Trash Detection in Surface Waters Using YOLOv8

2024 11th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE) | 979-8-3503-8930-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITACEE62763.2024.10762773

Muhammad Rafly Arjasubrata
School of Computing
Telkom University
Bandung, Indonesia
raflyarj@student.telkomuniversity.ac.id

Akmal Muzakki Bakir
School of Computing
Telkom University
Bandung, Indonesia
akmalmzkki@student.telkomuniversity.ac.id

Gloria Natasya Irene S.
School of Computing
Telkom University
Bandung, Indonesia
glorianatasya@student.telkomuniversity.ac.id

Sabrina Adinda Sari
School of Computing
Telkom University
Bandung, Indonesia
sabrinaadindasaridua@student.telkomuniversity.ac.id

Mahmud Dwi Sulistiyo
School of Computing
Telkom University
Bandung, Indonesia
mahmuddwis@telkomuniversity.ac.id

Abstract—Aquatic environments are increasingly polluted by various types of trash, such as bottles, plastics, and cans, which accumulate on water surfaces and pose severe threats to marine habitats. The increasing pollution necessitates proactive monitoring of waterways in coastal cities to detect and mitigate the presence of trash before it reaches the ocean. However, effective detection of floating trash remains challenging due to variations in trash types, sizes, and viewing angles. This study aims to develop and evaluate YOLOv8 models for detecting floating trash in surface waters from multiple perspectives to monitor pollution levels and quantify trash. We trained and assessed YOLOv8 architectures for single-class trash detection using the FloW-Img and WaterTrash datasets, capturing both USV and aerial perspectives to ensure comprehensive detection from various angles. The YOLOv8x model achieved the highest mAP50 scores, with 0.923 on the WaterTrash dataset and 0.834 on the FloW-Img dataset. Despite differences in model size, the performance gap was marginal, demonstrating YOLOv8's robustness. These results highlight the potential for further enhancements in detecting tiny and partially submerged trash, as well as distinguishing trash from reflections and waves.

Keywords—YOLOv8, Trash Detection, Object Detection, Surface Water Pollution

I. Introduction

Scattered trash poses a significant challenge, as improper disposal exacerbates pollution in communities near water bodies. The amount of trash in waterways can easily be observed by the abundant presence of floating trash. This pollution severely impacts ecosystems by altering marine habitats, reducing light penetration, and decreasing oxygen levels [1]. One of the most concerning aspects is the high concentration of plastic waste. High concentrations of plastic obstruct the respiratory passages and stomachs of numerous marine species [2]. For instance, in 2010, 215 million metric tons of plastic trash were generated, with 4.8 to 12.7 million metric tons entering the ocean from 192 coastal countries [3]. This cumulative ocean plastic is projected to reach 8.78 million kg per year by 2050 [4]. Microplastic debris (<1 mm), constituting about 80% of all stranded plastic, poses long-term threats to the biosphere, potentially releasing harmful monomers linked to cancer and reproductive abnormalities in humans, rodents, and invertebrates [5]. Therefore, public awareness and collective action in trash management are crucial to preventing these negative impacts.

Previously, assessing trash pollution in aquatic environments relied on manual visits, which lacked the efficiency and scalability offered by automated detection methods. This traditional approach, while providing some level of monitoring, was fraught with challenges including labor-intensive processes, which significantly limited the ability to perform widespread monitoring. While monitoring cameras have been utilized to some extent, they are often inadequate as they fail to detect and quantify pollution levels automatically. This manual approach is labor-intensive and prone to human error, limiting the ability to perform continuous and widespread monitoring. Research has been conducted on classifying trash types in controlled environments [6], [7]. Trash detection on roadways has also been carried out successfully, aiding in assessing and mapping cleanliness along surveyed roads [8]. However, detecting trash in waterways has faced challenges in identifying small-sized debris, like microplastics, deformed and half-submerged trash [9].

To address these challenges, a YOLO-based architecture known for its fast identification with relatively high accuracy [10]−[13], particularly the YOLOv8, is chosen. The YOLOv8 architecture employs both Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) to enhance the fusion of multi-scale features, thereby improving detection accuracy and robustness, especially for small objects [14] − [18]. The trained YOLOv8 models are used for detecting floating trash in surface waters from multiple perspectives to monitor pollution levels and quantify the amount of trash from different angles relative to the water surface.

