Deep Learning Approach to Real-Time Underwater Trash Detection with YOLOv10-Nano

Tomi Heri Julianus Todingan, Imanuel Kutika, Vicky Nolant Setyanto Lahimade, Alwin M. Sambul,
Oktavian A. Lantang, Muhamad Dwisnanto Putro
Master Program of Informatics, Postgraduate Program, Sam Ratulangi University, Manado, North Sulawesi, Indonesia
e-mails: tomitodingan026@student.unsrat.ac.id, imanuelkutika026@student.unsrat.ac.id,
vickylahimade026@student.unsrat.ac.id, asambul@unsrat.ac.id, oktavian_lantang@unsrat.ac.id,
dwisnantoputro@unsrat.ac.id

Received: 25 April 2025; revised: 27 May 2025; accepted: 17 June 2025

Abstract — **The accumulation of underwater trash presents a significant threat to marine environments and long-term sustainability, necessitating innovative technological approaches for effective monitoring and cleanup operations. Traditional detection methods often face challenges such as limited visibility and high costs. Convolutional Neural Networks (CNNs) in Deep learning offers an effective solution for real-time object detection in the challenging conditions of underwater settings. This research proposes the application of YOLOv10-Nano, a modern CNN-based architecture, tailored for the real-time identification of underwater trash. YOLOv10-Nano is chosen for its enhanced accuracy and low-latency performance compared to earlier YOLO models. It develops a strong balance between processing rate and detection precision, making it well-suited for deployment in actual scenarios. It is trained using the Trash-ICRA19 dataset, comprises with a wide range of underwater trash images, providing a strong foundation for underwater trash detection in real-time. Evaluation indicates that YOLOv10-Nano achieves 46.4% mAP50(mean Average Precision) and operates at a frame rate of 20.8 FPS (frames per second). These outcomes underline the effectiveness of YOLOv10-Nano as a lightweight yet capable tool for real-time underwater trash detection. It supports scalable and automated monitoring efforts, contributing to cleaner and more sustainable marine ecosystems.**

Key words— **Deep Learning; Underwater Trash; YOLOv10-Nano; CNN**

I. INTRODUCTION

The underwater trash has become a severe threat to marine ecosystems, biodiversity, and human livelihoods. The accumulation of trash on the ocean has escalated due to increased industrial activities, poor waste management, and urban runoff. These pollutants not only degrade the physical habitat but also introduce toxic substances that can be ingested by marine organisms. Marine pollution contributes to entanglement, habitat derivation, and digestion of waste by marine species, and the disruption of oceanic food chains [1], [2].

Efficient monitoring and detection of underwater trash are essential components in supporting global marine conservation and cleanup efforts [3]. However, conventional methods relying on manual visual inspection by divers or remotely operated vehicles (ROVs) are costly, time-consuming, and ineffective in handling large-scale monitoring [4].

Recent advances in deep learning, particularly computer vision, have opened various possibilities for automated environmental monitoring in underwater environments. CNN has demonstrated remarkable results in object recognition and image classification tasks, including underwater imagery with complex backgrounds and varying lighting conditions [5], [6], [7].

Among various approaches, the YOLO (You Only Look Once) series has made notable advancements by maintaining a strong compromise between accuracy in detection and processing speed [8], [9].

Numerous studies have tackled the challenge of underwater waste detection through the deep learning techniques. For instance, B. Xue et al. [10] developed an underwater object detection framework based on YOLOv3, trained on a combination of synthetic and real underwater images. In a similar vein, R. Jain et al. [11] explored the use of YOLOv8, EfficientDet-D0, Mask R-CNN, and YOLACT for recognizing submerged marine litter, achieving commendable detection performance, although their real-time efficiency remained limited. Other researchers [12], [13], [14] have introduced specialized CNN architectures tailored for underwater conditions. However, many of these models compromise either on inference speed or generalization capability. Despite ongoing efforts, the majority of current solutions either lack the necessary speed for real-time deployment or fail to maintain robustness under diverse underwater environments.

