MDFE-Net: Multi-level Degradation-aware Feature Enhancement Network

Yi Liu, Na Liu ✉, Long Yang, Xufei Zhuang, Wenhong Wu, and Guiping Liu ✉
Inner Mongolia University of Technology, 49 Aimin Street Xincheng District, Hohhot 010051, P. R. China
csnaliu@imut.edu.cn, **csguiping_liu@163.com**

***Abstract*—Detecting small-scale defects in complex industrial environments remains a challenging problem in computer vision, primarily due to progressive feature degradation during downsampling, background clutter interference, and insufficient multi-scale feature fusion. To overcome these obstacles, we propose MDFE-Net, a Multi-level Degradation-aware Feature Enhancement Network designed to enhance the robustness of small-defect representations by explicitly addressing feature degradation and enabling efficient multi-scale feature interactions. MDFE-Net consists of three key components: a degradation-aware feature enhancement module to strengthen weak small-object representations, a background-aware attention mechanism to suppress background noise, and a structure-preserving multi-scale fusion strategy to alleviate scale inconsistency. Notably, MDFE-Net is detector-agnostic, allowing for seamless integration into existing one-stage detection frameworks. We conduct experiments on several challenging datasets, including two wind turbine blade defect datasets with complex industrial backgrounds, as well as additional datasets from other visual domains characterized by small targets and cluttered scenes. These experiments assess both in-domain and cross-domain generalization. The experimental results demonstrate that MDFE-Net consistently improves detection accuracy for small and densely packed targets while maintaining high computational efficiency across diverse scenarios.**

***Index Terms*—small defects, feature degradation, multi-scale features, attention learning, and industrial inspection.**

I. Introduction

Detecting small-scale defects in complex industrial environments is a crucial task for ensuring safety, reliability, and efficiency in manufacturing processes. Small defects are often precursors to larger failures, making early detection essential to prevent costly breakdowns and enhance automated inspection systems [4] [26]. However, small-scale defect detection is fraught with challenges due to feature degradation during downsampling, background interference, and scale inconsistency across different defect sizes. These issues are especially prominent in industrial inspection tasks, where defects are small, densely distributed, and often hidden in cluttered environments.

Current solutions to small-scale defect detection have drawn attention to the importance of high-resolution images and deeper network architectures. Many existing methods aim to address the problem by enhancing detection performance through multi-scale representations or using deeper backbones to capture more detailed features [25] [6]. While these approaches have made progress, they fail to effectively tackle the progressive feature degradation during downsampling. As feature maps are reduced, fine-grained details essential for small defect detection are lost, leading to poor performance in complex settings. Additionally, background interference—such as noise and textures—remains a significant obstacle, causing traditional techniques to struggle with separating small defects from background elements [17] [5].

Furthermore, scale inconsistency in detection is another major challenge. In industrial settings, defects appear at varying scales within the same image, and existing models struggle to handle this variation. While multi-scale fusion techniques have been proposed, they often fail to preserve fine details of small defects across scales [20].

To address this gap, we propose **MDFE-Net**—a novel detector-agnostic framework. Our core insight involves explicitly defining and modeling feature degradation as a progressive phenomenon during forward propagation, followed by proactive intervention and enhancement at multiple downsampling stages. By systematically integrating degradation-aware mechanisms, background suppression, and structure preservation, this framework aims to directly reinforce the network's capacity to represent subtle defects.

The principal contributions of this work are threefold:

- **Unified Modeling and Enhancement of Feature Degradation**: We propose a unified perspective that models the attenuation of small-object features in deep networks as a progressive process. A multi-level enhancement mechanism is designed to actively compensate for and amplify feature responses of small defects during feature pyramid construction.
- **Modular MDFE-Net Framework**: We introduce three key innovations: (1) Degradation-Aware Feature Enhancement (DAFE) Module to mitigate information loss during downsampling; (2) Background-Aware Attention Mechanism (BAA) to suppress complex background interference and highlight foreground defects; (3)Structure-Preserving Multi-Scale Fusion (SPMSF) Strategy to optimize cross-scale feature aggregation while preserving the geometric integrity of small targets.
- **Comprehensive Experimental Validation**: Extensive experiments on challenging industrial defect datasets (e.g., wind turbine blade inspection) and cross-domain benchmarks demonstrate significant improvements in pre-

cision and recall. Results confirm the method's effectiveness, generalization capability, and practical applicability.

