FloW: A Dataset and Benchmark for Floating Waste Detection in Inland Waters

Yuwei Cheng*[1,2], Jiannan Zhu*[1,4], Mengxin Jiang[1,2], Jie Fu[3], Changsong Pang[1,4], Peidong Wang[1,2]
Kris Sankaran[3], Olawale Onabola[3], Yimin Liu[2], Dianbo Liu[3], Yoshua Bengio[3]
[1]ORCA-Uboat, [2]Tsinghua University, [3]Mila-Québec AI Institute, [4]Northwestern Polytechnical University

Abstract

Marine debris is severely threatening the marine lives and causing sustained pollution to the whole ecosystem. To prevent the wastes from getting into the ocean, it is helpful to clean up the floating wastes in inland waters using the autonomous cleaning devices like unmanned surface vehicles. The cleaning efficiency relies on a high-accurate and robust object detection system. However, the small size of the target, the strong light reflection over water surface, and the reflection of other objects on bank-side all bring challenges to the vision-based object detection system. To promote the practical application for autonomous floating wastes cleaning, we present FloW[†]*, the first dataset for floating waste detection in inland water areas. The dataset consists of an image sub-dataset FloW-Img and a multimodal sub-dataset FloW-RI which contains synchronized millimeter wave radar data and images. Accurate annotations for images and radar data are provided, supporting floating waste detection strategies based on image, radar data, and the fusion of two sensors. We perform several baseline experiments on our dataset, including vision-based and radar-based detection methods. The results show that, the detection accuracy is relatively low and floating waste detection still remains a challenging task.*

1. Introduction

In recent years, the marine wastes problem has arisen increasing concern of researchers and the general public. The wastes, especially the plastics inundating into the oceans are severely threatening the marine lives and causing sustained pollution to the whole ecosystem [24] (as illustrated in Figure 1(a)). However, according to [50], only 15% of the marine wastes are floating over the ocean surface, while 15% remain in the water column and 70% subside on the seafloor. While there are many projects launched to clean the floating marine wastes, it is essential to prevent wastes from getting into the ocean. Land-based sources, as opposed to marine-based sources, are considered the dominant input of wastes into oceans [20]. Thus, cleaning floating wastes in inland water areas, such as canals, rivers, lakes, and bays (as illustrated in Figure 1(b)), is a vital approach to slow the growing trend of marine wastes and water pollution.

(a) Marine Wastes (b) Inland Water Wastes

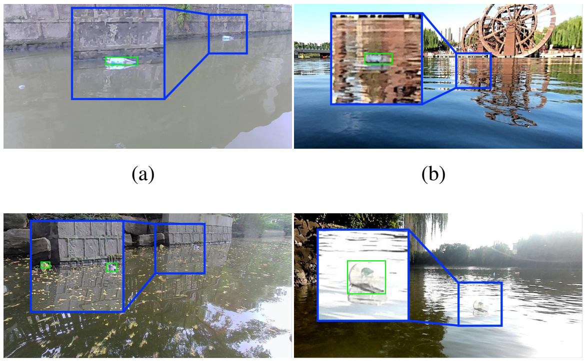
(c) Cleaning Manually (d) Cleaning Autonomously

Figure 1. Severe water pollution and cleaning methods.

Traditional strategies to cope with the floating wastes in inland waters mainly rely on manpower as shown in Figure 1(c). However, manual cleaning is low-efficient and costly when there are a large amount of floating wastes or the polluted water area is pretty broad [48]. What's more, the potential risks like the toxic gas and liquid from the drainage systems [49] and accidental drowning, limit the scenes of manual cleaning. Therefore, efficient autonomous floating waste cleaning methods are in great and urgent demand.

Recently, unmanned surface vehicles (USVs) have attracted a lot of attention for their usage in various missions. In particular, USVs have shown great potential in environmental applications, especially the floating waste cleaning [1, 48] (as shown in Figure 1(d)) since USVs can enter some hard-to-reach areas and operate independently. [17] has done a survey on river plastic pollution capture devices and listed 40 of them for comparison. Compared to some stationary devices, a good cleaning USV can be considered

*Equal contribution.
†FloW dataset as well as the development tools are publicly available at: https://github.com/ORCA-Uboat/FloW-Dataset.



(a) (b)

(c) (d)

Figure 2. (a) The floating waste occupies small image area. (b) The reflection of other objects makes it difficult to identify the bottle. (c) Harmless floating materials like fallen leaves disturb the visual system. (d) The reflection of strong light makes the plastic cup nearly invisible.

as the all-inclusive device that can guide (via booms and targeted driving), remove, contain, and transport the debris on water surfaces [17].

To achieve efficient and reliable autonomous cleaning of USVs, a real-time and accurate floating waste detection system is indispensable. With the development of computer vision and deep learning, the data of appearance about the objects provided by the visual system can be better utilized. Therefore, floating wastes detection relying on visual information can be the most cost-effective solution.

However, in inland waters, floating wastes like bottles and drink cans are of small size. Most of the floating wastes occupy small image areas, which increases the difficulties of accurate detection. For object detection based on deep neural network, compared to large objects, small objects usually lack appearance information and the high-level features of small objects are not discriminative, making it hard to distinguish them from the background and achieve accurate localization [2, 22]. The accuracy for small object detection is much lower than medium and large object detection.