The recent release of YOLOv10 introduces architectural optimizations that improve both speed and accuracy. YOLOv10 integrates lightweight modules and advanced feature fusion strategies, which suitable for high-end performance applications in constrained environments [15], [16]. Compared to its predecessors, YOLOv10 outperforms in real-time inference while maintaining competitive accuracy, which is essential for on-site underwater trash detection via ROVs or AUVs (Autonomous Underwater Vehicles) [17].

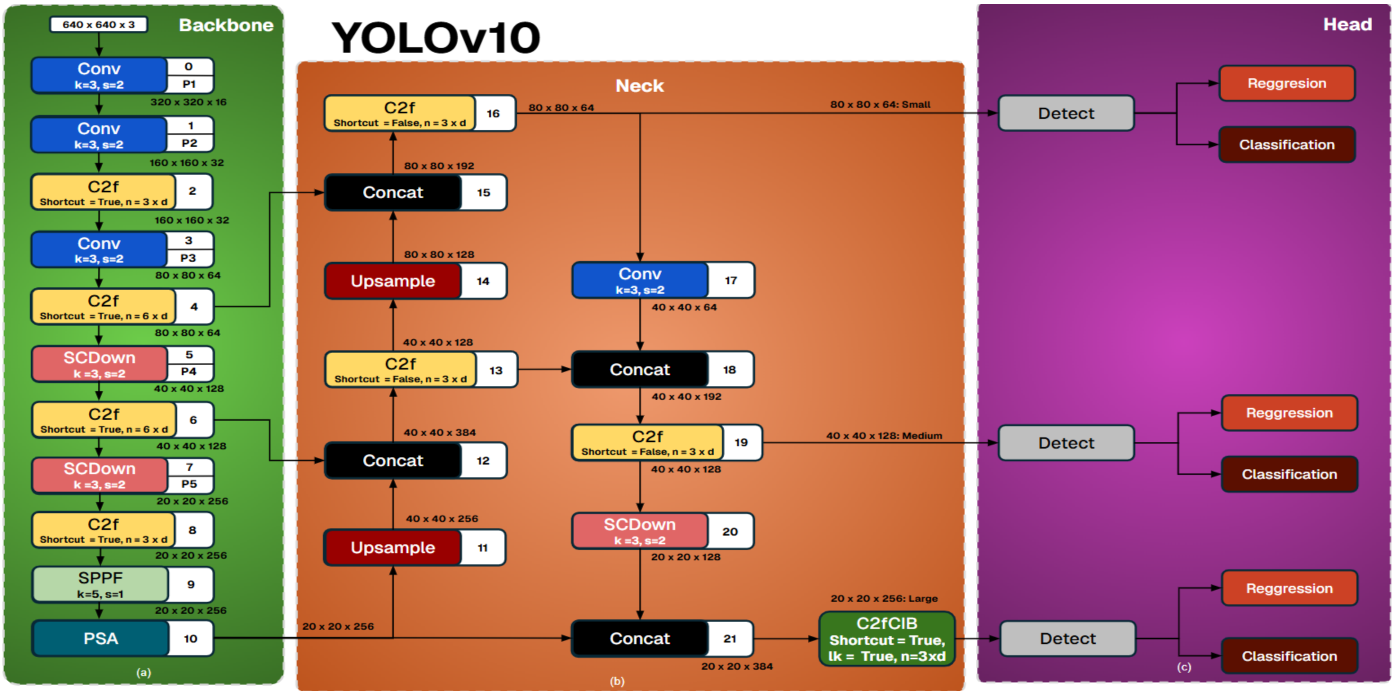

Figure 1. YOLOv10 Architectur

To the best of our knowledge, no prior studies have evaluated the performance of YOLOv10 for underwater trash detection, especially using a domain-specific dataset such as Trash-ICRA19.

This research aims to fill that gap by implementing and evaluating YOLOv10 for real-time underwater trash detection using the Trash-ICRA19 dataset. The Trash-ICRA19 dataset offers diverse samples of annotated underwater trash under realistic conditions, making it an ideal benchmark for evaluating detection models [18].

The novelty of this work lies in

1. Applying YOLOv10 on the underwater trash detection domain.
2. Evaluating its real-time performance on a benchmark dataset.
3. Analyzing the potential of YOLOv10 as a scalable, accurate, and low-latency solution for environmental monitoring.

