

Article

CoastCor-Net: A Wind Turbine Blade Defect Detection Network for Coastal Environments

Jiawei Xiang ^{1,†} , Xinyu Wan ^{2,†}  and Shoudong Ni ^{1,*}¹ College of Mechanical and Power Engineering, Nanjing Tech University, Nanjing 211816, China; Jiawei@njtech.edu.cn² School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, Nibong Tebal 14300, Penang, Malaysia; wanxinyu@student.usm.my

* Correspondence: nsd@njtech.edu.cn

† These authors contributed equally to this work.

Abstract

Coastal wind turbines operate under severe salt spray, high humidity, and wind-driven erosion, which accelerate coating degradation and corrosion-induced cracking. In such environments, corrosion defects exhibit blurred boundaries, weak textures, and significant scale variations, challenging object detectors in small-target localization and precise boundary regression. To address these limitations, this study proposes CoastCor-Net, an enhanced YOLOv11-based framework that improves spatial–semantic alignment, boundary representation, and channel–spatial dependency modeling. The architecture integrates three complementary modules to enhance boundary sensitivity, spatial–semantic consistency, and cross-channel interaction: a Decoding-Driven Enhancement Block, a Complementary Feature Alignment Module, and a Channel-Transposed Coordinate Attention module. Extensive experiments on the Wind Turbine Blade Damage Dataset show that CoastCor-Net achieves 84.7% mAP@0.5 and 54.1% mAP@0.5:0.95, surpassing YOLOv13n by 3.2 percentage points in mAP@0.5 and improving AP_{damage} by 5.2 percentage points. The framework also demonstrates strong robustness under composite coastal perturbations. These findings highlight the practical effectiveness of structured multi-level feature enhancement for reliable and high-precision blade inspection in complex coastal environments.

Keywords: wind turbine; corrosion; object detection; YOLO; attention mechanism

1. Introduction

With the continuous expansion of offshore wind power capacity, wind turbines in coastal and nearshore areas have become an important component of the clean energy system [1,2]. However, the marine environment accelerates blade-coating corrosion, surface deterioration, and crack growth, thereby affecting structural safety and power-generation efficiency [3,4]. If blade defects are not detected and addressed promptly, they may aggravate structural fatigue and even lead to serious safety accidents [5]. Therefore, constructing a high-precision and highly robust visual detection method to achieve automatic identification and localization of corrosion defects on coastal wind turbine blades is of great engineering significance and practical value [6].

With the widespread adoption of UAV inspection technology, high-resolution blade images can be collected in batches within a short period, significantly reducing the risks associated with manual high-altitude operations [7]. However, the surge in image quantity also makes manual screening inefficient and costly. In recent years, deep learning-based



Academic Editor: Cecilia Bartuli

Received: 15 February 2026

Revised: 2 March 2026

Accepted: 5 March 2026

Published: 16 March 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

object detection methods have achieved significant progress in the field of industrial defect inspection [8–10]. Single-stage detectors, such as the YOLO series, are widely used in real-time inspection scenarios due to their concise architecture, fast inference speed, and ease of deployment. Their objective is to achieve stable recognition and precise localization of multiple types of defects while maintaining real-time performance. However, compared with conventional industrial surface inspection scenarios, coastal environments pose more stringent challenges to visual detection.

First, salt spray scattering and high-humidity environments significantly reduce image contrast and introduce fog-like scattering noise, weakening the clarity of local texture boundaries. In this context, corrosion commonly observed in the dataset usually manifests as localized dark depressions or pit-like textures, which are small in scale and irregular in shape, and occur more frequently in salt-rich environments, posing challenges to the model's ability to perceive weak high-frequency features.

Second, blade erosion is typically caused by long-term impacts from raindrops, hail, and other factors, accompanied by material stress accumulation and coating degradation, and manifests as rough edges, local peeling, and fragmented textures. Such defects are often confused with background coating textures or illumination reflections, making it difficult for the model to distinguish real defects from natural texture variations [4].

Third, cracks usually appear as slender, low-contrast, cross-scale structures, with widths far smaller than their lengths, often spanning multiple texture regions. In addition, fiber textures, coating particles, and stain interference on the blade surface further cause the detector to suffer from missed detections or poorly fitted bounding boxes in small-object recognition and bounding box regression.

Through systematic observation and experimental analysis of coastal blade defect images, we find that the performance bottlenecks of current detection models in this scenario mainly stem from three capability limitations: shallow spatial details are difficult to effectively transmit to deep semantic representations during progressive downsampling, resulting in localization drift of small-scale corrosion or fine crack targets during feature fusion; the boundaries of erosion and crack regions are often blurred, and the model lacks mechanisms to strengthen local edge structures, making precise regression difficult; the dependency relationship between channel features and spatial features is insufficiently modeled, leading to significant fluctuations in class confidence under complex background interference.

To address the above issues, this study proposes CoastCor-Net, a multi-scale dense prediction detection network for coastal salt spray environments. The method enhances detection performance through three complementary mechanisms: spatial–semantic alignment, dilated boundary enhancement, and channel-transposed attention, implemented via CFAM, DDEB, and CTCA within a unified convolutional framework. The three modules form a collaborative optimization mechanism at the levels of feature alignment, boundary enhancement, and dependency modeling, thereby improving the detection accuracy and robustness of corrosion, erosion, and crack defects in complex environments.

An experimental platform is constructed on the public dataset Wind Turbine Blade Damage Dataset, and comparative validation is conducted around different module combinations and multiple mainstream detection methods [11]. The experimental results demonstrate that the proposed improved framework outperforms existing detection methods in terms of detection accuracy, recall, and boundary stability, especially showing significant advantages in fine-grained corrosion and erosion target recognition.

The main contributions of this paper can be summarized in three aspects: First, an improved single-stage object detection framework for coastal wind turbine blade corrosion detection is constructed to achieve collaborative optimization of spatial and semantic fea-

tures; Second, three key modules—CFAM, DDEB, and CTCA—are proposed and integrated to systematically address problems such as feature mismatch, limited receptive field, and insufficient channel dependency modeling; Third, the effectiveness and synergistic gain of each module are verified through systematic comparative experiments and ablation analysis. Through the above improvements, CoastCor-Net demonstrates superior accuracy and stability over existing detection methods in complex coastal wind turbine blade defect detection tasks, providing an engineerable visual inspection solution for offshore wind turbine intelligent inspection and structural health monitoring.

The remainder of this paper is organized as follows: Section 2 introduces related research progress; Section 3 describes the overall detection framework and key module design; Section 4 presents the experimental results; Section 5 discusses the findings; Section 6 concludes the paper and outlines future work.

2. Related Work

Blade coating-related defects, including aging, blistering, peeling, and corrosion, are characterized by weak textures, blurred boundaries, significant scale variations, and complex background interference.

In the field of wind turbine blade surface defect detection, studies commonly adopt UAV inspection images as data sources and enhance sensitivity to small-scale and weak-texture defects by improving feature fusion and attention mechanisms [12]. To address the problem of missed detections of blade surface defects under complex backgrounds, Liu et al. enhanced the interaction between multi-scale semantic and detailed information by introducing attention mechanisms and bidirectional feature fusion, thereby improving the detection rate of blade surface defects and robustness under occlusion scenarios [13]. Furthermore, in the broader context of industrial surface defect inspection, Meng et al. proposed a coordinate modeling attention module to capture long-range dependencies while maintaining spatial interpretability, facilitating direct integration into convolutional backbones to improve performance in fine-grained defect localization tasks [14]. In machine vision evaluation of wind turbine blades, Xu et al. emphasized automated diagnostic processes for multiple types of damage, combining signal processing and feature engineering to enhance usability across operating conditions, thus providing another reference approach for engineering deployment [15].

