

Article

Axial Motion Prior-Guided Robust Detection of Surface Defects in Hoist Wire Ropes

Yupeng Wang ¹, Boxuan Shang ², Changkuan Liu ¹, Wenbin Sun ³, Xueyi Zhao ³, Shupeng Sun ¹ , Ning Jiang ⁴ and Mengchao Zhang ^{1,3,*} 

¹ College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, Qingdao 266590, China; wyp3212@sdust.edu.cn (Y.W.); lushtreesun@163.com (S.S.)

² College of Ocean Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

³ Libo Heavy Industries Science and Technology Co., Ltd., Tai'an 271000, China

⁴ College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China; jeanine425@163.com

* Correspondence: zhangmc1995@sdust.edu.cn

Abstract

Reliable detection of surface defects in hoist wire ropes is essential for the safe operation of hoisting systems. Because of the stranded rope structure, early broken wires and local scars often appear as weak, small-scale features embedded in repetitive textures. Axial rope motion introduces directional blur during image acquisition, further weakening defect boundaries and increasing missed detections and localization errors. To address these challenges, we propose RMP-YOLOv8n, an axial motion prior-guided detector based on YOLOv8n. The method incorporates a rope motion prior-guided high-frequency enhancement module (RMP-HFEM), which uses axial edge and motion high-pass responses to guide shallow feature enhancement, together with a P2 branch that retains spatial detail for small defects. All models were trained exclusively on clear images from the training split, while paired axial motion-blur subsets generated exclusively from the held-out test images were used only for robustness evaluation. Experimental results on the wire-rope defect dataset show that, under severe axial motion blur, the mAP@50 of the clear-trained YOLOv8n decreases from $94.87 \pm 0.12\%$ to $48.31 \pm 0.11\%$, whereas RMP-YOLOv8n retains $90.62 \pm 0.08\%$ mAP@50 and $87.28 \pm 0.15\%$ recall across three training seeds, demonstrating substantially stronger robustness to axial motion blur. The proposed method provides technical support for the stable deployment of online hoist wire-rope inspection systems under high-speed operation, motion blur, and weak small-defect features.

Keywords: hoisting wire rope; defect detection; motion prior; machine vision



Academic Editors: Rui Araújo and Alexandru Stanciu

Received: 23 June 2026

Revised: 30 July 2026

Accepted: 1 August 2026

Published: 5 August 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.

1. Introduction

Wire ropes are critical load-bearing components in mine hoists, cranes, ropeways, elevators, and other continuous transportation systems [1]. They undertake traction transmission, load transportation, and safety protection under alternating loads, repeated bending, frictional wear, corrosion, vibration, and impact [2]. During long-term service, surface defects such as broken wires, local scars, wear marks, corrosion spots, and abnormal strand textures may gradually appear [3–5]. At the early stage, these defects are usually manifested as tiny breakpoints, local bright regions, weak edges, or slight texture discontinuities. If they are not detected in time, local surface damage may further develop

into strand degradation or strength loss, threatening the safe operation of hoisting systems. Therefore, reliable online perception of wire-rope surface defects is important for intelligent monitoring and safety maintenance of hoisting equipment.

Existing wire-rope inspection methods mainly include manual inspection, magnetic flux leakage testing, ultrasonic testing, acoustic emission monitoring, and machine vision inspection [6–9]. Manual inspection usually relies on touching, visual observation, and caliper measurement. Although simple, it is labor-intensive, discontinuous, and strongly affected by the experience of inspectors. Magnetic flux leakage and ultrasonic methods can reflect internal damage or physical-property changes, but they usually depend on sensor arrangement, signal conditioning, and professional interpretation [10,11]. Acoustic emission monitoring can capture dynamic damage signals, but it is sensitive to environmental noise and operating-state variation [12]. Compared with these methods, machine vision inspection is non-contact, intuitive, and easy to integrate with industrial cameras and edge computing devices [13,14]. It can provide direct spatial localization of visible surface defects and is therefore attractive for online inspection.

