# Image-Guided LiDAR Denoising for Robust 3D Object Detection in High-Speed Marine ADAS

Vatsal Srivastava
*Honda R&D Co., Ltd.*
351-0024 Saitama, Japan
vatsal_srivastava@jp.honda

E P Sreerag
*Honda R&D Co., Ltd.*
351-0024 Saitama, Japan
sreerag_ep@jp.honda

Itoshi Naramura
*Honda R&D Co., Ltd.*
351-0024 Saitama, Japan
itoshi_naramura@jp.honda

***Abstract*—Advanced Driver Assistance Systems (ADAS) for high-speed marine vessels face severe perception challenges from vessel-induced artifacts like wash, wake, and spray. These dynamic disturbances introduce critical noise into 3D LiDAR point clouds, severely degrading object detection performance. While image-guided filtering can mitigate this, we systematically demonstrate that models trained on standard open-source 2D segmentation benchmarks also fail under these conditions. To address this, we present our complete multi-modal perception architecture, which was used for both data collection and real-world testing. Within this framework, we first overcome the visual vulnerability through target-domain finetuning of the segmentation network. We then propose a robust point cloud denoising pipeline that utilizes the 2D segmentation output to actively filter noisy artifacts from the 3D LiDAR data. By combining target-domain data with this cross-modal pipeline, we demonstrate a significant improvement in overall object detection accuracy.**

***Index Terms*—High-speed boats, ADAS, maritime autonomy, semantic segmentation, object detection**

## I. Introduction

According to the U.S. Coast Guard's 2024 Recreational Boating Statistics, there were 3,887 boating accidents resulting in 556 deaths and thousands of injuries, with operator inattention, improper lookout, and operator inexperience ranking among the top primary contributing factors [5]. Because human error directly or indirectly contributes to over 80% of maritime navigational incidents, the recreational boating industry is increasingly looking toward Advanced Driver Assistance Systems (ADAS) to mitigate these risks. ADAS technology can make boating significantly safer and more enjoyable by providing active collision warnings and docking assistance, effectively acting as an intelligent co-pilot for both novice and experienced boaters.

The transition of autonomous systems from highly structured terrestrial environments to the chaotic, unstructured domain of open seas represents one of the foremost challenges in modern robotics. Marine environments are characterized by dynamic water surfaces, extreme lighting variations including aggressive sunglint, heavy occlusions from fog and sea spray, and a distinct lack of static background structures. This fundamentally breaks traditional perception algorithms designed for urban driving. During high-speed navigation of boats, these challenges are compounded by severe vessel-induced artifacts—namely wash, wake, and spray. These dynamic disturbances introduce critical noise into 3D LiDAR point clouds, severely degrading object detection performance - both in terms of false positives, and inaccurate bounding box estimates.

To address these multi-modal challenges, this paper presents a complete perception architecture designed for both high-speed and low-speed marine ADAS, which was deployed for real-world data collection and testing. Within this overarching framework, our core contribution is a real-time LiDAR point cloud denoising pipeline that uses water-surface segmentation on RGB images from extrinsically calibrated cameras (one for each LiDAR) to remove points arising from water-surface disturbances. This yields a cleaner point cloud for robust 3D object detection, which is a key downstream component of our ADAS pipeline for detecting nearby obstacle vessels.

To ensure the reliability of this visual filter, we first evaluate the vulnerabilities of open-source image segmentation benchmarks during high-speed navigation in marine environments. Because standard global evaluation metrics often mask severe worst-case failures, we use conditional tail-performance metrics, $\mathrm{CVaR}_{10}[\mathrm{IoU}_{\mathrm{water}}]$ and $\mathrm{CVaR}_{10}[\mathrm{FNR}_{\mathrm{water}}]$ - evaluating the bottom decile performance on a target-domain test set. Finally, we demonstrate how LiDAR-based object detection accuracy improves significantly when using denoised point clouds, specifically highlighting the enhancements in $\mathrm{AP}_{\mathrm{boat}}$ scores, pre- and post-denoising.