In corrosion detection, corrosion recognition methods based on single-stage object detection frameworks have attracted extensive attention in marine engineering and industrial structural fields, particularly with continuous improvements around the YOLO series models. To address issues such as weak textures, small scales, and irregular boundaries of corrosion regions in complex environments, researchers have explored feature enhancement, loss optimization, and attention modeling from multiple perspectives. Yu et al. proposed an improved YOLOv5-GOLD-NWD model for corrosion detection on coated metal surfaces. By introducing normalized Wasserstein distance loss and structural optimization strategies, the localization accuracy of small-scale corrosion regions was improved, and the effectiveness of the model was validated in coastal metal surface corrosion assessment tasks [16]. Subsequently, Yu et al. further constructed an improved detection framework integrating EfficientViT, NWD, and channel attention mechanisms for corrosion recognition and coating performance evaluation in marine exposure environments, emphasizing the importance of lightweight structures and multi-scale feature enhancement under complex salt spray conditions [17]. These two studies indicate that optimizing loss functions and introducing efficient attention structures can effectively improve the accuracy and robustness of corrosion detection. Cheng and Kang proposed a detection and grading method based on an improved YOLOv10 series architecture for hydraulic metal structure corrosion

detection and classification. While maintaining real-time performance, the method enhanced the recognition capability for irregular corrosion regions [18]. This study combined corrosion detection with corrosion grade discrimination, providing a new perspective for quantitative assessment of corrosion states. Chen et al. focused on metal surface corrosion grade recognition and proposed an improved YOLOv8 model. By integrating attention mechanisms and data augmentation strategies, precise classification and localization of multi-level corrosion states were achieved. This work demonstrated that collaborative modeling of channel and spatial information is crucial for improving detection stability in fine-grained corrosion classification scenarios [19]. Yu et al. also conducted a systematic comparative analysis of multiple improved YOLOv5 models in coastal corrosion detection scenarios, indicating that different attention modules and loss designs have significant impacts on recall and boundary fitting performance for small-scale corrosion targets, thus providing experimental evidence for subsequent structural optimization [20].

Existing corrosion detection methods primarily focus on three directions: (1) strengthening spatial and channel feature representations through attention mechanisms [21], (2) enhancing the recognition capability of small-scale corrosion regions through multi-scale feature fusion [22], and (3) balancing accuracy and real-time performance through structural optimization and lightweight design. However, in the scenario of coastal wind turbine blade corrosion, problems such as mismatch between spatial details and semantic information, insufficient cross-scale information interaction, and unstable boundary localization still persist. Therefore, existing detection frameworks require systematic structural improvements to achieve high-precision corrosion detection under complex salt spray environments.

3. Materials and Methods

3.1. Overall Architecture

The proposed CoastCor-Net is constructed based on structural improvements to the YOLOv11 single-stage object detection framework. The overall network still adopts a multi-scale dense prediction mechanism, while three enhancement modules—DDEB, CFAM, and CTCA—are embedded at key positions in the backbone and detection head to improve the detection accuracy and robustness of blade corrosion, erosion, and crack defects under coastal salt spray environments. In practical deployment, the UAV-captured images are transmitted to a remote workstation for online processing, where the proposed network performs defect detection in near real-time.

UAV inspection images first pass through the Stem convolutional layer for initial feature extraction and downsampling, producing base feature maps. The network then enters the backbone stage, which is composed of multiple stacked Conv and C3k2 structures, generating multi-scale feature representations through progressive downsampling. Compared with the original YOLOv11 architecture, this study embeds DDEB into the C3k2 module of Stage1, replacing the original Bottleneck structure to enhance early boundary and fine-grained texture representation capability. In Stage2 and Stage3, the original C3k2 modules are replaced with CFAM to achieve complementary alignment between spatial details and high-level semantics. Finally, a CTCA module is inserted before the first Conv2d layer of both the classification branch and regression branch in the detection head, strengthening the coupled modeling ability of channel and spatial features, and improving class confidence stability and boundary regression accuracy. Figure 1 illustrates the overall architecture of CoastCor-Net.

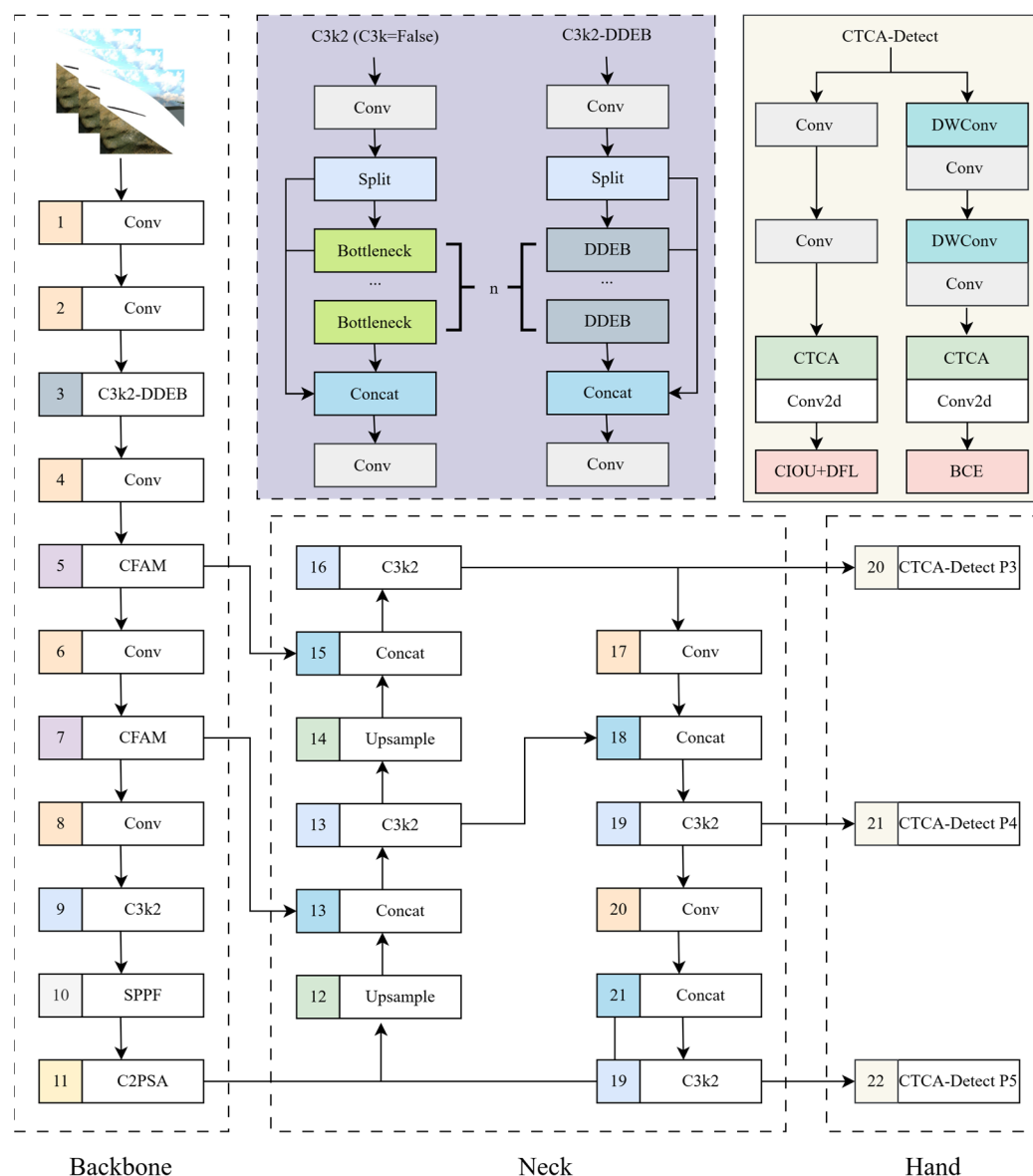


Figure 1. Overall architecture of CoastCor-Net.

3.2. DDEB Design

In the overall network, Stage1 is mainly responsible for extracting shallow spatial features, and its output retains relatively rich edge and texture information. However, under salt spray scattering and low-contrast conditions, corrosion pit-like textures and fine cracks often exhibit blurred boundaries. The receptive field of traditional convolution kernels is limited, making it difficult to simultaneously account for local sharpening and contextual modeling. To enhance the representation capability of corrosion and crack boundaries in coastal environments, this paper embeds the DDEB module into the C3k2 module of Stage1 in the YOLOv11 backbone. Its structure is shown in Figure 2.

Let the input feature be $X \in \mathbb{R}^{C \times H \times W}$. The input is first normalized and mapped by convolution, and then multi-scale spatial responses are extracted through three parallel convolutions with different dilation rates. The spatial enhancement process can be uniformly expressed as

$$Y_1 = X + \mathcal{S}(X) \quad (1)$$

where $\mathcal{S}(\cdot)$ denotes a composite mapping function including multi-scale depthwise dilated convolutions, Spatial Gating (SG), and Spatial-Channel Attention (SCA) [23]. The multi-

scale depthwise dilated convolutions adopt dilation rates of 1, 4, and 9 to simultaneously model local textures, erosion band structures, and cross-scale crack information. SG generates spatial response weights to suppress salt spray scattering noise regions, while SCA further models the dependency between channel and spatial dimensions, thereby enhancing the response intensity to real defect regions.

