remote sensing

MDPI

Article

A Benchmark Evaluation of Intelligent Identification Models for Marine Outfalls

Li Yang [1], Haolan Zhou [1,*], Sile Li [2], Shicheng Zhao [1] and Ruisheng Yang [1]

[1] College of Water Conservancy and Civil Engineering, South China Agricultural University, Guangzhou 510642, China; yangli@stu.scau.edu.cn (L.Y.); zhaoshicheng2022@163.com (S.Z.); yangruisheng@stu.scau.edu.cn (R.Y.)
[2] Guangzhou Panyu Urban Drainage Management Co., Ltd., Guangzhou 511450, China; lisile@gzpanyudrainage.com
* Correspondence: haolan_zhou@scau.edu.cn; Tel.: +86-150-1806-3777

Highlights

What are the main findings?

- A benchmark evaluation of YOLOv8n/v9t/v10n/v11n and lightweight RT-DETR is conducted on an NDWI-based multi-source remote sensing dataset for marine outfall detection.
- YOLOv8n achieves the best overall performance with the highest mAP50 (77.0%), while pipe detection remains the key bottleneck due to small size and weak features.

What are the implications of the main findings?

- The NDWI fusion strategy effectively enhances water–land contrast and improves detection robustness in complex coastal environments.
- The results provide a practical model selection reference for UAV-based real-time marine outfall monitoring and edge deployment.

Abstract

Monitoring marine outfalls is crucial for mitigating coastal pollution and protecting marine environments. Current methods rely mainly on manual inspection and satellite remote sensing interpretation, which are inefficient, inaccurate, and inadequate for large-scale real-time monitoring. Although UAV visible-light imagery has been introduced for marine outfall detection, challenges remain, including insufficient and diverse target features, small multi-scale target detection difficulties, and complex background interference. To address these limitations, this study systematically benchmarks mainstream object detection models (YOLOv8n, YOLOv9t, YOLOv10n, YOLOv11n, and RTDETR-light) on a dedicated multi-source remote sensing fusion dataset that we constructed for marine outfalls along Zhanjiang's southern coast, incorporating NDWIs. Our comparative experiments evaluate the models' effectiveness in this challenging scenario. Experimental results indicate that YOLOv8n is the most balanced model for marine outfall detection, achieving 84.1% precision, 68.6% recall, 77% mAP50, and an F1 score of 0.75. This benchmark provides empirical evidence and practical model selection criteria for intelligent marine outfall monitoring, thereby offering a reference framework for researchers and engineers in related fields.

Keywords: marine outfall detection; multi-source remote sensing fusion; NDWI; YOLO; RTDETR

Check for updates

Academic Editor: Mark Bourassa

Received: 4 March 2026
Revised: 4 May 2026
Accepted: 6 May 2026
Published: 8 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Marine ecosystems are crucial to global environmental sustainability [1]. As a maritime nation with an extensive coastline and diverse marine outfalls, China faces significant challenges in monitoring and managing these systems [2]. In recent years, increasing land-based pollution—especially agricultural nonpoint source and urban sewage discharge—has severely degraded nearshore water quality and ecological functions [3,4]. Marine outfalls act as the primary pathway for land-based pollutants entering oceans [5].

Current outfall monitoring still depends heavily on manual patrols and satellite remote sensing, hampered by low efficiency, limited coverage, and poor small-target identification [6]. Traditional methods struggle to achieve real-time, precise, and systematic monitoring, particularly for concealed and widely dispersed outfalls. Rapid advances in drone-based remote sensing and artificial intelligence have driven the gradual introduction of vision-based automated detection methods. Drones capture centimeter-level high-resolution imagery, overcoming the spatial resolution limitations of satellite remote sensing and adapting well to complex coastal environments [7].

Intelligent detection of marine outfalls originated from traditional image processing and remote sensing interpretation. However, visible-light-only detection faces persistent challenges: outfall types exhibit significant morphological differences; targets are small and vulnerable to interference from waves, shadows, and vegetation. Variable coastal weather further degrades imaging quality. Therefore, stable detection cannot be achieved using RGB data alone. Multispectral imaging enhances spectral contrast between water bodies and land features by incorporating bands such as near-infrared and red edge, helping distinguish outfalls from backgrounds [8].

Powered by UAV and deep learning advances, UAV-image-based outfall detection has become a research hotspot. UAVs' high maneuverability, high resolution, and flexible deployment enable effective monitoring of hard-to-cover waters. Meanwhile, deep learning significantly enhances detection automation and accuracy through its powerful feature extraction and recognition.

Meanwhile, deep learning object detection algorithms continue to evolve. The YOLO series is widely adopted for its strong speed-accuracy balance, critical for real-time detection. With continued YOLO iterations, more studies have incorporated them into outfall detection. Huang et al. [9] proposed the GDCNN-outfalls, which embeds a channel attention mechanism into Faster R-CNN, achieving a 2.1% accuracy improvement. Xu et al. [10] developed a drone-edge computing system based on YOLOv5, fusing high-resolution UAV imagery with low-latency edge processing to boost monitoring efficiency. These approaches demonstrate the YOLO series' advantage in balancing speed and accuracy on high-resolution drone imagery.

In addition, multispectral data and feature enhancement mechanisms have been progressively introduced to enhance model robustness. Huang et al. [11] leveraged multispectral reflectance features to enrich input information and enhanced challenging sample recognition with an improved loss function. Yu et al. [12] designed a lightweight CNN-Transformer hybrid model that maintains high accuracy while reducing computational overhead. Huang et al. [13] proposed AFENet, which incorporates an attention-guided mechanism to strengthen the perception of critical outfall regions and achieves 75.6% accuracy.

Notably, RT-DETR, proposed by Baidu [14], is the first real-time end-to-end Transformer detector in the DETR family. Its hybrid encoder and query selection balance accuracy and speed, providing a new approach for marine outfall detection. However, high computational cost and slow convergence limit its practical deployment.

However, a discernible research gap persists. While individual model improvements are frequently reported, systematic comparison and comprehensive evaluation of main-

stream detection models on a unified, diverse multi-source remote sensing dataset are lacking. Most studies refine a single architecture, leaving the relative efficacy, robustness, and practical suitability of state-of-the-art models like YOLOv8, YOLOv9, YOLOv10, YOLOv11, and RT-DETR for marine outfall detection underexplored under consistent experimental conditions.

To address this gap, we compare five advanced models—YOLOv8n, YOLOv9t, YOLOv10n, YOLOv11n, and RTDETR-light—for intelligent identification of marine outfalls. The research is based on a custom multi-source remote sensing fusion dataset that integrates visible and multispectral imagery from Zhanjiang's southern coast. The specific objectives are:

(1) To construct a finely annotated marine outfall dataset by fusing visible and multispectral data with the NDWI;
(2) To systematically benchmark the accuracy, robustness, and computational efficiency of the five selected models;
(3) To provide empirically grounded model selection criteria and engineering references for intelligent outfall monitoring systems.

The paper is organized as follows: Section 2 details the dataset composition, experimental methods, and evaluation metrics; Section 3 presents model detection accuracy and performance with in-depth analysis; Section 4 discusses the significance of research findings and proposes optimal model selection. Section 5 concludes the paper by summarizing the core findings, their implications, and potential future work.

2. Materials and Methods

2.1. Dataset and Data Preprocessing

2.1.1. Composition of the Dataset

This study focuses on Zhanjiang's southern coast for marine outfall imagery collection. Located in southwestern Guangdong, Zhanjiang has flat terrain, extensive coastlines, and numerous marine outfalls, making it an ideal study area. Consequently, we selected 718 marine outfall locations in southern Zhanjiang as data collection targets, employing a DJI Phantom 4 RTK drone (manufactured by SZ DJI Technology Co., Ltd., Shenzhen, China) to gather 1657 sets of outfall imagery. All images were collected at a 100 m above ground level (AGL) flight altitude, with a corresponding ground sample distance (GSD) of 2.74 cm/pixel. The DJI Phantom 4 RTK drone's visible and multispectral imaging system comprises one visible light camera and five multispectral cameras (blue, green, red, red edge, and near-infrared bands). The visible light camera covers the 400–700 nm visible spectrum, and the center wavelengths of the multispectral bands are as follows: blue (450 nm), green (560 nm), red (650 nm), red edge (730 nm), and near-infrared (840 nm). All images have a resolution of 1600 × 1300 pixels. Visible images are stored at 24-bit depth, and multispectral bands at 16-bit depth. Under normal illumination, captured images achieve a signal-to-noise ratio (SNR) exceeding 35 dB, ensuring image quality meets subsequent fusion and detection requirements. During simulated complex environment data augmentation, SNR decreases but remains above 20 dB to replicate real-world monitoring scenarios. These images were first used to construct an initial annotated dataset. After strict train/validation/test splitting, NDWI-based multi-source fusion and environmental simulation augmentation were independently applied to each subset. This processing expanded the final dataset to 4971 images, with all augmented versions kept within their respective subsets to avoid data leakage. The study area is shown in Figure 1; purple symbols indicate marine outfall locations.

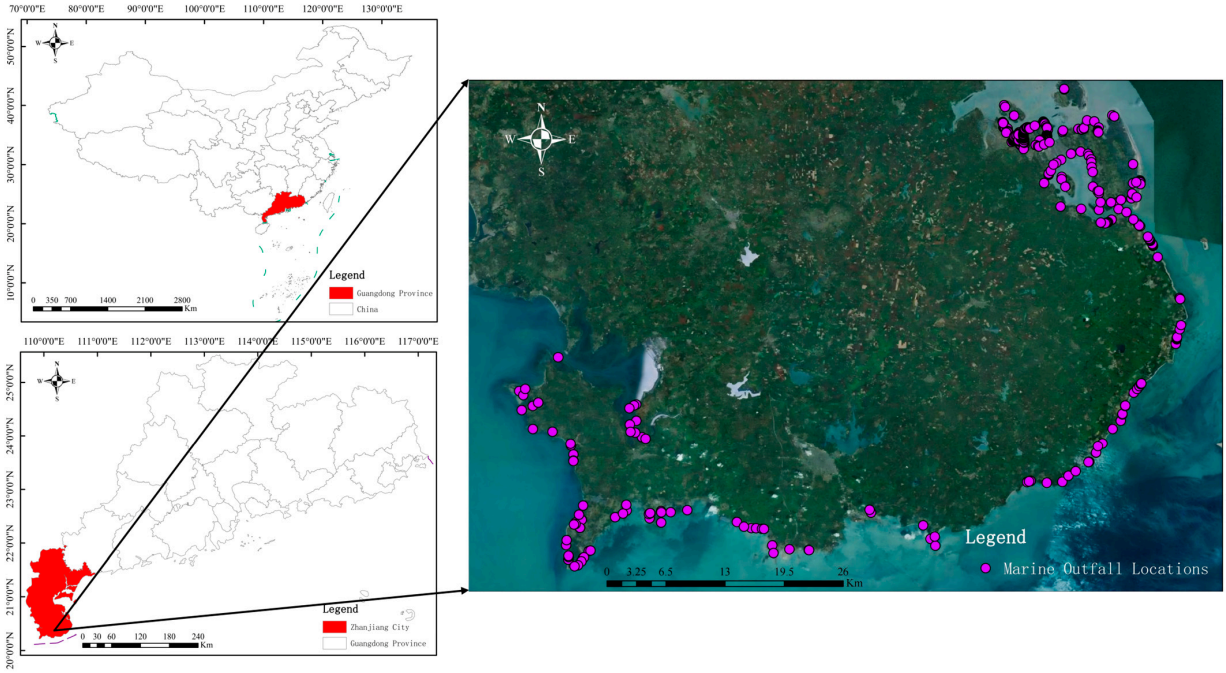

Figure 1. Zhanjiang South Coast Marine Outfall Locations.