II. Methodology

A. Overall Architecture of MDFE-Net

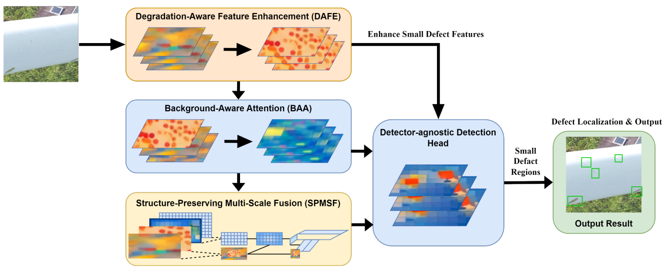

Fig. 1. Overall Architecture of MDFE-Net

The overall architecture of MDFE-Net is designed to enhance small-defect representations in complex industrial environments by addressing feature degradation, background clutter, and scale inconsistency in deep neural networks. Rather than proposing a task-specific detector, MDFE-Net is formulated as a detector-agnostic feature enhancement framework, enabling seamless integration into existing one-stage detection pipelines without modifying the detection head. This allows MDFE-Net to improve feature quality while preserving the efficiency and generality of mainstream detection architectures.

As illustrated in Fig. 1, an input image is first processed by a backbone network to extract multi-level feature maps at different spatial resolutions. During downsampling, small-defect information is progressively weakened, a key challenge in small-scale object detection [15] [10]. To mitigate this, MDFE-Net introduces enhancement mechanisms across multiple feature levels, reinforcing degraded representations throughout the feature hierarchy.

MDFE-Net consists of three complementary modules: **(1) Degradation-Aware Feature Enhancement (DAFE)** applied to intermediate feature maps, strengthening small-defect cues during downsampling; **(2) Background-Aware Attention (BAA)** to suppress irrelevant background and enhance defect-background discrimination, especially in complex industrial environments; **(3) Structure-Preserving Multi-Scale Fusion (SPMSF)**, which improves feature interaction across scales and maintains small-defect structures.

These components work together in a multi-level enhancement framework, enabling feature information to be progressively refined and fused. By modeling degradation awareness, background suppression, and multi-scale interaction, MDFE-Net effectively addresses the main challenges of small-scale defect detection. Importantly, the architecture adds minimal computational overhead and preserves the detection head, ensuring practical applicability for real-world industrial inspection systems.

B. Degradation-Aware Feature Enhancement (DAFE) Module

In deep neural networks, small-scale defects often suffer from progressive feature degradation as feature maps undergo multiple downsampling operations, such as pooling and strided convolutions. While these operations are essential for capturing global context and expanding the receptive field, they often result in the loss of fine-grained information critical for detecting small objects. As the network deepens, the feature maps become increasingly abstract, further diminishing the ability to retain detailed defect representations. Consequently, small defects are often overwhelmed by background noise, and their subtle features are lost, especially in complex industrial environments with dense and cluttered backgrounds.

To address this challenge, the **Degradation-Aware Feature Enhancement (DAFE)** module is introduced. The core idea of the DAFE module is to explicitly counteract feature degradation by enhancing weak, small-defect cues across the network's layers. Unlike traditional feature enhancement techniques that operate only at the final output layer, DAFE operates at intermediate feature levels, where degradation is most pronounced, ensuring that small defects are preserved throughout the network.

The DAFE module works in two key steps:

1) Feature Enhancement: The first step involves identifying the degraded small-defect features and selectively enhancing them. This is achieved through an **adaptive feature weighting mechanism** that adjusts the importance of different feature channels based on their relevance to small-defect detection. Features showing signs of degradation (such as low contrast or high background interference) are given higher weights, while irrelevant features are suppressed. This adaptive weight function can be represented as:

$$w_i = \frac{1}{1 + e^{-\alpha(I_i - I_{threshold})}} \tag{1}$$

where w_i is the weight for feature i, I_i is the intensity or importance score of feature i, $I_{threshold}$ is a pre-defined threshold indicating degraded features, and α controls the sensitivity to degradation. Features with higher degradation are assigned greater weights to improve their contribution to small-defect detection. To avoid background noise over-amplification from fixed hyperparameters in high-dynamic industrial images, $I_{threshold}$ and α are adaptively calculated via the mean μ_I and standard deviation σ_I of I_i in the feature map: $I_{threshold} = \mu_I - 0.8\sigma_I, \alpha = 1.2\exp(\sigma_I/\mu_I)$. The scalars 0.8 and 1.2 are calibrated on HBDD-E and DTU-Drones, enabling adaptive adjustment to varying background complexity and illumination, and precise differentiation between degraded defect features and noise.

This allows the network to emphasize fine-grained details and enhance the discriminability of small defects across the network layers.

