

Abstract

Autonomous driving on water surfaces plays an essential role in executing hazardous and time-consuming missions, such as maritime surveillance, survivors rescue, environmental monitoring, hydrography mapping and waste cleaning. This work presents WaterScenes, **the first multi-task 4D radar-camera fusion dataset** for autonomous driving on water surfaces. Equipped with a 4D radar and a monocular camera, our Unmanned Surface Vehicle (USV) proffers **all-weather solutions for discerning object-related information**. In addition to basic perception tasks, such as object detection, instance segmentation and semantic segmentation, we also provide annotations for free-space segmentation and waterline segmentation. Leveraging the **multi-task and multi-modal data**, we demonstrate that **4D radar-camera fusion can considerably improve the accuracy and robustness** of perception on water surfaces, especially in adverse lighting and weather conditions. WaterScenes dataset is public on <https://waterscenes.github.io>.

Contributions

⇒ We present WaterScenes, the first multi-task 4D radar-camera fusion dataset on water surfaces, which offers **data from multiple sensors, including a 4D radar, monocular camera, GPS, and IMU**. It can be applied in **six perception tasks, including object detection, instance segmentation, semantic segmentation, free-space segmentation, waterline segmentation, and panoptic perception**.

⇒ Our dataset covers **diverse time conditions** (daytime, nightfall, night), **lighting conditions** (normal, dim, strong), **weather conditions** (sunny, overcast, rainy, snowy) and **waterway conditions** (river, lake, canal, moat).

⇒ We provide **2D box-level and pixel-level annotations for camera images, and 3D point-level annotations for radar point clouds**. We also offer a **toolkit** for WaterScenes that includes: pre-processing, labeling, projection and visualization, assisting researchers in processing and analyzing our dataset.

⇒ We build corresponding benchmarks and evaluate popular algorithms for object detection, point cloud segmentation, image segmentation, and panoptic perception. Experiments demonstrate the **advantages of radar perception on water surfaces, particularly in adverse lighting and weather conditions**.

Research Model 1



Figure 2. Sample images in WaterScenes. Radar points are projected onto the image plane as colored dots.

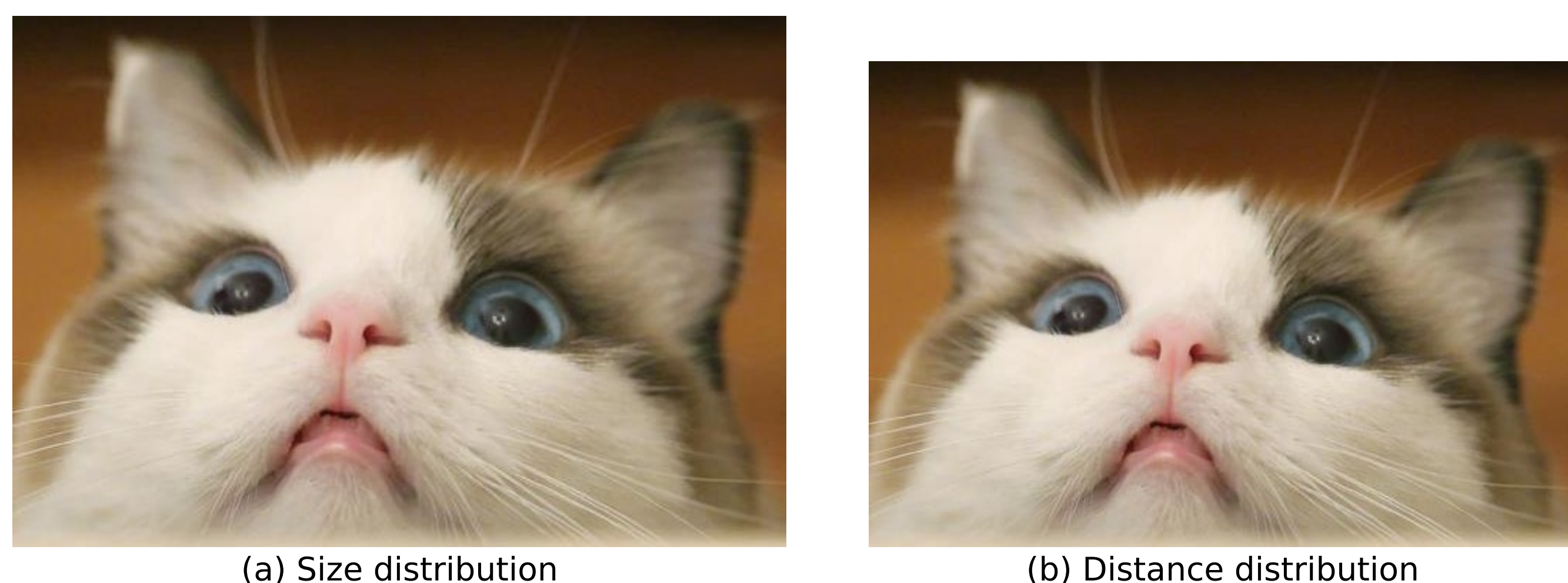


Figure 3. Statistics of objects in WaterScenes. (a) Wide range of object size. (b) Wide distribution of object distance.