After completing spatial enhancement, the module enters the gated feed-forward stage. To improve nonlinear representation capability while maintaining computational efficiency, channel expansion and gating mechanisms are adopted for feature recombination, and the overall formulation is

$$Y = Y_1 + \mathcal{F}(Y_1) \quad (2)$$

where $\mathcal{F}(\cdot)$ denotes the gated feed-forward mapping function. This branch first performs a 1×1 convolution to expand the channel dimension, then completes feature selection and fusion through a gating unit, and finally compresses back to the original number of channels and forms a second residual connection with the input.

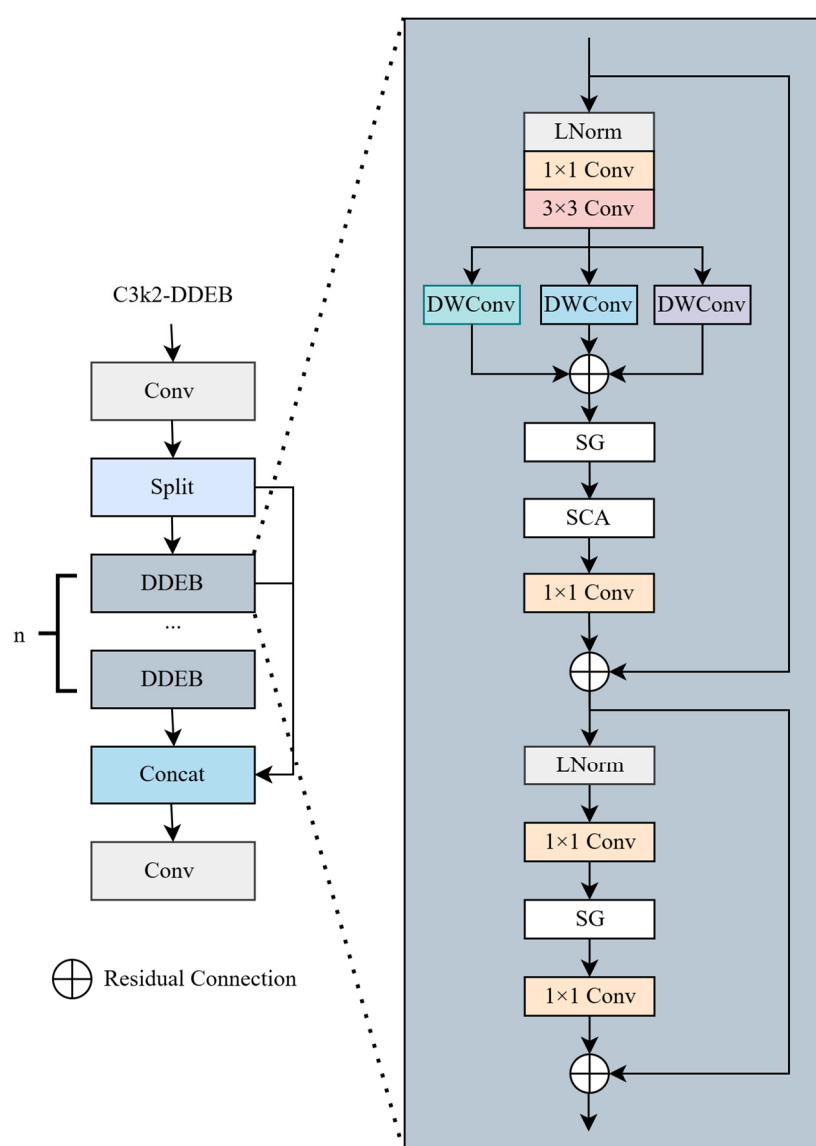


Figure 2. Network architecture of the DDEB module.

Combining the above two-stage structure, the overall mapping of DDEB can be simplified as

$$Y = X + \mathcal{F}(X + \mathcal{S}(X)) \quad (3)$$

This dual-residual structure ensures training stability while significantly enhancing boundary sensitivity at the shallow stage. Since salt spray and humidity in coastal environments reduce image contrast, corrosion and crack regions often present blurred boundaries and fine-grained texture degradation. By enlarging the receptive field and introducing a spatial–channel collaborative modeling mechanism, DDEB effectively alleviates the problems of localization drift for small targets and poorly fitted boundaries.

3.3. CFAM Design

To alleviate the spatial–semantic imbalance of corrosion defects in deep features under coastal environments, this paper replaces the original C3k2 structures in Stage2 and Stage3 of the backbone network with CFAM (Figure 3). The module follows the overall design philosophy of channel splitting, directional transformation, complementary mapping, and fused output. By explicitly constructing the interaction between spatial and semantic dual branches, it achieves effective alignment and compensation of shallow spatial information toward deep semantic features.

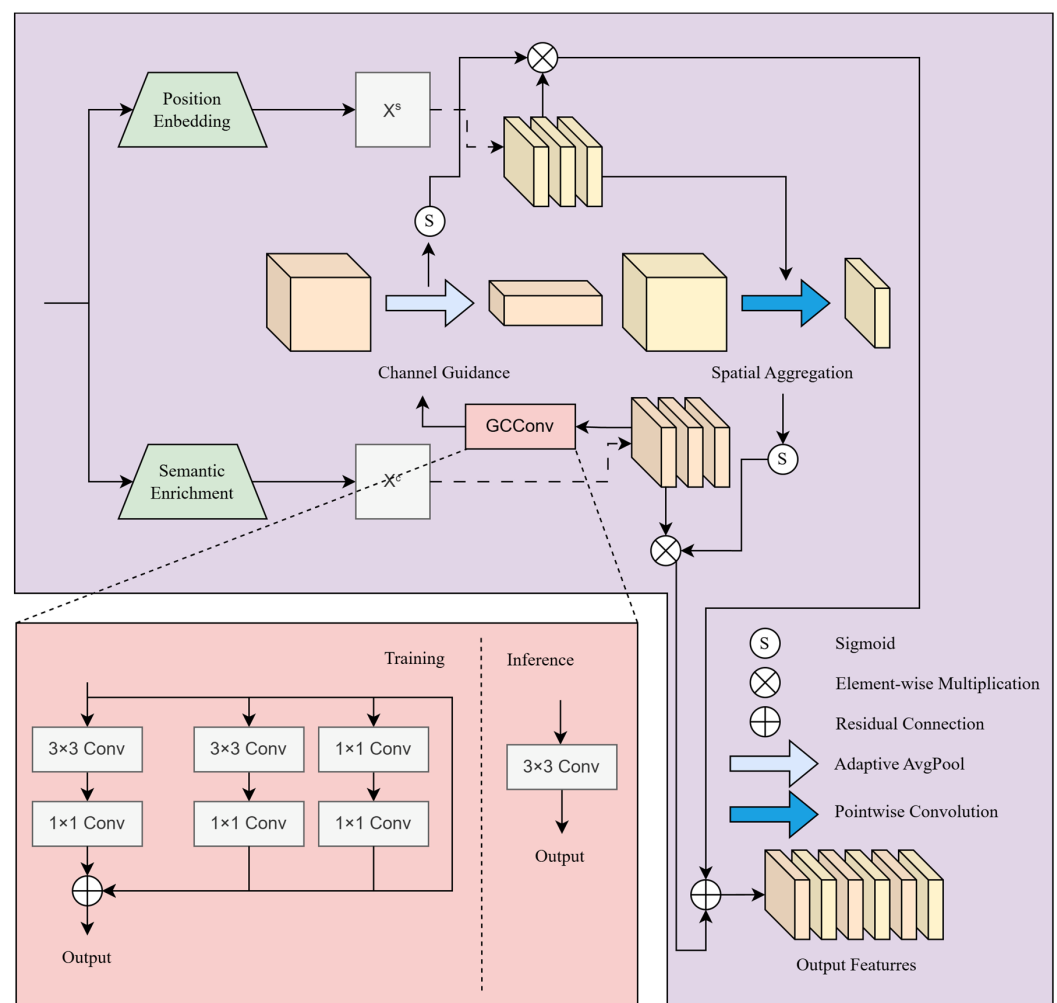


Figure 3. Network architecture of the CFAM module.