The YOLOv8 architecture was assessed on two datasets. The first is the FloW-Img dataset [19], which comprises annotated images of domestic trash floating on inland water captured from the perspective of an Unmanned Surface Vehicle (USV). The second dataset is the WaterTrash dataset [9], which in cludes images of trash floating in urban water channels captured from a top-down perspective. Both datasets were tested independently for each model size to gauge the performance of detecting floating trash on the water surface from both USV and a top-down perspectives, ensuring the model's versatility in detecting from various perspective relative to the water surface, as seen in Fig. 1.

(a) (b)

Fig. 1. (a) Prediction from YOLOv8x from USV perspective. (b) Prediction from YOLOv8x from a top-down perspective

The subsequent sections of the paper are structured as follows: Section II reviews related works in trash classification and detection. Section III outlines the methodology used in this study. Section IV details the experimental setup, including the dataset, environment, training hyperparameters, and the metrics used for evaluation, and presents the experimental results and findings. Finally, Section V concludes the paper, summarizing key findings and suggesting potential directions for future research.

II. Related Work

Significant advancements have been made in automated trash classification within controlled environments using the TrashNet dataset [7]. Vo et al. [20] achieved 95% accuracy on TrashNet with their DNN-TC method based on RexNet [21]. Similarly, Masand et al. [22] proposed ScrapNet, which used transfer learning from EfficientNet B3 [23] and also achieved 95% accuracy. These findings highlight the effectiveness of deep learning for developing and optimizing trash detection models

Trash detection in public areas, such as roadways and parks, has seen successful implementation. Das et al. [24] developed a model for trash detection in outdoor scenarios, achieving mAP_{50} scores of 60.4% on the TACO dataset [25] and 88.4% on the PlastOpol dataset [26], using various YOLOv5 scales trained on datasets from Bangladesh streets and Openlittermap [27]. AquaVision by Panwar et al. [28] detects and classifies pollutants in oceans, trained on TrashNet [7], TACO [25], and AquaTrash datasets, based on the Retina-Net architecture [29]. These studies demonstrate the effectiveness of trash detection methods across diverse environments using a range of deep learning and convolutional models and datasets.

Tharani et al. [30] developed a model for detecting visible trash floating on urban water surfaces using YOLOv3 [31], augmented with an attention layer to enhance small trash detection. Despite these efforts, the model exhibited limitations in detecting small-sized trash, achieving AP scores of 48.1% and 31.2% on easy and hard datasets, respectively. Li et al. [10] proposed a method to improve small target detection in river floating trash by integrating a refined FPN and enhancing coordinate attention within the YOLOv5 architecture. The modified YOLOv5 model achieved a 15.7% accuracy improvement and an 83% reduction in error detection rate. YOLO-Float [32], a modification of YOLOv5 incorporating a low-level representation enhancement module (LREM) and an attentional-fusion module (AFM), achieved an mAP_{50} of 83.3% on the FloW-Img test dataset. Notably, this method primarily focuses on enhancing accuracy rather than detection speed. S. Xu et al. [11] proposed a method by incorporating the Space-to-Depth-Conv-Se (SPDCS) module and introducing the Spatial Pyramid Pooling Augmentation (SPPAUG) module, an enhancement of the Spatial Pyramid Pooling Fast (SPPF) module, applied to YOLOv5 called YOLOW. The improved model achieves an mAP_{50} of 82.1% on the FloW-Img test dataset.

The research above describes various efforts related to trash detection in surface water. However, no research has specifically trained and tested models for detecting floating trash in surface water from both USV and top-down perspectives.

III. Method

YOLOv8, released by Ultralytics in January 2023, is a state-of-the-art model for object detection with three main components: backbone, neck, and head. The backbone, based on a modified CSPDarknet53 [33], extracts features using 53 convolutional layers and cross-stage partial connections. The neck uses FPN [14] and PAN [17] to merge multi-scale feature maps, which enhances the model's ability to detect small objects by balancing accuracy and speed. The head includes convolutional and fully connected layers with a decoupled design, optimizing classification and regression tasks independently. YOLOv8 uses an anchor-free detection method predicting object centers. The flow and interaction of the YOLOv8 architecture are shown in Fig. 2.

