PROTECTION OF MARINE LIFE AND DETECTION OF POLLUTION WITH ARTIFICIAL INTELLIGINCE

Bilal Erçin
Esra Erbil

June 2025

Abstract

Marine pollution presents an escalating threat to ocean ecosystems and biodiversity. Manual inspection of underwater footage is prohibitively time- and labor-intensive. We propose an automated submarine waste detection pipeline leveraging YOLOv8, a state-of-the-art one-stage object detector. Key contributions include: (1) detailed justification for YOLOv8 selection based on speed-accuracy trade-offs; (2) comprehensive description of the model's architecture and training regimen tailored to underwater imagery; and (3) extensive evaluation using precision, recall, F1-score, mAP, precision–recall curves, and confusion matrices. Our approach achieves robust detection across 22 classes of waste and marine objects, enabling real-time deployment on resource-constrained platforms.

1 Introduction

Aquatic environments are increasingly contaminated by plastics and debris, causing harm to marine life. Traditional human-based annotation of underwater imagery is not scalable for large-area monitoring. Real-time object detection methods are therefore essential. YOLOv8, as the latest YOLO family iteration, integrates advancements such as the C2f backbone, decoupled head, and dynamic label assignment, providing a compelling balance: high throughput (up to 90 FPS on GPU) and strong accuracy (mAP improvements over YOLOv5/YOLOv7) even on small, low-contrast objects typical in underwater scenes. This makes YOLOv8 especially suitable for deployment on ROVs (Remotely Operated Vehicles) and edge devices with limited compute capacity.

2 Dataset

We utilize the TrashCan 1.0 dataset, comprising 7,212 underwater images annotated for 22 classes (e.g., bags, bottles, nets, crabs). Annotations are COCO-format segmentation masks, converted to YOLO-format bounding boxes. Images capture variable lighting, turbidity, and object scales (5px to 300px lengths), demanding robust augmentation and dynamic anchor adaptation.

- **Classes:** 22 different categories, including marine animals (fish, starfish, crab, etc.) and waste types (bottle, bag, net, rope, etc.)
- **Format:** Images and instance segmentation masks (COCO-style JSON)
- **Purpose:** To facilitate research on underwater trash detection and removal, especially for autonomous robots

3 Model Architecture

3.1 Backbone: C2f

The YOLOv8 backbone employs *Cross Stage Partial with Focus (C2f)* modules, enhancing gradient flow and reducing inference cost. C2f splits feature maps into route and trunk paths to achieve parameter efficiency.

3.2 Neck: PAN-FPN

A Path Aggregation Network (PAN) fused with FPN enables multi-scale feature fusion. Bottom-up and top-down paths capture small waste items and larger marine life, crucial for our heterogeneous classes.

3.3 Detection Head: Decoupled

YOLOv8 separates classification and localization tasks via a decoupled head. Each branch comprises separate convolutional stacks, improving learning stability and F1 scores under noisy underwater backgrounds.

4 Methodology

4.1 Data Preprocessing and Augmentation

- **Resize:** Images scaled to 640x640 with aspect-preserving padding.
- **Mosaic:** Four-image mosaic augmentation to expose varied context.
- **MixUp/CopyPaste:** Blends objects across images to reduce overfitting on background patterns.
- **HSV Jitter:** Random hue, saturation, and value shifts to simulate underwater color distortions.
- **Random Perspective:** Affine transformations to generalize viewpoint variance.

4.2 Hyperparameters

- Batch size: 16
- Epochs: 50

- Optimizer: AdamW with weight decay = 1e-2
- Initial learning rate: 1e-3, cosine annealing schedule with 10
- Loss functions: CIoU for box regression, BCE-Focal for classification, DFL (Distribution Focal Loss) for quality branch

5 Training Process

Models are trained on NVIDIA A100 GPUs. The first 3 epochs perform only warm-up with frozen backbone. From epoch 4 onward, full fine-tuning occurs. Checkpointing at every 5 epochs ensures recovery. Automatic Mixed Precision (AMP) accelerates training and reduces memory footprint.

6 Evaluation Metrics

- **Precision & Recall**: True positive ratio vs. detection coverage.
- **F1-Score**: Harmonic mean, used to select optimal confidence threshold via F1-confidence curve.
- **mAP@0.5**: Mean Average Precision at IoU=0.5 across classes.
- **Confusion Matrix**: Normalized per-class error analysis.