Let the input feature be $X \in \mathbb{R}^{C \times H \times W}$. CFAM first performs proportional splitting along the channel dimension, dividing the feature into two parts:

$$(X_s, X_c) = \text{Split}(X) \quad (4)$$

where $X_s \in \mathbb{R}^{\alpha C \times H \times W}$ focuses on preserving spatial positional information, and $X_c \in \mathbb{R}^{(1-\alpha)C \times H \times W}$ emphasizes semantic representation capability, with α denoting the splitting coefficient. Since Stage2 and Stage3 already possess a certain degree of semantic abstraction capability, a “spatial-priority in deeper layers” proportional strategy is adopted, allowing deeper layers to retain more spatial details to enhance localization stability.

In the directional transformation stage, the spatial branch adopts pointwise convolution to preserve the original structural information and avoid excessive destruction of low-level textures; the semantic branch adopts a 3×3 convolution to enhance cross-channel correlation modeling capability. The two branches produce

$$X'_s = \phi_s(X_s), X'_c = \phi_c(X_c) \quad (5)$$

The module then enters the complementary mapping stage. Unlike simple concatenation or weighted summation, CFAM establishes bidirectional dependency through dual interaction mechanisms of channel guidance and spatial guidance.

First, channel mapping is performed on the semantic branch. A channel-wise convolution and global average pooling are applied to X'_c to obtain a channel weight vector $\omega_c \in \mathbb{R}^{C \times 1 \times 1}$. After Sigmoid activation, it is mapped to the spatial branch, achieving the modulation effect of “strong semantics guiding weak spatial features.”

Subsequently, spatial mapping is constructed on the spatial branch. Through GCConv and a normalization layer, a spatial attention map $\omega_s \in \mathbb{R}^{1 \times H \times W}$ is obtained. After Sigmoid processing, it is mapped to the semantic branch, achieving the effect of “strong structural information inversely correcting semantics” [24].

The final output feature is expressed as

$$X_{CFAM} = (X'_c \odot \omega_s) + (X'_s \odot \omega_c) \quad (6)$$

where $\odot$ denotes element-wise multiplication. This expression only retains the core complementary relationship without introducing additional complex formulations, ensuring structural conciseness and implementability.

From a structural perspective, CFAM maintains a lightweight convolution-based implementation and does not introduce global self-attention computation. Therefore, the computational complexity mainly stems from two lightweight convolutions and one global pooling operation, resulting in extremely low additional parameter overhead. Its core advantage lies in embedding shallow spatial localization information into deep semantic representations through weight modulation, thereby enhancing boundary fitting capability and small-target recall for fine-grained defects such as corrosion, erosion, and cracks.

CFAM is positioned at Stage2 and Stage3, serving as a mid-level alignment module. It forms front-back collaboration with DDEB in Stage1: the former is responsible for spatial-semantic structural alignment, while the latter focuses on deblurring and boundary enhancement. After their coordination, the network maintains stable localization and classification performance under complex environments.

3.4. CTCA Design

To enhance the modeling capability of channel-spatial coupling relationships at the detection head stage, this paper embeds the CTCA module before the convolutions of

the two decoupled branches in the detection head (Figure 4). While maintaining the lightweight characteristics of the convolutional framework, this module establishes long-range inter-channel dependencies through cross-dimensional mapping and interactively couples them with spatial features, thereby stabilizing class confidence and suppressing background interference.

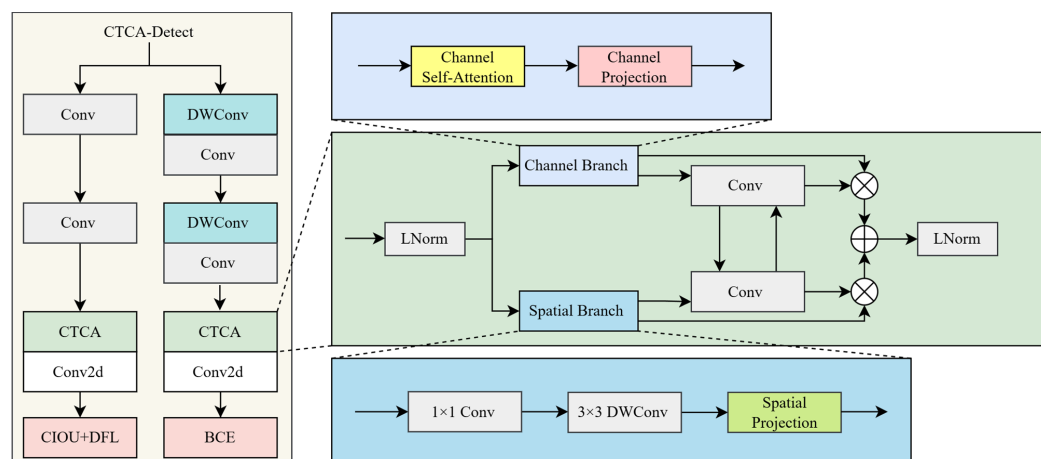


Figure 4. Network architecture of the CTCA module.

Let the input feature from the feature fusion layer be $X \in \mathbb{R}^{H \times W \times C}$. It is first normalized by LNorm to obtain $\tilde{X}$. Unlike conventional spatial self-attention that unfolds computation along the HW dimension, CTCA adopts a channel-transposition strategy, rearranging the feature as $X_c \in \mathbb{R}^{C \times (HW)}$, that is, treating channels as token units and constructing correlation modeling along the channel dimension.

Subsequently, linear projections are applied to generate queries, keys, and values:

$$Q_c = W_q X_c, K_c = W_k X_c, V_c = W_v X_c \quad (7)$$

Attention computation is then performed along the channel dimension:

$$Y_c = \text{Softmax} \left(\frac{Q_c^T K_c}{\sqrt{d}} \right) V_c \quad (8)$$

where d is the scaling factor. This operation captures the global response relationships among different channels, enabling the detection head to obtain cross-semantic dependency modeling capability before entering the classification and regression branches.

To reduce computational complexity and avoid noise amplification caused by excessive globalization, CTCA does not directly output Y_c but instead introduces a dual-projection coupling structure. First, the Channel Branch extracts enhanced channel features Y_{cp} , while the Spatial Branch extracts spatially enhanced features Y_{sp} from the original feature. The two are cross-weighted through a Sigmoid gating function:

$$Y_{ctca} = Y_{cp} \cdot \sigma(Y_{sp}) + \text{DWConv}(V_c) \cdot \sigma(Y_{cp}) \quad (9)$$

where $\sigma(\cdot)$ denotes the Sigmoid activation function, and DWConv represents depthwise separable convolution. This structure realizes a bidirectional coupling mechanism of “channel guiding spatial features and spatial feedback to channels.”

The module is embedded before the Conv2d layers of the two decoupled branches in the detection head, enabling the classification branch to obtain more stable discriminative features and the regression branch to obtain more consistent structural responses, thereby

improving the localization accuracy of slender and low-contrast defects such as corrosion, erosion, and cracks. Compared with conventional spatial attention mechanisms, its cross-dimensional coupling strategy can significantly alleviate class confidence oscillation under complex coastal backgrounds.

3.5. Dataset

The experimental evaluation uses the UAV-based wind turbine inspection dataset constructed by Shihavuddin [11]. The dataset contains 2995 high-resolution images (586×371 pixels), capturing key components such as blades, rotors, and towers. To ensure robust generalization, the dataset is divided into training, validation, and test sets according to a ratio of 8:1:1.

The annotated defects are categorized into dust and damage. As shown in Figure 5, the defects exhibit diverse forms. The dust category mainly consists of large-scale targets, whereas the damage category is characterized by small-scale and irregular features, such as corrosion, cracks, and surface peeling. Improving the detection accuracy of these damage targets is the primary objective of this study.



Figure 5. Dataset examples.

4. Results

4.1. Experimental Environment and Data Settings

To ensure the reproducibility and fairness of the reported experimental results, all models were trained and evaluated under identical hardware and software settings. The experimental platform and training parameter settings are shown in Table 1.

The dataset was divided according to an 8:1:1 ratio to ensure independence among training, validation, and testing processes. The evaluation metrics include Precision, Recall, mAP@0.5, and mAP@0.5:0.95. Among them, mAP@0.5 is used to measure the overall detection capability of the model, while mAP@0.5:0.95 evaluates localization stability under different IoU thresholds, thereby comprehensively reflecting the adaptability of the model to complex coastal corrosion scenarios.