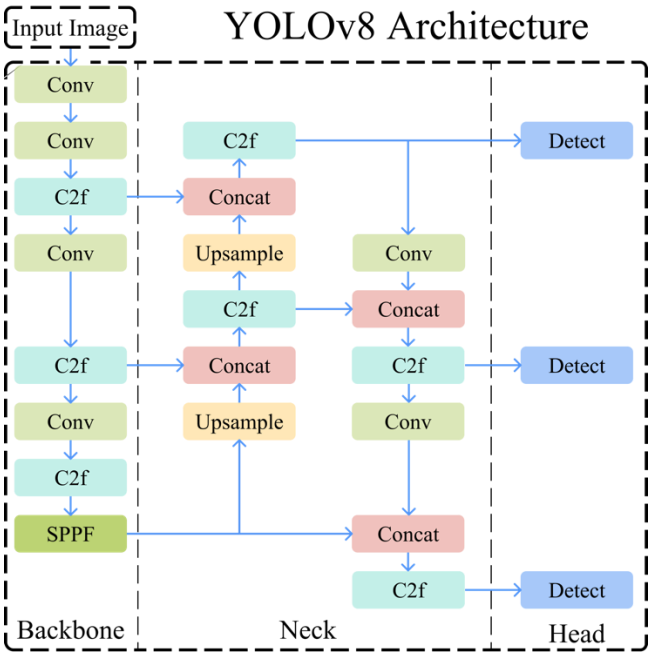

Fig. 2. Architecture block of YOLOv8 [34]

YOLOv8 utilizes both sigmoid and SoftMax functions for object detection and classification, with sigmoid predicting the object's presence in each bounding box and SoftMax predicting the object's class. These predictions are combined to detect and classify objects in images. YOLOv8 employs various loss metrics, including classification loss using Binary Cross-Entropy, bounding box loss using Distributed-Focus Loss (DFL) [35], and Complete Intersection over Union (CIoU) for efficient localization and size prediction. The architecture includes five scaled versions: nano (**n**), small (**s**), medium (**m**), large (**l**), and extra-large (**x**), which vary in the number of layers and the number of neurons. Larger networks achieve higher accuracy but require more computational resources and are prone to overfitting, as shown by the positive relationship between inference time,

model size, and AP on the MSCOCOval2017 dataset [36] presented in TABLE I.

TABLE I. PERFORMANCE METRICS OF DIFFERENT YOLOV8 ARCHITECTURES.

Model	MAP 50-95	Params (M)	FLOPs (B)	Speed (MS) A100 TensorRT
YOLOv8n	37.3	3.2	8.7	0.99
YOLOv8m	44.9	11.2	28.6	1.20
YOLOv8s	50.2	25.9	78.9	1.83
YOLOv8l	52.9	43.7	165.2	2.39
YOLOv8x	53.9	68.2	257.8	3.53

IV. EXPERIMENT AND RESULT

A. Dataset

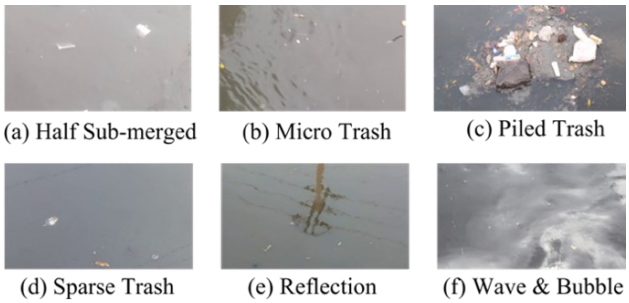

Fig. 3. Example of images displaying a diverse range of difficult situations found in waterways

Two labeled datasets are used as benchmark for detecting trash in surface water. The WaterTrash dataset [9] includes images from drainage canals collected at various times, presenting challenges due to differences in water hue and transparency. The FloW-Img dataset [19] consist of images of floating trash observed from a USV in inland water. Both datasets serve as distinct testing sets to evaluate model performance across different perspectives and camera angles. The perspective difference is shown in Fig. 4. To further improve model performance and generalization, additional datasets are utilized. The first one is UAVVaste dataset [37] contains 772 images with 3,716 hand-labeled trash items gathered using a UAV in urban and nature environments, aiding the model in recognizing small-scale trash. The second dataset is TACO [25], sourced through crowd-sourcing, offers a diverse range of trash captured from a pedestrian point of view, enhancing the model's ability to recognize various shapes and patterns of deformed domestic trash.

Fig. 4. (a) Image from FloW-Img dataset [19] taken from USV. (b) Image from WaterTrash dataset [9] presenting an aerial perspective

Each dataset is split into an 80:20 ratio for train and validation sets. The FloW-Img and WaterTrash datasets have predefined test sets on which the model will be assessed independently. For the train data of WaterTrash and FloW-Img, the train sets of each dataset are shuffled and further divided into an 80:20 ratio to create separate train and validation sets. Similarly, the remaining datasets are shuffled and divided into 80:20 ratios for train and validation splits. Instance counts for each dataset are provided in TABLE II.

TABLE II. NUMBER OF INSTANCES IN EACH SET FOR EACH DATASET.