Besides the small size, the complex real-world scenes of inland waters also cause challenges for the vision-based detection methods. For example, the reflections of objects on the bank-side as well as other harmless floating materials would disturb the visual detection system. The strong light reflection on the water surface may lead to overexposed images and make the waste invisible. Examples of the aforementioned problems are shown in Figure 2. Therefore, vision-based floating waste detection remains an unresolved problem for practical applications.

Complementary sensors can increase the adaptability of autonomous driving system. With the flourishing millimeter wave circuit technology and advanced signal processing techniques, 77 GHz millimeter wave radar has been widely used in perception system of autonomous driving for detection of objects and obstacles [40]. Compared to camera, radar is more robust to weather and lighting condition, and has the ability to reveal the target in a further range. In addition, radar data informs not only the location but also the Doppler velocity of the target [38]. For floating waste detection, radar provides a complement to image data.

It can be seen that public datasets can spur relevant studies and enable comparability for different strategies. For vision-based object detection, advances have been achieved through exploration of the famous dataset like COCO [28] and PASCAL VOC [11]. To draw attention to the floating waste cleaning using USVs in inland waters and help researches on small object detection over water surfaces, we introduce FloW, the first dataset for floating waste detection from USVs' view in inland waters which consists of an image sub-dataset FloW-Image (FloW-Img) and a multimodal sub-dataset FloW-Radar-Img (FloW-RI). The FloW-Img contains 2,000 images with 5,271 labeled floating wastes. Small objects account for more than half of the labeled objects. Beside the annotated images, we also provide 200 independent video sequences consisting of more than 20,000 image frames without annotations to support researches on floating wastes tracking on water surfaces. In addition, one highlight of our dataset is the FloW-RI, which contains 4,000 frames of synchronized and calibrated images and millimeter wave radar data of floating wastes. The FloW dataset is collected under diverse illumination and wave conditions, containing wastes of various appearances at different ranges and view angles. We evaluate some well-used algorithms on our dataset to illustrate the challenges of floating waste detection over water surface.

This work mainly contributes in following aspects:

- To our knowledge, FloW is the first dataset for floating wastes detection collected from the view of USVs in real-world inland waters under various conditions. The dataset can attract attention to floating waste problems in inland waters and help the development and comparability of inland waters floating waste detection algorithms.
- More than half of the labeled floating wastes in the image sub-dataset FloW-Img are regarded as small objects, which can spur studies on small objects detection over water surface.
- The sub-dataset FloW-RI contains continuous sequences of synchronized image and millimeter wave radar data with accurate annotations, supporting researches on floating waste detection based on radar data and fusion of image and radar data.
- We benchmark some well-used object detection al-

gorithms on our dataset to illustrate the problem in vision-based floating waste detection in inland waters.

2. Related Work

2.1. Vision-based Object Detection on Water Surface

Large progress has been made in the field of object detection in images, especially algorithms based on deep learning, which are usually categorized into two types: the region proposed based methods like Faster R-CNN [47], FPN [26] and the regression/classification based methods like YOLO [45], SSD [29].

For object detection over water surfaces, methods based on traditional image segmentation and saliency are firstly applied to this task [31, 6, 19]. These methods may not be robust to different environmental or lighting conditions, and different targets (i.e. lands or vessels) are indistinguishable to the model [7]. Researchers have explored using deep learning methods for object detection on water surfaces, especially the maritime object detection. Newly proposed models like Faster R-CNN [47], Mask R-CNN [16] etc. are applied to maritime object detection directly or with modification to increase the model's real-time performance and detection accuracy [34, 57, 56]. However, the differences between the marine and inland water environments and target sizes pose new challenges for our floating waste detection tasks. Some works focus on detecting the floating wastes using unmanned aerial vehicles or surveillance camera [36, 52]. Different from their works which utilize top-view images for waste monitoring, our works help to support floating waste detection for cleaning USVs.

2.2. Object Detection based on Millimeter Wave Radar

In the past few years, 77 GHz millimeter wave radar is widely used in autonomous driving. Object detection based on radar becomes an important solution for robust perception system of autonomous vehicles. For inland USVs, millimeter wave radar serves as a novel sensor for object detection.

The constant false alarm rate (CFAR) detector and the its variants such as the cell averaging (CA)-CFAR and ordered statistic (OS)-CFAR are regarded as the classical radar detector for millimeter wave radar [40]. For classical radar object detection, the detector mainly concentrates on detecting the presence or absence of targets rather than generating the semantic information about the targets (e.g., clutters, floating bottles or stones).

In recent year, deep-learning based object detection using millimeter wave radar has attracted increasing attention. These methods can be used for specific target detection, such as pedestrians and vehicles. There are two main presentation formats of radar data used in these works. One is the final output of the classic radar signal processing pipeline: the radar point clouds [9, 33, 35]. The other one is the medium data generated in the radar signal processing pipeline [39, 25, 30], such as the Rang-Dopple matrix. Range-Azimuth matrix, and radar cube. Danzer *et al.* [9] present an approach for 2D car detection solely using sparse radar point clouds based on the PointNet architecture. In [30], Major *et al.* propose a deep learning based vehicle detection method that operates on radar Range-Azimuth-Doppler tensors. Nabati and Qi [35] present a region proposal network for object detection in autonomous vehicles. Lim *et al.* [25] use radar and camera early fusion for vehicle detection. Researches on specific target detection using millimeter wave radar main aim at applications in autonomous driving. Using the radar for small object detection on water surfaces still remains an unexplored task.