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$AP = \int_0^1 P(R) dR \quad (12)$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i \quad (13)$$

Table 1. Experimental platform and training parameter settings.

Parameter	Configuration
CPU	Intel Core i5-12400F
GPU	NVIDIA GeForce RTX 3090
Deep learning framework	PyTorch 1.12.1
CUDA version	11.8
Input resolution	640 × 640
Batch size	32
Epochs	300
Optimizer	SGD
Momentum	0.937
Initial learning rate	0.01
Weight decay	0.0005

4.2. Comparison with Mainstream Detection Models

To verify the effectiveness of the proposed method, CoastCor-Net is compared with mainstream object detection models under the same training strategy. The results are shown in Table 2.

Table 2. Performance comparison of mainstream models.

Model	Params (M)	FLOPs (G)	mAP@0.5 (%)	AP_damage (%)	AP_dirt (%)
Faster-RCNN [25]	45.1	148.9	54.1	28.6	79.5
SSD [26]	23.0	28.5	65.9	49.3	82.4
RT-DETR-L [27]	32.0	103.4	55.5	52.2	58.8
YOLOv5n	1.8	4.1	80.1	69.4	90.8
YOLOv7-tiny [28]	2.8	6.2	80.1	69.5	90.8
YOLOv8n	3.0	8.1	80.4	67.4	93.4
YOLOv11n	2.6	6.3	81.0	68.7	93.3
YOLOv13n	2.4	6.1	81.5	68.7	94.4
Hyper-YOLO [29]	3.9	11.1	81.0	69.2	92.9
DMR-YOLO [12]	3.1	7.3	82.2	70.6	93.7
CoastCor-Net	3.6	9.4	84.7	73.9	95.6

All experiments were conducted under fixed random seeds to ensure reproducibility. The performance improvements were consistently observed across repeated runs.

From Table 2, traditional two-stage methods show insufficient performance in complex coastal environments, particularly in achieving relatively low accuracy in fine-grained corrosion target detection. Lightweight YOLO-series models demonstrate relatively stable overall performance; however, they still suffer from insufficient recall for small-scale corrosion and fine crack targets.

The proposed method achieves 84.7% in mAP@0.5, representing an improvement of 3.2 percentage points over YOLOv13n and a 5.2 percentage point increase in AP_damage, indicating that the proposed multi-module collaborative structure significantly enhances corrosion region recognition capability.

Figure 6 illustrates the relationship between parameter scale and detection performance (mAP@0.5) for different models. YOLOv5, YOLOv8, and YOLOv11 all exhibit performance improvements as parameter scale increases, but the growth rate gradually

diminishes. The improvement from the n to s variants is relatively significant, whereas the increase from m to l becomes noticeably smaller, indicating that simply enlarging model scale does not yield proportional performance gains. Under similar parameter budgets, YOLOv11 achieves overall better performance than YOLOv5 and YOLOv8, suggesting that structural optimization is more effective than merely stacking parameters. The proposed CoastCor-Net maintains stable leading performance under a comparable parameter budget, with a reasonable and controlled advantage margin, demonstrating that its performance improvement primarily stems from structural enhancement and feature augmentation mechanisms rather than excessive parameter increase.

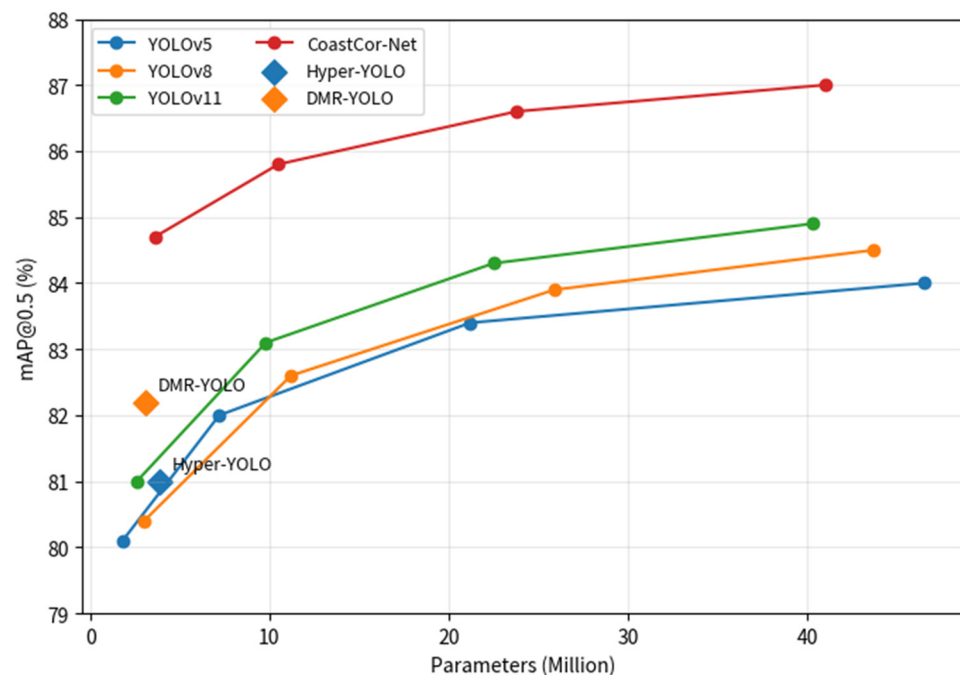


Figure 6. Performance comparison under different parameter scales.

4.3. Overall Module Ablation Study

To verify the independent contribution of the three core modules, experiments were conducted by adding each module individually to the baseline model. The results are shown in Table 3.

Table 3. Overall module ablation results.

Model	DDEB	CFAM	CTCA	mAP@0.5 (%)
Baseline (YOLOv11n)	×	×	×	81.0
+DDEB	✓	×	×	82.6
+CFAM	×	✓	×	82.9
+CTCA	×	×	✓	82.3
+DDEB + CFAM	✓	✓	×	83.8
Full Model	✓	✓	✓	84.7

✓ denotes that the module is included in the model, whereas × denotes that the module is not included.

The results indicate that all three modules individually improve detection accuracy, among which CFAM contributes the most significant performance gain. When combined, the three modules achieve the maximum improvement, demonstrating a clear synergistic effect.

First, after introducing DDEB, mAP@0.5 increases from 81.0% to 82.6%, representing an improvement of 1.6 percentage points. Multi-scale dilated convolutions and spatial enhancement structures effectively enlarge the receptive field and enhance responsiveness

to fine cracks and erosion boundaries. In coastal salt spray environments, corrosion regions often exhibit blurred boundaries and low contrast. Without boundary enhancement mechanisms, shallow features are prone to localization drift. Therefore, the enhancement effect of DDEB at the early stage is evident.

Second, when only CFAM is introduced, performance increases to 82.9%, yielding a gain of 1.9 percentage points, which is the largest improvement among the single modules. This phenomenon indicates that under complex background interference, the mismatch between spatial details and deep semantics is one of the main bottlenecks limiting detection accuracy. CFAM establishes bidirectional dependency through channel guidance and spatial inverse correction mechanisms, enabling shallow localization information to effectively compensate deep semantic representation, thereby significantly improving recall for small corrosion targets.

When only the CTCA module is added, mAP@0.5 increases to 82.3%. Although the improvement is relatively smaller, its contribution is mainly reflected in stabilizing confidence scores and improving boundary regression consistency at the detection head stage. By modeling cross-channel dependencies through channel transposition, CTCA enables the classification and regression branches to obtain more stable feature representations under complex backgrounds. Therefore, its contribution can be regarded as back-end optimization enhancement.

Further analysis of the combined experiment shows that when both DDEB and CFAM are introduced, mAP@0.5 reaches 83.8%, further improving over the single-module cases. This indicates a clear synergistic relationship between shallow boundary enhancement and mid-level feature alignment. DDEB provides high-quality spatial detail foundations, while CFAM effectively transfers this information to deep semantic representations, forming a functional closed loop across different structural stages.

The three modules form a structured collaborative mechanism across the shallow layer, mid-level layer, and detection head stage of the network, thereby systematically improving the stability and accuracy of corrosion and crack detection under complex coastal environments.