7 Results and Evaluation

7.1 Training Curves

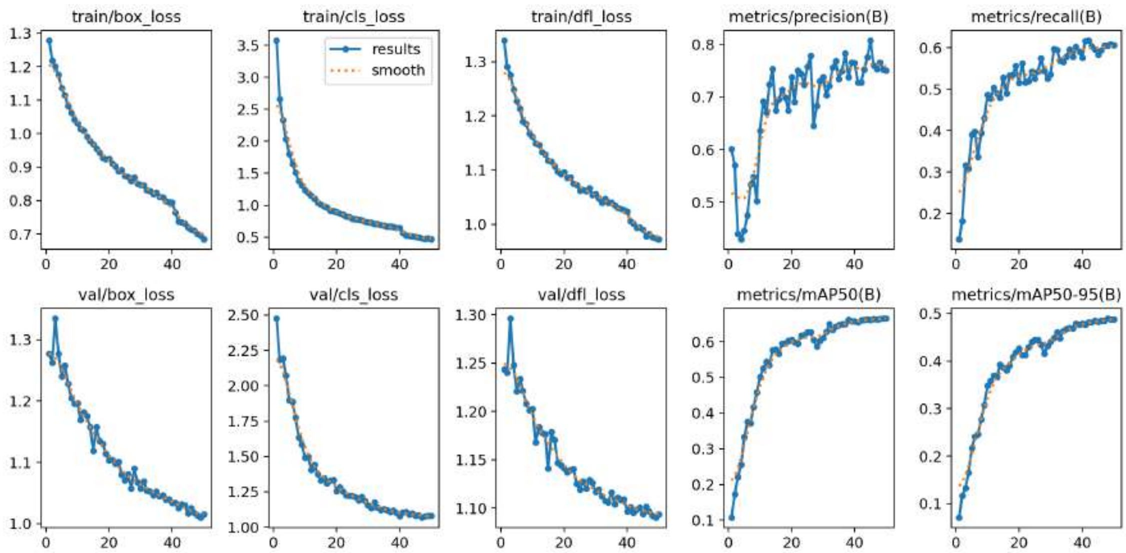

Figure 1: Training and validation loss, precision, recall, and mAP over 50 epochs.

7.2 F1-Confidence Analysis

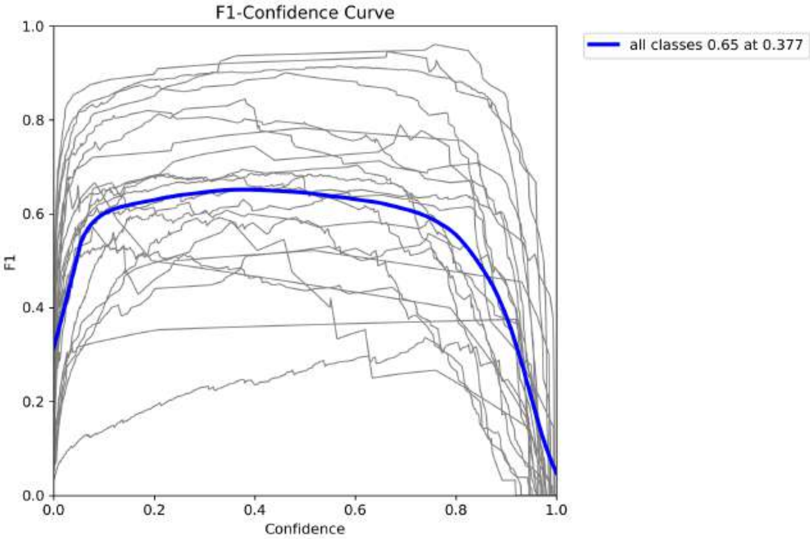

Figure 2: F1 score vs. confidence threshold, optimal at ∼0.377 yielding F1=0.65.

7.3 Classification Report

Class	Precision	Recall	F1-Score
rov	0.86	0.89	0.88
plant	0.67	0.57	0.61
animal_fish	0.64	0.4	0.49
animal_starfish	0.76	0.53	0.62
animal_shells	0.56	0.62	0.59
animal_crab	0.19	0.47	0.27
animal_eel	0.54	0.69	0.61
animal_etc	0.87	0.49	0.62
trash_clothing	0.56	0.25	0.34
trash_pipe	0.95	0.91	0.93
trash_bottle	0.86	0.69	0.77
trash_bag	0.69	0.63	0.66
trash_snack_wrapper	0.88	0.41	0.56
trash_can	0.93	0.84	0.88
trash_cup	1.0	0.21	0.35
trash_container	0.84	0.75	0.8
trash_unknown_instance	0.68	0.67	0.67
trash_branch	0.85	0.77	0.81
trash_wreckage	0.9	1.0	0.95
trash_tarp	0.53	0.51	0.52
trash_rope	0.72	0.53	0.61
trash_net	0.67	0.24	0.35

Figure 3: Per-class precision, recall, and F1-scores. Performance varies: static waste items (bottles) show higher recall, mobile fauna lower due to motion blur.