2.3. Related Dataset

Researchers from the computer vision communities have paid efforts to waste detection to deal with the environmental issues. Datasets for vision-based waste detection have been published. [55] present the TrashNet dataset containing classes of common trashes like glass, paper, etc. for classifying garbage into recycling. Wang *et al.* [53] present their dataset UAV-BD for the real-world waste bottle detection on the ground in unmanned aerial vehicles (UAV) images. The recently published TACO dataset [42] contains multi-class annotated photos of wastes in the wild. [54] build an indoor RGBD waste dataset called MJU-Waste and propose a deep learning based method for waste segmentation. Beside, the Deep-sea Debris Database [12] collects images and images of trash in realistic underwater environments. Based on these datasets, works have been done for vision-based waste detection [14, 10, 18]. However, as shown in Figure 3, being valuable for waste classification or monitoring though, the application scenes of these datasets differ a lot from the inland water surfaces environments. To our knowledge, FloW is the first dataset addressing the floating waste detection in the complex scene of real-world inland waters. Collected from the view of USVs, works upon FloW dataset can directly support the cleaning USVs.

As for object detection on water surfaces, the Singapore Marine Dataset (SMD) [41] and the benchmark built on it [34] have spurred works for maritime object detection for USVs [21, 44]. The MODD2 [3] provides stereo videos for marine USVs obstacle detection. Besides, [15] present a dataset including categories of riverside, vessel, etc. for USVs' safe navigation in inland waters. These datasets mainly contain targets like ships and vessels, which are relatively big and not easily affected by the environmental influences.

In addition to the relevant visual datasets, driven by the

Figure 3. FloW dataset in comparison with other dataset. Our dataset is the first one aimed at the floating waste detection from the view of USVs in the complex real-world inland waters environment. (Some pictures are scaled or cut for alignment.)

increasing trends in applications of millimeter wave radar, datasets for object detection based on radar have also been published, which are mainly built for autonomous driving. Astyx has published a dataset containing annotated high-resolution 3D automotive radar point clouds for deep learning based object detection [32]. The nuScenes dataset [4] includes radar data collected on public roads with multi-class annotations. The CARRADA dataset presented in [38] is a dataset of synchronized camera and radar recordings with range-angle-Doppler annotations for object detection, semantic segmentation, and object tracking. While the millimeter wave radar is becoming popular in autonomous driving for object detection, using it for waste detection in inland waters is a novel and meaningful attempt.

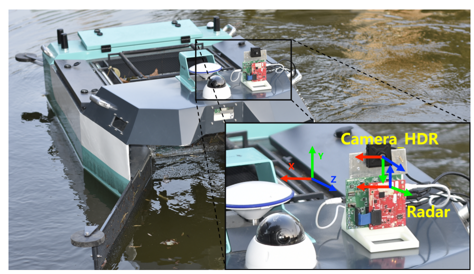


Figure 4. Our acquisition platform is equipped with two cameras and a millimeter wave radar. The directions of the sensors used in collecting the FloW-RI are marked in different color (X-red, Y-green, Z-blue).

3. Dataset

Detail information about the FloW dataset will be discussed in this section.

3.1. Acquisition Platform

We make use of our cleaning USV equipped with cameras and millimeter wave radar as the collection platform (shown in Figure 4) to collect images of floating bottles in the real-world inland waters thereby meeting the requirements of autonomous cleaning using USVs.

As can be seen, two cameras are equipped in front of the vehicle facing forward recording the cleaning USVs' driving scene in inland waters. Considering that in the real-world scenes of inland waters, there are dynamic ranges which may extend further than the normal camera can perceive, one camera is set in high dynamic range (HDR) model to balance the lighting for clearer images. The HDR camera (with AR0230 chip) captures images with resolution of 1280×720 while the other camera provides images of resolution 1280×640 pixels. Two cameras are enabled under different environments for the variety of image data and provide independent sets of images to avoid overlapping.

The radar used in our work is a Texas Instruments 77 GHz frequency modulate continuous wave (FMCW) radar AWR1843 with an array of 3 transmitting antennas and 4 uniformly spaced receiving antennas. The radar is also equipped on the vehicle facing forward, in order to collect complementary information of the floating wastes. Detail information about the radar and relevant parameters setting will be discussed in the next part.

3.2. Data Collection and Processing

When collecting the dataset, the variety of collection scenes, lighting conditions, number of objects per frame, the appearance of objects, view angles, and target ranges are considered to increase the diversity of samples.

During a period of 3 months, our USV drove both manually and automatically in different real-world inland waters, to record video sequences of floating wastes under various lighting conditions. The floating wastes in our dataset mainly include plastic bottles and drink cans with different appearances. In each water area, we set a route for the vehicle to repeatedly sail across the surface. As the vehicle get closer to the object from different directions, the variety of object distribution, ranges, and view angles can be ensured. The vehicle is also driven manually around the floating wastes to increase the diversity of wastes distribution. The collection processes of FloW-Img and FloW-RI are independent and there is no overlap between the two sub-datasets.

FloW-Img Dataset. With the collected raw video sequences (with a sample frequency of 10 Hz), the data for the image dataset are generated by firstly extracting every 30th image from part of the raw video sequences. Then images of high similarity and some blurred images are excluded manually. The video sequences in the FloW-Img dataset are cut from other raw sequences. There is no overlap between the annotated images and the video sequences.

FloW-RI Dataset. For the multimodal dataset collection, the wave parameters of FMCW radar and the signal processing procedure for radar raw data are chosen to provide effective information of the floating wastes in inland waters.