2.1.2. Data Preprocessing

Collected marine outfalls are categorized into four types: sluice, gully, pipe, and culvert (Figure 2). All images were manually annotated with category labels and location information for each outfall. The annotated images and labels were first shuffled and then split into training (70%), validation (20%), and test (10%) sets, before any data fusion or augmentation. NDWI-based multi-source fusion and data augmentation were performed independently on each subset. In this way, all augmented and fused images derived from the same original image remain within the same split, preserving the standard 70-20-10 partition. The final dataset contains 4971 images after data augmentation and NDWI fusion. It includes 11,815 annotated marine outfall label instances, distributed as follows: sluice (3910), gully (2296), pipe (3270), and culvert (2339).

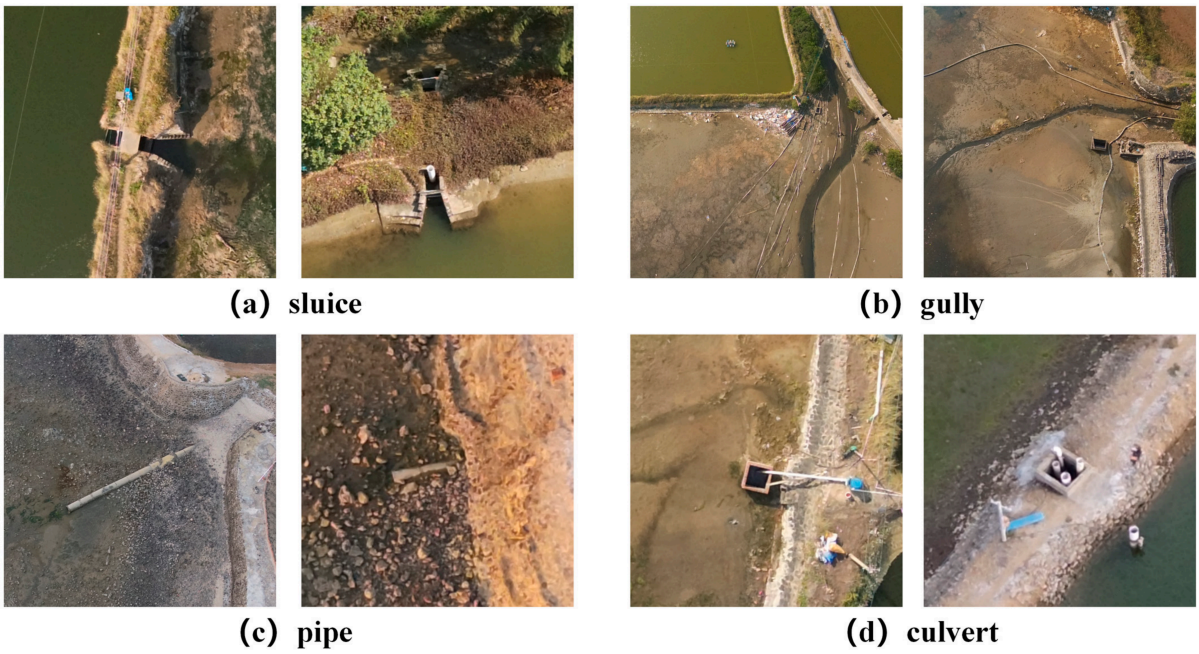

Figure 2. Types of Marine Outfalls.

2.1.3. Data Augmentation

Zhanjiang's tropical-subtropical coastal climate brings frequent winds, rain, fog, and uneven lighting, degrading UAV image quality and detection stability. To improve robustness and avoid data leakage, data augmentation was applied independently to each split (training, validation, test) after dataset partitioning. All derivatives of an original image were retained within the same subset. We implemented random environmental simulations using the imgaug library as follows: Motion blur—ZoomBlur (intensity 1, application probability 0.2); Fog—fog corruption (intensity 1, application probability 0.5); Rainfall—raindrop size 0.05–0.10, rain intensity 0.05–0.1, application probability 0.5; Backlight—pixel values multiplied by a random coefficient in [0.7, 1.3], application probability 0.5. Each operation was applied randomly and independently, such that multiple effects (e.g., rainfall and fog) could co-occur. The environmental simulation is shown in Figure 3. Environmental simulation offers greater practicality than conventional augmentation, enhancing model robustness in real-world applications.

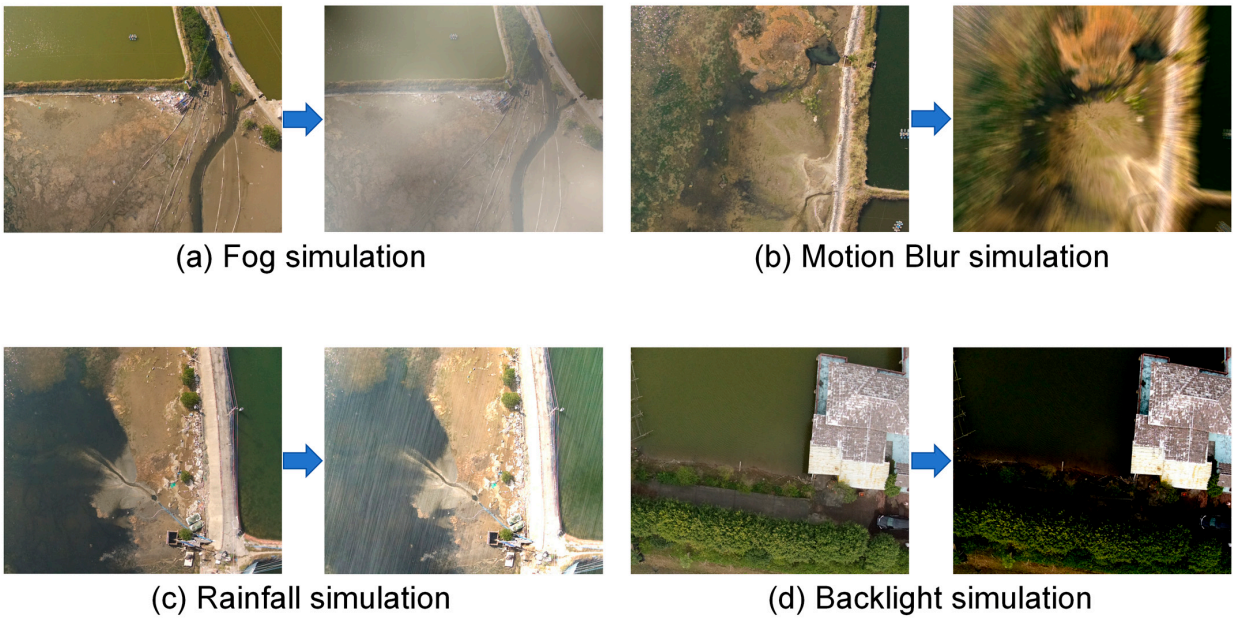

Figure 3. Environmental Simulation Series.

2.2. NDWI and NDWI-Based Fusion Procedure

2.2.1. NDWI

The Normalized Difference Water Index (NDWI), proposed by McFeeters [15] in 1996, indicates wetland hydrological conditions. In remote sensing imagery, NDWI delineates open water features and augments their visual prominence [16]. It exploits the high reflectance of water bodies in the green band and strong absorption in the near-infrared band. The index is calculated as:

$$\mathrm{NDWI} = \frac{(\mathrm{Green} - \mathrm{NIR})}{(\mathrm{Green} + \mathrm{NIR})} \quad (1)$$

Green and NIR denote the green and near-infrared bands, respectively. The NDWI typically ranges from −1 to 1, with higher values indicating richer water body information. Negative values arise when near-infrared reflectance exceeds green reflectance, as in exposed soil or water bodies [17].

By enhancing water–land contrast, NDWI reduces background interference from vegetation and soil, helping detection models focus on hydrologically relevant outfall features.

2.2.2. NDWI-Based Fusion Procedure

NDWI-based fusion combined the visible-light image alongside the near-infrared (NIR) greyscale image. The 16-bit NIR image was converted to 8-bit and to floating-point to match the visible image and ensure accurate NIR channel extraction. Three channels were extracted from the visible light image: red, green, and blue. The NDWI channel was computed from the green and NIR channels and normalised to an 8-bit greyscale image (0–255). This NDWI channel was then combined with the red and blue channels to form a new three-channel fusion image, with the final neural network input organized as Channel 1—Red, Channel 2—NDWI, Channel 3—Blue. The fusion process is illustrated in Figure 4. NDWI-based fusion was applied separately to the training, validation, and test sets after dataset splitting, without any cross-contamination between subsets.

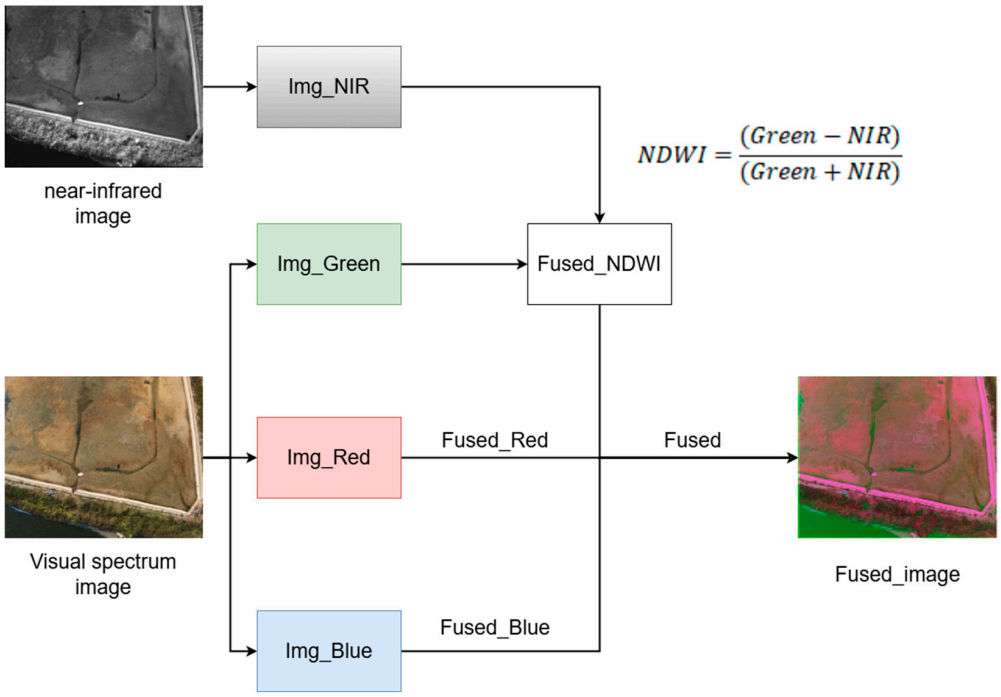

Figure 4. NDWI-based fusion process.

2.3. Marine Outfall Monitoring Model

2.3.1. YOLOv8

YOLOv8, released by Ultralytics in 2023 [18], introduces comprehensive optimizations over YOLOv5, including replacing the C3 module with C2F to improve feature reuse and gradient propagation efficiency [19], adopting an anchor-free design to simplify detection and enhance generalization [20], and incorporating Mosaic data augmentation with adaptive training strategies to boost multi-scale adaptability [21]. Its architecture retains the classic YOLO backbone network, neck network, and detection head [22] (Figure 5). It achieves a favorable balance between accuracy and inference speed.