4.4. Comparison of Different Backbone Networks

To analyze the adaptability of different feature extraction structures to coastal corrosion scenarios, we replaced only the backbone network while keeping the detection head and training strategy consistent to evaluate structural adaptability. The results are shown in Table 4.

Table 4. Comparison of different backbone networks.

Backbone	Params (M)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	FPS
CSPDarknet [30]	3.5	83.1	51.2	118
MobileNetV3 [31]	2.9	80.9	48.6	124
PartialNet [32]	3.3	83.8	52.4	145
Improved Backbone (Ours)	3.6	84.7	54.1	135

From Table 4, although the lightweight backbone MobileNetV3 has fewer parameters, it performs noticeably worse on the mAP@0.5:0.95 metric, indicating weaker boundary regression stability under high IoU thresholds. CSPDarknet and PartialNet show relatively stable overall detection capability, but still suffer from insufficient recall for fine-grained targets under complex salt spray backgrounds.

In contrast, the improved backbone proposed in this paper achieves the best results on both mAP@0.5 and mAP@0.5:0.95, with more pronounced improvement in the high-IoU range. These results suggest that the constructed feature enhancement and alignment

mechanisms not only improve detection accuracy but also enhance localization consistency. Since coastal corrosion and crack targets typically exhibit slender and blurred boundary characteristics, the performance gain in the high-IoU range demonstrates substantial improvement in boundary fitting capability. Although the model size has increased slightly, the FPS remains high and does not affect the real-time performance of the detector.

Considering both accuracy and parameter scale, the proposed backbone achieves superior feature representation efficiency while maintaining a lightweight structure, reflecting better structural adaptability.

4.5. Convolution Structure Comparison Experiment

To verify the advantages of the multi-scale dilated convolution structure in complex coastal environments, comparative experiments were conducted by replacing only the convolution type in Stage1 while keeping the overall network structure consistent. The experimental results are shown in Table 5.

Table 5. Comparison results of convolution structures.

Convolution Type	mAP@0.5 (%)	Recall (%)
Standard Conv	84.0	78.6
DWConv	84.6	79.4
SWConv	84.2	81.5
Multi-scale Dilated Conv	84.7	83.8

From Table 5, the performance difference between standard convolution and DWConv in overall detection accuracy is relatively small, but the improvement in recall is limited, indicating that single-scale convolutions are insufficient for modeling cross-scale defect structures. SWConv enhances spatial representation capability through structural reorganization, leading to improved performance, but it still does not achieve the optimal result.

The multi-scale dilated convolution achieves the highest performance in both mAP@0.5 and Recall, demonstrating that enlarging the receptive field and integrating responses from different scales play a critical role in modeling corrosion pit-like textures and fine crack structures. In particular, the significant improvement in recall suggests this structure effectively reduces the probability of missed detections, which is of important engineering significance for practical inspection tasks.

4.6. Comparison of Different Attention Mechanisms in the Detection Head

To verify the effectiveness of the proposed CTCA module at the detection head stage, comparative experiments were conducted by replacing only the attention mechanism before the detection head while keeping the Backbone and Neck structures completely consistent. All models were trained and tested under the same training strategy. The comparison results are shown in Table 6.

Table 6. Comparison results of different attention mechanisms.

Attention Mechanism	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Recall (%)
No Attention	83.8	50.8	78.6
SE [33]	84.1	51.6	79.8
CBAM [34]	83.8	52.4	80.9
ECA [35]	84.4	52.0	80.3
Coordinate Attention [36]	84.3	52.9	81.4
Non-Local [37]	83.6	53.1	81.8
CTCA (Ours)	84.7	54.1	83.8

4.7. Robustness Evaluation Under Composite Coastal Perturbation

To evaluate the robustness of the proposed model under practical coastal imaging disturbances, a composite augmentation setting was constructed. In this setting, Gaussian noise, brightness variation, and image rotation were simultaneously applied to the test images to approximate sensor interference, illumination instability, and UAV attitude variation encountered during offshore inspection (Figure 7). All transformations were conducted while preserving the original annotations to ensure fair comparison.

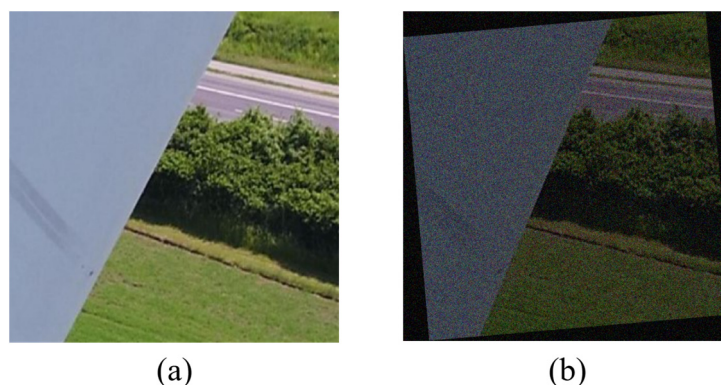


Figure 7. Example of composite disturbance augmentation: (a) original image; (b) augmented image.

Table 7 presents the mAP@0.5 results under clean and composite disturbance conditions. As expected, all models exhibit performance degradation when subjected to compounded perturbations, indicating sensitivity to environmental variability. YOLOv8n decreases from 80.4% to 77.2%, and YOLOv11n drops from 81.0% to 78.0%, corresponding to a performance reduction of approximately 3%. In contrast, CoastCor-Net maintains 82.6% mAP@0.5 under the same composite disturbance, showing a relatively smaller degradation.

Table 7. Performance Comparison of Different Models Under Clean and Composite Disturbance Conditions.

Model	mAP@0.5 (%)	
	Clean	Composite (Noise + Brightness + Rotation)
YOLOv8n	80.4	77.2
YOLOv11n	81.0	78.0
CoastCor-Net	84.7	82.6

These results suggest that the proposed spatial alignment and channel coupling mechanisms contribute to more stable feature representation under complex visual perturbations. From an engineering perspective, the enhanced robustness of CoastCor-Net indicates improved adaptability to real-world coastal UAV inspection scenarios.

4.8. Visualization Results Analysis

To further validate the capability of CoastCor-Net in recognizing fine-grained corrosion and erosion defects under complex coastal environments, this paper presents comparative visualization results between the baseline model and different improved structures, as shown in Figure 8. From left to right, the figure displays the original image, the detection results of YOLOv5, YOLOv11, and the proposed model, respectively. The red dashed regions indicate enlarged areas.

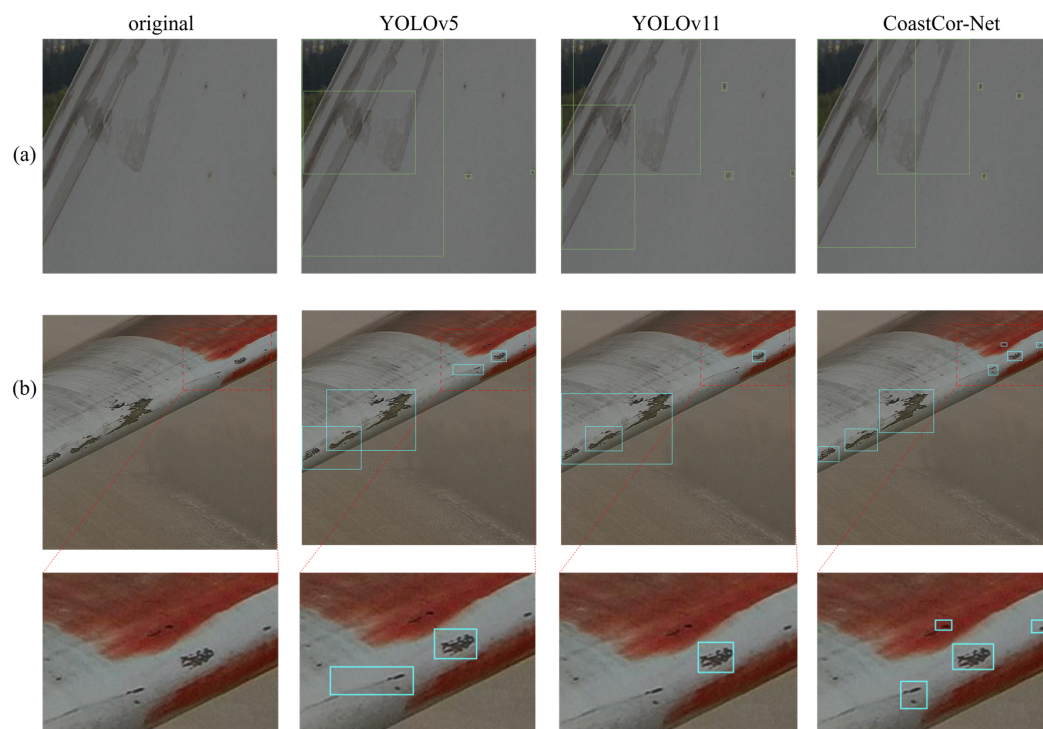


Figure 8. Visualization results of different models. (a) dust; (b) damage.