2) Contextual Refinement: In the second step, the enhanced feature maps undergo a **contextual refinement process** to ensure that the enhanced features are not only accurate but also contextually relevant. This process involves the use of

local attention mechanisms that guide the network's focus to defect-relevant regions, reducing the influence of irrelevant background noise. The attention map can be formulated as:

$$A_i = \frac{exp(F_i)}{\sum_j exp(F_j)} \tag{2}$$

where A_i is the attention weight for feature i, F_i is the feature value for the region of interest, and the sum in the denominator normalizes the attention map across all regions. This mechanism allows the network to focus more on the defect area and suppress irrelevant background information.

C. Background-Aware Attention Mechanism (BAA)

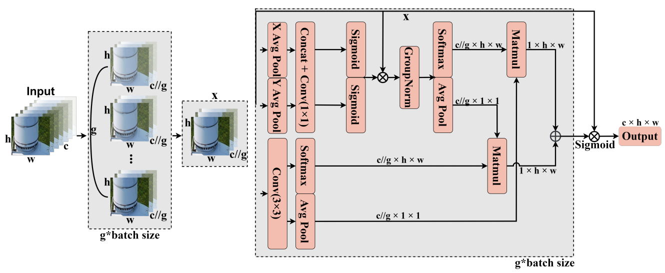

Fig. 2. The architecture of Background-Aware attention.

In complex industrial inspection scenarios, small-scale defects are often embedded in backgrounds with strong textures, repetitive patterns, and illumination variations. Such background clutter can produce high-response activations in deep feature maps, which can overwhelm subtle defect cues and degrade detection performance. Existing attention mechanisms mainly emphasize feature saliency but often fail to explicitly model background interference, making them less effective in scenarios where background responses are visually dominant [24] [2].

To address this, BAA explicitly suppresses background-dominated responses while preserving defect-relevant features. Unlike traditional attention modules that focus solely on channel-wise or spatial saliency, BAA jointly models global background statistics and local defect cues, enabling more discriminative feature refinement. This design is inspired by attention mechanisms that refine feature maps by balancing foreground and background information for enhanced detection, such as the Efficient Multi-scale Attention (EMA) mechanism used in state-of-the-art detection frameworks [14].

Building upon the principles of EMA, we upgrade this approach to the BAA module. While EMA primarily focuses on multi-scale feature fusion and foreground-background separation, BAA extends this by introducing a stronger emphasis on background awareness. This allows BAA to suppress background noise better and more accurately extract defect-related features. The upgraded BAA module retains the structure and formulation of the EMA module, which is demonstrated in the following sections.

As shown in Fig. 2, given an input feature map $F \in \mathbb{R}^{C\times H\times W}$, BAA first extracts a global context representation through a context modeling branch, which captures the overall distribution of feature responses and reflects background characteristics. This global descriptor provides a coarse yet informative summary of background patterns prevalent in the scene. In parallel, a spatial projection branch is employed to preserve localized defect responses, ensuring that fine-grained spatial cues associated with small defects are retained.

The outputs of the global context branch and the spatial branch are subsequently fused to generate a background-aware attention map A, which adaptively modulates the input feature map according to both background statistics and local defect saliency. The refined feature representation is obtained as:

$$F' = F \odot A \tag{3}$$

where $\odot$ denotes element-wise multiplication. The specific implementation workflow is illustrated in Fig. 2. Through this mechanism, feature responses dominated by background clutter are effectively suppressed, while defect-relevant regions are selectively emphasized.

By explicitly incorporating background awareness into the attention modeling process, BAA enhances the purity and discriminability of feature representations. This mechanism works synergistically with the Degradation-Aware Feature Enhancement (DAFE) module to further improve the robustness of small-defect features under complex background conditions, providing a solid foundation for subsequent multi-scale feature fusion.

D. Structure-Preserving Multi-Scale Fusion (SPMSF)

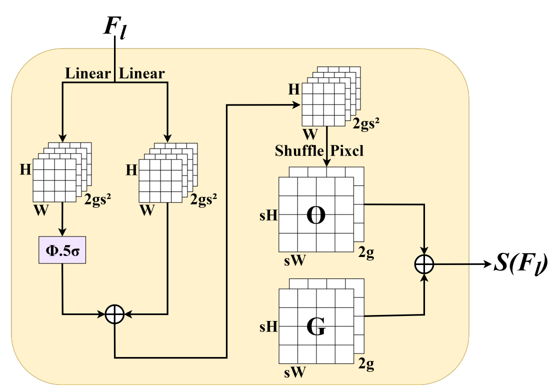

Fig. 3. Structure-Preserving Multi-Scale Fusion (SPMSF) strategy

Multi-scale feature fusion plays a critical role in detecting objects with large scale variations. However, in small-scale defect detection, conventional fusion strategies often introduce scale inconsistency and structural distortion, particularly when low-resolution semantic features are directly aggregated with high-resolution structural features. While such fusion can enhance semantic richness, it may inadvertently weaken the continuity and integrity of fine-grained defect structures, which are essential for accurate localization and recognition of small defects [12] [22] [1].