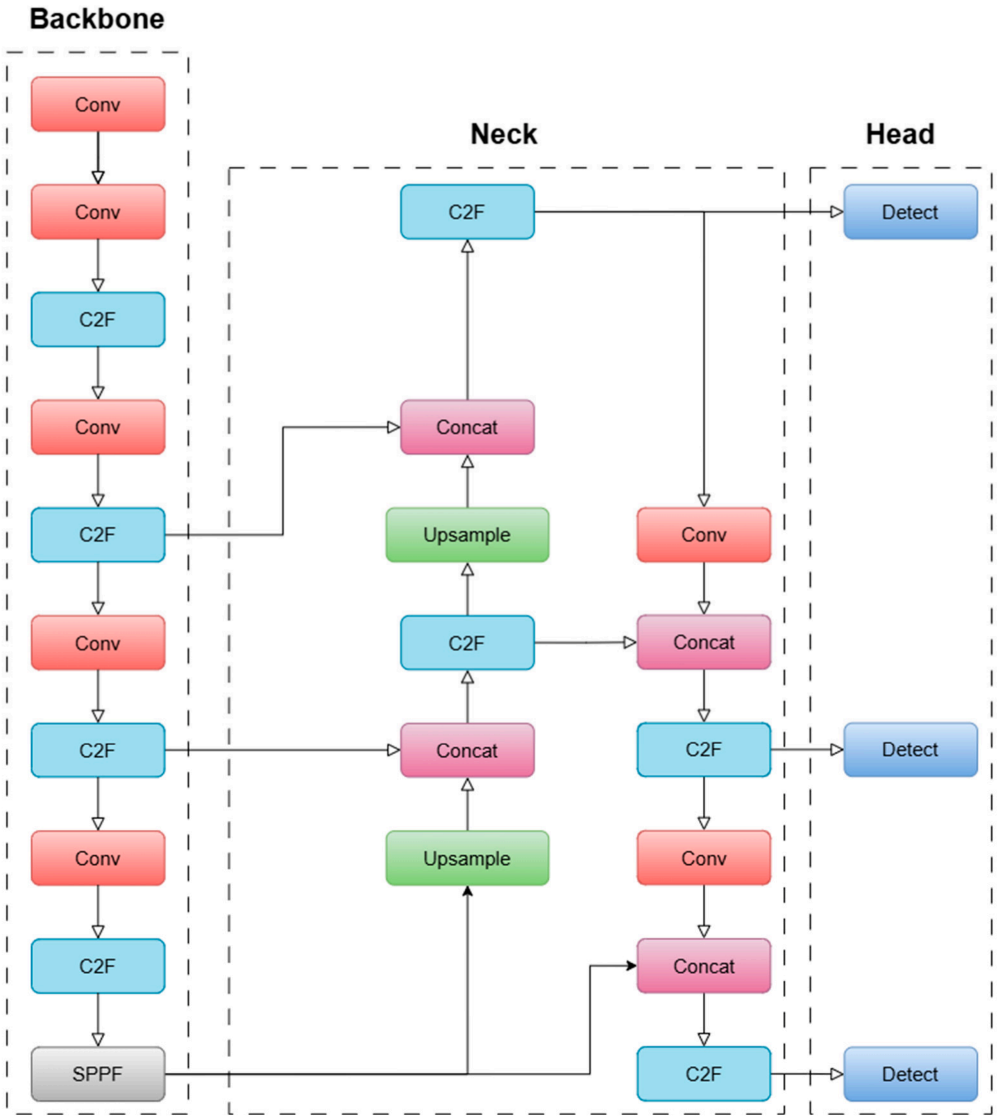

Figure 5. YOLOv8 Architecture.

2.3.2. YOLOv9

In February 2024, Wang et al. [23] proposed YOLOv9, which mitigates information loss in deep networks by introducing programmable gradient information and a generalized efficient layer aggregation network [24]. The PGI network architecture (Figure 6) preserves more semantic information through auxiliary supervision branches, enhancing small object detection [25]. The GELAN architecture combines the strengths of CSPNet and ELAN for efficient feature fusion and gradient flow [26]. Maintaining YOLOv8's anchor-free framework [27], YOLOv9 further improves detection stability and generalization performance.

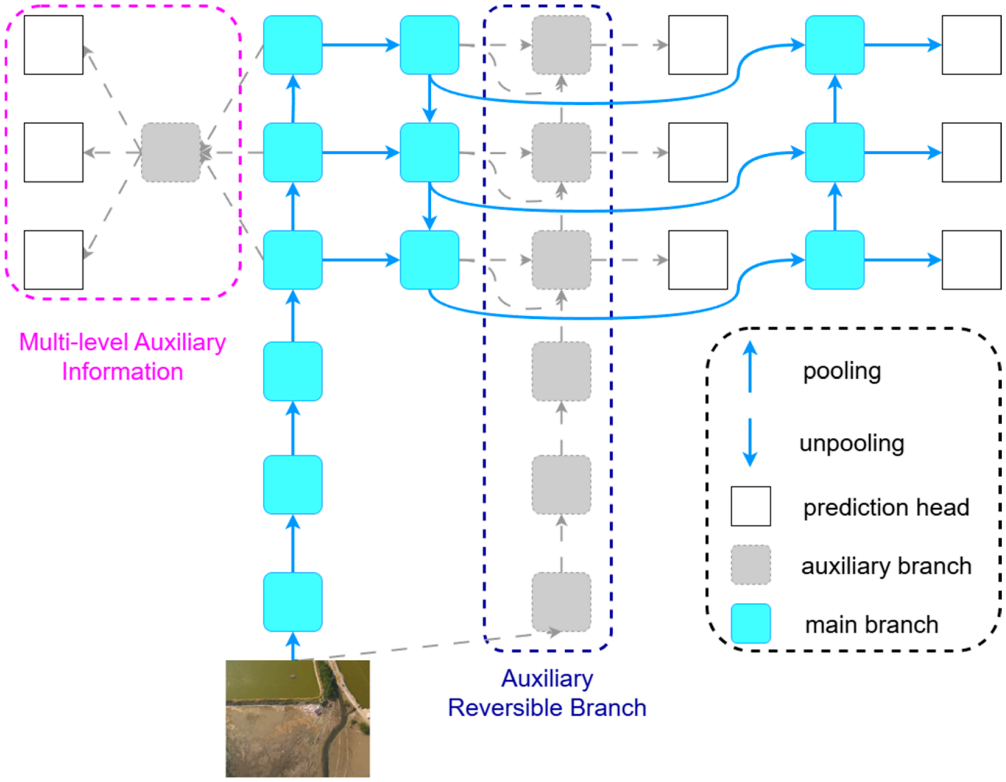

Figure 6. PGI Network Architecture.

2.3.3. YOLOv10

YOLOv10 was released in May 2024 by Wang et al. [28] from Tsinghua University (architecture in Figure 7), employs a dual-label assignment strategy using both "one-to-many" and "one-to-one" heads for training and inference [29]. This approach reduces reliance on post-processing non-maximum suppression, thereby lowering inference latency. Additionally, the model incorporates a path aggregation network and parallel segmentation attention modules to enhance multi-scale feature fusion and key-region perception [30,31]. This significantly improves real-time performance while maintaining high accuracy.

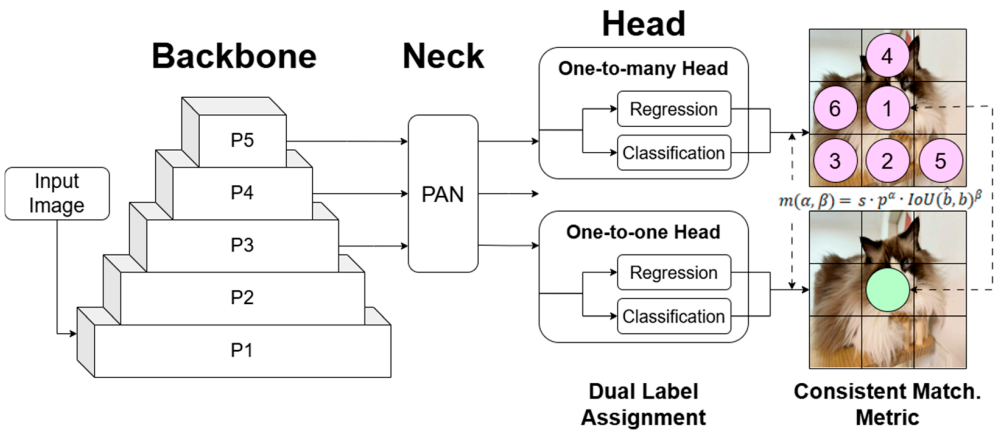

Figure 7. YOLOv10 Architecture.

2.3.4. YOLOv11

YOLOv11, released by Ultralytics in September 2024 [32]. YOLOv11 architecture (Figure 8) further streamlines and optimizes the YOLOv8 architecture: it replaces C2f blocks with C3k2 blocks to reduce parameters and computational overhead [33]; it integrates cross-stage partial spatial attention modules within the neck to enhance perception of

target spatial distribution [34]. It simultaneously supports extended tasks such as instance segmentation and pose estimation, demonstrating strong cross-domain adaptability.

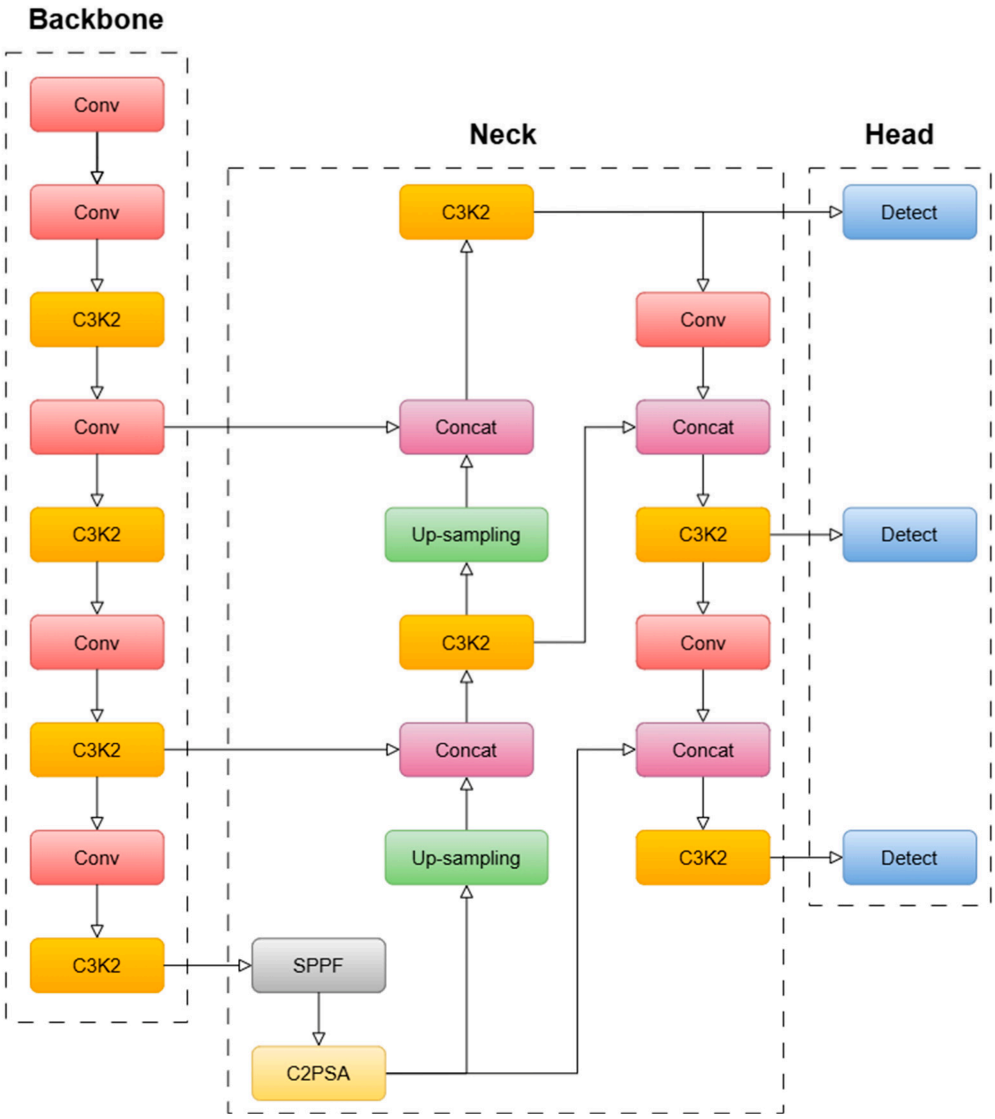

Figure 8. YOLOv11 Architecture.