## II. Related Work

Compared to ground plane removal from LiDAR point clouds in terrestrial ADAS, in marine environments, the detection and removal of water surface from LiDAR is a significantly harder problem. The highly dynamic nature of the water surface in the open sea causes variation in 6D pose of the ego vessel, even when "static". For high-speed vessels, the problem further transitions from 2D surface fitting to a 3D volumetric one, since wash, wake, and spray generate dense artifacts that extend well above the nominal water surface.

A recent approach in water-surface removal attempted to use plane-fitting on near-field calm water surface for a USV [3], but such methods are ill-suited to high-speed open-sea marine scenes, where water surface is highly dynamic and has 3D artifacts. In contrast, our pipeline utilizes semantic information from images, yielding a cleaner point cloud for

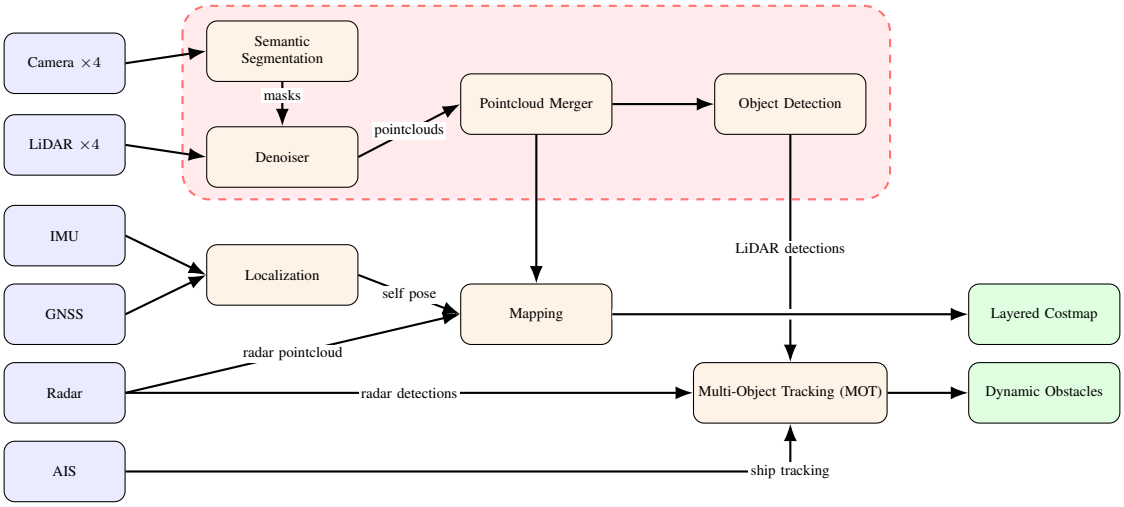



Fig. 1. Our proposed multi-modal perception pipeline: sensors are denoted in blue, software modules in orange and the final outputs in green. The primary focus of this paper is highlighted in red.

downstream detection. The performance of our pipeline thus remains unaffected by vessel speed or the severity of water surface disturbance.

## III. Open-Source Datasets

Accurate water-surface segmentation is a fundamental prerequisite for safe marine ADAS. However, open-source marine datasets with sufficiently detailed and consistent annotations remain scarce. Although several strong benchmarks exist for low-speed vessels, their ability to generalize to the dynamic operating conditions of high-speed boats remains underexplored.

For this study, we considered MOANA [6], Pohang Canal [7] with labels from PoLaRIS [8], MODD [9], and the LaRS benchmark [2]. MOANA, MODD, and PoLaRIS do not provide the dense pixel-level annotations required for our segmentation training. The LaRS benchmark offers substantial scene diversity and exceptionally high-quality annotations, but it still lacks sufficient coverage of the severe water-surface disturbances observed in our target domain.

Progress in LiDAR-based object detection for marine environments also depends heavily on high-quality datasets. The Pohang Canal Dataset provides recordings from three LiDARs in maritime scenes, and PoLaRIS extends it to around 10,000 labelled ships. However, because PoLaRIS only has labels of one target boat even in scenes with multiple boats, we could not use it directly for detector training.