For FMCW radar, a frame is composed of N_{chirp} linear frequency-modulated chirps, whose frequency varies linearly with time between the minimum frequency f_{min} and the maximum frequency f_{max}. Thus, each chirp can be characterized by the duration T_{chirp} and the bandwidth $B = f_{\text{max}} - f_{\text{min}}$. Considering a uniform linear antenna array, the echo signal at the lth receiver for the target located at (R, θ) can be expressed as follows:

$$d(l,n,p) \approx \exp\left\{ j \cdot 2\pi \left[\left(\frac{2 \cdot R \cdot B}{c \cdot T_{\text{chirp}}} + f_d \right) \frac{n}{f_s} + \frac{f_{min} \cdot l \cdot d \cdot \sin\theta}{c} + f_d \cdot p \cdot T_{\text{chirp}} + \frac{2 \cdot f_{min} \cdot R}{c} \right] \right\} + \omega(l,n,p), \tag{1}$$

where n is the index of fast-time samples within each chirp, p is the index of slow-time samples across different chirps, c is the speed of light, f_d denotes the Doppler shift, f_s denotes the sampling rate, d denotes the spacing between adjacent antenna elements, and ω is noise. As the performance of radar detection is affected by the wave parameters, given the requirement for floating wastes detection, the parameters are designed as listed in Table. 1.

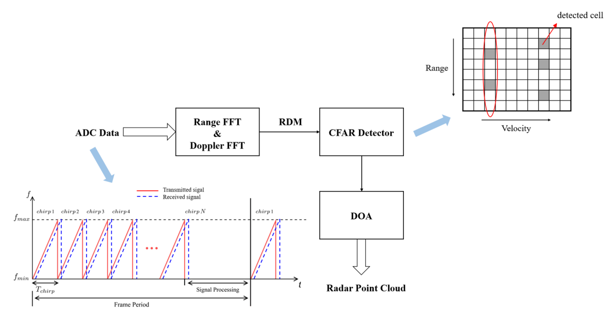


Figure 5. Conventional FMCW radar signal processing chain.

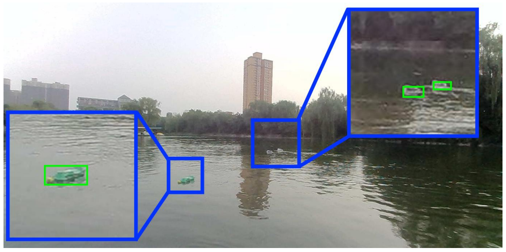
Table 1. Radar waveform parameters and performance data.

Parameter	Value	Performance	Value
f_{min}	77 GHz	Range Resolution	0.0349m
f_{max}	80.76 GHz	Maximum Range	14.5m
T_{chirp}	133.33us	Velocity Resolution	0.03m/s
N_{chirp}	128	Maximum Velocity	4.10m/s

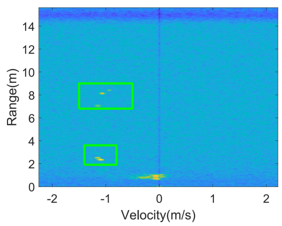
The raw data sampled by Analog-to-Digital (ADC) can only be used after the signal processing procedures. Figure 5 shows the conventional FMCW radar signal processing chain. The sampled echoes are first transferred to the range-Doppler matrix (RDM) via 1D (range) FFT and 2D (Doppler) FFT [38]. Then, the signals with stronger energy in the RDM are detected as targets. Finally, in the direction of arrival (DOA) processing block, the azimuth and elevation to each target is estimated by digital beamforming (DBF).

In the FloW-RI, we provide the data with preliminary signal processing, that is RDM. To increase the spatial resolution of radar, we use the time division multiplexing-multiple input multiple output (TDM-MIMO) technique. Hence, our radar system can reach the equivalent of eight virtual receiving antennas for azimuth estimation and four virtual antennas for pitch estimation. Therefore, for each frame of radar detection, we provide RDM of 12 virtual receiving antennas. The RDM is a complex matrix and each cell consists of a real component and an imaginary component. Besides, to make the radar data more intuitive, we provide 4D radar point clouds data which are generated by constant false alarm rate (CFAR) detection and DBF based on the 12 RDM of each frame.

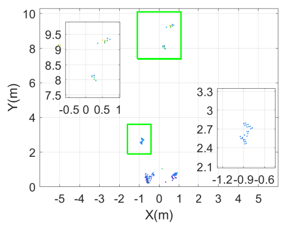
In the FloW-RI, the millimeter wave radar data and the images are synchronized by recording the timestamp of each frame. The sample frequencies of both sensors are 10 Hz. As the speed of the USVs in inland waters is relatively low (no more than 2 m/s), the synchronization accuracy



(a) Image of floating wastes.



(b) Radar RDM.



(c) Radar point clouds (top view).

Figure 6. In this frame, there are three floating bottles visible in the image, radar RDM, and radar point clouds. Green boxes frame the three bottles in all three data formats. Point clouds of bottles are zoomed in and are shown aside.

can meet the requirement for real-time floating waste detection in inland waters. The extrinsic parameters between the radar and camera are provided in the dataset. With the temporal and spatial aligned radar and image data, the fusion of two sensors is ensured. A frame of data including image, radar RDM, and radar point clouds is shown in Figure 6.