Table 1. Statistics of each category in WaterScenes.

Category	Pier	Buoy	Sailor	Ship	Boat	Vessel	Kayak	Total
Frames	25,787	3,769	3,613	19,776	9,106	9,362	366	54,120
Objects	121,827	16,538	8,036	34,121	10,819	11,092	374	202,807
Percentage	60.07%	8.15%	3.96%	16.82%	5.33%	5.47%	0.18%	100%
Points	8.45	14.53	4.75	81.23	38.51	80.32	6.72	33.50
Power (dB)	13.68	17.88	12.15	14.40	14.14	13.52	10.12	13.70
Velocity (m/s)	0.08	0.09	0.79	1.08	0.40	2.21	0.88	0.79



Figure 1. Sensor suite for our USV and coordinate system of each sensor.

Research Model 2

We evaluate SOTA models on WaterScenes dataset. For object detection, we utilize YOLOv8 for camera-based detection and an early fusion method based on YOLOv8 modules for fusion-based detection. The evaluation results highlight the effectiveness of 4D radar-camera fusion for improving object detection accuracy, particularly in adverse lighting and weather conditions.

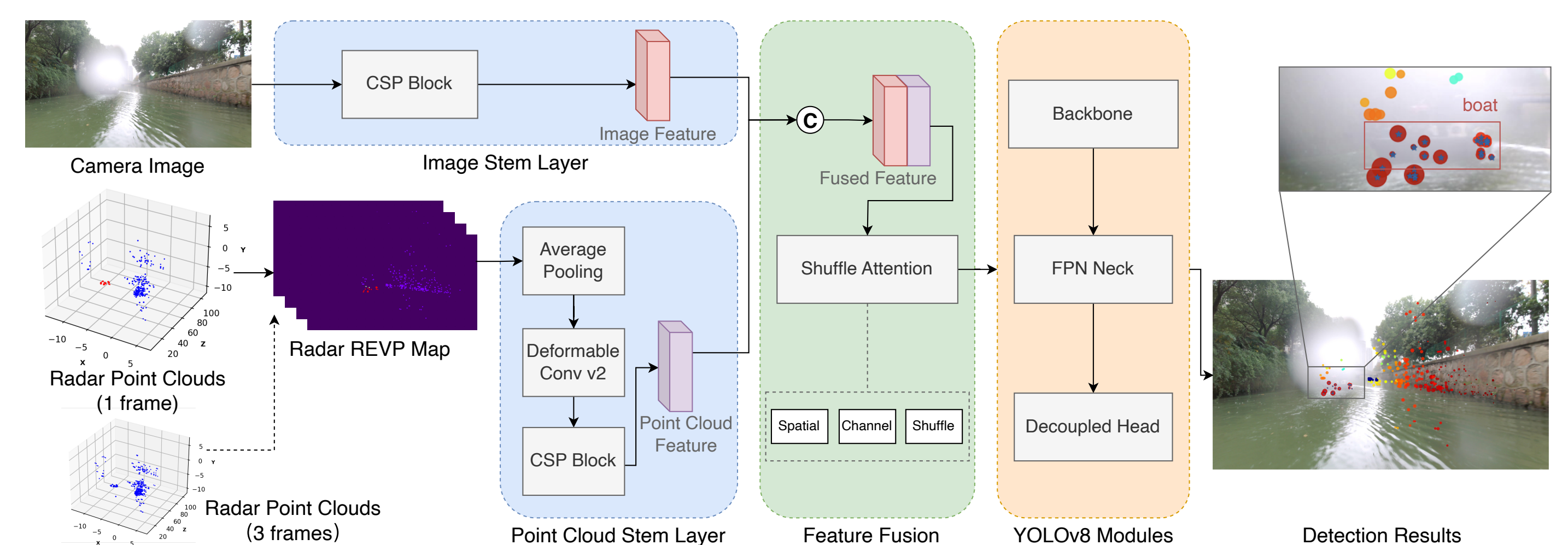


Figure 4. Radar-camera fusion network for the detection benchmark on WaterScenes dataset. Camera images and radar point clouds are fed into the stem layers for feature extraction. Subsequently, the extracted features are concatenated along the channel dimension and processed by the attention mechanism before forwarding into YOLOv8-M modules. As a result, the fusion-based network successfully detects the boat in scenarios where the camera is occluded by waterdrops.