## IV. Hardware Setup

In this section we describe the hardware setup that we designed and used for real-world testing and data collection, on-board a high-speed boat. We had the following devices on-board:

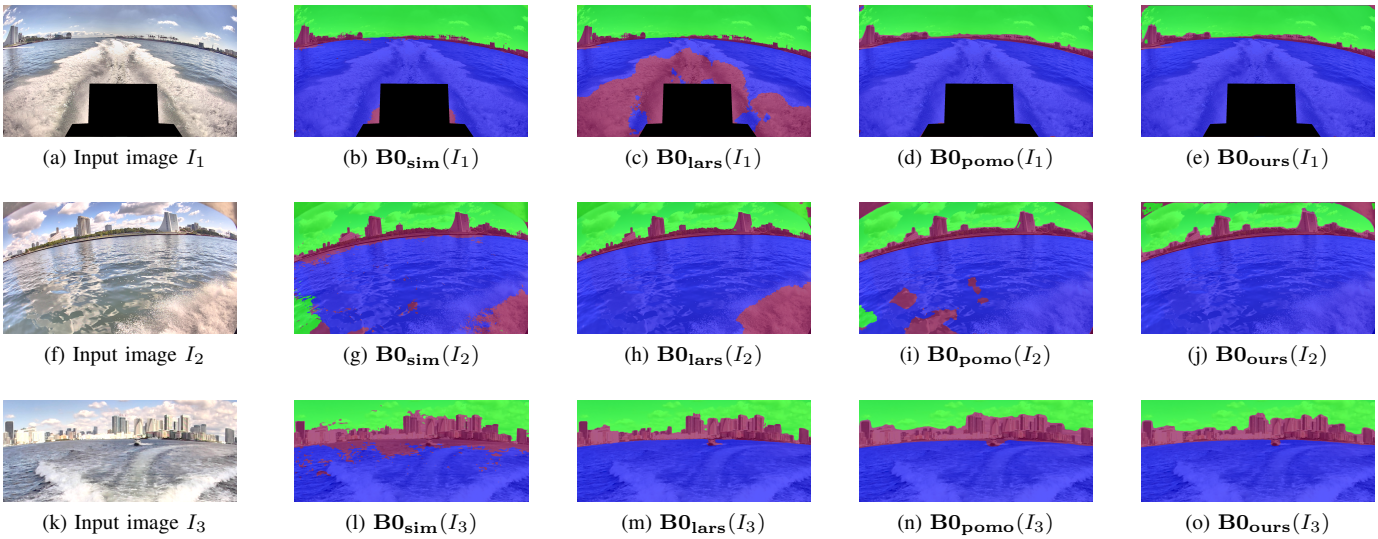
- **LiDAR:** Four solid state LiDARs (front, rear, port, starboard) providing pointclouds at 10Hz each.
- **Camera:** Four FHD cameras (front, rear, port, starboard) aligned with corresponding LiDARs, also operating at 10Hz each.
- **Radar:** A marine radar operating at 0.4Hz.
- **State Sensors:** Self position and orientation estimation using IMU and GNSS at 50 Hz.
- **Communications:** An Automatic Identification System (AIS) receiver.
- **GPU:** NVIDIA RTX 3090.

## V. Our Multi-Modal Perception Pipeline

Figure 1 details our overall perception pipeline. We estimate self position and orientation using IMU and GNSS. Four paired camera-LiDAR modules are extrinsically calibrated to provide a 360-degree field of view. Additionally, a marine radar enables long-range detection and mapping, while AIS enables tracking of vessels.

A single compute unit (24GB NVIDIA RTX 3090) handles all processing. Camera images undergo semantic segmentation via a TensorRT-optimized SegFormer B0 network [1] (averaging approx. 40 ms for four frames in parallel), classifying pixels into Background, Sky, and Water. Raw LiDAR point clouds are synchronized and projected onto these 2D masks inside the Denoiser; points falling on Water or Sky pixels are removed. The filtered point cloud is then merged and fed into a CenterPoint-based 3D object detector [4] to detect obstacle boats (averaging approx. 40 ms).

To build an actionable navigational space, the denoised LiDAR data generates a high-resolution, near-field obstacle costmap, while radar data generates a macro-scale costmap robust to weather disturbances. These representations are fused into a unified layered costmap. Finally, a Multi-Object Tracking module maintains dynamic obstacle tracks using combined data from the LiDAR detections, Radar, and AIS broadcasts.