2.3.5. RT-DETR

RT-DETR is the first Transformer-based real-time end-to-end detection model proposed by Baidu [14] in July 2023 (architecture in Figure 9). RT-DETR employs an efficient hybrid encoder for multi-scale features: the AIFI module enables intra-layer interactions through self-attention, while the CCFM module achieves cross-scale fusion via convolutions [35]. An IoU-aware query selection mechanism ensures high initial classification and localization quality, while supporting dynamic adjustment of decoder layers to balance speed and accuracy [36]. Despite its strong performance in structured scenarios, high computational cost and convergence time remain key practical challenges.

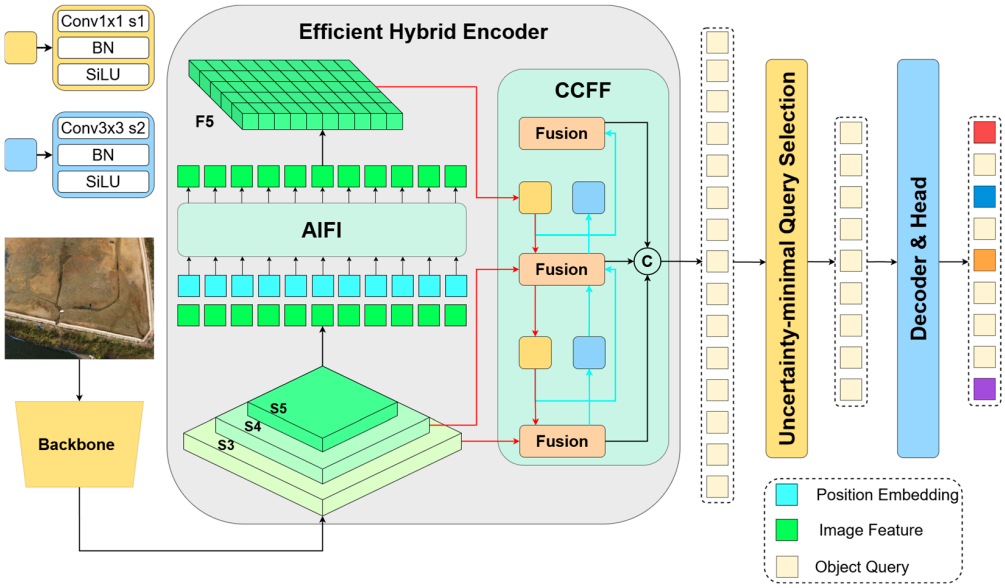

Figure 9. RT-DETR Architecture.

2.4. Model Evaluation

To evaluate the model's detection accuracy and overall performance, five fundamental metrics were employed: precision, recall, mAP, critical success index, and F1 score. Precision (P) measures accuracy, preventing false positives. Recall (R) measures completeness, preventing false negatives. mAP measures the overall precision-recall balance, serving as the primary metric for model comparison. Critical success index (CSI) measures overall detection skill at a fixed decision threshold, penalizing both missed detections and false alarms equally. The F1 score represents a balanced measure of precision and recall, capturing their trade-off. The formulas for these metrics are as follows:

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$mAP = \frac{\sum_{i=0}^{n} AP(i)}{n} \tag{4}$$

$$CSI = \frac{TP}{TP + FP + FN} \tag{5}$$

$$F1Score = 2 \times \frac{P \times R}{P + R} \tag{6}$$

Here, true positives, false positives, and false negatives are denoted as TP, FP, and FN, respectively; n represents the total number of categories, and AP denotes the average precision across all categories. It measures per-class detection performance by balancing precision and recall across thresholds. It is computed as the area under the precision-recall (PR) curve using discrete numerical integration. The formula for AP is as follows:

$$AP = \sum_{i=1}^{n-1} (R_{i+1} - R_i) P_{inter}(R_{i+1}) \tag{7}$$

here, n represents the total number of points on the PR curve, and R_1, R_2, …, R_n are the recall values at the first interpolation points of the precision interpolation segments, arranged in ascending order.

3. Results

3.1. Training Details

Experimental environment and hyperparameters are listed in Tables 1 and 2.

Table 1. Experimental Environment Configuration.

Name	Model
Operating System	Windows 10 64-bit
CPU	AMD Ryzen 5 PRO 4650 G
GPU	NVIDIA GeForce RTX 3060 12 GB
Python version	Python-3.10.16
Pytorch version	Pytorch-2.5.1

Table 2. Hyperparameter tuning.

Hyperparameter Options	Settings
epoch	200
lr0	0.01
lrf	0.01
patience	20
batch	4
optimizer	SGD

To comprehensively evaluate accuracy and real-time performance under identical experimental configurations, we employ YOLOv8n, YOLOv9t, YOLOv10n, YOLOv11n, and RTDETR-light for marine outfall detection. RTDETR-light is a lightweight variant of RT-DETR (15.5 M parameters), created by adjusting the width and depth scaling factors of RTDETR-l from [1.00, 1.00, 1024] to [0.33, 0.25, 1024] as a representative Transformer-based detector for architectural comparison. Model architecture features are detailed in Table 3.

Table 3. Deep Learning Model Architecture Features.

Models	YOLOv8	YOLOv9	YOLOv10	YOLOv11	RT-DETR
Model variant	yolov8n	yolov9t	yolov10n	yolov11n	rtdetr-light
Params	3,011,628	2,006,188	2,708,600	2,590,620	15,524,440
Layers	225	917	385	319	457
GFLOPS	8.2	7.9	8.4	6.4	37.9

Params denotes the number of model parameters; more parameters indicate greater model complexity and higher video memory (VRAM) during inference training. Layers refer to the number of layers within the model, reflecting model depth; deeper models can learn more intricate hierarchical features. GFLOPS denotes the number of floating-point operations for a single model inference, representing computational complexity. On identical hardware, a lower GFLOPS value theoretically indicates faster inference speeds.

3.2. Results Analysis

3.2.1. Comparison of Model Accuracy

Table 4 presents the quantitative results of marine outfall detection, using precision, recall, mAP, and CSI to evaluate and compare model accuracy.

Table 4. Model Accuracy Assessment for Marine Outfall Detection.

Models	P/%	R/%	mAP50/%	mAP(50:95)/%	CSI/%
YOLOv8n	84.1	68.6	77.0	48.8	60.7
YOLOv9t	80.2	63.7	72.6	45.1	55.1
YOLOv10n	79.7	66.0	73.1	46.4	56.5
YOLOv11n	81.4	66.4	74.0	45.7	57.6
RTDETR-light	76.5	66.1	73.2	46.1	54.9

Table 4 shows that the YOLOv8n demonstrates excellent overall performance, achieving an mAP50 of 77.0% and significantly outperforming other comparison models in the outfall detection task. Specifically, YOLOv8n achieved a precision of 84.1%, a recall of 68.6%, and an mAP (50:95) of 48.8%, along with the highest critical success index (60.7%). The higher CSI confirms its superior overall detection capability when considering both false negatives and false positives. Although its mAP50 advantage over other models is modest (3–4%), it demonstrates that its architecture—particularly the C2F module for enhancing gradient flow and the anchor-free design—is well-suited for the diverse scales and shapes present in our fusion dataset. Notably, the latest YOLOv10n and YOLOv11n did not outperform YOLOv8n, suggesting that innovations designed for general-purpose benchmarks—such as the NMS-free design in YOLOv10 and the C3k2 module in YOLOv11—may not directly translate into performance improvements in specialized domains like marine outfall detection, which involve distinct target features and background complexity. Despite its higher computational cost, RTDETR-light still achieves a competitive recall (66.1%), demonstrating the potential of Transformer-based architectures for this task; however, its relatively low CSI (54.9%) reflects a trade-off in which higher recall is accompanied by more false positives, likely due to difficulties in making fine-grained distinctions within cluttered coastal scenes.

Table 5 presents the per-category detection results across four outfall types: sluice, gully, pipe, and culvert. In terms of mAP50, YOLOv8n demonstrates the most balanced performance, particularly excelling in culvert (86.3%) and sluice (85.7%) detection. YOLOv9t and YOLOv10n show slight declines in certain categories, while RTDETR-light, despite achieving the highest recall (57.3%) for pipes, falls significantly below the YOLO series in mAP50 for gullies. This indicates limited detection stability for certain structurally simple or irregularly shaped objects.

Sluices and culverts—which have relatively regular shapes and rich features—performed exceptionally well across all models. Notably, YOLOv8n surpassed 85% mAP50 in both categories, with CSI reached 71.2% (sluices) and 72.9% (culverts). This indicates strong recognition capabilities for targets with distinct geometric structures and textural features. Gullies typically have low contrast with their surroundings and exhibit highly variable shapes. YOLOv8n and YOLOv11n achieved mAP50 scores of 77.9% and 73.1%, with relative CSIs of 60.4% and 55.8%. In contrast, RTDETR-light achieved only 59.5% mAP50 and a CSI of 39.3% for gullies. This suggests that the Transformer architecture may be less effective than CNN-based YOLO variants at extracting local features in complex backgrounds.

Per-category analysis revealed the primary challenge faced by all models: the pipe category. All five models exhibited significantly lower precision, recall, mAP50, and CSI for pipes than for sluices, gullies, and culverts—a key finding. Pipes are typically small, texture-poor, and often blend in with the shoreline or man-made structures, yielding ambiguous visual features. For example, YOLOv8n's pipe precision (74.9%) is far below its culvert precision (90.7%), with recall dropping to 47.6%, and CSI to 41.0%. Although RTDETR-light achieved the highest pipe recall (57.3%) and mAP50 (64.2%), this came at the cost of the lowest precision (68.0%) and a CSI of only 45.1%. This trade-off indicates that

while RT-DETR's global attention mechanism better locates faint pipes, it is more prone to false positives from linear features such as paths or drainage lines. The consistently low CSI values for pipes highlight that improving pipe detection is a key bottleneck for advancing overall system performance.

Table 5. Model Detection of Accuracy Results for Different Types of Marine Outfalls.