From the overall detection results, under long-distance perspectives, all models are able to localize the main structural regions. However, there are significant differences in the number of detected fine-grained corrosion regions and localization accuracy. The baseline model shows weak responses to small-scale corrosion points, resulting in missed detections, and some predicted bounding boxes deviate from the true boundaries. In the medium-scale enlarged region, corrosion patches exhibit irregular shapes with blurred edges and are mixed with coating background textures. The baseline model only identifies some prominent corrosion patches in this region and fails to adequately respond to small-scale fragmented damage near the edges. In contrast, the proposed model demonstrates stronger capability in capturing corrosion details, with more sufficient detection boxes and tighter boundary fitting to the actual defect regions.

To further explore the interpretability of the proposed model, we conducted Grad-CAM visualization analysis to compare the feature activation regions among YOLOv8, YOLOv11, and CoastCor-Net. As shown in Figure 9, the activation responses of YOLOv8 and YOLOv11 were relatively scattered, with some attention being diverted to non-defective areas. In contrast, CoastCor-Net exhibited a more concentrated and structurally consistent activation pattern around the corrosion and crack regions, especially near the blurred boundaries and fine defect areas. This demonstrates that the collaborative enhancement of the spatial alignment and channel coupling mechanisms enables the model to capture more discriminative defect features while suppressing background interference. The visualization results further validate the effectiveness of CFAM, DDEB, and CTCA in improving the reliability of defect localization.

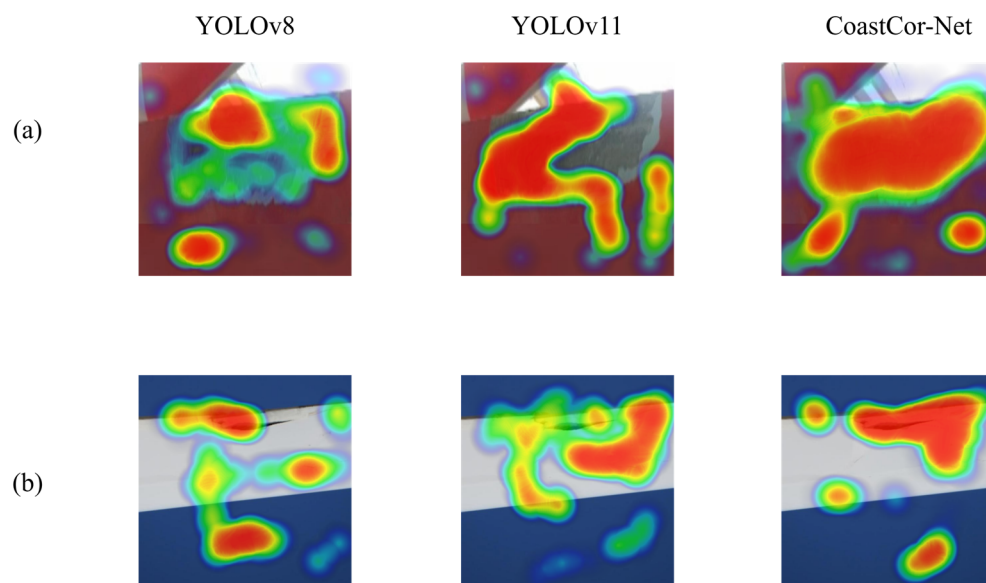


Figure 9. Grad-CAM visualization comparison of YOLOv8, YOLOv11, and CoastCor-Net. rad-CAM visualization comparison of YOLOv8, YOLOv11, and CoastCor-Net. (a) Sheet-like corrosion; (b) Strip corrosion.

5. Discussion

This study focuses on the complexity of wind turbine blade corrosion detection under coastal salt spray environments and constructs a multi-module collaborative enhancement detection framework. CoastCor-Net consistently outperforms mainstream lightweight detection models on both mAP@0.5 and mAP@0.5:0.95 metrics, with more pronounced improvements in fine-grained defect recognition for the “damage” category. Under complex low-contrast environments, standard convolution structures fail to fully capture blurred corrosion boundaries, whereas spatial enhancement and channel coupling mechanisms effectively mitigate this limitation.

From a structural perspective, the DDEB module expands the receptive field through multi-scale dilated convolutions at the shallow stage, making the model more sensitive to the spatial distribution of slender cracks and irregular corrosion patches. After embedding into Stage1, the module strengthens boundary information before feature downsampling, laying a foundation for subsequent semantic representation. Ablation experiments show that introducing DDEB alone brings stable performance improvement, indicating that shallow boundary enhancement is particularly critical in coastal environmental scenarios. CFAM achieves complementary alignment between spatial and semantic features through a bidirectional guidance mechanism. Experimental results demonstrate that this module yields the largest gain when added individually, suggesting that the primary performance bottleneck under complex background interference stems from spatial–semantic imbalance. By explicitly constructing cross-branch modulation relationships, the model maintains more stable localization capability in deeper layers, thereby improving recall for small-scale defects.

At the detection head stage, the CTCA module models cross-channel dependencies through channel transposition, enabling the classification and regression branches to obtain more consistent responses under high salt spray backgrounds. Although its individual gain is relatively modest, it plays an important role in stabilizing back-end outputs within the overall structure. In particular, the improvement in the mAP@0.5:0.95 metric reflects enhanced boundary regression accuracy.

From the perspective of scaling trends, as the number of model parameters increases, all models exhibit diminishing marginal performance gains. Compared with traditional

YOLO-series models whose performance gradually saturates with scale expansion, the proposed model maintains higher accuracy under similar parameter budgets, indicating that structural optimization is more effective than simple parameter stacking. This characteristic is of practical engineering significance in resource-constrained UAV inspection scenarios.

Beyond visual detection performance, the proposed framework can support structural health monitoring (SHM) of offshore wind turbines by providing quantitative information on defect location, size, and category. These outputs can be integrated into fatigue assessment and reliability evaluation models to estimate structural risk and remaining service life, thereby assisting condition-based maintenance decisions. In large-scale wind farms, defect severity and spatial distribution can further be used to prioritize inspection and repair tasks, improving resource allocation efficiency and reducing operational risk.

However, the proposed method still has certain limitations. The model training relies on high-quality annotated data. Although the multi-module structure controls computational cost, it still introduces a slight increase in inference overhead compared with extremely lightweight models. Future research may further optimize deployment efficiency by integrating lightweight convolution alternatives or knowledge distillation strategies.

6. Conclusions

To address the problems of blurred boundaries, spatial–semantic imbalance, and insufficient channel dependency modeling in wind turbine blade corrosion, erosion, and crack detection under coastal salt spray environments, this study proposes CoastCor-Net, a multi-scale dense prediction detection framework. By introducing three enhancement modules—DDEB, CFAM, and CTCA—at key positions in the backbone network and detection head, the proposed method achieves collaborative optimization of shallow boundary enhancement, mid-level feature alignment, and back-end channel coupling modeling.

Experimental results on the Wind Turbine Blade Damage Dataset demonstrate that the proposed model outperforms mainstream single-stage detection methods on both mAP@0.5 and mAP@0.5:0.95 metrics, especially achieving higher recall and more stable localization performance in fine-grained target recognition for the “damage” category. Meanwhile, under the premise of maintaining a lightweight structure, the model’s inference speed satisfies the requirements of real-time UAV inspection.

Further ablation experiments verify the independent contributions and synergistic gains of the three modules, demonstrating that structural improvements are more effective than simple model scale expansion. The analysis of the parameter scale–performance relationship also reveals a clear diminishing marginal return trend, highlighting the parameter efficiency advantage of the proposed method.