This study aids the development of practical and deployable systems for marine debris detection and can support broader efforts in automated ocean cleanup and sustainability initiatives.

II. METHODS

To address the complex challenge of detecting underwater trash in real-time, this study adopts a deep learning-based approach by utilizing the YOLOv10 object detection architecture in conjunction with the Trash-ICRA19 dataset. This combination is selected to fulfill the critical requirements of accuracy, speed, and adaptability in underwater environments. The model is trained on Trash-ICRA19 to learn from diverse examples of submerged waste under realistic conditions, ensuring its relevance and generalization. In addition, appropriate training and inference configurations are applied to optimize the model's performance and assess its real-time detection capability. This methodological approach enables a practical evaluation of YOLOv10's potential as a fast, lightweight, and reliable solution for automated underwater trash monitoring.

A. Yolov10

As illustrated in Fig. 1, YOLOv10 represents a cutting-edge one-stage object detection framework that incorporates multiple architectural advancements to improve both accuracy and computational efficiency, especially for real-time recognition tasks in challenging scenarios. This model is based on the groundwork laid by earlier YOLO versions. yet brings key advancements that significantly elevate its capabilities.

One of the major breakthroughs in YOLOv10 is the elimination of Non-Maximum Suppression (NMS) in its post-processing pipeline. Traditional NMS, while effective at removing redundant bounding boxes, is computationally expensive and often eliminates true positives especially in scenarios with overlapping objects. YOLOv10 circumvents this by employing a new matching strategy in both training and inference, which replaces NMS and enhances detection reliability.

The core of this strategy involves one-to-one matching, where each ground truth label is matched to a single predicted box using optimal assignment algorithms like the Hungarian method. This ensures high-quality predictions and stable training dynamics. Furthermore, one-to-many matching is also used during training, allowing multiple candidate predictions to learn from a single object instance. This approach is especially beneficial in cluttered or occluded scenes—common challenges in underwater environments.

Another architectural advancement of YOLOv10 is the incorporation of large-kernel convolution in the Contextual Information Block (CIB). Large kernels are designed to capture

TABEL I
TRASH-ICRA19 DATASET CONFIGURATION

Parameters	Setup
Training Data	5,720 images
Validation Data	820 images
Testing Data	1,145 images

TABEL II
TRAINING CONFIGURATION

Parameters	Setup
Platform	Kaggle
GPU	P100
Image Size	640 x 640 Pixels
Epochs	300
Batch Size	32
Optimizer	Stochastic Gradient Descent (SGD)
Learning Rate	0,01

long-range dependencies and broader spatial contexts, which are critical when objects vary in scale or are partially obscured by marine elements like sediment or aquatic vegetation. The CIB uses these kernels to extract rich contextual features, improving the network's understanding of the surroundings and boosting its accuracy in diverse visual conditions.

YOLOv10 also integrates PSA (Parameter-free Self-Attention), a mechanism that enhances feature representation without introducing extra trainable parameters. PSA recalibrates feature maps by highlighting important spatial information and suppressing irrelevant noise, enabling the model to concentrate on the most relevant regions of the image. This is especially useful in underwater scenes where lighting variations and textures can introduce substantial background noise.

YOLOv10 was selected for this study due to its balanced trade-off between speed and accuracy. Compared to YOLOv5 and YOLOv8, YOLOv10 maintains competitive frame-per-second (FPS) performance while achieving superior precision. Its architecture is optimized for GPU-based real-time applications, making it well-suited for deployment in remotely operated vehicles (ROVs). These characteristics make YOLOv10 a compelling choice for real-time marine debris detection and monitoring, capable of accurately identifying small, transparent, and irregularly shaped trash objects in challenging underwater conditions.