Model	Class	P/%	R/%	mAP50/%	mAP(50:95)/%	CSI/%
YOLOv8n	sluice	87.4	79.4	85.7	56.4	71.2
	gully	83.6	68.5	77.9	51.7	60.4
	pipe	74.9	47.6	58.0	32.3	41.0
	culvert	90.7	78.9	86.3	54.8	72.9
YOLOv9t	sluice	83.1	75.6	81.6	53.1	65.5
	gully	78.4	60.6	70.9	44.3	51.9
	pipe	71.4	44.2	53.9	30.3	37.6
	culvert	87.7	74.3	83.9	52.7	67.2
YOLOv10n	sluice	85.5	78.5	84.7	55.3	69.2
	gully	77.7	61.0	70.0	47.0	51.9
	pipe	72.1	47.2	54.7	30.4	39.9
	culvert	83.5	77.2	83.0	52.9	67.1
YOLOv11n	sluice	82.8	76.8	83.6	52.3	66.3
	gully	80.4	64.7	73.1	47.8	55.8
	pipe	74.0	46.6	54.6	28.7	40.1
	culvert	88.5	77.7	84.6	53.9	70.5
RTDETR-light	sluice	84.4	80.2	85.2	55.0	69.8
	gully	70.3	47.1	59.5	38.0	39.3
	pipe	68.0	57.3	64.2	37.7	45.1
	culvert	83.0	79.7	84.0	53.6	68.5

From YOLOv8n to YOLOv11n, no generation has demonstrated a clear, comprehensive improvement over its predecessor, indicating that balancing lightweight design and accuracy remains an evolving challenge. YOLOv8n leads across most categories, likely due to optimized training strategy or structural design. RTDETR-light closely matches YOLOv8n on mAP50:95, demonstrating its robustness across multiple IoU thresholds. However, its noticeable weakness in single-category detection, particularly for the gully, suggests that Transformer-based detectors may require more targeted data augmentation or feature fusion mechanisms when handling category imbalance and weakly featured objects.

To visualise detection performance across conditions, six test samples were selected, representing visible light, NDWI fusion, and data augmentation detection. The results are shown in Figures 10–14. The samples contain: (1) NDWI fusion—one pipe; (2) visible light—one sluice, one gully, one pipe; (3) fog simulation—one culvert; (4) motion blur—one sluice, two culverts; (5) rainfall simulation—one sluice; (6) backlight simulation—two gullies. To visualise the detection results of the marine outfall detection model, a threshold of 0.5 was applied; bounding-box labels denote outfall type, and the accompanying number gives the confidence score, ranging from 0.5 to 1.

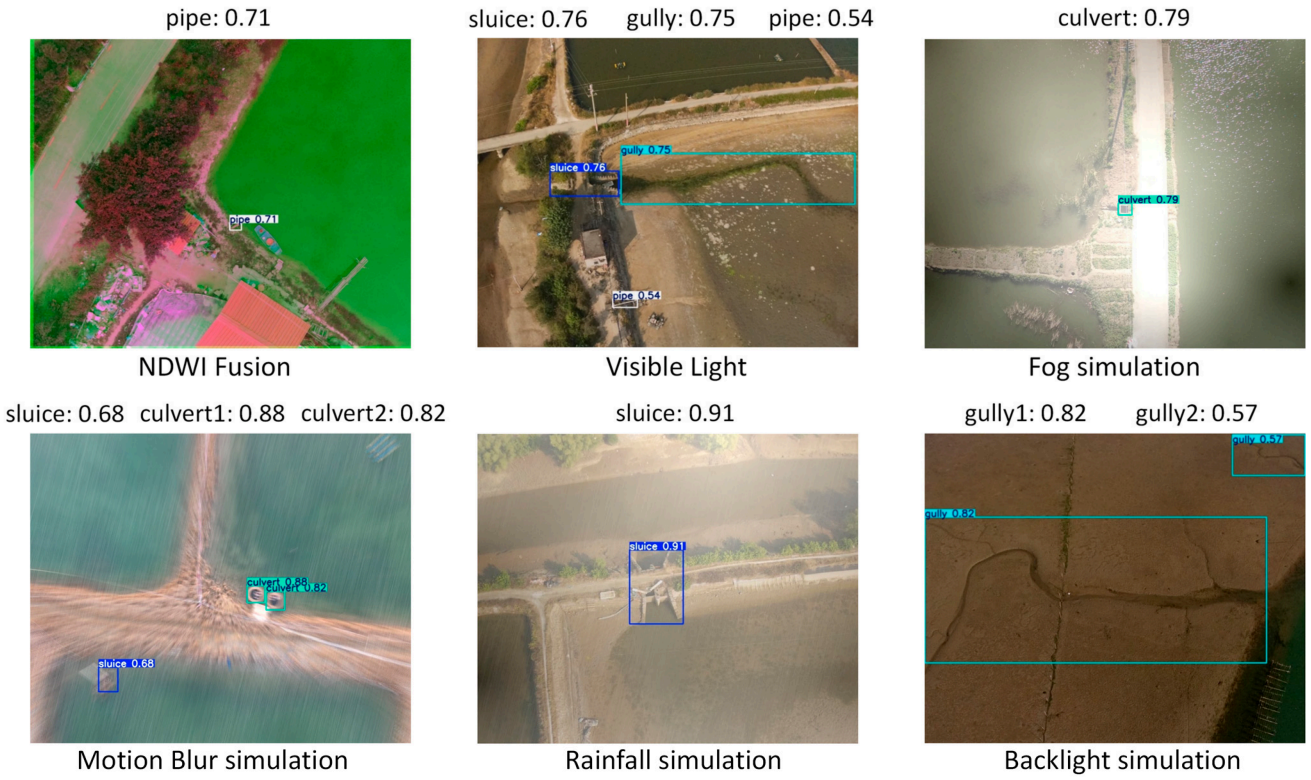

Figure 10. YOLOv8n Model Detection Results for Marine Outfalls.

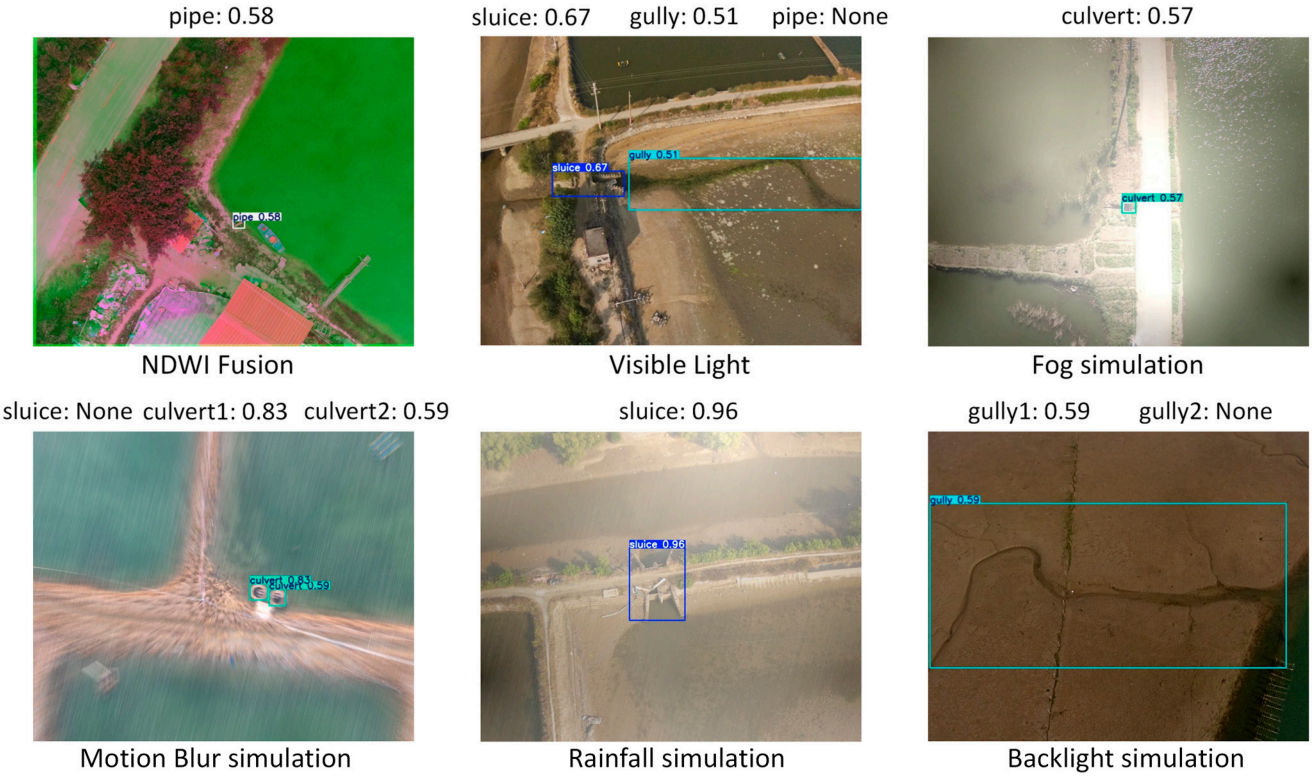

Figure 11. YOLOv9t Model Detection Results for Marine Outfalls.

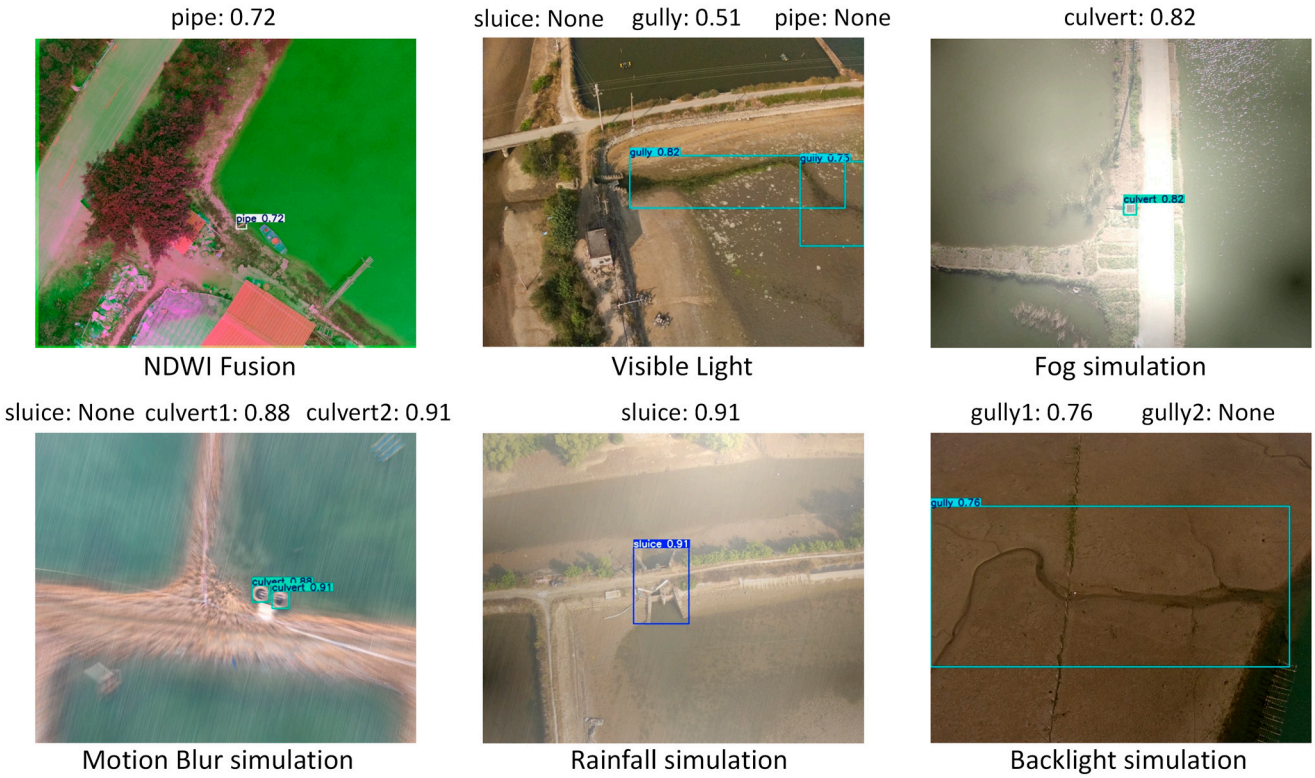

Figure 12. YOLOv10n Model Detection Results for Marine Outfalls.