To address this limitation, MDFE-Net introduces a Structure-Preserving Multi-Scale Fusion (SPMSF) strategy, which aims to facilitate effective cross-scale interaction while explicitly maintaining the structural characteristics of small defects. The core motivation is that, unlike generic objects, small industrial defects are primarily characterized by local structural patterns, such as edges, cracks, or spot-like textures. Preserving these structural cues during feature aggregation is therefore crucial for robust detection.

As illustrated in Fig. 3, the proposed fusion strategy operates on a set of multi-level feature maps extracted at different resolutions. Instead of directly merging features through naive up-sampling and summation, SPMSF maintains structure-preserving pathways for each feature level, allowing local structural information to be retained before and during the fusion process. These pathways serve as identity-guided references that help prevent structural distortion caused by scale transformations.

During fusion, feature maps from different levels are adaptively aggregated in a scale-consistent manner, where each feature level contributes according to its structural relevance. Formally, the fused representation can be expressed as:

$$F_{\text{fusion}} = \sum_{l} w_l \cdot S(F_l) \tag{4}$$

where F_l denotes the feature map at the l-th level, w_l represents the adaptive fusion weight, and $S(\cdot)$ denotes a structure-preserving transformation that aligns feature scales while maintaining local continuity. Unlike conventional fusion operators, $S(\cdot)$ emphasizes structural consistency rather than purely semantic alignment.

By integrating structure preservation into the multi-scale fusion process, SPMSF effectively mitigates scale inconsistency and enhances the coherence of small-defect representations across feature hierarchies. When combined with the degradation-aware feature enhancement module and the background-aware attention mechanism, the proposed fusion strategy enables MDFE-Net to progressively refine and consolidate feature representations, resulting in improved robustness and accuracy for small-scale defect detection in complex industrial scenarios.

III. Experiment

A. Datasets and Experimental settings

Datasets

- Huadian BladeDefect Dataset - Extra (HBDD-E). HBDD-E is the primary dataset used in this study, constructed in collaboration with China Huadian Group for wind turbine blade defect detection. As an extension of the original HBDD dataset, which comprises 4,875 images captured from wind farms in Ningdong, Guyuan, and Haiyuan, Ningxia, China, between 2019 and 2023, it is further supplemented with 437 images collected in 2025 from wind farms in Barkol and Yiwu counties, Hami, Xinjiang, China, resulting in a total of 5,312 images with a uniform resolution of 586×371 pixels. HBDD-E covers diverse real-world landforms, including Gobi deserts, loess hills, mountain woodlands, and desert steppes, and exhibits considerable variations in defect appearance, illumination, and background complexity that closely resemble actual detection scenarios. The dataset is annotated with nine typical blade defects, namely craze, hide craze, surface attach, corrosion, surface corrosion, surface eye, surface injury, surface oil, and thunderstrike. Most defect regions account for less than 5% of the image area, and sample instances and the target size distribution are illustrated in Fig. 4.

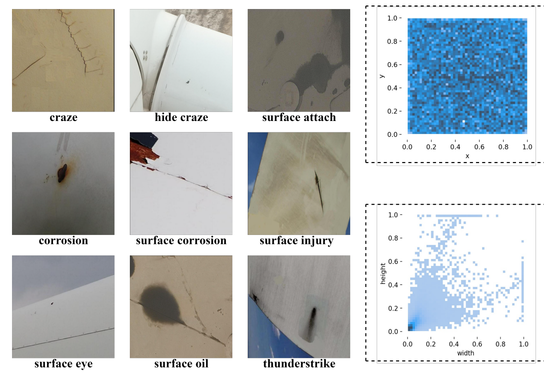

Fig. 4. Sample images and defect size distribution of the HBDD-E dataset.