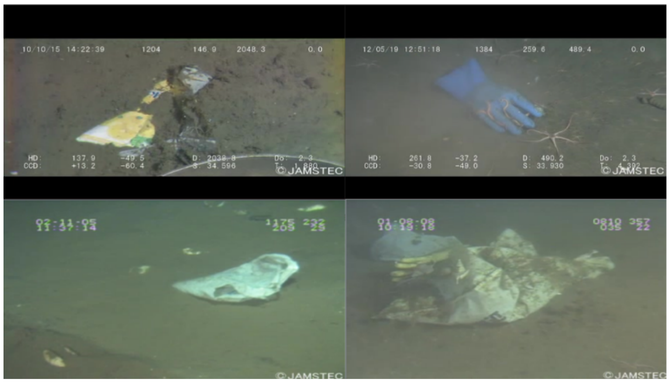

Figure 2. Dataset Sample

TABEL V
MODEL COMPARISON

Model	GFLOPs(G)	Parameter(M)	mAP50(%)
YOLOv5-nano [19]	7.2 G	2,5 M	41.7 %
YOLOv8-nano [20]	8.2 G	3,01 M	45.5 %
YOLOv10-nano	**8.4 G**	**2,7 M**	**46.4 %**

GHz

B. Dataset

The Trash-ICRA19 [18] dataset is a specialized underwater image dataset developed for marine debris detection. It contains RGB images captured by remotely operated vehicles (ROVs) during real-world marine missions from JAMSTEC. The dataset comprises a diverse collection of 5,720 training images, 820 validation images, and 1,145 testing images, as summarized in Table I.

The dataset includes a diverse variety of visibility levels, backgrounds, and lighting conditions. Each image is annotated with bounding boxes in YOLO format, facilitating supervised training for object detection models.

This dataset was selected because of its domain-specific focus and realistic representation of underwater trash, which is essential for training a model to generalize effectively in field deployments as shown in Fig 2.

The diversity of object types and underwater conditions captured in the dataset allows the YOLOv10 model to learn meaningful patterns and features that are directly relevant to marine debris detection.

Using Trash-ICRA19 ensures that the resulting system is optimized for the types of environments and object appearances encountered in practical underwater scenarios.

C. Training Configuration

This section outlines the configuration utilized to train the YOLOv10 model for detecting underwater trash. The training process was performed on the Kaggle cloud environment utilizing a P100 GPU. To align with the model's input requirements, all training images were adjusted to a resolution of 640 × 640 pixels. The model was trained over 300 epochs with a batch size of 32. The Stochastic Gradient Descent (SGD) optimizer was utilized with an initial learning rate of 0.01. These hyperparameter values were chosen to maintain a balance between stable training dynamics and efficient computational performance. A comprehensive overview of the training and evaluation configurations is presented in Table II.

D. Inference Configuration

For the inference phase, the trained YOLOv10 model is evaluated on a local machine using a CPU-based setup.

This setup is selected to simulate real-world deployment scenarios where access to high-end GPUs may not be available. Running inference on limited hardware helps assess the model's practicality for field applications, such as underwater trash monitoring using low-power edge devices.

The inference environment uses the Ubuntu operating

TABEL VI
MODEL COMPARISON

Model	GFLOPs(G)	Parameter(M)	FPS
YOLOv5-nano	7.2 G	2,5 M	21.32
YOLOv8-nano	8.2 G	3,01 M	18.65
YOLOv10-nano	**8.4 G**	**2,7 M**	**20.8**

system, with Python 3.9.20 as the compiler and PyTorch 2.0 as the deep learning framework. The system runs on a 12th Gen Intel® Core™ i7-12700 CPU operating at 2.10 GHz.Table III summarizes the hardware and software configuration used during inference.

III. RESULT AND DISCUSSION

This part of the paper highlights the experimental findings obtained from implementing the YOLOv10-nano model for identifying underwater trash. The discussion is organized to address various critical aspects, such as assessing the model's performance, comparing it with earlier YOLO variants, evaluating the dataset used, and analyzing the model's inference speed in different scenarios. These elements collectively serve to demonstrate the reliability and real-world applicability of the proposed method.

A. Evaluation on Model

The YOLOv10-nano model's performance is assessed using the mean Average Precision (mAP) metric at a 0.5 Intersection over Union (IoU) threshold. As indicated in Table 4, the model achieves an mAP50 of 46.4% on the Trash-ICRA19 dataset. This indicates that nearly half of the model's predictions correctly overlap with the ground truth objects by at least 50%, demonstrating its competence in recognizing various types of underwater debris. This score reflects the model's effectiveness in identifying underwater waste even under challenging conditions, including murky waters, varying object sizes, and poor visibility. Such environmental variability often hampers the performance of object detectors, yet YOLOv10-nano maintains consistent detection results across diverse samples in the dataset.