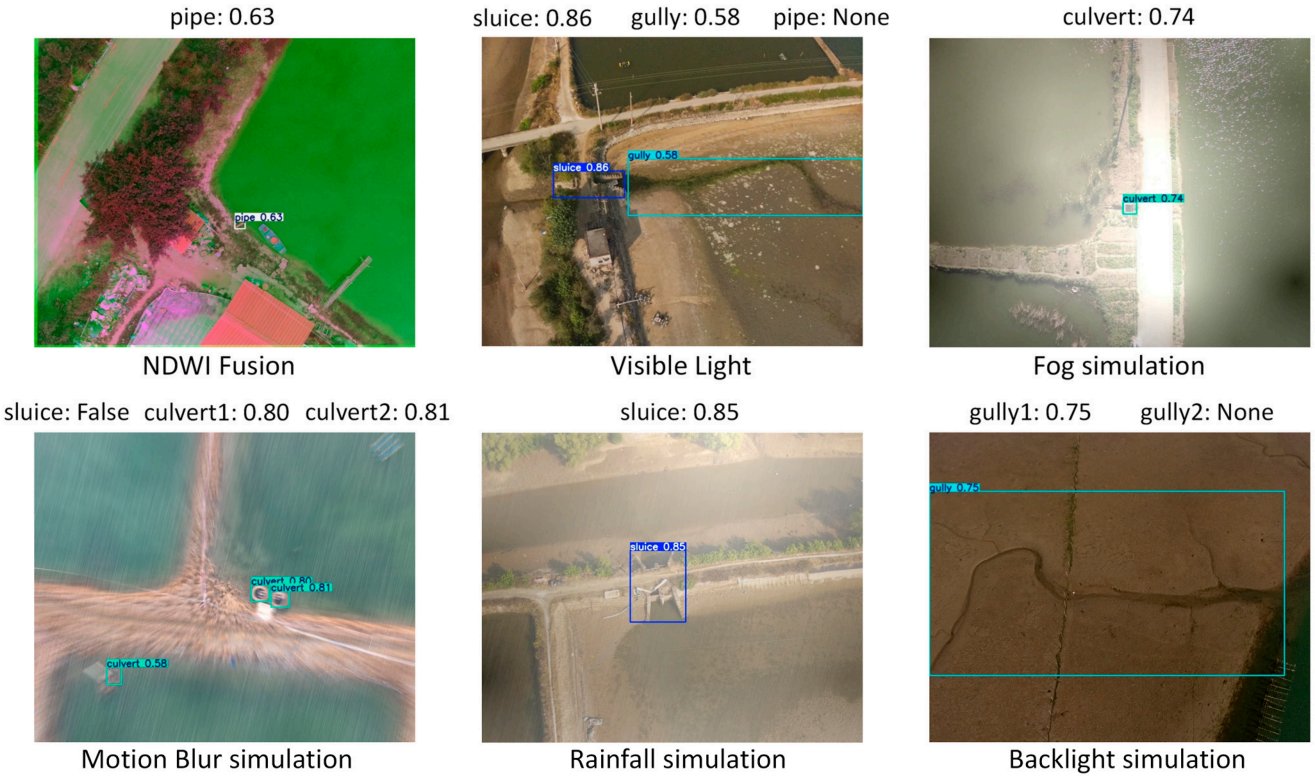

Figure 13. YOLOv11n Model Detection Results for Marine Outfalls.

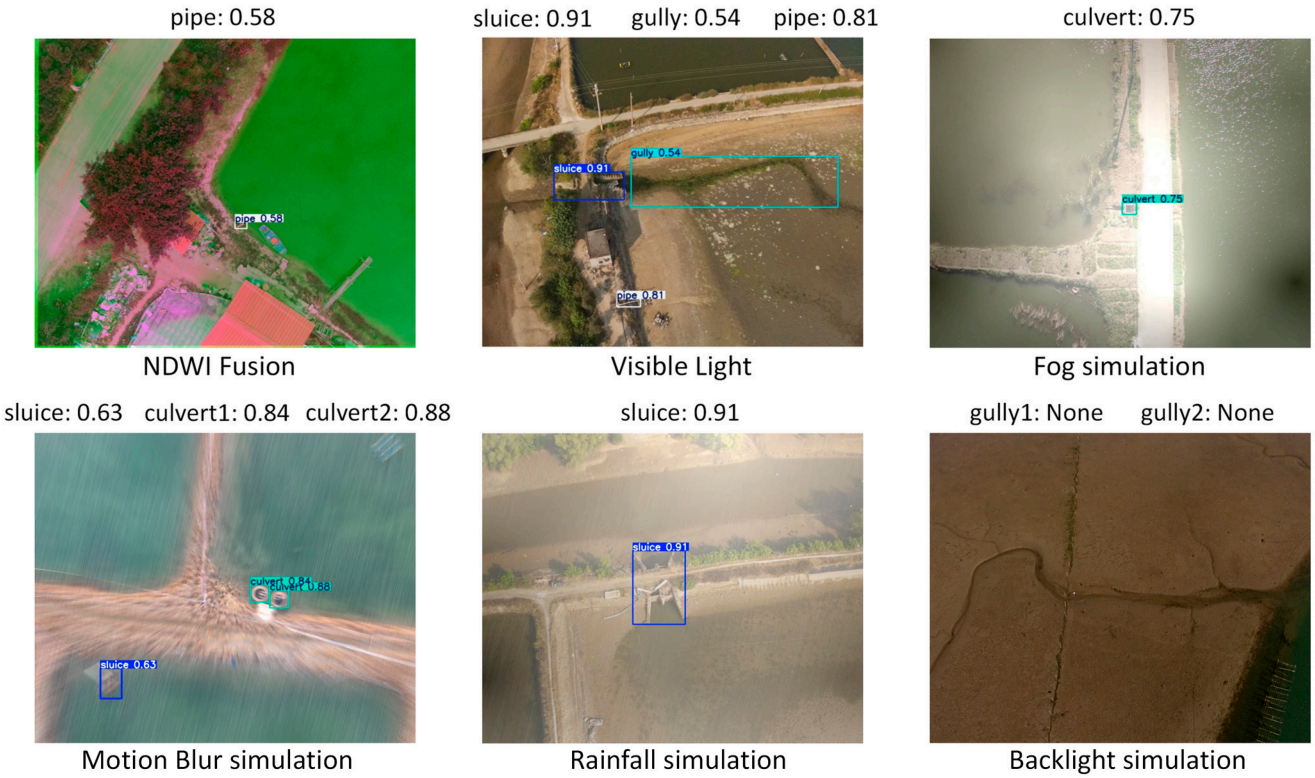

Figure 14. RTDETR-light Model Detection Results for Marine Outfalls.

For the first sample (pipe), YOLOv10n achieved the highest confidence score (0.72), closely followed by YOLOv8n (0.71). The second sample contained three outfall types. Only YOLOv8n and RTDETR-light accurately detected all three types. RTDETR-light achieved maximum confidence scores of 0.91 (sluice) and 0.81 (pipe). YOLOv10n missed the sluice and pipe but achieved the highest gully confidence score (0.82). For the third sample (culvert), YOLOv10n recorded the highest confidence score (0.82). In the fourth sample, only YOLOv8n and RTDETR-light detected all three outfalls; the other three models failed to correctly identify the sluice, where YOLOv11n misclassified it as a culvert. This likely arises because the key visual distinction—intricate structures and auxiliary equipment—is easily weakened or erased by motion blur, causing a sluice to appear as a simple opening with trailing effects that visually converge with a culvert. For the fifth sample (sluice), YOLOv9t achieved the highest confidence score (0.96). In the final sample (gullies), only YOLOv8n detected both gullies, with confidence scores of 0.82 and 0.57. Overall, YOLOv8n demonstrated the highest detection accuracy across the sample images, detecting every outfall type without a single miss and maintaining the highest overall detection confidence levels.

3.2.2. Comparison of Model Performance

With patience initially set to 20 iterations, all models completed 200 training iterations. Table 6 compares the training times required for each model.

As shown in Table 6, YOLOv8n exhibits the shortest training (264.24 min) and inference (2.6 ms) times, reflecting its robust model architecture. RTDETR-light, owing to its substantial parameter count and relatively complex structure, requires the longest training (1152.72 min) and inference (12.5 ms) times, rendering its real-time performance significantly inferior to the YOLO series.

Table 6. Model detection performance.

Models	Complete Epochs	Train Time (min)	Time per Epoch (min)	Inference Time (ms)
YOLOv8n	200	264.24	1.3212	2.6
YOLOv9t	200	459.48	2.2974	3.4
YOLOv10n	200	357.12	1.7856	3.5
YOLOv11n	200	287.4	1.437	2.8
RTDETR-light	200	1152.72	5.7636	12.5

Figure 15 presents the training and validation loss results for the five models over 200 training iterations. For the four YOLO models, training and validation losses decline steadily, improving both object detection and classification tasks. YOLOv8n exhibits the lowest initial classification training loss (4.571), followed by YOLOv11n (4.774) and YOLOv9t (5.412), both converging to approximately 0.7 by the end of training. YOLOv8n's low initial classification loss and smooth convergence indicate that its feature extractor, pre-trained on large-scale datasets, transfers well to our domain-specific fused imagery. In contrast, YOLOv10n exhibits a notably higher initial classification loss of 11.26, converging to about 1.4 upon training completion, which may explain its slightly lower final performance despite architectural innovations. Box loss and dfl_loss follow a similar downward trend, indicating that model training improved bounding-box accuracy predictions and object region focus. The close alignment of training and validation losses across all YOLO models confirms robust training without overfitting, validating our data augmentation and regularization strategies.