3.3. Annotation

FloW-Img Annotation. The images in FloW-Img are labeled by using the famous tool LabelImg [51] to draw a bounding box covering the floating wastes in each frame with a single-label. The annotation procedure is time-consuming as the size of floating wastes in a far distance is pretty small, making it hard to find the boundary of the object. In this case, we hire some experienced annotators in order to ensure efficiency as well as the quality of annotation. Before labeling, the following annotation rules are unified to avoid ambiguity: All floating wastes captured in different directions should be labeled. The reflection of the bottles is not included in the labeled area nor regarded as an independent object. All of the annotations are repeatedly verified and corrected to ensure the annotation quality.

FloW-RI Annotation. The annotation for the multi-modal dataset is based on the images and radar RDM. It is hard to annotate the RDM directly as the RDM does not directly reflect the distribution of targets as the image does. Moreover, the ground truth Doppler velocity of the target is hard to get. In this case, as illustrated in Figure 7, following steps are taken for annotation in RDM. (1) First,

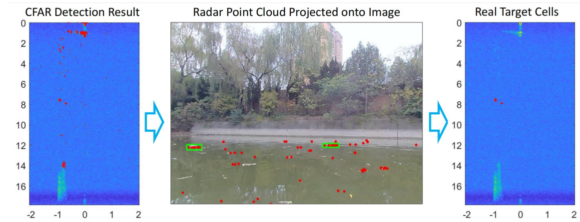


Figure 7. The annotation processes. In the first image, the red points on the RDM image show the cells detected by the CFAR detector with a low threshold. The second image shows the corresponding points projected on the RGB image, and the real target points are annotated (framed by the green boxes). Last image shows the final real target cells in red.

we set a relatively low detection threshold for RDM and use DOA estimation to generate 4D radar point clouds. (2) Then, with the extrinsic calibration result between camera and radar, we project the point clouds in one frame onto the corresponding image. For the projection of radar point clouds, we refer to projection method used in [8]. (3) With all the point clouds projected onto the image, the real target points are manually annotated referring to the image. The points projected on the bottles are considered as real target points. (4) Finally, as there is a one-to-one matching between the radar points and cells in RDM, the labels for radar point clouds are transformed back to annotations for cells in RDM.

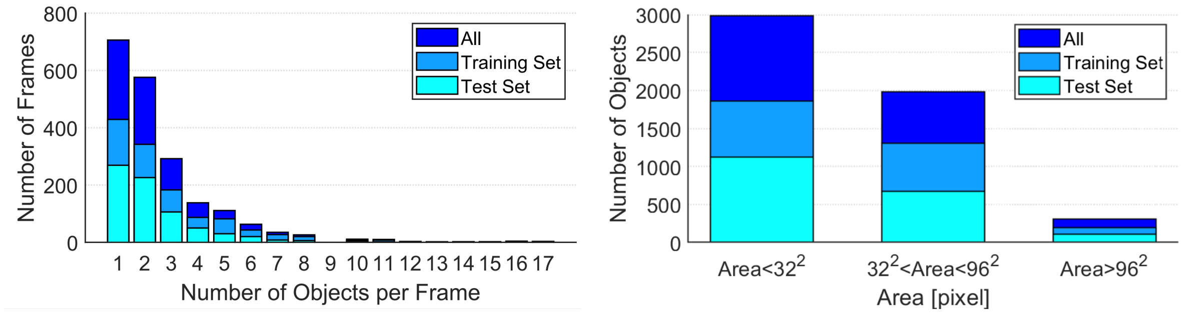
3.4. Dataset Split and Statistics

FloW-Img. The image dataset contains 2,000 images with 5,271 labeled floating wastes. We follow a 6:4 split for the training and test set respectively. We randomly select 1,200 images as the training set and the rest as the test set.

The distributions of the number of objects in one frame and the occupied area of labeled objects for the whole dataset, the training set, and the test set are shown in Figure 8. As can be seen, the distributions of different sizes of targets in training and test data are approximately matching. Small objects (area $< 32 \times 32$) account for the largest proportion in our dataset, which makes the detection more challenging.

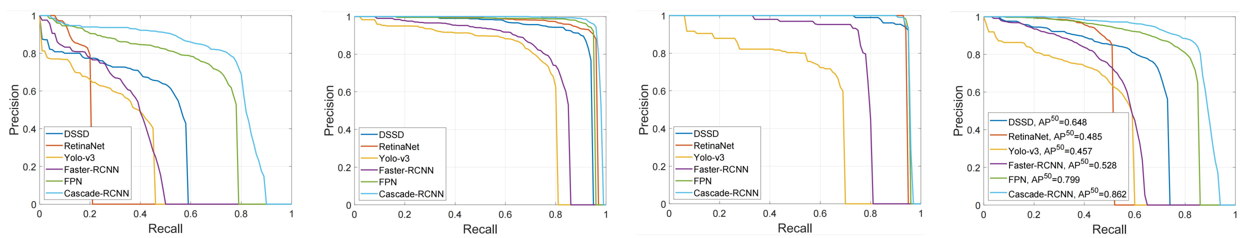
Besides, we also provide 200 video sequences without annotations in the FloW-Img dataset. The lasting time of the sequences varies from several seconds to one minute and the number of targets contained in each sequence varies from 1 to 8. The video sequences can be used to support researches on floating wastes tracking.

FloW-RI. The FloW-RI contains 4,000 frames of synchronized images and radar RDM data, as well as the corresponding radar point clouds of each frame for intuitiveness. All the frames are from 21 continuous sequences without sampling to encourage further studies on our dataset, e.g.