- DTU-Drones Dataset (DTU-Drones). For comparative validation, this paper employs the DTU-Drones dataset [18] as the benchmark, which was collected from 2017 to 2018 during routine inspections of Nordic Tank wind turbines at the DTU Wind Farm in Roskilde, Denmark; it contains 662 blade images captured under both daytime and nighttime conditions and across various weather scenarios, and is publicly available in the Mendeley Data repository. Since the original public dataset is unannotated, we manually annotated it using the LabelMe tool, following the same annotation protocol as for the HBDD-E dataset. This unified labeling process ensures annotation consistency and prevents biases arising from different annotation strategies.
- DIOR Dataset (DIOR). To further assess cross-domain generalization, additional experiments are conducted on the DIOR remote sensing dataset [9], which contains 23,463 images with 192,472 annotated instances across 20 object categories. DIOR features complex backgrounds and a high proportion of small objects, making it suitable for evaluating the transferability of degradation-aware feature enhancement beyond industrial defect detection.

Experimental settings and Evaluation criteria

All experiments were implemented using PyTorch 1.13.1 and conducted on a workstation equipped with an Intel Xeon Platinum 8352V CPU and an NVIDIA GeForce RTX 4090 GPU (24 GB). MDFE-Net was integrated into a standard one-stage object detection framework and trained end-to-end.

Given the inherent difficulties in industrial defect detection, such as dense small objects and complex backgrounds, together with the batch size of 16, this study uniformly sets the training epoch to 700 to ensure thorough optimization and stable convergence of model parameters. The optimizer, initial learning rate, weight decay, and data augmentation strategies all follow the default configurations of the baseline model. For fair comparison, all baseline methods and ablation variants are trained under identical settings, with no additional hyperparameter tuning performed for competing approaches. During training, model convergence is monitored via real-time validation metrics to effectively mitigate the risk of overfitting.

To comprehensively evaluate the performance of the proposed WTB defect detection network, we adopt several widely used metrics: Precision(P), recall(R), mean Average Precision (mAP), mAP@0.5 and mAP@0.5 : 0.95, number of Parameters, and Giga Floating Point Operations ($GFLOPs$). Among these, $GFLOPs$ quantifies the number of floating-point operations the model requires, in billions (10^9), serving as a fundamental measure of computational complexity and throughput to evaluate hardware or algorithmic efficiency.

B. Comparative experimental results and analysis

To demonstrate the detector-agnostic nature of MDFE-Net, we integrate it into the YOLOv11 framework, utilizing the YOLOv11 [8] detection head without any modifications. Then, we compare the proposed MDFE-Net with a representative set of state-of-the-art object detection methods covering different detection paradigms. The baselines include classical detectors such as the single-stage SSD [11], the keypoint-based CenterNet [3], and the two-stage Faster R-CNN [16]. In addition, several recent one-stage detectors with strong performance on small-object detection are selected, including YOLOv7 [21], YOLOv8 [7], YOLOv11 [8], and YOLOv12 [19], where YOLOv11 is adopted as the primary baseline framework. To further assess robustness under complex background conditions, LSC-YOLO [23] is included as a specialized small-object detection method. Moreover, DERT [13], a transformer-based detector leveraging global self-attention, is incorporated for comparison. All baseline methods are trained using the default configurations reported in their original publications to ensure fair comparison.

Table I presents a quantitative comparison of MDFE-Net with state-of-the-art detection methods on the HBDD-E dataset. Overall, MDFE-Net achieves the best performance across all evaluation metrics, obtaining 92.3% precision, 88.6% recall, 94.6% mAP@0.5, and 69.8% mAP@0.5:0.95, while maintaining a moderate model size and computational cost. Compared with the strongest baseline YOLOv12, MDFE-Net improves mAP@0.5 and mAP@0.5:0.95 by +1.3% and +2.9%, respectively, indicating its superior capability in detecting small-scale defects under complex industrial backgrounds.

- **Comparison with classical detection frameworks.** Classical detectors such as SSD ($\mathbf{41.9}\%$), CenterNet ($\mathbf{36.7}\%$), and Faster R-CNN ($\mathbf{32.3}\%$) exhibit limited performance on HBDD-E, with $mAP@0.5:0.95$ values significantly lower than modern detectors. These methods rely on coarse feature representations and lack effective mechanisms to preserve fine-grained information during downsampling, making them inadequate for detecting small and low-contrast defects. In contrast, MDFE-Net significantly outperforms these models, demonstrating that explicitly modeling feature degradation and enhancing weak features is crucial for reliable defect detection in real-world wind turbine blade inspection scenarios.
- **Comparison with advanced one-stage detectors.** Recent YOLO-based detectors show substantially improved performance. While YOLOv8 achieves the highest mAP@0.5 : 0.95 ($\mathbf{69.9}\%$), MDFE-Net demonstrates competitive performance ($\mathbf{69.8}\%$) while achieving superior precision ($\mathbf{92.3}\%$ vs $\mathbf{90.9}\%$) and the highest mAP@0.5 ($\mathbf{94.6}\%$). Among them, YOLOv8 and YOLOv12 achieve strong results with mAP@0.5 exceeding $\mathbf{92}\%$. However, their performance gains mainly stem from generic multi-scale feature aggregation, which remains insufficient for addressing progressive feature degradation caused by deep downsampling. By contrast, MDFE-Net consistently surpasses these models, particularly in mAP@0.5 : 0.95, demonstrating that degradation-aware feature enhancement and structure-preserving multi-scale fusion provide complementary benefits beyond conventional architectural refinements.
- **Comparison with attention-based and specialized methods.** Compared with LSC-YOLO, which enhances local spatial context to improve small-object detection, MDFE-Net achieves higher precision and mAP, suggesting that mitigating feature degradation is more effective than solely strengthening local context. Furthermore, although the transformer-based DERT benefits from global self-attention and achieves competitive recall, its performance on HBDD-E is limited by insufficient preservation of fine-grained spatial details for very small defects. The superior performance of MDFE-Net over DERT highlights the advantage of combining background-aware attention with degradation-aware feature modeling, which is better suited to small-scale defect detection in complex industrial environments.

