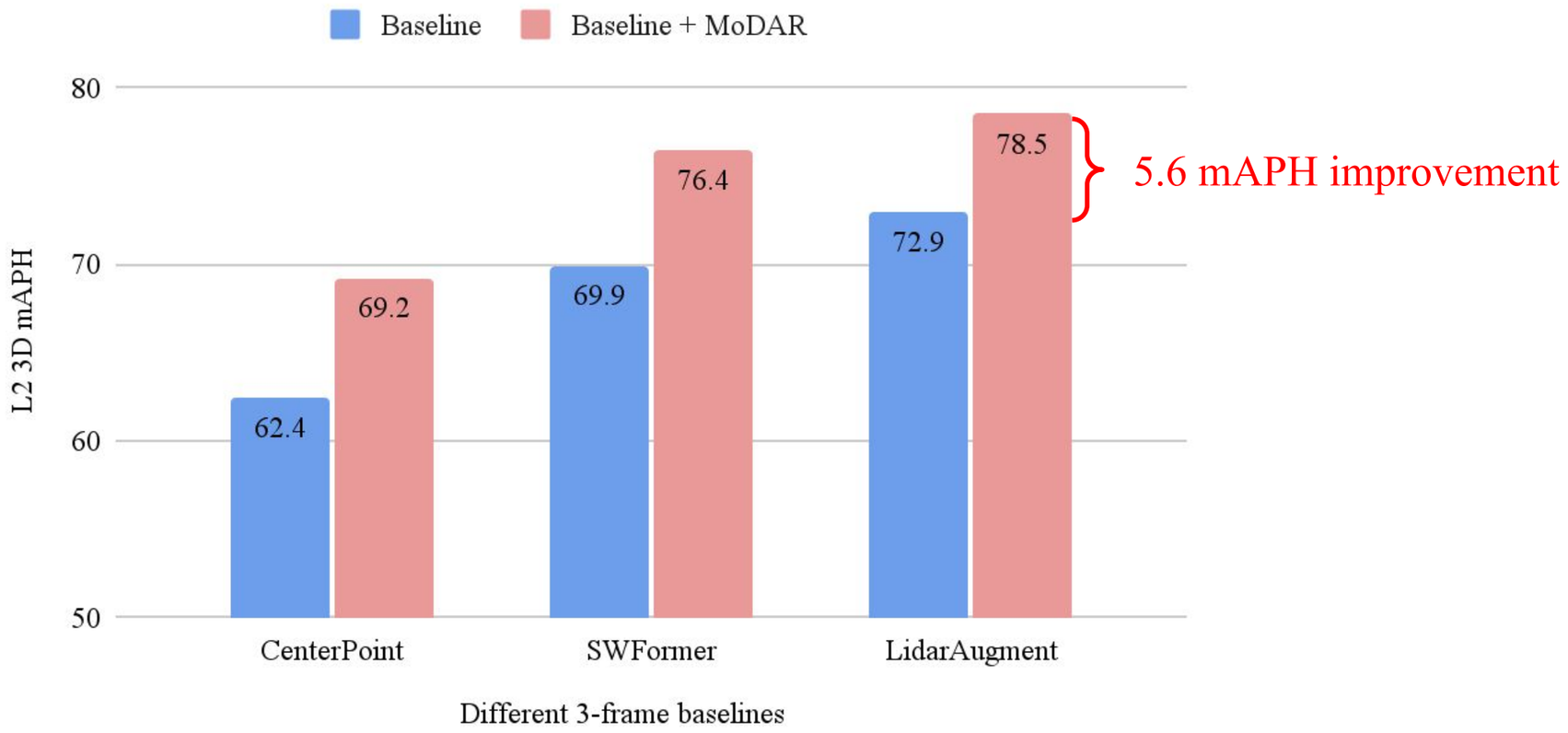
L2 3D mAPH evaluation on the WOD validation set.



[CenterPoint] Yin, Tianwei, Xingyi Zhou, and Philipp Krahenbuhl. "Center-based 3d object detection and tracking." CVPR. 2021.
[SWFormer] Sun, Pei, et al. "Swformer: Sparse window transformer for 3d object detection in point clouds." ECCV. 2022.
[LidarAugment] Leng, Zhaoqi, et al. "LidarAugment: Searching for Scalable 3D LiDAR Data Augmentations." ICRA. 2023.

MoDAR achieves SOTA results

L2 3D mAPH evaluation on the WOD validation set.



[CenterPoint] Yin, Tianwei, Xingyi Zhou, and Philipp Krahenbuhl. "Center-based 3d object detection and tracking." CVPR. 2021.
[SWFormer] Sun, Pei, et al. "Swformer: Sparse window transformer for 3d object detection in point clouds." ECCV. 2022.
[LidarAugment] Leng, Zhaoqi, et al. "LidarAugment: Searching for Scalable 3D LiDAR Data Augmentations." ICRA. 2023.

Effects of different object state features

The metric is L2 3D APH on the WOD val set. Base model is CenterPoint.