(a) (b)

Figure 8. The dataset split and data distribution of the FloW-Img. (a) The distribution of object number in one frame. (b) The distribution of occupied area per object.



(a) Small (Area $< 32^2$) (b) Medium ($32^2 <$ Area $< 96^2$) (c) Large (Area $> 96^2$) (d) All

Figure 9. PR (precision-recall) curves of different object detection algorithms on the test set of FloW-Img (IoU threshold = 0.5). AP^{50} denotes the average precision with the IoU threshold at 0.50.

Table 2. Detection result on test set of FloW-Img using different object detection algorithms.

	DSSD [13]	RetinaNet [27]	YOLO-v3 [46]	Faster R-CNN [47]	FPN [26]	Cascade R-CNN [5]
$mmAP$	0.275	0.249	0.128	0.184	0.334	0.434
FPS	28.6	7.6	23.2	9.3	7.4	3.9

* $mmAP$ denotes the average precision under IoU=.50:.05:.95.

the object tracking. For the FloW-RI, 16 sequences are selected as the training set and the other 5 sequences are selected as the test set.

4. Experiment and Evaluation

4.1. Experiment on FloW-Img

Vision-based Object Detection Baseline. We evaluate the performance of 6 well-used algorithms for deep-learning based object detection on our dataset, including one-stage algorithms DSSD [13], RetinaNet [27], YOLO-v3 [46], and two-stage algorithms Faster R-CNN [47], FPN [26], as well as multi-stage Cascade R-CNN [5]. In addition to the detection accuracy, as the real-time performance is important to autonomous cleaning for USVs, we also recorded the frames per second (FPS) for each algorithm. The experiments are done on the same machine with GeForce GTX 1070 GPU. The precision-recall curves of the results of the above-mentioned algorithms are shown in Figure 9. The average detection accuracy and the FPS for each algorithm are shown in Table 2.

Table 3. The comparison of $mmAP$ on small objects of UAV-BD and FloW-Img dataset.

Dataset	Cascade R-CNN	FPN	DSSD
UAV-BD [53]	0.491	0.481	0.527
FloW-Img	0.293	0.213	0.121

We also carry out experiments on relevant dataset to make a comparison and to illustrate the distinctions of our dataset. We use three algorithms to test on the UAV-BD [53] dataset. The comparison of $mmAP$ on small objects of UAV-BD and our dataset is shown in Table 3.

As can be seen, using the same detection algorithms, the detection accuracy on small objects in our dataset is

Table 4. Result on FloW-Img test set. Images from different datasets are used to train the modal.

Dataset	$mmAP$
UAV-BD [53]	0.058
TACO [42]	0.069
MJU-Waste [54]	0.032

much lower than on UAV-BD. We find that, in addition to the small size, there are three other challenges in floating waste detection on our dataset. First, the strong light reflection on water surfaces would shade targets. Also, the water waves can also obscure targets. These two reasons can cause miss detection. Besides, reflections of objects on bank cause false alarm as current method can not distinguish real targets and bank object reflections. The above reasons cause challenges for current vision-based method. On our dataset, among the above algorithms, the Cascade R-CNN performs best on detection accuracy on our dataset, being more robust to the size of the objects. However, the FPS of the Cascade R-CNN is relatively bad and may not meet the real-time requirements of floating waste detection for USVs.

To verify the importance of our dataset for the task of floating waste detection, we also use other relevant datasets to train the modal and test the modal on the data of real-world floating bottles. The experiment is based on Cascade R-CNN (Resnet101 as backbone). We use the UAV-BD dataset, TACO dataset and MJU-Waste dataset (only the images contain bottles or drink cans are selected for our use) respectively to re-train the modal and test on the FloW-Img test set. As shown in Table 4, the detection accuracy on the real-world floating bottles is pretty low.

4.2. Experiment on FloW-RI

For the multimodal dataset FloW-RI, we carry out experiments on image and radar data respectively, as well as on the fusion of image and radar data.

For vision-based detection, we use the Cascade R-CNN (Resnet101 as backbone, re-trained on the training set of FloW-RI) to detect floating wastes in images of FloW-RI test set.

For radar-based object detection, we evaluate the detection accuracy on both the radar RDM images and radar point clouds. For detection in radar RDM images, we also use the Cascade R-CNN (Resnet101 as backbone) like we use for detection in images. For detection in radar point clouds, we carry out the experiment using the VoteNet [43]. The VoteNet is an end-to-end 3D object detection network based on a synergy of deep point set networks and Hough voting.

For the image-radar fusion based object detection, we test two algorithms proposed in [37] and [23].

Table 5. Detection result on FloW-RI.

Data Format	Method	AP^{50}
RGB Image	Cascade R-CNN [5]	0.640
Radar RDM Image	Cascade R-CNN [5]	0.898
Radar Point Cloud	VoteNet [43]	0.405
Point Cloud + Image	CRF-Net [37]	0.754
Point Cloud + Image	Li *et al.* [23]	0.677

* The evaluation for radar point cloud are done in 2-D spaces.

The results on the test set are shown in Table 5. As can be seen, when using the same detection algorithm, the detection performance on radar RDM is better than on images. Radar data show good potential to be used for floating waste detection in inland waters. In addition, it can also be seen that, the fusion based method out-performs the vision based methods in some degree. However, these two fusion-based methods and most of other proposed methods mainly aim at object detection in driving scenes on roads. We think that, for small object detection on water surfaces, there are still some improvements worth to be done.