C. Ablation experimental results and analysis

In order to evaluate the performance of the proposed MDFE-Net, comprehensive ablation experiments were conducted to assess the importance and effectiveness of each individual module. The experimental results are summarized in Table II. Compared to the baseline model, YOLOv11, our proposed model demonstrated significant improvements through the integration of novel modules.

- First, the Degradation-aware defects enhancement module effectively enhanced the feature extraction capabilities by modeling the degradation process, improving the network's ability to capture small target defects. This structural enhancement resulted in a $\mathbf{1.7}\%$ increase in Precision (P) and a $\mathbf{3.1}\%$ improvement in Recall (R). Con-

TABLE I
COMPARATIVE EXPERIMENTAL RESULTS FOR DIFFERENT DETECTION MODELS (HBDD-E)(%).

NetWork	P	R	mAP		$Params$	$GFLOPs$
			(@0.5)	**(@0.5:0.95)**		
SSD [11]	70.6	58.4	69.6	41.9	24.02	61.06
Faster R-CNN [16]	57.5	54.5	58.1	32.3	28.32	940.95
CenterNet [3]	68.3	53.2	63.1	36.7	32.65	70.32
DERT [13]	76.3	75.2	82.8	55.1	7.02	15.80
YOLOv7 [21]	66.5	52.0	60.2	33.7	**6.03**	**13.20**
YOLOv8 [7]	90.9	88.2	94.1	**69.9**	11.15	28.32
YOLOv11 [8]	89.8	85.5	92.1	67.6	9.45	21.72
YOLOv12 [19]	89.6	87.1	93.3	67.8	9.32	21.43
LSC-YOLO [23]	87.8	85.2	92.0	65.7	8.73	27.54
MDFE-Net	**92.3**	**88.6**	**94.6**	69.8	9.97	21.95

currently, the mean Average Precision at 0.5 ($mAP@0.5$) and mAP at multiple thresholds ($mAP@0.5:0.95$) increased by $\mathbf{1.9\%}$ and $\mathbf{2.2\%}$, respectively.

- Secondly, the Background-aware attention mechanism (BAA) contributed significantly by suppressing background noise and enhancing the foreground features. This mechanism was essential in improving the model's robustness, particularly in complex environments with significant background interference.
- Lastly, when combined with our Structure-preserving multi-scale fusion strategy (SPMSF), which optimally integrates multi-scale features, the model achieved a synergistic performance boost. The combination of these three modules led to an increase of $\mathbf{2.8\%}$ in Precision (P), $\mathbf{3.6\%}$ in Recall (R), $\mathbf{2.7\%}$ in $mAP@0.5$, and $\mathbf{3.3\%}$ in $mAP@0.5{:}0.95$, highlighting the effectiveness of the integrated modules in enhancing small target detection capabilities.

These results clearly demonstrate the synergistic effect of integrating the Degradation-aware defects enhancement module, Background-aware attention mechanism, and Structure-preserving multi-scale fusion strategy. The collaborative integration of these components significantly improves the model's performance across a range of detection metrics, illustrating their crucial roles in addressing the challenges of small target detection in complex industrial scenarios.

TABLE II
ABLATION EXPERIMENTS (HBDD-E)(%).