Furthermore, YOLOv10-nano is designed with efficiency in mind, utilizing only 8.4 GFLOPs and comprising 2.7 million parameters. These design choices make it a strong candidate for real-time operations on platforms with constrained hardware capabilities, such as underwater robots and embedded systems. The model's ability to combine accuracy with a lightweight architecture showcases its suitability for underwater object detection scenarios.

To further assess YOLOv10-nano's capability, a comparison is made with YOLOv5-nano [19] and YOLOv8-nano [20], which are both lightweight detection models widely used in real-time applications. This comparison uses the same Trash-ICRA19 dataset to ensure a fair evaluation across all models, with the detailed outcomes presented in Table 5. This side-by-side analysis helps determine how each model balances trade-offs between speed, model size, and detection accuracy in the same underwater context.

YOLOv5-nano shows the lowest computational cost at 7.2 GFLOPs and 2.5 million parameters, achieving an mAP50 of 41.7%. While it is efficient, its detection performance is significantly lower than that of YOLOv10-nano. YOLOv10-nano offers a considerable improvement in accuracy with an mAP50 of 46.4%, despite only a small increase in computation (8.4 GFLOPs) and parameters (2.7M). This highlights the advantage of YOLOv10's improved backbone and matching strategies in boosting accuracy while maintaining efficiency.

Meanwhile, YOLOv8-nano achieves 45.5% mAP50 with 8.2 GFLOPs and a larger model size of 3.01M parameters. Although it has slightly lower computational demand than YOLOv10-nano, it still uses more parameters and provides lower accuracy. This shows that YOLOv10-nano develops a greater balance between accuracy and model complexity, indicating that it serves as a more streamlined and capable solution for detecting underwater waste.

B. Evaluation on Dataset

The Trash-ICRA19 dataset plays an important role in training and evaluating the model's performance to detect underwater trash. This dataset consists of three main object classes: Plastic, ROV (Remotely Operated Vehicle), and BIO (Biological entities). Among these, plastic is the most important class, as it represents the dominant type of underwater trash that poses serious environmental threats. Unlike organic waste, plastic materials do not decompose easily and can remain in the ocean for hundreds of years, harming marine ecosystems and wildlife.

The ability to detect various forms of plastic trash is essential for developing effective marine conservation technologies. As shown in Fig 3, the proposed YOLOv10-nano model demonstrates strong capabilities in identifying plastic waste in a wide range of shapes, sizes, and conditions. The model successfully detects individual plastic items, bundled trash, transparent plastics, and even partially buried or covered plastic debris under different underwater lighting and visibility conditions.

This capability reflects the strength of the architecture trained on the dataset, which provides diverse real-world scenarios and object variations. The detection of ROV and BIO classes is also supported, ensuring that the model can differentiate between actual trash and non-trash objects such as marine equipment or marine life, reducing false positives during underwater operations.

C. Runtime Efficiency

Speed is one of the most important factors in real-time applications, especially for underwater trash detection, where fast response and continuous monitoring are crucial. The inference speed of different YOLO models was evaluated on a CPU-based local environment, and the findings are presented in Table VI. This setup reflects realistic deployment conditions where high-end GPUs may not be available, such as in embedded systems or autonomous underwater vehicles.

YOLOv10-nano achieves an inference speed of 20.8 FPS, which is slightly lower than YOLOv5-nano with 21.32 FPS, but still fast enough for real-time performance. YOLOv8-nano, in comparison, performs the slowest at 18.65 FPS. This shows that YOLOv10-nano is competitive in terms of speed, maintaining a good balance between processing speed and computational cost (8.4 GFLOPs and 2.7M parameters).

The small drop in FPS relative to YOLOv5-nano is offset by the significant gain in detection accuracy, making YOLOv10-nano a more practical choice when both precision and responsiveness are required. This result demonstrates that even though YOLOv10-nano runs slightly slower than YOLOv5-nano, it provides significantly better precision, making it more effective for underwater trash detection tasks where accurate object identification is essential. Its runtime efficiency, combined with improved detection performance, positions YOLOv10-nano as a viable solution for integration into real-time underwater surveillance systems where hardware resources are limited but detection reliability is critical.