Dataset Name	Instance		
	Train	*Validation*	*Test*
WaterTrash [9]	28489	7122	10193
FloW-IMG [19]	2598	650	2023
UAVVaste [37]	2415	638	-
TACO [25]	3824	961	-

B. Experiment

The experiment was conducted using an NVIDIA SXM A100 equipped with 40GB of VRAM and utilized CUDA version 12.3.0. Ultralytics version 8.0.217 and PyTorch version 2.1.1 were employed. The Python environment for this experimental setup operated on version 3.10.6.

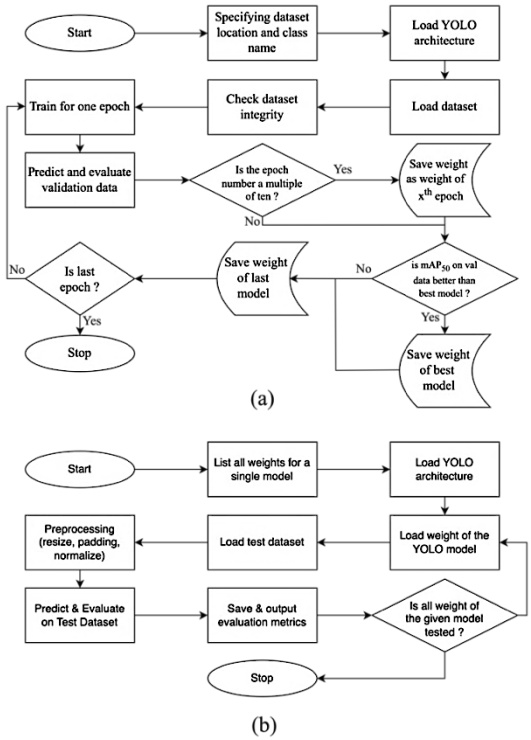

Fig. 5. (a) Flowchart of the training phase. (b) Flowchart of the testing phase.

The experiment consists of two phases: training and testing. Each YOLOv8 architecture size is trained independently with the same hyperparameters. In the training phase, depicted in Fig. 5a, the class is designated as "Trash" and dataset paths are documented in the YAML file. Validation data is assessed at each epoch with weights saved every ten epochs to track the training progress and prevent overfitting. The best model weights are based on the highest mAP_{50} score. In the testing phase, the YOLOv8 architecture

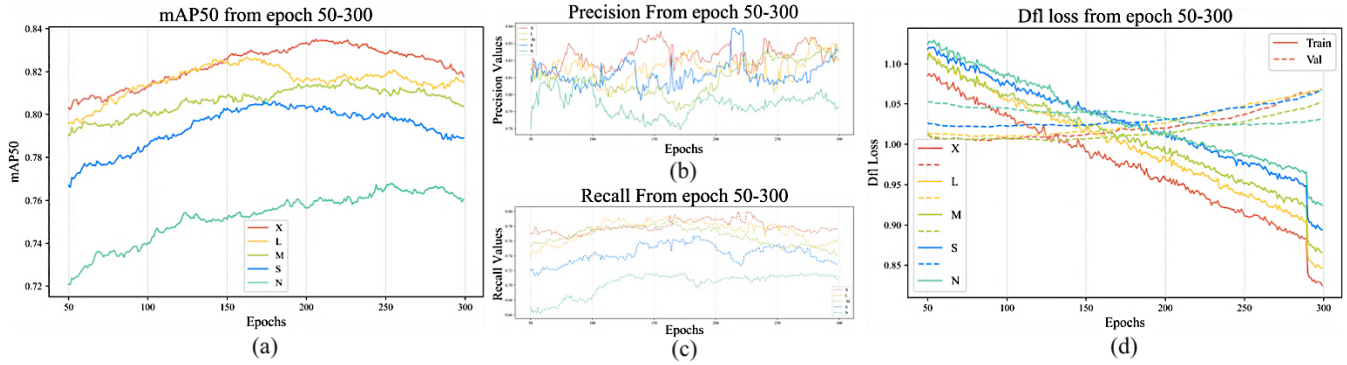

Fig. 6. Evaluation metric trends during training process for all YOLOv8 variants from epoch 50 to 300. (a) Trends in mAP_{50} scores on the validation dataset. (b) Trends in precision scores on the validation dataset. (c) Trends in recall scores on the validation. (d) Dfl Loss trends on train and validation data

is instantiated with specific model weights, and the test dataset undergoes preprocessing. Predictions and evaluations are conducted on each FloW-Img and WaterTrash dataset with results summarizing evaluation metrics. The described testing phase is depicted in Fig. 5b. Each YOLOv8 model is trained with an input size of 640×640, a batch size of 48 for 300 epochs, using the Stochastic Gradient Descent (SGD) optimizer with a momentum of 0.937 and a decay of 0.0005. The IoU threshold for valid bounding box prediction is set to 0.7 with a maximum detection of 300 objects per frame. mAP_{50} is used as the main evaluation metric for its comprehensive measure of localization accuracy with an IoU threshold of 0.5 to detect trash regardless of its exact location. Testing is conducted on saved weights every ten epochs as well as on the best and final weights to assess potential overfitting.