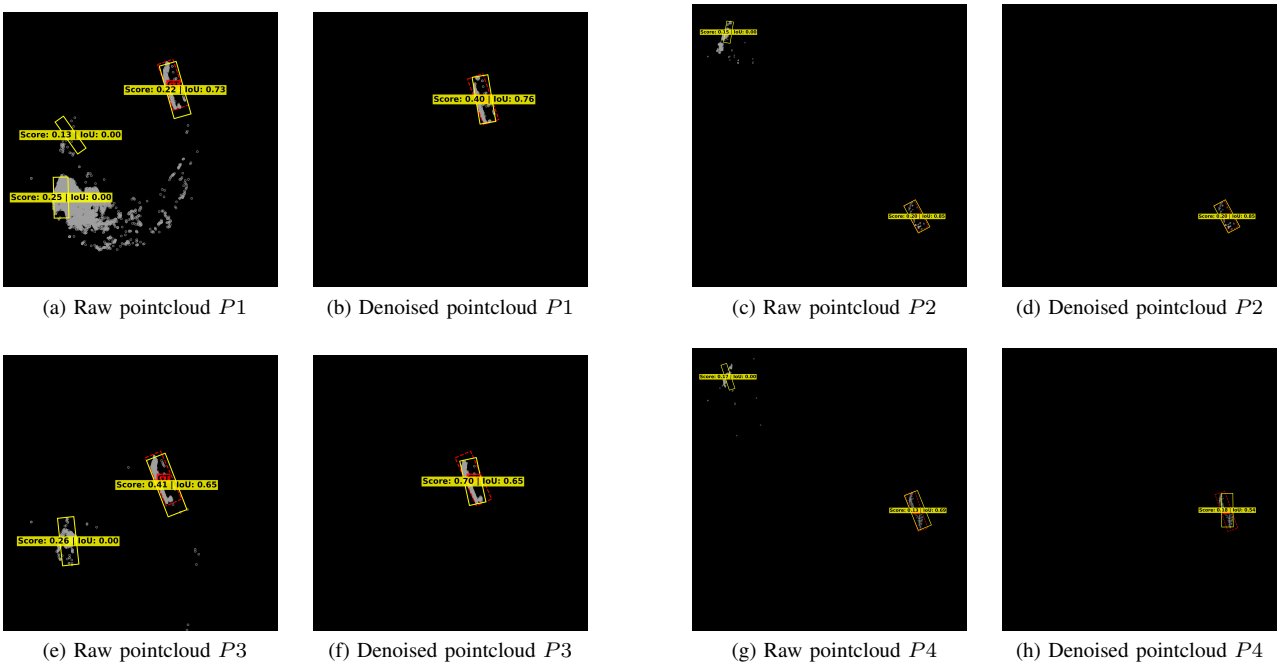


(a) Input image $I_1$ (b) $\mathbf{B0_{sim}}(I_1)$ (c) $\mathbf{B0_{lars}}(I_1)$ (d) $\mathbf{B0_{pomo}}(I_1)$ (e) $\mathbf{B0_{ours}}(I_1)$

(f) Input image $I_2$ (g) $\mathbf{B0_{sim}}(I_2)$ (h) $\mathbf{B0_{lars}}(I_2)$ (i) $\mathbf{B0_{pomo}}(I_2)$ (j) $\mathbf{B0_{ours}}(I_2)$

(k) Input image $I_3$ (l) $\mathbf{B0_{sim}}(I_3)$ (m) $\mathbf{B0_{lars}}(I_3)$ (n) $\mathbf{B0_{pomo}}(I_3)$ (o) $\mathbf{B0_{ours}}(I_3)$

Fig. 2. Qualitative comparison of SegFormer output on different training sets. Rows show input image and predictions from $\mathbf{B0_{sim}}$, $\mathbf{B0_{lars}}$, $\mathbf{B0_{pomo}}$, and $\mathbf{B0_{ours}}$.