Figure 3. Model Evaluation on Dataset

IV. CONCLUSION

This study addresses the challenge of underwater trash detection, which remains a significant environmental issue due to the persistent presence of non-degradable materials like plastic waste. By utilizing the YOLOv10-nano model and the Trash-ICRA19 dataset, the research demonstrates that deep learning object detection methods are effective in identifying various underwater trash types in complex visual conditions. The YOLOv10-nano model achieves a notable detection performance with an mAP50 of 46.4%, and its ability to distinguish objects such as plastics, ROVs, and biological entities reflects its strength in understanding contextual underwater scenes. This highlights the model's generalization capability even in the presence of noise, blur, and occlusions common in underwater environments. Its precision remains the highest among the models compared, and the inference speed of 20.8 FPS shows it is capable of near real-time operation even on limited hardware. This balance between accuracy and computational efficiency underscores its practicality for deployment in embedded systems or remotely operated platforms. These findings confirm that the proposed method is suitable for real-world implementation, particularly in scenarios where computational efficiency and accuracy are both critical.

The results of this study also highlight the potential of the one-to-one and one-to-many matching strategy used in YOLOv10, which contributes to better object localization without relying on non-maximum suppression (NMS). This mechanism reduces duplicate detections and enhances boundary accuracy, making the model more reliable in detecting overlapping or densely clustered debris. This methodological advancement supports further research into lightweight, real-time object detection systems suitable for underwater robotics, environmental monitoring, and automated marine survey operations.

Based on these findings, future work may explore extending this approach to multi-class marine waste detection with more diverse datasets and real-time deployment using edge devices or autonomous underwater vehicles (AUVs). Additional improvements in model training strategies, data augmentation for underwater conditions, and the integration of temporal analysis across video frames may enhance detection reliability in dynamic environments. Moreover, further collaboration with marine researchers and environmental organizations could support large-scale application and refinement of this technology.

ACKNOWLEDGEMENT

Special thanks to all members of the AIVision research group for their contributions in making this work possible.

REFERENCE

[1] S. Lincoln *et al.*, "Marine litter and climate change: Inextricably connected threats to the world's oceans," *Sci. Total Environ.*, vol. 837, 2022, doi: 10.1016/j.scitotenv.2022.155709.

[2] J. Soto-Navarro, G. Jordá, M. Compa, C. Alomar, M. C. Fossi, and S. Deudero, "Impact of the marine litter pollution on the Mediterranean biodiversity: A risk assessment study with focus on the marine protected areas," *Mar. Pollut. Bull.*, vol. 165, 2021, doi: 10.1016/j.marpolbul.2021.112169.

[3] F. Chenillat, T. Huck, C. Maes, N. Grima, and B. Blanke, "Fate of floating plastic debris released along the coasts in a global ocean model," *Mar. Pollut. Bull.*, vol. 165, 2021, doi: 10.1016/j.marpolbul.2021.112116.

[4] T. O. Fossum, Ø. Sture, P. Norgren-Aamot, I. M. Hansen, B. C. Kvisvik, and A. C. Knag, "Underwater autonomous mapping and characterization of marine debris in urban water bodies," 2022, [Online]. Available: http://arxiv.org/abs/2208.00802

[5] N. Nnamoko, J. Barrowclough, and J. Procter, "Solid Waste Image Classification Using Deep Convolutional Neural Network," *Infrastructures*, vol. 7, no. 4, 2022, doi: 10.3390/infrastructures7040047.

[6] G. Ramkumar, G. Anitha, M. Suresh Kumar, M. Ayyadurai, and C. Senthilkumar, "An effectual underwater image enhancement using deep learning algorithm," *Proc. - 5th Int. Conf. Intell. Comput. Control Syst. ICICCS 2021*, pp. 1507–1511, 2021, doi: 10.1109/ICICCS51141.2021.9432116.

[7] J. Chen and M. J. Er, "Dynamic YOLO for small underwater object detection," *Artif. Intell. Rev.*, vol. 57, no. 7, p. 165, 2024, doi: 10.1007/s10462-024-10788-1.