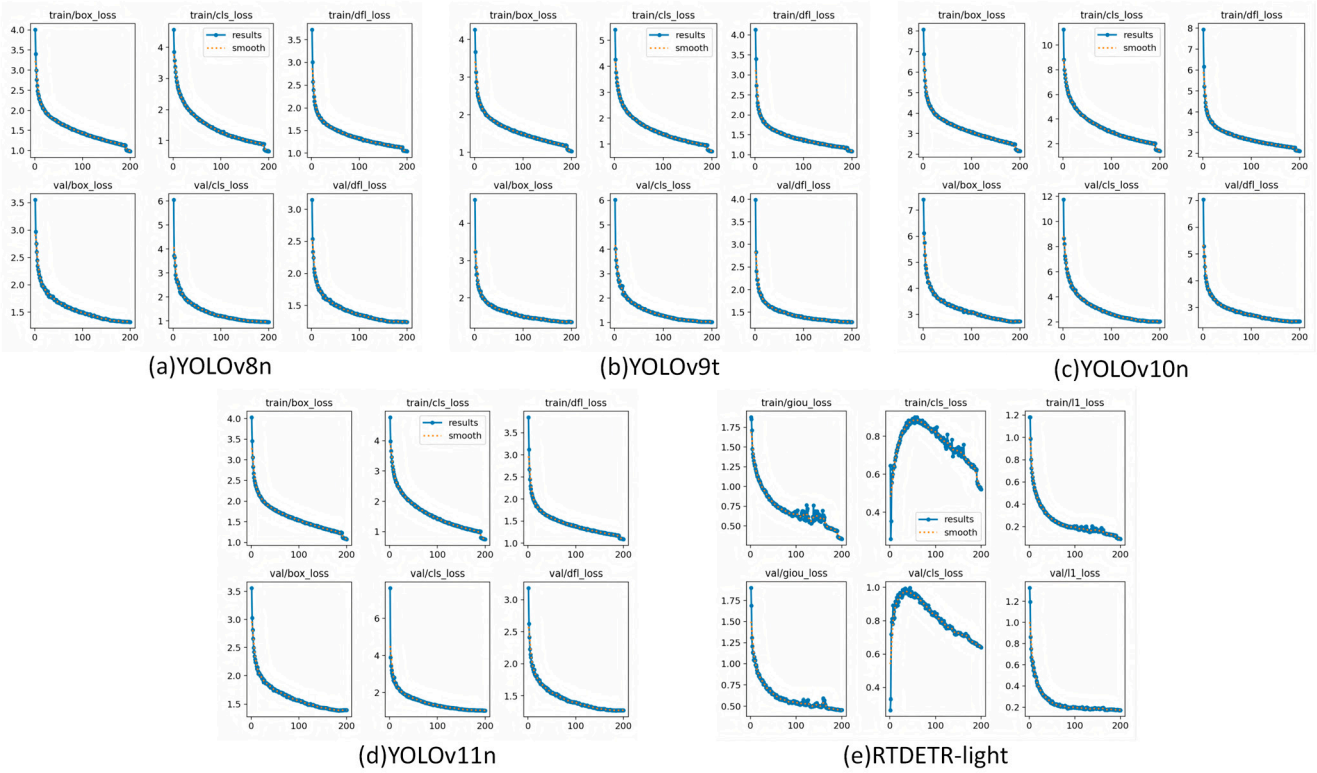

Figure 15. Model Training Loss and Validation Loss Results.

Unlike the YOLO series, RTDETR-light employs giou_loss and l1_loss. Both losses exhibit a steady decline during training and validation, but convergence is noticeably slower (taking more epochs to plateau), reflecting the general challenge of training Transformer-

based detectors from a cooler start in data-limited scenarios. Overall, YOLOv8n exhibits the lowest training and validation losses among all models, demonstrating superior convergence and robustness.

The reliability of the training and validation results was evaluated by examining the performance curves and confidence intervals of the marine outfall detection models. Figure 16 presents the F1 confidence curves for each model.

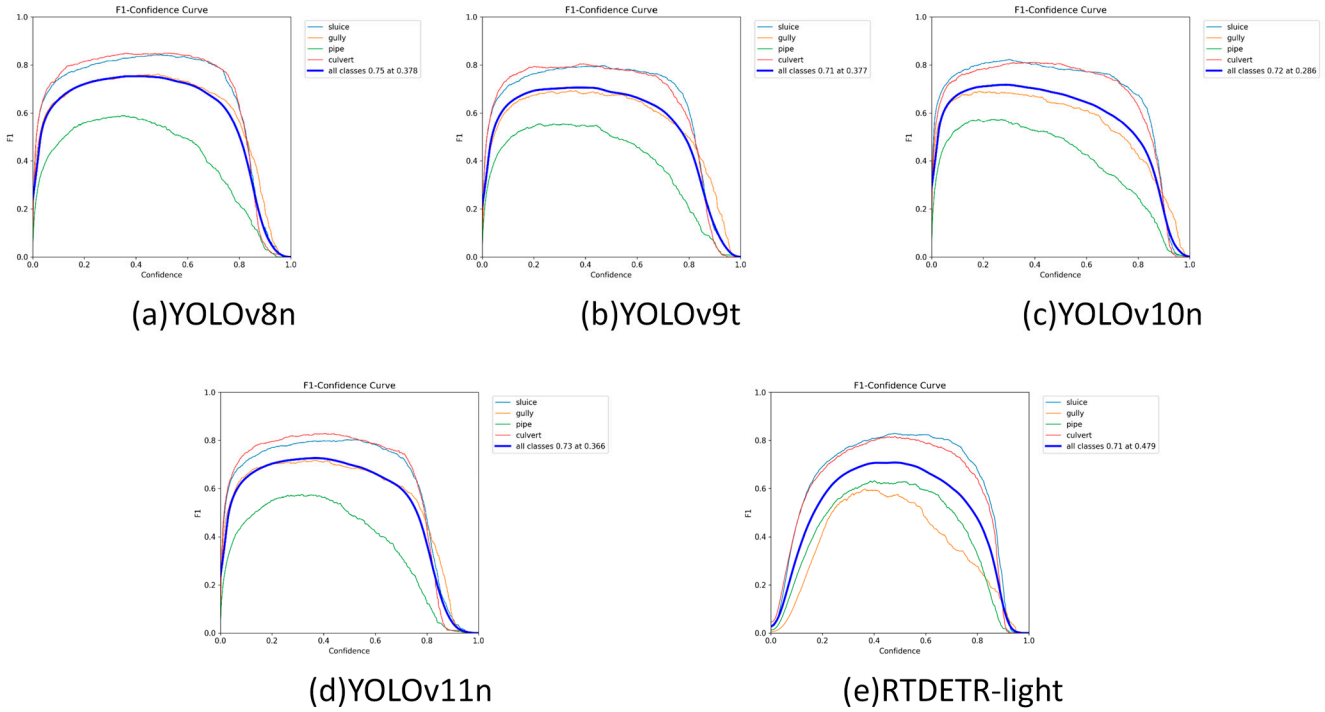

Figure 16. F1 Confidence Curves of the Model in Marine Outfall Detection.

The confidence threshold at the peak of the F1-confidence curve represents the optimal default under most balanced requirements. YOLOv10n achieves its highest F1 score (0.72) at a threshold of 0.286, indicating its strong performance at lower confidence thresholds. RTDETR-light achieves its highest F1 score of 0.71 at a threshold of 0.479, indicating superior performance at higher confidence levels. YOLOv8n possesses the highest F1 score (0.75) among the five models, signifying the best precision-recall balance and superior overall model performance.

4. Discussion

4.1. Comprehensive Model Evaluation

This study systematically compared five mainstream object detection models for marine outfall identification. YOLOv8n achieved the highest mAP50 (77.0%), precision (84.1%), recall (68.6%), CSI (60.7%), and F1-score (0.75), demonstrating superior overall performance. This can be attributed to its balanced architecture tailored for complex, small-target detection. Replacing the C3 module with the C2F module enhances gradient flow and feature representation. Its anchor-free paradigm further simplifies the detection pipeline and improves cross-scale generalization in coastal imagery. Although newer YOLO variants introduced innovations such as PGI and dual-head assignment, they did not outperform YOLOv8n on this task. This suggests that for marine outfall detection—with limited data variability relative to general object detection—architectural stability and efficient multi-scale feature fusion may be more advantageous than the latest advances aimed at extreme speed or parameter reduction. While RTDETR-light achieved competitive recall for certain

categories (e.g., pipes), its performance was hampered by higher computational complexity and longer convergence time, consistent with known limitations of Transformer-based models in data-constrained domains. The proposed RTDETR-light (15.5 M) narrows the substantial parameter gap with the original RTDETR-l (32.8 M), yet its size and computational complexity remain greater than those of the YOLO model. Overall, this comparison highlights architectural differences and demonstrates the potential of RTDETR-light for practical deployment.

4.2. Impact of Multi-Source Data Fusion on Detection Performance

Integrating NDWI into a multi-source fusion strategy proved pivotal for enhancing detection robustness. By increasing water–land spectral contrast, the fused imagery effectively suppressed interference from vegetation, soil, and shadows—common false-positive sources in coastal scenes. The consistently high detection confidence across diverse environmental simulations demonstrates that this preprocessing step yields a more invariant feature representation. This addresses a critical limitation of RGB-only imagery, where outfall signatures are often obscured by spectral confusion. Thus, the fusion strategy is not merely data enhancement but a targeted solution to the inherent spectral ambiguity, steering the model's focus toward hydrologically relevant features.

4.3. Model Lightweighting and Real-Time Analysis

For UAV or edge deployment, the accuracy–efficiency trade-off is paramount. YOLOv8n achieved the optimal balance, combining the shortest training and inference times with the highest accuracy. Its low GFLOPS and parameter count suit resource-constrained real-time systems. In contrast, despite its elegant end-to-end design, RTDETR-light's substantially higher computational cost (37.9 GFLOPS) and slower inference (12.5 ms) render it impractical for latency-critical real-time monitoring. This finding reinforces that in applied remote sensing, model selection must prioritise deployment feasibility alongside benchmark accuracy. YOLOv8n's efficiency makes it a more viable candidate for operational marine outfall monitoring.

4.4. Analysis of Performance Limitations for Specific Marine Outfall Types

Table 5 reveals significant performance disparities among marine outfall types. Models achieved satisfactory accuracy for sluices, gullies, and culverts (YOLOv8n attained mAP50 > 77% for all three), whereas pipe detection remained consistently poor. The best-performing YOLOv8n reached only 47.6% recall and 58.0% mAP50 for pipes. This likely stems from pipes being linear, narrow structures that present fewer distinctive visual features than larger outlets, making their signatures easily obscured by complex coastal backgrounds. Even in high-resolution UAV imagery, pipes are among the smallest targets, and general-purpose detectors struggle with such fine linear features. NDWI fusion enhances land–water contrast and benefits outfalls connected to large water bodies, but pipes with minimal or intermittent discharge may generate weak NDWI signals, limiting the fusion's advantage.

The low pipe detection rate indicates that some outfalls may be missed under suboptimal conditions; however, this does not negate the method's overall applicability, as the models reliably identify the majority of dominant outfall structures. For critical pipe detection, RTDETR-light—lower overall mAP—offers relatively stronger pipe recall (57.3%), indicating that Transformer-based architectures can better capture contextual and structural relationships in challenging categories.

4.5. Limitations and Generalizability

This study developed a dataset and evaluated models on Zhanjiang's southern coast, providing an effective technical framework and model selection basis for intelligent marine outfall identification. However, the experiments rely primarily on a single geographic region. Although environmental simulations enhanced robustness to local variations, the models were not validated on independent datasets from regions with distinctly different geographic and climatic characteristics. This partly limits the universality of the findings, and further testing is required.

The study also demonstrates partial generalization potential through NDWI integration in a multi-source remote sensing fusion dataset. As a standardized method that enhances water body features while suppressing vegetation and soil interference, NDWI's physical principles apply broadly across coastal environments, suggesting transferability of the NDWI-based strategy. The four defined outfall types share common engineering characteristics in morphology and function, representing typical coastal discharge facilities. The selected YOLO and RT-DETR models inherently possess strong feature learning capabilities, enabling rapid adaptation to new backgrounds through fine-tuning on regional labelled data.