(a) Raw pointcloud $P1$ (b) Denoised pointcloud $P1$ (c) Raw pointcloud $P2$ (d) Denoised pointcloud $P2$

(e) Raw pointcloud $P3$ (f) Denoised pointcloud $P3$ (g) Raw pointcloud $P4$ (h) Denoised pointcloud $P4$

Fig. 3. Qualitative comparison of detector output on raw and denoised pointclouds in bird's eye view (BEV). Point clouds are shown in gray, ground truth bounding boxes are shown in red, detections are shown in yellow, along with confidence and IoU for true positives.

## VI. Experiments & Evaluation Metrics

This section presents a two-stage empirical evaluation of our proposed LiDAR denoising pipeline. First, we analyze the "visual bottleneck" by evaluating how well state-of-the-art semantic segmentation models trained on open-source maritime benchmarks generalize zero-shot to our target domain, and we compare them with a model fine-tuned on target-domain data. We then evaluate the downstream impact of our cross-modal denoising strategy. By comparing 3D object detection performance on raw and denoised point clouds, we show that image-guided filtering is necessary for reliable object detection in high-speed marine environments.

### *A. Semantic Segmentation: Synthetic vs. Open-source vs. Target Domain Data*

We established four training configurations using the following source datasets: (1) synthetic data, (2) a combined MOANA–Pohang dataset, (3) the LaRS benchmark, and (4) an in-house training set. To isolate the effect of the training data, we used the same SegFormer-B0 architecture in all trials.

All models were evaluated on an in-house test set, `Ours_heavywake`, containing 463 images. This dataset consists of open-sea imagery dominated by vessel-induced wash, wake, and spray - phenomena that are impacted by vessel speed and hull dynamics. Robust perception under these conditions is essential for our pipeline, particularly for downstream noise removal.

Analysis of existing open-source datasets reveals a substantial distribution shift with respect to these high-speed artifacts. LaRS and Pohang contain very few instances of wash or spray, suggesting that they are better suited to Unmanned Surface Vehicles (USVs) or low-speed vessels. Although MOANA includes examples of bow spray, its exclusive use of forward-facing cameras prevents it from capturing aft wake and broadside wash, leaving an important gap in the training distribution for multi-camera setups on high-speed vessels.

Standard semantic segmentation metrics often mask episodic failures in continuous video feeds. A model might achieve a high global score by correctly segmenting open, calm water while completely failing during the brief seconds a camera is obscured by heavy spray.

To capture both nominal performance and these worst-case failure modes, we report standard global metrics alongside tail-performance metrics. The standard metrics include mean Intersection over Union ($\mathrm{mIoU}$), water-class IoU ($\mathrm{IoU}_{\text{water}}$), and water-class True Positive Rate ($\mathrm{FNR}_{\text{water}}$). To quantify worst-case robustness, we utilize Conditional Value at Risk (CVaR) metrics, specifically $\mathrm{CVaR}_{10}[\mathrm{IoU}_{\text{water}}]$ and $\mathrm{CVaR}_{10}[\mathrm{FNR}_{\text{water}}]$. These are computed by extracting the 10% of test frames with the lowest respective scores, and calculating the final pixel-wise metric exclusively over this degraded subset.

### B. 3D Object Detection: Raw vs. Denoised Point clouds

We evaluate the effectiveness of our denoising pipeline by comparing CenterPoint-based 3D object detector [4] performance on raw and denoised point clouds. Due to the absence of open source benchmarks for LiDAR-based object detection in marine scenes, the evaluation is performed on an in-house test set, `Ours_wakenoise`, containing extrinsically calibrated, LiDAR-Camera data with 3D bounding box annotations.

We prepare two sets of point cloud data from the test set - (1) raw unfiltered point clouds (2) parallel, we feed data from the test set into our denoising pipeline as is - camera images undergo semantic segmentation, and the segmentation masks are fed to the Denoiser along with the raw point clouds. We again evaluate performance of the detector on denoised point clouds thus generated, and compare the two results. The matching criteria used for bounding boxes is Bird's Eye View IoU, denoted as $\mathrm{IoU}_{\mathrm{bev}}$. Performance is reported using $\mathrm{AP}_{\text{boat}}$ at $\mathrm{IoU}_{\mathrm{bev}}$ thresholds $0.2, 0.5, 0.7$, and Precision–Recall curves with finer thresholds.