5. Conclusion

In this paper, we introduce FloW, the first dataset for floating waste detection in inland waters. The FloW consists of two sub-datasets, FloW-Img and FloW-RI. The FloW-Img contains 2,000 annotated images and 200 video sequences without annotation. The FloW-RI is a multimodal dataset which contains 4,000 frames of synchronized images and millimeter wave radar data. We evaluate the performance of some classical vision-based object detection algorithms and radar detectors on our dataset. The experiment results show that for vision-based object detection, the robustness of the model needs to be improved to meet the requirement of practical applications. By publishing the dataset, we hope that the accuracy of floating wastes detection could be improved and more attention could be paid to wastes cleaning in the water areas. In the future, we will collect other kinds of floating wastes for more categories in some polluted waters for public use, to stimulate the practical applications of autonomous cleaning in inland waters.

References

[1] Abir Akib, Faiza Tasnim, Disha Biswas, Maeesha Binte Hashem, Kristi Rahman, Arnab Bhattacharjee, and Shaikh Anowarul Fattah. Unmanned floating waste collecting robot. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, pages 2645–2650. IEEE, 2019. 1

[2] Yancheng Bai, Yongqiang Zhang, Mingli Ding, and Bernard Ghanem. Sod-mtgan: Small object detection via multi-task generative adversarial network. In *Proceedings of the Euro-*

pean Conference on Computer Vision (ECCV), pages 206–221, 2018. 2

[3] Borja Bovcon, Janez Perš, Matej Kristan, et al. Stereo obstacle detection for unmanned surface vehicles by imu-assisted semantic segmentation. *Robotics and Autonomous Systems*, 104:1–13, 2018. 3

[4] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11621–11631, 2020. 4

[5] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6154–6162, 2018. 7, 8

[6] Tom Cane and James Ferryman. Saliency-based detection for maritime object tracking. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 18–25, 2016. 3

[7] Tom Cane and James Ferryman. Evaluating deep semantic segmentation networks for object detection in maritime surveillance. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE, 2018. 3

[8] Simon Chadwick, Will Maddern, and Paul Newman. Distant vehicle detection using radar and vision. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8311–8317. IEEE, 2019. 6

[9] Andreas Danzer, Thomas Griebel, Martin Bach, and Klaus Dietmayer. 2d car detection in radar data with pointnets. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 61–66. IEEE, 2019. 3

[10] Sukmawati Nur Endah, Ilman Nabil Shiddiq, et al. Xception architecture transfer learning for garbage classification. In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, pages 1–4. IEEE, 2020. 3

[11] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 2

[12] Japan Agency for Marine Earth Science and Technology. Deep-sea debris database. http://www.godac.jamstec.go.jp/catalog/dsdebris/e/index.html. 3

[13] Cheng-Yang Fu, Wei Liu, Ananth Ranga, Ambrish Tyagi, and Alexander C Berg. Dssd: Deconvolutional single shot detector. *arXiv preprint arXiv:1701.06659*, 2017. 7

[14] Michael Fulton, Jungseok Hong, Md Jahidul Islam, and Junaed Sattar. Robotic detection of marine litter using deep visual detection models. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 5752–5758. IEEE, 2019. 3

[15] Wided Hammedi, Metzli Ramirez-Martinez, Philippe Brunet, Sidi-Mohamed Senouci, and Mohamed Ayoub Messous. Deep learning-based real-time object detection in inland navigation. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2019. 3

[16] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 3

[17] Olivia K Helinski, Cara J Poor, and Jordyn M Wolfand. Ridding our rivers of plastic: A framework for plastic pollution capture device selection. *Marine Pollution Bulletin*, 165:112095, 2021. 1, 2

[18] Jungseok Hong, Michael Fulton, and Junaed Sattar. Trashcan: A semantically-segmented dataset towards visual detection of marine debris. *arXiv preprint arXiv:2007.08097*, 2020. 3

[19] Xianli Jin, Peiwen Niu, and Linfeng Liu. A gmm-based segmentation method for the detection of water surface floats. *IEEE Access*, 7:119018–119025, 2019. 3

[20] Laurent CM Lebreton, Joost Van Der Zwet, Jan-Willem Damsteeg, Boyan Slat, Anthony Andrady, and Julia Reisser. River plastic emissions to the world's oceans. *Nature communications*, 8:15611, 2017. 1

[21] Sung-Jun Lee, Myung-Il Roh, Hye-Won Lee, Ji-Sang Ha, Il-Guk Woo, et al. Image-based ship detection and classification for unmanned surface vehicle using real-time object detection neural networks. In *The 28th International Ocean and Polar Engineering Conference*. International Society of Offshore and Polar Engineers, 2018. 3

[22] Jianan Li, Xiaodan Liang, Yunchao Wei, Tingfa Xu, Jiashi Feng, and Shuicheng Yan. Perceptual generative adversarial networks for small object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1222–1230, 2017. 2

[23] Liang-qun Li and Yuan-liang Xie. A feature pyramid fusion detection algorithm based on radar and camera sensor. In *2020 15th IEEE International Conference on Signal Processing (ICSP)*, volume 1, pages 366–370. IEEE, 2020. 8

[24] Wai Chin LI, HF Tse, and Lincoln Fok. Plastic waste in the marine environment: A review of sources, occurrence and effects. *Science of the Total Environment*, 566:333–349, 2016. 1

[25] Teck-Yian Lim, Amin Ansari, Bence Major, Daniel Fontijne, Michael Hamilton, Radhika Gowaikar, and Sundar Subramanian. Radar and camera early fusion for vehicle detection in advanced driver assistance systems. 3