Components	P	R	mAP	
			(@0.5)	**(@0.5:0.95)**
YOLOv11 (Baseline)	89.8	85.5	92.1	67.6
+ DAFE	91.4	88.1	93.8	69.1
+ BAA	89.8	86.7	93.0	68.0
+ SPMSF	90.9	85.8	92.5	68.6
+ DAFE + BAA	91.9	88.3	94.1	69.4
+ DAFE + SPMSF	92.1	87.6	93.9	69.5
+ BAA + SPMSF	91.2	86.8	93.7	68.7
MDFE-Net (Full)	**92.3**	**88.6**	**94.6**	**69.8**

D. Generalization Discussion

To further evaluate the generalization capability of the proposed MDFE-Net, we conduct experiments on two additional datasets, namely DTU-Drones and DIOR, which differ significantly from the primary HBDD-E dataset in terms of data acquisition conditions, object distribution, and background complexity.

TABLE III
GENERALIZATION PERFORMANCE COMPARISON ON DTU-DRONES AND DIOR DATASETS (%).

Dataset	**Method**	P	R	mAP	
				(@0.5)	**(@0.5:0.95)**
DTU-Drones	YOLOv7	66.1	52.1	60.2	33.8
	YOLOv8	89.4	87.3	93.4	68.2
	YOLOv11	87.7	85.2	91.5	65.9
	YOLOv12	88.5	85.2	91.8	66.0
	MDFE-Net	**91.2**	**88.1**	**93.9**	**69.7**
DIOR	YOLOv11	84.2	76.2	84.9	64.5
	YOLOv12	87.1	78.8	85.5	66.7
	MDFE-Net	**89.3**	**82.1**	**87.6**	**68.6**

Generalization on DTU-Drones. As reported in Table III, MDFE-Net consistently outperforms all compared YOLO-based baselines on the DTU-Drones dataset. Specifically, MDFE-Net achieves a precision of 91.2%, arecall of 88.1%, an $mAP@0.5$ of 93.9%, and an $mAP@0.5{:}0.95$ of 69.7%. Compared with the strongest baseline YOLOv12, MDFE-Net yields an improvement of $+2.3\%$ in $mAP@0.5$ and $+5.6\%$ in $mAP@0.5:0.95$. These results demonstrate the superior robustness of MDFE-Net under varying inspection conditions, such as changes in illumination, viewpoints, and background complexity, which are common in real-world wind turbine blade inspections.

The performance gains can be attributed to the degradation-aware feature enhancement module, which alleviates the loss of fine-grained details during the downsampling process, thereby improving the detection of small and weak defect patterns. In addition, the background-aware attention mechanism helps suppress irrelevant responses caused by complex blade textures and environmental noise, leading to higher detection precision. arising from complex blade textures and environmental noise, thereby improving

Generalization on DIOR. To further assess cross-domain generalization, MDFE-Net is evaluated on the DIOR dataset, a large-scale remote sensing benchmark characterized by diverse object categories, dense layouts, and complex backgrounds. As shown in Table III, MDFE-Net achieves 89.3% precision, 82.1%recall, 87.6% $mAP@0.5$, and 68.6% $mAP@0.5:0.95$, outperforming YOLOv12 by $+2.1\%$ in $mAP@0.5$ and $+1.9\%$ in $mAP@0.5:0.95$.

Despite the substantial domain gap between industrial defect detection and remote sensing object detection, MDFE-Net maintains a clear performance advantage, indicating the strong transferability of the proposed feature enhancement strategy. The structure-preserving multi-scale fusion module plays a critical role in maintaining consistent object representations across scales, particularly for detecting small, densely distributed objects in aerial images.

IV. Conclusion

This paper presents MDFE-Net, a multi-level degradation-aware feature enhancement network designed to address the challenges of small-object detection under complex industrial backgrounds. By integrating three complementary modules, degradation-aware feature enhancement, background-aware attention, and structure-preserving multi-scale fusion, MDFE-Net enhances small-defect detection while suppressing irrelevant background noise. Extensive experiments on wind turbine blade and remote sensing datasets show that MDFE-Net outperforms state-of-the-art methods, achieving superior precision, recall, and mAP while maintaining efficiency. These results highlight the robustness and generalization capability of the proposed method, offering a practical solution for real-world industrial inspection tasks.

Acknowledgment

This work was funded by the 2022 Inner Mongolia Talent Support Project (Grant No. DC2300001440), the Science Research Foundation of Inner Mongolia University of Technology (Grant No. DC2300001262), and the Inner Mongolia Natural Science Foundation (Grant Nos. 2024MS06017, 2025MS06036). We thank Ran Guo and Min Xiong for their efforts in the formatting refinement of this manuscript, and we are also grateful to all reviewers and editors for their valuable comments. The authors acknowledge the use of large language model technology (DeepSeek) for language polishing and editing assistance in the preparation of this manuscript. All scientific content and technical contributions remain the sole responsibility of the authors.