## VII. Results and Discussion

### A. Segmentation

The quantitative results of our cross-domain evaluation are summarized in Table I. Analysis of the global performance metrics initially suggests that zero-shot transfer from open-source datasets yields an acceptable baseline. For instance, the model trained exclusively on the LaRS benchmark ($\mathbf{B0_{lars}}$) achieves a global $\mathrm{IoU}_{\text{water}}$ of 96.85% and maintains a low global $\mathrm{FNR}_{\text{water}}$ of 2.00% when evaluated on our heavy wash dataset.

However, evaluating the tail performance exposes severe vulnerabilities in models lacking exposure to extreme target-domain artifacts. Notably, despite the well-documented sim-to-real domain gap, the worst-case performance of the real-world $\mathbf{B0_{lars}}$ model degrades to numerical levels nearly identical to the purely synthetic $\mathbf{B0_{sim}}$ model. While $\mathbf{B0_{lars}}$ maintains high average accuracy, its conditional false negative rate ($\mathrm{CVaR}_{10}[\mathrm{FNR}_{\text{water}}]$) spikes to 14.39%, closely mirroring the 15.39% worst-case error rate recorded by $\mathbf{B0_{sim}}$ in the bottom decile.

The $\mathbf{B0_{pomo}}$ model, trained on the combined Pohang and MOANA datasets, demonstrates improved zero-shot robustness, achieving a $\mathrm{CVaR}_{10}[\mathrm{IoU}_{\text{water}}]$ of 94.80% and reducing the tail FNR to 4.68%. This improvement is likely attributable to the limited presence of bow spray in the MOANA training images. Nevertheless, a nearly 5% worst-case error rate remains suboptimal for downstream denoising operations.

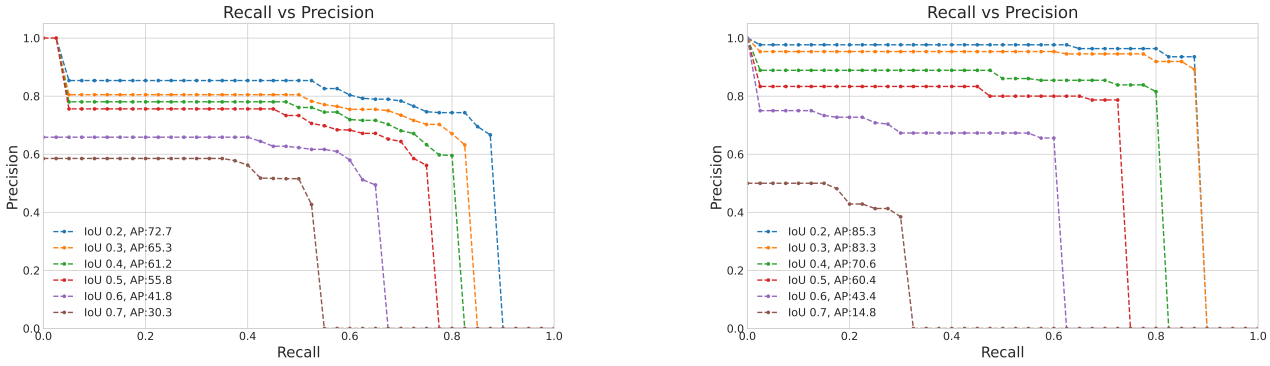
The inclusion of target-domain data in the $\mathbf{B0_{ours}}$ model resolves these temporal failures. While the global $\mathrm{IoU}_{\text{water}}$ improves to 99.09%, the most significant gain is observed in the conditional tail metrics. $\mathbf{B0_{ours}}$ maintains a $\mathrm{CVaR}_{10}[\mathrm{IoU}_{\text{water}}]$ of 97.77%, proving that the network successfully learned to segment water surface seamlessly through heavy broadside wash and aft wake. Furthermore, the $\mathrm{CVaR}_{10}[\mathrm{FNR}_{\text{water}}]$ is reduced to a negligible 0.75%. By using target-domain images for the training distribution, our perception pipeline transitions from a system that is accurate on average to one that remains structurally robust in the worst-case, satisfying the prerequisites for downstream noise removal.