C. Results and Analysis

All trained YOLOv8 models show a positive correlation between model size and mAP_{50} scores on both validation and test datasets, indicating that larger models tend to perform better, as depicted in Fig. 6a and Fig. 7. However, in the FloW-Img test results, each model peaks in mAP_{50} score before 300 epochs and then declines, suggesting potential overfitting, this can be seen in Fig. 7a. Fluctuations in precision as depicted in Fig. 6b indicate instability in identifying true positives, while recall consistently improves, as shown in the graphs in Fig. 6c, albeit with occasional misidentifications, such as shown in Fig. 8. Overfitting is further supported by the DFL Loss graph shown in Fig. 6d, which flattens before epoch 150 and then rises, indicating difficulty in generalizing to unseen data. This underscores the model's overfitting tendencies and limited generalization.

TABLE III. ERFORMANCE COMPARISON ON FLOW-IMG TEST DATASET ON THE BEST MODEL.

Model	mAP_{50}	Precision	Recall	Epoch
YOLOv8x	**0.834**	0.833	0.784	210
YOLOv8l	0.826	0.819	**0.791**	150
YOLOv8m	0.815	0.811	0.771	210
YOLOv8s	0.804	0.815	0.756	180
YOLOv8n	0.766	0.794	0.713	250
YOLO-Float [32]	0.833	-	-	-
STE-YOLO [38]	0.832	-	-	-
YOLOW [11]	0.821	**0.870**	0.751	-
YOLOv7-CA [13]	0.811	-	-	-

However, Fig. 7b reveals an intriguing trend in the mAP_{50} score of the models on the WaterTrash test dataset, where the score continues to increase until late epochs. The peak mAP_{50} scores for the X, L, and S variants of the YOLOv8 models are observed at 270 epochs, while the M and N counterparts reach their peaks at 280 and 300 epochs, respectively. This prolonged increase in mAP_{50} scores suggests that the models are learning to detect trash instances more effectively over time, with a particular focus on the characteristics present within the WaterTrash dataset. This implies that the models are becoming increasingly adept at identifying patterns specific to the WaterTrash dataset, potentially at the expense of generalizability. Consequently, the overfitting observed during the training process may be attributed to the model's tendency to learn and find patterns of trash instances from the WaterTrash dataset, potentially neglecting other crucial aspects of trash detection. The continued improvement indicates the model's ability to adapt, and the YOLOv8x reaches the highest mAP_{50} scores on the FloW-Img test dataset as depicted in TABLE III.

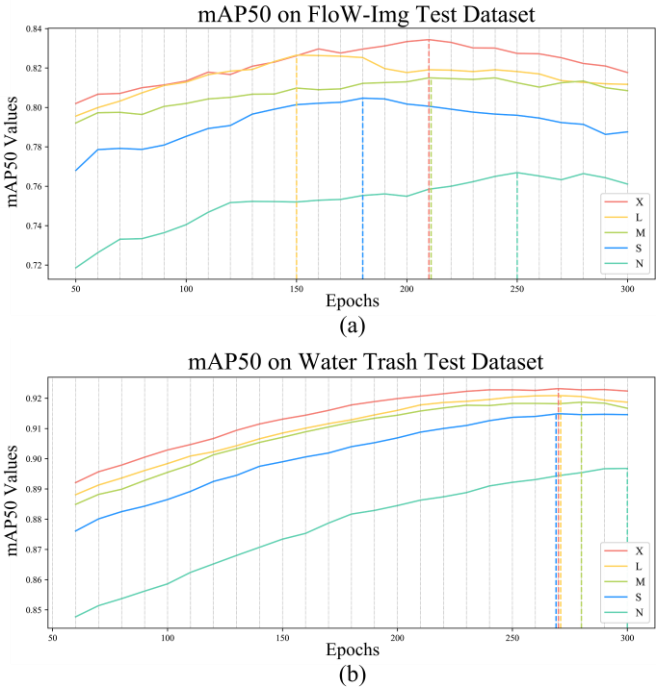

Fig. 7. The mAP_{50} scores at every 10-epoch interval from epoch 50 to 300. (a) mAP_{50} scores on the FloW-Img test dataset. (b) mAP_{50} scores on the WaterTrash test dataset