Table 2. Comparison of object detection performance using camera-based and fusion-based methods. In the Modalities column, C denotes the modality from the camera sensor, R denotes the modality from the 4D radar sensor, n-frame(s) denotes the accumulation of n-frame radar point clouds.

Modalities	mAP @50	Adverse Lighting	Adverse Weather
C	84.4	74.9	79.5
C + R ₁ -frame	88.0 (3.6↑)	80.1 (5.2↑)	82.4 (2.9↑)
C + R ₃ -frames	88.8 (4.4↑)	82.1 (7.2↑)	84.2 (4.7↑)

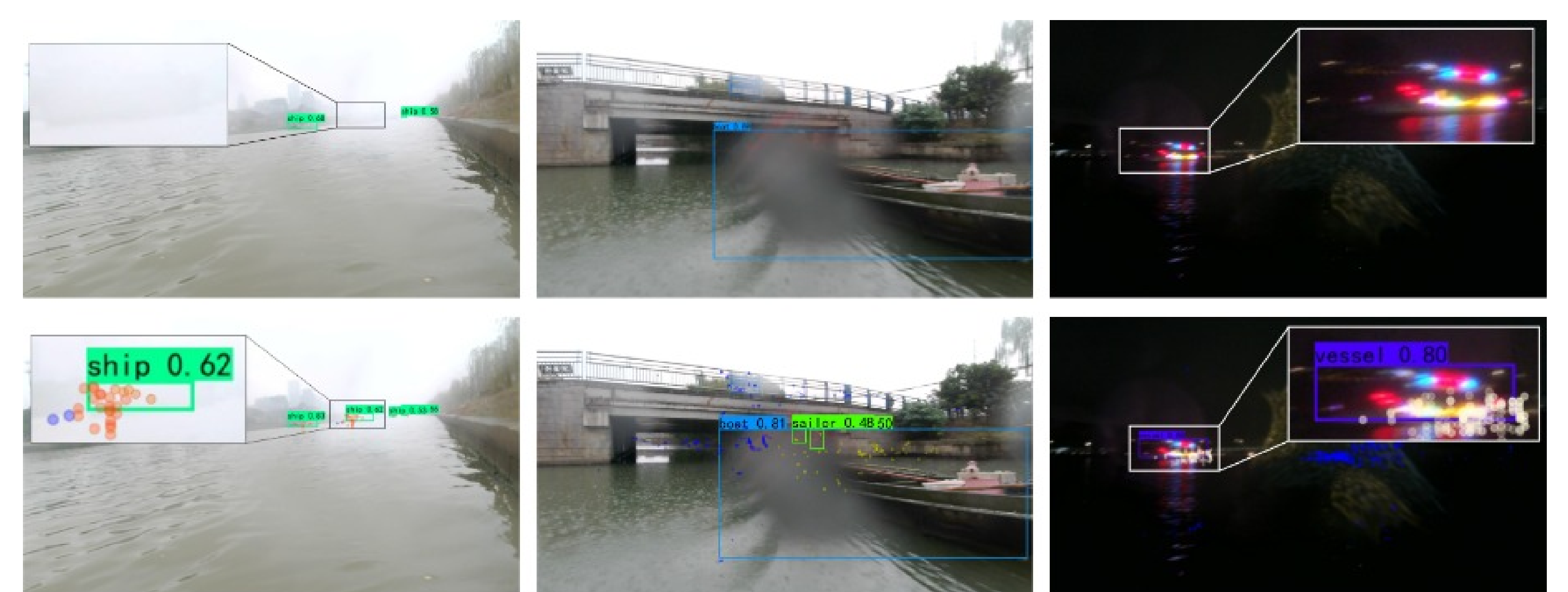


Figure 5. Visualization of object detection results on WaterScenes. The first and second rows present camera-based YOLOv8-M and fusion-based YOLOv8-M, respectively.

Conclusion

This work proposed a novel 4D radar-camera dataset named WaterScenes for six perception tasks on water surfaces. With the complementary advantages of radar and camera sensors, our dataset enables multi-attribute and all-weather perception of objects on water surfaces. Experimental results indicate that the 4D radar-camera combination is a robust solution for USVs on water surfaces. The presented WaterScenes offers a valuable resource for researchers interested in autonomous driving on water surfaces and aims to motivate novel ideas and directions for the development of water surface perception algorithms.

References

- [1] Yao S, Guan R, Wu Z, et al. WaterScenes: A Multi-Task 4D Radar-Camera Fusion Dataset and Benchmark for Autonomous Driving on Water Surfaces[J]. arXiv preprint arXiv:2307.06505, 2023.