Figure 2 shows the qualitative comparison of the models on three sample input images. The segmentation mask is overlayed onto the input image for visualization - red denotes background, blue denotes water, and green denotes sky. This helps us understand the areas where each of the model is prone to failure. Of the three sample inputs, Figs 2a and 2k are images captured from the rear camera on-board our vessel during high-speed navigation, and shows heavy wash and wake characteristic to our target domain. Fig 2k also shows that the models not finetuned on target domain data sometimes confuse the wake of other boats as background. Fig 2f was captured from a side camera, and shows dense spray.

Clearly, $\mathbf{B0_{sim}}$ struggles with accurately segmenting far-field regions, specifically along the sky-sea boundary at the horizon. In stark contrast, $\mathbf{B0_{lars}}$ performs exceptionally well near the horizon, and more generally near all class boundaries

TABLE I
SEGMENTATION RESULTS: SYNTHETIC VS. OPEN-SOURCE VS. TARGET DOMAIN DATA

| Model Name | Training Data | | | | | Test Data | Global Performance | | | Tail performance | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Synthetic | Pohang | MOANA | LaRS | Ours | | mIoU | $\text{IoU}_{\text{water}}$ | $\text{FNR}_{\text{water}}$ | $\text{CVaR}_{10}[\text{IoU}_{\text{water}}]$ | $\text{CVaR}_{10}[\text{FNR}_{\text{water}}]$ |
| $\mathbf{B0_{sim}}$ | ✓ | | | | | Ours_heavywake | 80.22 | 92.00 | 7.46 | 84.20 | 15.39 |
| $\mathbf{B0_{lars}}$ | | | | ✓ | | Ours_heavywake | 87.22 | 96.85 | 2.00 | 84.85 | 14.39 |
| $\mathbf{B0_{pomo}}$ | | ✓ | ✓ | | | Ours_heavywake | 91.12 | 98.65 | 0.82 | 94.80 | 4.68 |
| $\mathbf{B0_{ours}}$ | | ✓ | ✓ | | ✓ | Ours_heavywake | **93.54** | **99.09** | **0.10** | **97.77** | **0.75** |




(a) Pre-denoising

(b) Post-denoising

Fig. 4. Precision–Recall curves for $\mathbf{CP_{boat}}$ on Ours_wakenoise set before and after applying our proposed image-guided LiDAR denoising pipeline.

TABLE II
OBJECT DETECTION RESULTS: RAW VS POST-DENOISING POINT CLOUDS

| Method | Test Data | Denoised? | $\text{AP}_{\text{boat}}$ for diff. $\text{IoU}_{\text{bev}}$ matching thresholds | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | AP@0.2 | AP@0.3 | AP@0.4 | AP@0.5 | AP@0.6 | AP@0.7 |
| **CenterPoint** | Ours_wakenoise | ✗ | 72.7 | 65.3 | 61.2 | 55.8 | 41.8 | **30.3** |
| **CenterPoint** | Ours_wakenoise | ✓ | **85.3** | **83.3** | **70.6** | **60.4** | **43.4** | 14.8 |

- bacground-sky, sky-water, and background-water. However, it sees massive failures sometimes when there is wake, wash and spray in the scene. $\mathbf{B0_{pomo}}$ performs relatively better in this regard, at the cost of a reduced sharpness along the class boundaries. Inarguably, the best performing model - qualitatively as well, is B_inhouse, accurately segmenting both near-field and far-field regions.

### *B. Object Detection*

The quantitative object-detection results are summarized in Table II, and the corresponding precision–recall curves are shown in Fig. 4. Applying the proposed image-guided denoising pipeline substantially improves the performance of $\mathbf{CP_{boat}}$ across all $\text{IoU}_{\text{bev}}$ matching thresholds: AP@0.2 increases from 72.7 to 85.3, AP@0.4 increases from 61.2 to 70.6 and AP@0.6 increases from 41.8 to 43.4. When the matching theshold is made stricter, at $\text{IoU}_{\text{bev}} = 0.7$, we see a drop in AP@0.7 from 30.3 to 14.8. The reason for this drop is explained below.