The WaterTrash test results in TABLE IV highlight a smaller performance gap between the YOLOv8x and the YOLOv8n compared to the FloW-Img dataset. Specifically, the mAP_{50} score difference between the **x** and **n** variants is only 2.90%, with precision and recall differences at 2.64% and 5.60%, respectively. Notably, the results on the WaterTrash dataset show strong performance across all models, with the mAP_{50} exceeding 0.900 and all models achieving recall above 0.800 except for the n variant. This suggests robust detection capabilities, maintaining high accuracy despite variations in model size. The consistently high precision rates also indicate that the models are not only effectively identifying floating trash but also minimizing false positives, which is crucial for practical applications. The marginal performance differences among the models demonstrate that even the smaller models, such as YOLOv8n, can be highly effective in specific scenarios, providing a cost-effective and resource-efficient solution. Furthermore, the ability of these models to sustain high detection performance across both the FloW-Img and WaterTrash datasets underscores their versatility and reliability in diverse environmental settings. This comprehensive performance ensures that the models are well-suited for real-world applications, facilitating effective monitoring and assessment of water pollution.

TABLE IV. PERFORMANCE COMPARISON ON WATERTRASH TEST DATASET ON THE BEST MODEL.

Model	mAP_{50}	Precision	Recall	Epoch
Yolov8x	**0.923**	**0.910**	**0.868**	270
Yolov8l	0.920	0.907	0.860	270
Yolov8m	0.918	0.899	0.863	280
Yolov8s	0.914	0.889	0.851	270
Yolov8n	0.896	0.883	0.819	300

(a) (b)

Fig. 8. Sample of false positive predictions highlighted in orange attributed to reflections and waves on water bodies. (a) The prediction results from the YOLOv8x model weight at epoch 210. (b) Corresponding image with ground truth bounding boxes

TABLE V compares the performance of trained YOLOv8 variants at epochs 210 and 270, revealing minimal differences in mAP_{50} scores, with the most significant difference observed in YOLOv8x. Interestingly, YOLOv8n shows an improvement from epoch 210 to 270, suggesting that smaller models may require more iterations for optimal performance. Furthermore, the results indicate that while larger models like YOLOv8x and YOLOv8l achieve high performance early on, smaller models can catch up with sufficient training. TABLE VI highlights the WaterTrash dataset's better performance improvement in later epochs compared to the FloW-Img dataset. This indicates that the trained models can maintain their performance on the FloW-Img dataset while significantly improving on the WaterTrash dataset, achieving a balanced performance across datasets. These findings underscore the resilience and adaptability of the YOLOv8 models, making them highly suitable for real-world applications. The ability of these models to improve over time with continued training suggests potential benefits for long-term deployment in environmental monitoring systems. The model's ability to retain high detection performance across both datasets demonstrates their potential to handle diverse and challenging environments effectively, ensuring reliable monitoring and assessment of water polution.

TABLE V. PERFORMANCE COMPARISON ON TWO EPOCH POINTS ON FLOW-IMG TEST DATASET.

Model	Epoch					
	210			*270*		
	mAP_{50}	***Precision***	***Recall***	***mAP_{50}***	***Precision***	***Recall***
Yolov8x	**0.834**	**0.833**	**0.784**	0.825	0.831	0.772
Yolov8l	0.819	0.816	0.783	0.813	0.816	0.765
Yolov8m	0.815	0.811	0.711	0.812	0.820	0.757
Yolov8s	0.800	0.819	0.741	0.792	0.810	0.743
Yolov8n	0.756	0.794	0.706	0.763	0.794	0.714

TABLE VI. PERFORMANCE COMPARISON ON TWO EPOCH POINTS ON WATERTRASH TEST DATASET.