[8] M. Hussain, "YOLOv1 to v8: Unveiling Each Variant-A Comprehensive Review of YOLO," *IEEE Access*, vol. 12, pp. 42816–42833, 2024, doi: 10.1109/ACCESS.2024.3378568.

[9] S. Zhao, J. Zheng, S. Sun, and L. Zhang, "An Improved Yolo Algorithm for Fast and Accurate Underwater Object Detection," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4079287.

[10] B. Xue *et al.*, "An Efficient Deep-Sea Debris Detection Method Using Deep Neural Networks," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 12348–12360, 2021, doi: 10.1109/JSTARS.2021.3130238.

[11] R. Jain, S. Zaware, N. Kacholia, H. Bhalala, and O. Jagtap, "Advancing Underwater Trash Detection: Harnessing Mask R-CNN, YOLOv8, EfficientDet-D0 and YOLACT," *2nd Int. Conf. Sustain. Comput. Smart Syst. ICSCSS 2024 - Proc.*, pp. 1314–1325, 2024, doi: 10.1109/ICSCSS60660.2024.10625362.

[12] C. Huang *et al.*, "YOLO-MES: An Effective Lightweight Underwater Garbage Detection Scheme for Marine Ecosystems," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3552090.

[13] J. Jia, M. Fu, X. Liu, and B. Zheng, "Underwater Object Detection Based on Improved EfficientDet," *Remote Sens.*, vol. 14, no. 18, 2022, doi: 10.3390/rs14184487.

[14] J. Niu, S. Gu, J. Du, and Y. Hao, "Underwater Waste Recognition and Localization Based on Improved YOLOv5," *Comput. Mater. Contin.*, vol. 76, no. 2, pp. 2015–2031, 2023, doi: 10.32604/CMC.2023.040489.

[15] A. Wang *et al.*, "YOLOv10: Real-Time End-to-End Object Detection," 2024, [Online]. Available: http://arxiv.org/abs/2405.14458

[16] Q. Tian, Y. Huo, M. Yao, and H. Wang, "A method for detecting dead fish on large water surfaces based on improved YOLOv10," 2024,

[Online]. Available: http://arxiv.org/abs/2409.00388

[17] M. L. Ali and Z. Zhang, "The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection," *Computers*, vol. 13, no. 12, 2024, doi: 10.3390/computers13120336.

[18] M. Fulton, J. Hong, M. J. Islam, and J. Sattar, "Robotic detection of marine litter using deep visual detection models," *Proc. - IEEE Int. Conf. Robot. Autom.*, vol. 2019-May, pp. 5752–5758, 2019, doi: 10.1109/ICRA.2019.8793975.

[19] R. Khanam and M. Hussain, "What is YOLOv5: A deep look into the internal features of the popular object detector," 2024, [Online]. Available: http://arxiv.org/abs/2407.20892

[20] J. Zhu, T. Hu, L. Zheng, N. Zhou, H. Ge, and Z. Hong, "YOLOv8-C2f-Faster-EMA: An Improved Underwater Trash Detection Model Based on YOLOv8," *Sensors*, vol. 24, no. 8, 2024, doi: 10.3390/s24082483.

Tomi Heri Julianus Todingan was born in Kendari, on September 18, 2000. The author began his elementary education at Sekolah Dasar Pelangi, Kendari from 2006 to 2009, and then continued at Sekolah Dasar Negeri 126 Manado from 2010 to 2012. After completing elementary school, the author pursued his junior high school education at Sekolah Menengah Pertama Eben Haezar 2 Manado from 2012 to 2015. He then continued his studies at the senior high school level at Sekolah Menengah Atas Negeri 9 BINSUS Manado. The author pursued his undergraduate (Strata 1) education in the Informatics Study Program, Faculty of Engineering, at Sam Ratulangi University, Manado. Further deepening his scientific field, the author continued his graduate studies (Strata 2) in the Master Informatics Study Program, Postgraduate Student, at Sam Ratulangi University, Manado.

With a background in the AI Vision research group, the author has built a strong foundation in the field of computer vision, particularly in applying deep learning techniques to real-world visual detection challenges. This expertise becomes a key driver in supporting the research titled "Deep Learning Approach to Real-Time Underwater Trash Detection with YOLOv10-Nano."