Due to minute calibration errors or time synchronization gaps or tiny errors in segmentation, the denoiser can sometimes accidentally remove points that are not noise, but are true obstacle points. This happens particularly around object boundaries, and is a limitation of our system. However, this mainly affects the localization error of the detection. When using stricter matching thesholds like $\text{IoU}_{\text{bev}} = 0.7$, detections with higher localization error become false positives. Rather than a highly precise bounding box estimate, in our case, the detection of obstacles is of higher priority. Therefore, thesholds as high as $\text{IoU}_{\text{bev}} = 0.7$ are perhaps too strict for our use case.

Figure 3 shows qualitative comparison of detector outputs on the same LiDAR data, pre- and post-denoising which further demonstrates the effectiveness of our pipeline. The pointclouds are shown in gray, ground truth bounding boxes in red, and detections in yellow. In figures 3c and 3g, upper left region shows heavy wake and spray from our own boat, while the lower right region shows a moving obstacle boat far away from our own vessel. In figures 3a and 3e, there is dense wake and wash caused by a nearby obstacle boat.

The post-denoising precision–recall curve also expands markedly toward the upper-right region, indicating a consistent gain in precision and recall after wake-induced clutter is removed. Together, these results show that image-guided LiDAR denoising substantially improves marine 3D object detection under severe surface disturbances.

## VIII. Conclusion

This paper presented a complete multi-modal perception architecture for marine ADAS. Our system integrates image semantics, pointcloud geometry for reliable 3D object detection, tracking and layered costmap generation within a unified system. Within this broader architecture, our primary contribution was a robust image-guided LiDAR denoising pipeline designed for the severe wake, wash, and spray encountered during high-speed operation.

Our experiments showed that reliable denoising depends on reliable target-domain water segmentation. Segmentation models that appear strong under global metrics can still fail during the short worst-case intervals that matter most for downstream denoising. To counter this, target-domain supervision provides the robustness required for stable filtering. This enables the proposed cross-modal pipeline to remove water-induced clutter from raw point clouds during high-speed navigation. The denoising pipeline thus yields a cleaner geometric representation for downstream 3D object detection, mapping, and tracking.

Overall, the proposed system demonstrates both a practical LiDAR denoising strategy and a broader weather-resilient perception architecture for both low-speed and high-speed ADAS.

## Acknowledgment

The authors would like to express their gratitude to all members of the Marine ADAS project at the Research Center of Excellence, Honda R&D for their vital contributions towards the design and development of our system and real-world testing.

## References

[1] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12077–12090, 2021.

[2] L. Žust, J. Perš, and M. Kristan, "LaRS: A diverse panoptic maritime obstacle detection dataset and benchmark," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2023.

[3] "Object detection framework based on LiDAR and camera fusion for complex water environments," *Sensors*, 2024.

[4] T. Yin, X. Zhou, and P. Krähenbühl, "Center-based 3D object detection and tracking," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021.

[5] U.S. Coast Guard, "2024 Recreational Boating Statistics," U.S. Department of Homeland Security, Washington, DC, 2025.

[6] H. Jang, W. Yang, H. Kim, D. Lee, Y. Kim, J. Park, M. Jeon, J. Koh, Y. Kang, M. Jung, S. Jung, and A. Kim, "MOANA: Multi-radar dataset for maritime odometry and autonomous navigation application," *International Journal of Robotics Research*, vol. 0, no. 0, pp. 1–11, 2025, doi: 10.1177/02783649251354897.

[7] D. Chung, J. Kim, C. Lee, and J. Kim, "Pohang canal dataset: A multimodal maritime dataset for autonomous navigation in restricted waters," *International Journal of Robotics Research*, vol. 42, no. 12, pp. 1104–1114, 2023, doi: 10.1177/02783649231191145.

[8] J. Choi, D. Cho, G. Lee, H. Kim, G. Yang, J. Kim, and Y. Cho, "PoLaRIS dataset: A maritime object detection and tracking dataset in Pohang canal," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, 2025, pp. 13626–13632, doi: 10.1109/ICRA55743.2025.11128583.

[9] M. Kristan, V. Sulič Kenk, S. Kovačič, and J. Perš, "Fast image-based obstacle detection from unmanned surface vehicles," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 641–654, 2016, doi: 10.1109/TCYB.2015.2412251.