Nevertheless, direct model transfer still faces challenges. Differences in seawater color, turbidity, and tidal flat texture may affect NDWI calculation accuracy; outfall materials, scale, and environmental contrast vary across regions; and legal and planning factors may lead to differing outfall densities and distribution patterns. Therefore, to comprehensively evaluate model robustness and generalization, future efforts should focus on collecting marine outfall imagery from diverse coastal zones across China and globally—rocky shores, sandy shores, mudflats, urban shorelines, and industrial shorelines—to construct a more diverse cross-regional benchmark dataset. Cross-validation based on this dataset should enable quantitative assessment of generalization performance.

4.6. Practical Significance and Recommendations

This benchmark provides clear data-driven guidance for deploying intelligent marine outfall monitoring systems. For maximal accuracy and balance, YOLOv8n is recommended. It optimally balances precision and recall, offering efficient training and rapid inference, and performs stably across diverse outfall scenarios, making it a reliable general-purpose solution, especially suitable for real-time UAV applications. When minimizing missed pipe detections is the operational goal, RTDETR-light may be considered despite its higher computational cost. While its superior pipe recall may require more powerful onboard hardware, its lower accuracy necessitates manual verification steps to filter false positives. Research findings caution against blindly adopting the latest model versions (YOLOv10, v11) without task-specific validation. Their architectural optimizations have not translated into significant advantages in this domain, reaffirming the importance of empirical evaluation over chasing cutting-edge technology on generic benchmarks.

5. Conclusions

This study systematically benchmarks state-of-the-art real-time object detectors for marine outfall detection using a newly constructed multi-source remote sensing dataset fused with NDWI. YOLOv8n offered the best accuracy–efficiency balance, with the highest mAP50 (77.0%) and F1-score (0.75), and fast inference. The NDWI-based fusion strategy effectively enhanced water–land contrast and detection robustness in complex coastal environments.

All models struggled with pipes—the most challenging category owing to its small size, weak visual features, and high background similarity—indicating that small, low-feature target detection remains a key bottleneck in practical monitoring. Lightweight CNN-based

YOLO models proved more suitable for UAV edge deployment, whereas Transformer-based detectors showed higher recall for tiny targets at greater computational cost.

This work provides a reproducible benchmark and practical model selection reference for intelligent marine outfall monitoring, supporting the development of automated coastal pollution supervision. The multi-source fusion method and experimental framework may also serve as a reference for similar remote sensing detection tasks.

Future work could explore: (1) constructing a larger, cross-regional dataset covering diverse coastal landforms, climates, and outfall types to improve model generalization; (2) designing targeted improvements for small, low-feature pipes—such as specialized attention mechanisms, multi-scale feature enhancement, or tailored loss functions—to overcome this bottleneck; and (3) investigating lighter Transformer-based or hybrid CNN–Transformer architectures for a better accuracy–speed–cost balance in real-time UAV-based monitoring.

Author Contributions: Conceptualization, L.Y.; methodology, L.Y.; software, L.Y.; validation, L.Y., S.L., S.Z. and R.Y.; formal analysis, L.Y. and S.L.; investigation, L.Y., S.L., S.Z. and R.Y.; resources, H.Z.; data curation, L.Y.; writing—original draft preparation, L.Y.; writing—review and editing, H.Z., S.L., S.Z. and R.Y.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC was funded by Haolan Zhou.

Data Availability Statement: The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Sile Li was employed by Guangzhou Panyu Urban Drainage Management Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Ma, J.; Wu, Z.; Guo, M.; Hu, Q. Dynamic relationship between marine fisheries economic development, environmental protection and fisheries technological Progress—A case of coastal provinces in China. *Ocean Coast. Manag.* **2024**, *247*, 106885. [CrossRef]
2. Yuan, B.; Cui, Y.; An, D.; Jia, Z.; Ding, W.; Yang, L. Marine environmental pollution and offshore aquaculture structure: Evidence from China. *Front. Mar. Sci.* **2023**, *9*, 979003. [CrossRef]
3. Liang, H.; Zhang, J.; Zhang, J.; Zhang, P.; Deng, X.; Chen, J.; Wang, Z.; Long, C.; Lu, C.; Wang, D.; et al. Unveiling the eutrophication crisis: 20 years of nutrient development in Zhanjiang Bay, China. *Front. Mar. Sci.* **2024**, *11*, 1373716. [CrossRef]
4. Qian, X. Research on the coordinated development model of marine ecological environment protection and economic sustainable development. *J. Sea Res.* **2023**, *193*, 102377. [CrossRef]
5. Lincoln, S.; Chowdhury, P.; Posen, P.E.; Robin, R.; Ramachandran, P.; Ajith, N.; Harrod, O.; Hoehn, D.; Harrod, R.; Townhill, B.L. Interaction of climate change and marine pollution in Southern India: Implications for coastal zone management practices and policies. *Sci. Total Environ.* **2023**, *902*, 166061. [CrossRef] [PubMed]
6. Ling, M.; Cheng, Q.; Peng, J.; Jiang, L.; Wang, R. Retrieval Algorithm of Water Pollutant Concentration Based on UAV Remote Sensing Technology. *Mob. Inf. Syst.* **2022**, *2022*, 5017000. [CrossRef]
7. Cao, Z.; Kooistra, L.; Wang, W.; Guo, L.; Valente, J. Real-time object detection based on uav remote sensing: A systematic literature review. *Drones* **2023**, *7*, 620. [CrossRef]
8. Bondur, V.; Vorobyev, V.; Zamshin, V.; Serebryany, A.; Latushkin, A.; Li, M.; Martynov, O.; Hurchak, A.; Grinchenko, D. Monitoring anthropogenic impact on some coastal water areas of the Black Sea using multispectral satellite imagery. *Izv. Atmos. Ocean. Phys.* **2018**, *54*, 1008–1022. [CrossRef]
9. Huang, Y.; Wu, C.; Yang, H.; Zhu, H.; Chen, M.; Yang, J. An improved deep learning approach for retrieving outfalls into rivers from UAS imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4703814. [CrossRef]
10. Xu, H.; Huang, Q.; Yang, Y.; Li, J.; Chen, X.; Han, W.; Wang, L. UAV-ODS: A real-time outfall detection system based on UAV remote sensing and edge computing. In Proceedings of the 2022 IEEE International Conference on Unmanned Systems (ICUS), Guangzhou, China, 28–30 October 2022; pp. 1–9.
11. Huang, Y.; Wu, C. Evaluation of deep learning benchmarks in retrieving outfalls into rivers with UAS images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4703912. [CrossRef]

12. Yu, M.; Zhang, J.; Zhu, L.; Liang, S.; Lu, W.; Ji, X. An Intelligent System for Outfall Detection in UAV Images Using Lightweight Convolutional Vision Transformer Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 6265–6277. [CrossRef]
13. Huang, Q.; Fan, J.; Xu, H.; Han, W.; Huang, X.; Chen, Y. AFENet: Attention-guided feature enhancement network and a benchmark for low-altitude UAV sewage outfall detection. *Array* **2024**, *22*, 100343. [CrossRef]
14. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. Detrs beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2024; pp. 16965–16974.
15. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [CrossRef]
16. Teng, J.; Xia, S.; Liu, Y.; Yu, X.; Duan, H.; Xiao, H.; Zhao, C. Assessing habitat suitability for wintering geese by using Normalized Difference Water Index (NDWI) in a large floodplain wetland, China. *Ecol. Indic.* **2021**, *122*, 107260. [CrossRef]
17. Maashi, M.; Alzaben, N.; Negm, N.; Veeramani, V.; Begum, S.S.; Palaniappan, G. Forecasting land use changes in crop classification and drought using remote sensing. *J. Arid Land* **2025**, *17*, 575–589. [CrossRef]
18. Ramos, L.T.; Sappa, A.D. A comprehensive analysis of YOLO architectures for tomato leaf disease identification. *Sci. Rep.* **2025**, *15*, 26890. [CrossRef]
19. Wang, X.; Gao, H.; Jia, Z.; Li, Z. BL-YOLOv8: An improved road defect detection model based on YOLOv8. *Sensors* **2023**, *23*, 8361. [CrossRef]
20. Wu, T.; Dong, Y. YOLO-SE: Improved YOLOv8 for remote sensing object detection and recognition. *Appl. Sci.* **2023**, *13*, 12977. [CrossRef]
21. Yang, G.; Wang, J.; Nie, Z.; Yang, H.; Yu, S. A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention. *Agronomy* **2023**, *13*, 1824. [CrossRef]
22. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors* **2023**, *23*, 7190. [CrossRef]
23. Wang, C.-Y.; Yeh, I.-H.; Mark Liao, H.-Y. Yolov9: Learning what you want to learn using programmable gradient information. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2024; pp. 1–21.
24. Chien, C.T.; Ju, R.Y.; Chou, K.Y.; Chiang, J.S. YOLOv9 for fracture detection in pediatric wrist trauma X-ray images. *Electron. Lett.* **2024**, *60*, e13248. [CrossRef]
25. Hua, Z.; Aranganadin, K.; Yeh, C.-C.; Hai, X.; Huang, C.-Y.; Leung, T.-C.; Hsu, H.-Y.; Lan, Y.-C.; Lin, M.-C. A benchmark review of yolo algorithm developments for object detection. *IEEE Access* **2025**, *13*, 123515–123545. [CrossRef]
26. Alkhammash, E.H. A Comparative Analysis of YOLOv9, YOLOv10, YOLOv11 for Smoke and Fire Detection. *Fire* **2025**, *8*, 26. [CrossRef]
27. Yaseen, M. What is YOLOv9: An in-depth exploration of the internal features of the next-generation object detector. *arXiv* **2024**, arXiv:2409.07813.
28. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J. Yolov10: Real-time end-to-end object detection. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 107984–108011.
29. Hussain, M. Yolov5, yolov8 and yolov10: The go-to detectors for real-time vision. *arXiv* **2024**, arXiv:2407.02988.
30. Santos Júnior, E.S.d.; Paixão, T.; Alvarez, A.B. Comparative performance of YOLOv8, YOLOv9, YOLOv10, and YOLOv11 for layout analysis of historical documents images. *Appl. Sci.* **2025**, *15*, 3164. [CrossRef]
31. Bento, J.; Paixão, T.; Alvarez, A.B. Performance Evaluation of YOLOv8, YOLOv9, YOLOv10, and YOLOv11 for Stamp Detection in Scanned Documents. *Appl. Sci.* **2025**, *15*, 3154. [CrossRef]
32. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]
33. He, L.-h.; Zhou, Y.-z.; Liu, L.; Cao, W.; Ma, J.-h. Research on object detection and recognition in remote sensing images based on YOLOv11. *Sci. Rep.* **2025**, *15*, 14032. [CrossRef]
34. Vrsalovic, I.; Lerga, J.; Ivasic-Kos, M. A System for Real-Time Detection of Abandoned Luggage. *Sensors* **2025**, *25*, 2872. [CrossRef] [PubMed]
35. Kong, Y.; Shang, X.; Jia, S. Drone-DETR: Efficient small object detection for remote sensing image using enhanced RT-DETR model. *Sensors* **2024**, *24*, 5496. [CrossRef] [PubMed]
36. Park, G.; Lee, Y. Wildfire smoke detection enhanced by image augmentation with StyleGAN2-ADA for YOLOv8 and RT-DETR models. *Fire* **2024**, *7*, 369. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.