Model	Epoch					
	210			*270*		
	mAP_{50}	***Precision***	***Recall***	***mAP_{50}***	***Precision***	***Recall***
Yolov8x	0.920	0.889	0.863	**0.932**	**0.910**	**0.868**
Yolov8l	0.917	0.899	0.857	0.920	0.907	0.860
Yolov8m	0.915	0.893	0.855	0.918	0.897	0.861
Yolov8s	0.908	0.895	0.831	0.914	0.889	0.851
Yolov8n	0.886	0.878	0.798	0.894	0.878	0.816

V. CONCLUSION

The YOLOv8 models demonstrate exceptional proficiency in detecting floating trash on surface waters from both USV and areial perspectives, with the YOLOv8x achieving state-of-the-art mAP_{50} scores of 0.834 on the FloW-Img test dataset and 0.932 on the WaterTrash test dataset, with minimal performance variation among different model sizes. This indicates the outstanding accuracy, adaptability, and scalability of the YOLOv8 architecture. However, challenges remain in detecting tiny and -partially submerged trash and in distinguishing trash from reflections and waves. These limitations highlight the need for further research to enhance detection capabilities, especially in complex and variable water conditions. Therefore, future work should focus on refining detection algorithms and incorporating advanced filtering techniques to improve accuracy in these challenging scenarios

ACKOWLEDGEMENT

We extend our sincere gratitude to the Artificial Intelligence Laboratory and the Directorate of Research and Community Service of Telkom University, for technical and financial supports, respectively.

REFERENCES

[1] K. D. Ganvir, P. R. Nerkar, L. W. Ghate, and H. H. Bhagat, "The impact of water pollution and preliminary study on river trash collecting mechanism," *International Journal of Technical Research and Applications*, vol. 7, no. 1, pp. 85–87, 2019.

[2] M. F. Mukhtar *et al.*, "DEVELOPMENT of RIVER TRASH COLLECTOR SYSTEM," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1742-6596/1529/4/042029.

[3] J. R. Jambeck *et al.*, "Plastic waste inputs from land into the ocean," *Science (1979)*, vol. 347, no. 6223, pp. 768–771, Feb. 2015, doi: 10.1126/science.1260352.

[4] D. M.-C. Chen, B. L. Bodirsky, T. Krueger, A. Mishra, and A. Popp, "The world's growing municipal solid waste: trends and impacts," *Environmental Research Letters*, vol. 15, no. 7, p. 074021, Jul. 2020, doi: 10.1088/1748-9326/ab8659.

[5] M. A. Browne, T. Galloway, and R. Thompson, "Microplastic—an emerging contaminant of potential concern?," *Integr Environ Assess Manag*, vol. 3, no. 4, pp. 559–561, Oct. 2007, doi: 10.1002/ieam.5630030412.

[6] G. E. Sakr, M. Mokbel, A. Darwich, M. N. Khneisser, and A. Hadi, "Comparing deep learning and support vector machines for autonomous waste sorting," in *2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*, IEEE, Nov. 2016, pp. 207–212. doi: 10.1109/IMCET.2016.7777453.

[7] M. Yang and G. Thung, "Classification of trash for recyclability status," *CS229 project report*, vol. 2016, no. 1, p. 3, 2016.

[8] M. S. Rad *et al.*, "A Computer Vision System to Localize and Classify Wastes on the Streets," 2017, pp. 195–204. doi: 10.1007/978-3-319-68345-4_18.

[9] M. Tharani, A. W. Amin, F. Rasool, M. Maaz, M. Taj, and A. Muhammad, "Trash Detection on Water Channels," 2021, pp. 379–389. doi: 10.1007/978-3-030-92185-9_31.

[10] H. Li *et al.*, "Detection of Floating Objects on Water Surface Using YOLOv5s in an Edge Computing Environment," *Water (Basel)*, vol. 16, no. 1, p. 86, Dec. 2023, doi: 10.3390/w16010086.

[11] S. Xu, H. Tang, J. Li, L. Wang, X. Zhang, and H. Gao, "A YOLOW Algorithm of Water-Crossing Object Detection," *Applied Sciences*, vol. 13, no. 15, p. 8890, Aug. 2023, doi: 10.3390/app13158890.

[12] S. S. Firdauz, E. Rachmawati, and M. D. Sulistiyo, "Mask Detection on Motorcyclists Using YOLOv4," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 4, Mar. 2023, doi: 10.47065/bits.v4i4.2980.

[13] K. Li, Y. Wang, and Z. Hu, "Improved YOLOv7 for Small Object Detection Algorithm Based on Attention and Dynamic Convolution," *Applied Sciences*, vol. 13, no. 16, p. 9316, Aug. 2023, doi: 10.3390/app13169316.

[14] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 936–944. doi: 10.1109/CVPR.2017.106.

[15] J. Dang, X. Tang, and S. Li, "HA-FPN: Hierarchical Attention Feature Pyramid Network for Object Detection," *Sensors*, vol. 23, no. 9, p. 4508, May 2023, doi: 10.3390/s23094508.