References

[1] Xiaolong Chen, Yu Zhang, Hongwei Li, and Bin Zhou. Structure-preserving multi-scale fusion for fine-grained defect detection. *Pattern Recognition*, 145:109876, 2024.

[2] Zhiling Chen, Chen Chen, Shihua Lin, Yifan Wu, and Ya Yu. Coca: Fusing position embedding with collinear constrained attention for polyp segmentation. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2022.

[3] Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian. Centernet: Keypoint triplets for object detection. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV 2019)*, pages 6568–6577, 2019.

[4] Z. Fu, F. Zhang, X. Ren, B. Hao, X. Zhang, C. Yin, G. Li, and Y. Zhang. Le-yolo: Lightweight and efficient detection model for wind turbine blade defects based on improved yolo. *IEEE Access*, 2024.

[5] Lei He, Zunhui Yi, Chaoyang Chen, Ming Lu, Ying Zou, and Pei Li. Detail-preserving noise suppression post-processing for low-light image enhancement. *Displays*, 83:102738, 2024.

[6] Yanfei Jia, Guangda Chen, and Liquan Zhao. Defect detection of photovoltaic modules based on improved varifocalnet. *Scientific Reports*, 14(1):15170, 2024.

[7] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023.

[8] Glenn Jocher and Jing Qiu. Ultralytics yolo11, 2024.

[9] Ke Li, Gang Wan, and Gong Cheng. Dior, 2025.

[10] Fan Liu, Liang Yao, Chuanyi Zhang, Ting Wu, Xinlei Zhang, Xiruo Jiang, and Jun Zhou. Boost uav-based object detection via scale-invariant feature disentanglement and adversarial learning. *CoRR*, abs/2405.15465, 2024.

[11] Wei Liu, Dragomir Anguelov, and Dumitru. Erhan. Ssd: Single shot multibox detector. In B Leibe, J Matas, N Sebe, and M Welling, editors, *COMPUTER VISION-ECCV 2016, PT I*, volume 9905 of *Lecture Notes in Computer Science*, pages 21–37, 2016.

[12] Yang Liu, Jian Wang, Wei Zhang, and Chen Li. Feature pyramid fusion network for small defect detection. *IEEE Transactions on Industrial Informatics*, 18(9):6215–6225, 2022.

[13] Wenyu Lv, Yian Zhao, Shang Xu, and et al. DETRs beat YOLOs on real-time object detection. arXiv, 2023. Preprint.

[14] Daliang Ouyang, Su He, and Guozhong Zhang. Efficient multi-scale attention module with cross-spatial learning. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, Rhodes Island, Greece, June 2023. IEEE.

[15] T. Qiu, X. Li, Z. Li, et al. Locate, enhance and fuse: a progressively optimized network for camouflaged object detection. *Multimedia Tools and Applications*, 84(19):21117–21139, 2025.

[16] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.

[17] Wenqi Ren, Shifang Liu, Lin Ma, et al. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing*, pages 1–1, 2019.

[18] ASM SHIHAVUDDIN and Xiao Chen. Dtu - drone inspection images of wind turbine, 2018.

[19] Yunjie Tian, Qixiang Ye, and David Doermann. Yolov12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524*, 2025.

[20] Eleftherios G. Vourkos, Rafaela A. Agathokleous, Andreas S. Panayides, Soteris A. Kalogirou, and Eftychios G. Christoforou. Improving photovoltaics hotspot and dirt accumulation detection with computer vision: A systematic yolo model comparison based on field aerial images. *Sustainable Energy Technologies and Assessments*, 2025.

[21] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.

[22] Hao Wang, Xu Chen, Ming Liu, and Qijun Zhao. Scale-consistent feature alignment for multi-scale object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12345–12354, 2023.

[23] Jiaming Wang and Na Liu. Lsc-yolo: Small target defects detection model for wind turbine blade based on yolov9. In *Neural Information Processing*, pages 256–270, Singapore, 2026. Springer Nature Singapore.

[24] Jun Wei, Shuhui Wang, and Qingming Huang. F^3net: fusion, feedback and focus for salient object detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 12321–12328, 2020.

[25] Yang Zhao, Ke Zhan, Zhen Wang, and Wenzhong Shen. Deep learning-based automatic detection of multitype defects in photovoltaic modules and application in real production line. *Progress in Photovoltaics: Research and Applications*, 29(4):471–484, 2021.

[26] L. Zou, A. Chen, C. Li, X. Yang, and Y. Sun. Dcw-yolo: An improved method for surface damage detection of wind turbine blades. *Applied Sciences*, 14(19):8763, 2024.