Location	Size	Heading	Class	Scores	Veh. L2 APH	Ped. L2 APH
✗	✗	✗	✗	✗	64.2	51.9
✓	✗	✗	✗	✗	69.7	61.2
✓	✓	✗	✗	✗	71.2	62.7
✓	✓	✓	✗	✗	70.9	64.0
✓	✓	✓	✓	✗	71.3	64.1
✓	✓	✓	✓	✓	**73.2**	**65.0**

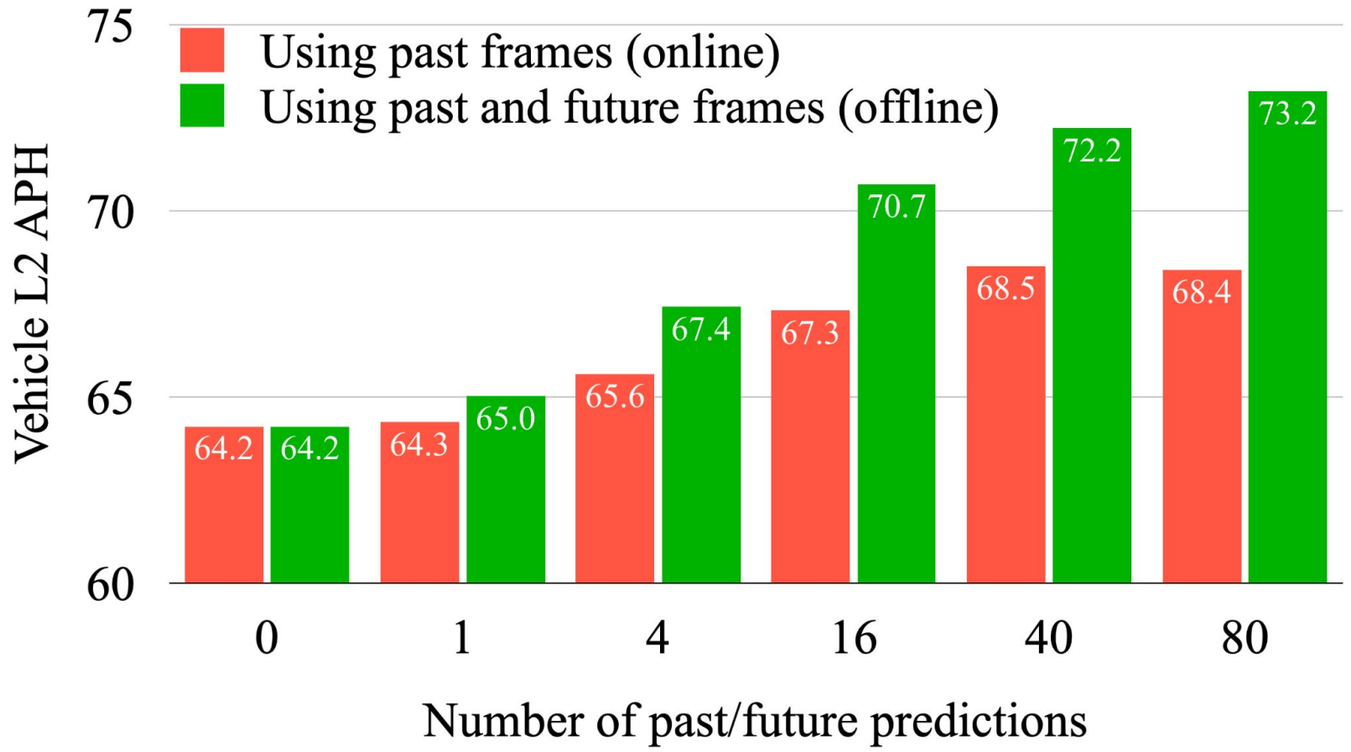
Distance breakdown

Distance breakdown for LiDAR-based and LiDARMoDAR based detection. The metrics are vehicle L2 3D APH on the WOD val set across different ground-truth depth ranges. The base detector is a 3-frame SWFormer.

Model	0-30m	30-50m	50m+	All
LiDAR only	90.4	69.7	45.6	69.7
LiDAR + MoDAR	**92.2**	**76.9**	**58.5**	**77.0**
	(+1.2)	(+7.2)	(+12.9)	(+7.3)

Ablation Study on Temporal Context Sizes

Effects of MoDAR temporal context sizes on 3D object detection. The metrics are vehicle L2 3D APH on the WOD val set. The (left) red bars are the performance of models using only MoDAR points generated from the past frames, while the (right) green bars are the performance of models that use the same number of past and future frames for MoDAR point generation.



Conclusion and Future Work

Conclusion and Future Work

- We proposed MoDAR, a virtual sensor modality that uses motion forecasting to propagate object states from past and future frames to a target frame.
- The MoDAR points generated from a point cloud sequence can be fused with other sensor modalities such as LiDAR to achieve more robust 3D object detection especially for cases with low visibility (occluded) or in long range.
- The method is generic and applicable to improve any point-cloud based 3D detectors. Based on LidarAugment, we achieved new state of the art results on Waymo Open Dataset without ensemble when submitting the paper.

Conclusion and Future Work

- We proposed MoDAR, a virtual sensor modality that uses motion forecasting to propagate object states from past and future frames to a target frame.
- The MoDAR points generated from a point cloud sequence can be fused with other sensor modalities such as LiDAR to achieve more robust 3D object detection especially for cases with low visibility (occluded) or in long range.
- The method is generic and applicable to improve any point-cloud based 3D detectors. Based on LidarAugment, we achieved new state of the art results on Waymo Open Dataset without ensemble when submitting the paper.

Future Work

- For offboard purposes: Design offboard motion forecasting models to improve MoDAR.
- For onboard purposes: Study the interdependence between the detection and the motion forecasting models, towards end-to-end detection and prediction.