[16] Q. Yang, C. Zhang, H. Wang, Q. He, and L. Huo, "SV-FPN: Small Object Feature Enhancement and Variance-Guided RoI Fusion for Feature Pyramid Networks," *Electronics (Basel)*, vol. 11, no. 13, p. 2028, Jun. 2022, doi: 10.3390/electronics11132028.

[17] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 8759–8768. doi: 10.1109/CVPR.2018.00913.

[18] L. Zhou, X. Rao, Y. Li, X. Zuo, B. Qiao, and Y. Lin, "A Lightweight Object Detection Method in Aerial Images Based on Dense Feature Fusion Path Aggregation Network," *ISPRS Int J Geoinf*, vol. 11, no. 3, p. 189, Mar. 2022, doi: 10.3390/ijgi11030189.

[19] Y. Cheng *et al.*, "FloW: A Dataset and Benchmark for Floating Waste Detection in Inland Waters," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2021, pp. 10933–10942. doi: 10.1109/ICCV48922.2021.01077.

[20] A. H. Vo, L. Hoang Son, M. T. Vo, and T. Le, "A Novel Framework for Trash Classification Using Deep Transfer Learning," *IEEE Access*, vol. 7, pp. 178631–178639, 2019, doi: 10.1109/ACCESS.2019.2959033.

[21] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017.

[22] A. Masand, S. Chauhan, M. Jangid, R. Kumar, and S. Roy, "ScrapNet: An Efficient Approach to Trash Classification," *IEEE Access*, vol. 9, pp. 130947–130958, 2021, doi: 10.1109/ACCESS.2021.3111230.

[23] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, K. Chaudhuri and R. Salakhutdinov, Eds., in Proceedings of Machine Learning Research, vol. 97. PMLR, Feb. 2019, pp. 6105–6114. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html

[24] D. Das, K. Deb, T. Sayeed, P. K. Dhar, and T. Shimamura, "Outdoor Trash Detection in Natural Environment Using a Deep Learning Model," *IEEE Access*, vol. 11, pp. 97549–97566, 2023, doi: 10.1109/ACCESS.2023.3313166.

[25] P. F. Proença and P. Simões, "TACO: Trash Annotations in Context for Litter Detection," Mar. 2020.

[26] M. Córdova *et al.*, "Litter Detection with Deep Learning: A Comparative Study," *Sensors*, vol. 22, no. 2, p. 548, Jan. 2022, doi: 10.3390/s22020548.

[27] S. Lynch, "OpenLitterMap.com – Open Data on Plastic Pollution with Blockchain Rewards (Littercoin)," *Open Geospatial Data, Software and Standards*, vol. 3, no. 1, p. 6, Dec. 2018, doi: 10.1186/s40965-018-0050-y.

[28] H. Panwar *et al.*, "AquaVision: Automating the detection of waste in water bodies using deep transfer learning," *Case Studies in Chemical and Environmental Engineering*, vol. 2, p. 100026, Sep. 2020, doi: 10.1016/j.cscee.2020.100026.

[29] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *IEEE Trans Pattern Anal Mach Intell*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[30] M. Tharani, A. W. Amin, M. Maaz, and M. Taj, "Attention Neural Network for Trash Detection on Water Channels," Jul. 2020.

[31] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018.

[32] Y. Li, R. Wang, D. Gao, and Z. Liu, "A Floating-Waste-Detection Method for Unmanned Surface Vehicle Based on Feature Fusion and Enhancement," *J Mar Sci Eng*, vol. 11, no. 12, p. 2234, Nov. 2023, doi: 10.3390/jmse11122234.

[33] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A New Backbone That Can Enhance Learning Capability of CNN," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jun. 2020.

[34] Ultralytics, "Home," *Ultralytics YOLO Docs*, Jul. 05, 2024. https://docs.ultralytics.com/

[35] X. Li *et al.*, "Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, in NIPS'20. Red Hook, NY, USA: Curran Associates Inc., 2020.

[36] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," May 2014.

[37] M. Kraft, M. Piechocki, B. Ptak, and K. Walas, "Autonomous, Onboard Vision-Based Trash and Litter Detection in Low Altitude Aerial Images Collected by an Unmanned Aerial Vehicle," *Remote Sens (Basel)*, vol. 13, no. 5, p. 965, Mar. 2021, doi: 10.3390/rs13050965.

[38] J. Yu, H. Zheng, L. Xie, L. Zhang, M. Yu, and J. Han, "Enhanced YOLOv7 integrated with small target enhancement for rapid detection of objects on water surfaces," *Front Neurorobot*, vol. 17, Dec. 2023, doi: 10.3389/fnbot.2023.1315251.