Future work will focus on improving cross-scenario generalization capability and optimizing lightweight deployment, including the introduction of adaptive data augmentation strategies, cross-domain training mechanisms, and model compression techniques, to further enhance the practicality and stability of offshore wind power intelligent inspection systems.

Author Contributions: J.X. was responsible for the study’s conceptualization, methodology, software, formal analysis, investigation, and writing—original draft preparation. S.N. was responsible for resources, writing—review and editing, supervision, project administration, and funding acquisition. X.W. was responsible for validation, data curation, and visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available in a publicly accessible repository via the following link: <https://doi.org/10.3390/en12040676>.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ma, X.; Li, M.; Li, W.; Liu, Y. Overview of Offshore Wind Power Technologies. *Sustainability* **2025**, *17*, 596. [\[CrossRef\]](#)
2. Fontanella, A.; Fusetti, A.; Cioni, S.; Papi, F.; Muggiasca, S.; Persico, G.; Dossena, V.; Bianchini, A.; Belloli, M. Wake Development in Floating Wind Turbines: New Insights and an Open Dataset from Wind Tunnel Experiments. *Wind. Energy Sci.* **2025**, *10*, 1369–1387. [\[CrossRef\]](#)
3. Dong, X.H.; Yuan, T.J.; Ma, R.H. Corrosion Mechanism on Offshore Wind Turbine Blade in Salt Fog Environment. *Appl. Mech. Mater.* **2013**, *432*, 258–262. [\[CrossRef\]](#)
4. Alsaadi, M.; Mishnaevsky, L., Jr.; Tobin, E.F.; Devine, D.M. Advances in Smart Coating Technologies for Wind Turbine Blade Protection: A Focus on Self-Healing and Anti-Erosion Performance. *J. Mar. Sci. Eng.* **2025**, *13*, 2224. [\[CrossRef\]](#)
5. Sarode, D.; Saharkar, U.; Ahire, H.H.; Darade, M.M.; Chougule, S.M.; Dalvi, S.A.; Kurhade, A.S. Corrosion-Resistant Materials for Ocean Structures: Innovations, Mechanisms, and Applications. *Sustain. Mar. Struct.* **2025**, *7*, 63–87. [\[CrossRef\]](#)
6. Masita, K.; Hasan, A.N.; Shongwe, T.; Hilal, H.A. Deep Learning in Defect Detection of Wind Turbine Blades: A Review. *IEEE Access* **2025**, *13*, 98399–98425. [\[CrossRef\]](#)
7. Heo, S.-J.; Na, W.S. Review of Drone-Based Technologies for Wind Turbine Blade Inspection. *Electronics* **2025**, *14*, 227. [\[CrossRef\]](#)
8. Ryalat, M.; ElMoaqet, H.; AlFaouri, M. Design of a Smart Factory Based on Cyber-Physical Systems and Internet of Things towards Industry 4.0. *Appl. Sci.* **2023**, *13*, 2156. [\[CrossRef\]](#)
9. Fu, P.; Yuan, H.; He, J.; Wu, B.; Xu, N.; Gu, Y. IMTS-YOLO: A Steel Surface Defect Detection Model Integrating Multi-Scale Perception and Progressive Attention. *Coatings* **2026**, *16*, 51. [\[CrossRef\]](#)
10. Yin, Z.; Li, H.; Qi, B.; Shan, G. BBW YOLO: Intelligent Detection Algorithms for Aluminium Profile Material Surface Defects. *Coatings* **2025**, *15*, 684. [\[CrossRef\]](#)
11. Shihavuddin, A.; Chen, X.; Fedorov, V.; Nymark Christensen, A.; Andre Brogaard Riis, N.; Branner, K.; Bjorholm Dahl, A.; Reinhold Paulsen, R. Wind Turbine Surface Damage Detection by Deep Learning Aided Drone Inspection Analysis. *Energies* **2019**, *12*, 676. [\[CrossRef\]](#)
12. Shi, L.; Wang, S.; Zhao, J.; Kuang, Z.; Wang, L.; Ma, L.; Yang, H.; Wang, H. DMR-YOLO: An Improved Wind Turbine Blade Surface Damage Detection Method Based on YOLOv8. *Appl. Sci.* **2026**, *16*, 1333. [\[CrossRef\]](#)
13. Liu, Y.; Zheng, Y.; Shao, Z.; Wei, T.; Cui, T.; Xu, R. Defect Detection of the Surface of Wind Turbine Blades Combining Attention Mechanism. *Adv. Eng. Inform.* **2024**, *59*, 102292. [\[CrossRef\]](#)
14. Xiao, M.; Yang, B.; Wang, S.; Zhang, Z.; He, Y. Fine Coordinate Attention for Surface Defect Detection. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106368. [\[CrossRef\]](#)
15. Xu, B.; Li, Z.; Shen, X.; Shi, T.; Cai, X. Signal Processing and Feature Engineering for Automated Multi-Type Damage Assessment of Wind Turbine Blades via Machine Vision. *Mech. Syst. Signal Process.* **2025**, *241*, 113581. [\[CrossRef\]](#)
16. Yu, Q.; Han, Y.; Lin, W.; Gao, X. Detection and Analysis of Corrosion on Coated Metal Surfaces Using Enhanced YOLO v5 Algorithm for Anti-Corrosion Performance Evaluation. *J. Mar. Sci. Eng.* **2024**, *12*, 1090. [\[CrossRef\]](#)
17. Yu, Q.; Han, Y.; Huang, X.; Gao, X. Improved YOLO-Based Corrosion Detection and Coating Performance Evaluation Under Marine Exposure in Zhoushan, China. *J. Mar. Sci. Eng.* **2025**, *13*, 1842. [\[CrossRef\]](#)
18. Cheng, H.; Kang, F. Corrosion Detection and Grading Method for Hydraulic Metal Structures Based on an Improved YOLOv10 Sequential Architecture. *Appl. Sci.* **2024**, *14*, 12009. [\[CrossRef\]](#)
19. Chen, H.; Cao, Y.; Cao, S.; Piao, H. A Study of Corrosion-Grade Recognition on Metal Surfaces Based on Improved YOLOv8 Model. *Sensors* **2025**, *25*, 2630. [\[CrossRef\]](#)
20. Yu, Q.; Han, Y.; Gao, X.; Lin, W.; Han, Y. Comparative Analysis of Improved YOLO v5 Models for Corrosion Detection in Coastal Environments. *J. Mar. Sci. Eng.* **2024**, *12*, 1754. [\[CrossRef\]](#)
21. Xiao, Y.; Xu, T.; Xin, Y.; Li, J. FBRT-YOLO: Faster and Better for Real-Time Aerial Image Detection. *Proc. AAAI Conf. Artif. Intell.* **2025**, *39*, 8673–8681. [\[CrossRef\]](#)
22. Feijoo, D.; Benito, J.C.; Garcia, A.; Conde, M.V. DarkIR: Robust Low-Light Image Restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 18184–18194. [\[CrossRef\]](#)
23. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [\[CrossRef\]](#) [\[PubMed\]](#)

24. Yang, G.; Wang, Y.; Shi, D.; Wang, Y. Golden Cudgel Network for Real-Time Semantic Segmentation. In Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 10–17 June 2025; pp. 25367–25376. [\[CrossRef\]](#)
25. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016*; Proceedings, Part I 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
27. 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 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 16965–16974. [\[CrossRef\]](#)
28. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475. [\[CrossRef\]](#)
29. Feng, Y.; Huang, J.; Du, S.; Ying, S.; Yong, J.-H.; Li, Y.; Ding, G.; Ji, R.; Gao, Y. Hyper-YOLO: When Visual Object Detection Meets Hypergraph Computation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 2388–2401. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Wang, C.-Y.; Liao, H.-Y.M.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I.-H. CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 390–391. [\[CrossRef\]](#)
31. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324. [\[CrossRef\]](#)
32. Zhao, H.; Li, X.; Zhang, Y.; Chen, T.; Wang, Z. PartialNet: Revisiting Channel Sparsity for Efficient Deep Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 2207–2216. [\[CrossRef\]](#)
33. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [\[CrossRef\]](#)
34. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In *Computer Vision—ECCV 2018*; Springer: Cham, Switzerland, 2018; pp. 3–19. [\[CrossRef\]](#)
35. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 11531–11539. [\[CrossRef\]](#)
36. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722. [\[CrossRef\]](#)
37. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-Local Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2018; pp. 7794–7803. [\[CrossRef\]](#)

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.