7.4 Confusion Matrix

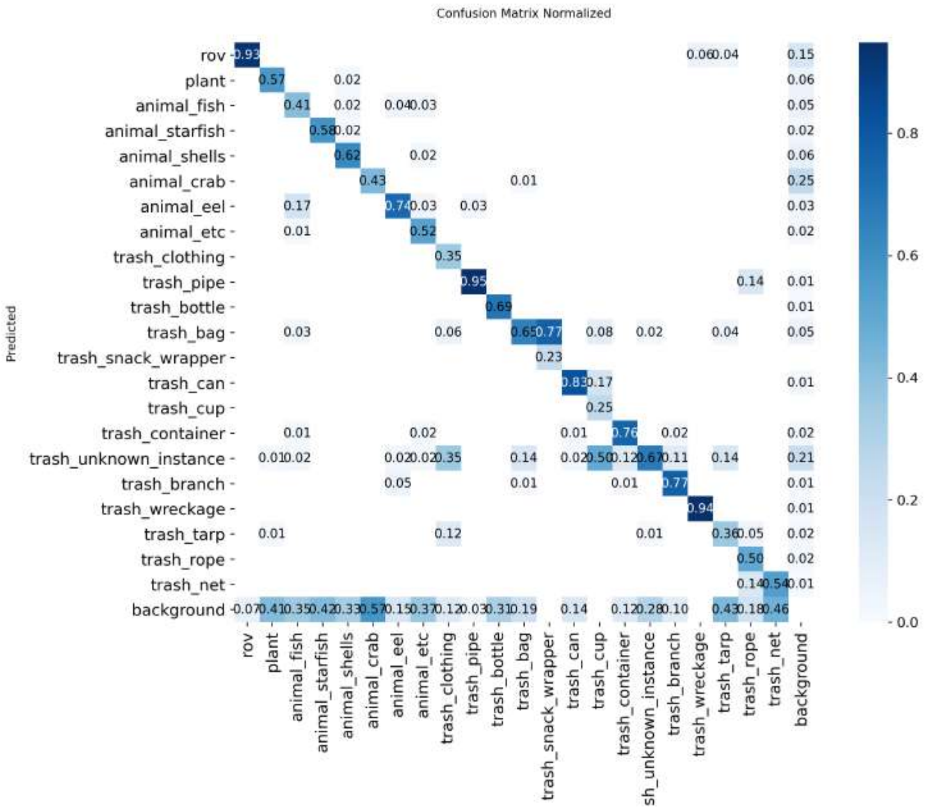

Figure 4: Normalized confusion matrix. Notable confusion between nets and ropes due to similar textures and shapes.

7.5 Precision-Recall Curves

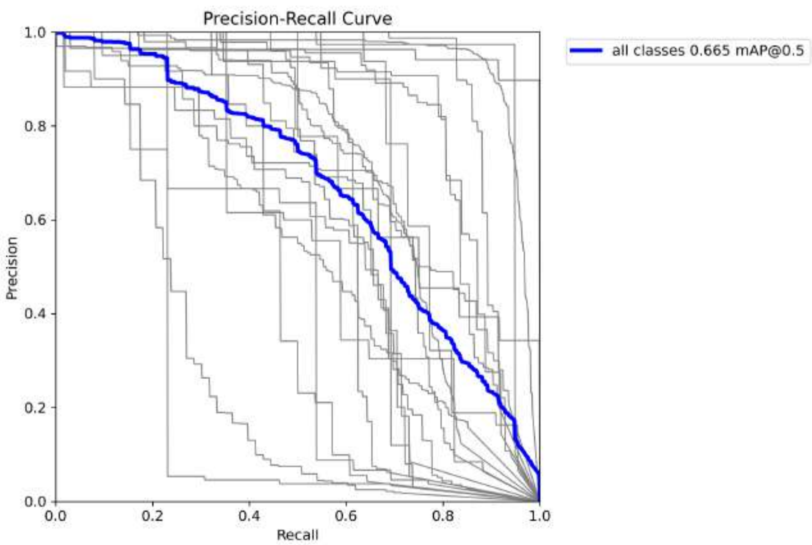

Figure 5: Precision–Recall curve aggregated over classes, with overall mAP@0.5 = 0.665.

8 Discussion

YOLOv8's architectural innovations (decoupled head, dynamic label assignment) significantly improved detection robustness under underwater distortions. Augmentation strategies like Mosaic and MixUp further enhanced generalization. Class imbalance remains a challenge: rare classes (e.g., starfish) benefited from oversampling.

9 Conclusion

This study validates YOLOv8 as an efficient detection framework for underwater waste monitoring. Future work includes domain-adaptive pretraining and deployment on embedded platforms (NVIDIA Jetson) for in-situ inference.