[26] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 3, 7

[27] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 7

[28] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 2

[29] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C

Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016. 3

[30] Bence Major, Daniel Fontijne, Amin Ansari, Ravi Teja Sukhavasi, Radhika Gowaikar, Michael Hamilton, Sean Lee, Slawomir Grzechnik, and Sundar Subramanian. Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019. 3

[31] Jorge S Marques, Alexandre Bernardino, Gonçalo Cruz, and Maria Bento. An algorithm for the detection of vessels in aerial images. In *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 295–300. IEEE, 2014. 3

[32] Michael Meyer and Georg Kuschk. Automotive radar dataset for deep learning based 3d object detection. In *2019 16th European Radar Conference (EuRAD)*, pages 129–132. IEEE, 2019. 4

[33] Michael Meyer and Georg Kuschk. Deep learning based 3d object detection for automotive radar and camera. In *2019 16th European Radar Conference (EuRAD)*, pages 133–136. IEEE, 2019. 3

[34] Sebastian Moosbauer, Daniel Konig, Jens Jakel, and Michael Teutsch. A benchmark for deep learning based object detection in maritime environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019. 3

[35] Ramin Nabati and Hairong Qi. Rrpn: Radar region proposal network for object detection in autonomous vehicles. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 3093–3097. IEEE, 2019. 3

[36] Guanchong Niu, Jie Li, Sheng Guo, Man-On Pun, Leo Hou, and Lujian Yang. Superdock: a deep learning-based automated floating trash monitoring system. In *2019 IEEE international conference on robotics and biomimetics (ROBIO)*, pages 1035–1040. IEEE, 2019. 3

[37] Felix Nobis, Maximilian Geisslinger, Markus Weber, Johannes Betz, and Markus Lienkamp. A deep learning-based radar and camera sensor fusion architecture for object detection. In *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–7. IEEE, 2019. 8

[38] Arthur Ouaknine, Alasdair Newson, Julien Rebut, Florence Tupin, and Patrick Pérez. Carrada dataset: Camera and automotive radar with range-angle-doppler annotations. *arXiv preprint arXiv:2005.01456*, 2020. 2, 4, 5

[39] Andras Palffy, Jiaao Dong, Julian FP Kooij, and Dariu M Gavrila. Cnn based road user detection using the 3d radar cube. *IEEE Robotics and Automation Letters*, 5(2):1263–1270, 2020. 3

[40] Sujeet Milind Patole, Murat Torlak, Dan Wang, and Murtaza Ali. Automotive radars: A review of signal processing techniques. *IEEE Signal Processing Magazine*, 34(2):22–35, 2017. 2, 3

[41] Dilip K Prasad, Deepu Rajan, Lily Rachmawati, Eshan Rajabally, and Chai Quek. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: a survey. *IEEE Transactions on Intelligent Transportation Systems*, 18(8):1993–2016, 2017. 3

[42] Pedro F Proença and Pedro Simões. Taco: Trash annotations in context for litter detection. *arXiv preprint arXiv:2003.06975*, 2020. 3, 8

[43] Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep hough voting for 3d object detection in point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9277–9286, 2019. 8

[44] Dalei Qiao, Guangzhong Liu, Jun Zhang, Qiangyong Zhang, Gongxing Wu, and Feng Dong. M3c: Multimodel-and-multicue-based tracking by detection of surrounding vessels in maritime environment for usv. *Electronics*, 8(7):723, 2019. 3

[45] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 3

[46] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 7

[47] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015. 3, 7

[48] Niramon Ruangpayoongsak, Jakkrit Sumroengrit, and Monthian Leanglum. A floating waste scooper robot on water surface. In *2017 17th International Conference on Control, Automation and Systems (ICCAS)*, pages 1543–1548. IEEE, 2017. 1

[49] Bashir Yahaya Sanda and Isyaku Ibrahim. Causes, categories and control of water pollution. 1

[50] Kyriacos Themistocleous, Christiana Papoutsa, Silas Michaelides, and Diofantos Hadjimitsis. Investigating detection of floating plastic litter from space using sentinel-2 imagery. *Remote Sensing*, 12(16):2648, 2020. 1

[51] Tzutalin. Labelimg. Free Software: MIT License, 2015. https://github.com/tzutalin/labelImg. 6

[52] Colin van Lieshout, Kees van Oeveren, Tim van Emmerik, and Eric Postma. Automated river plastic monitoring using deep learning and cameras. *Earth and space science*, 7(8):e2019EA000960, 2020. 3

[53] Jinwang Wang, Wei Guo, Ting Pan, Huai Yu, Lin Duan, and Wen Yang. Bottle detection in the wild using low-altitude unmanned aerial vehicles. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 439–444. IEEE, 2018. 3, 7, 8

[54] Tao Wang, Yuanzheng Cai, Lingyu Liang, and Dongyi Ye. A multi-level approach to waste object segmentation. *Sensors*, 20(14):3816, 2020. 3, 8

[55] Mindy Yang and Gary Thung. Classification of trash for recyclability status. *CS229 Project Report*, 2016, 2016. 3

[56] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11):3212–3232, 2019. 3

[57] Zhiguo Zhou, Siyu Yu, and Kaiyuan Liu. A real-time algorithm for visual detection of high-speed unmanned surface vehicle based on deep learning. In *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, pages 1–5. IEEE, 2019. 3