However, visual detection of wire-rope surface defects is not a conventional object detection problem. The surface of a wire rope is composed of periodic strand textures, where abundant normal edges and high-frequency responses can be confused with defect boundaries. Early broken wires, local scars, and bright-spot defects are usually small in size and low in contrast and may only appear as weak local texture disturbances. As feature maps are down-sampled in deep detection networks, these small and weak defect cues can be compressed or lost. This makes small-defect localization and weak-feature extraction key challenges in wire-rope visual inspection.

YOLO-series detectors are widely used in industrial inspection because their single-stage architecture combines rapid inference with compact deployment [15–17]. Recent improvements include feature-pyramid fusion [18], attention-based sharpening [19], multi-branch feature aggregation [20,21], and shallow detection branches. These methods mainly optimize generic feature representation; they do not explicitly constrain enhancement according to a known degradation direction. On a wire-rope surface, this distinction matters because directional blur can suppress defect boundaries while periodic strand textures continue to produce strong responses. A motion prior tied to the equipment geometry provides a task-specific constraint that general-purpose modules do not supply.

Robustness studies of convolutional neural networks have shown that blur, noise, resolution loss, and illumination variation can substantially affect model predictions [22]. Corruption-based object-detection benchmarks likewise indicate that performance on clean images does not necessarily transfer to degraded imagery [23]. In online hoist wire-rope inspection, motion blur has a clearer physical origin than generic random corruption: the rope moves continuously along a mechanically constrained path, and the resulting blur is governed mainly by the projected axial speed, exposure time, and camera installation. Motion blur in this task is consequently treated as a predictable operating-condition degradation rather than an arbitrary image perturbation.

This observation changes the formulation of the problem. Wire-rope defect detection should be evaluated not only under clear-image conditions, but also as a robustness-oriented perception task under predictable directional degradation. Lei et al. [14] addressed a closely related problem for elevator hoisting steel belts using YOLO11-DED, in which dynamic up-sampling improves multi-scale feature reconstruction under motion blur. Unlike this sampling-optimization strategy, the present study uses the known axial motion direction to guide shallow feature enhancement, thereby linking the network response directly to the imaging characteristics of moving wire ropes.

Based on this analysis, this paper proposes RMP-YOLOv8n, an axial motion prior-guided robust detector for hoist wire-rope surface defects. The axial prior is incorporated at two levels. At the network level, RMP-HFEM uses directional edge and motion high-pass responses to guide shallow feature enhancement, while a P2 branch retains spatial detail for small defects. At the evaluation level, independent axial motion-blur subsets are constructed only from the held-out test split. The training and validation sets remain clear, allowing the experiments to examine whether the proposed structure generalizes to previously unseen directional degradation without introducing blurred samples into parameter optimization. The main contributions of this study are as follows:

1. A rope motion prior-guided high-frequency enhancement module, RMP-HFEM, is proposed. It introduces the known axial motion direction into shallow feature enhancement to improve the representation of defect residual information under blurred conditions.
2. A P2 small-object detection branch is introduced to preserve shallow spatial details and improve localization of tiny defects.
3. Comparative experiments with multiple YOLO-series networks, module ablation studies, and motion-blur tests with different intensities are conducted to analyze the detection performance and robustness of the proposed method under clear and motion-blurred conditions.

The remainder of this paper is organized as follows. Section 2 describes the dataset and task definition, followed by the proposed method in Section 3 and the experimental settings in Section 4. Section 5 presents the experimental results, while Section 6 discusses the main findings and limitations. Sections 7 and 8 provide the conclusions and outline future research directions, respectively.

2. Task Definition and Data Preparation

The wire-rope dataset used in this study was self-collected and contains 1318 images, 1512 annotated instances, and two defect categories: break and thunderbolt. Break mainly denotes broken wires and local fractures, whereas thunderbolt denotes bright spots, scars, or regions with marked texture disturbance. Representative samples are shown in Figure 1; All displayed images are from the focal dataset, and the target defects are highlighted by bounding boxes. Because the rope surface comprises periodic strand structures, defects appear as local abnormalities embedded in textured backgrounds rather than isolated objects with complete contours.

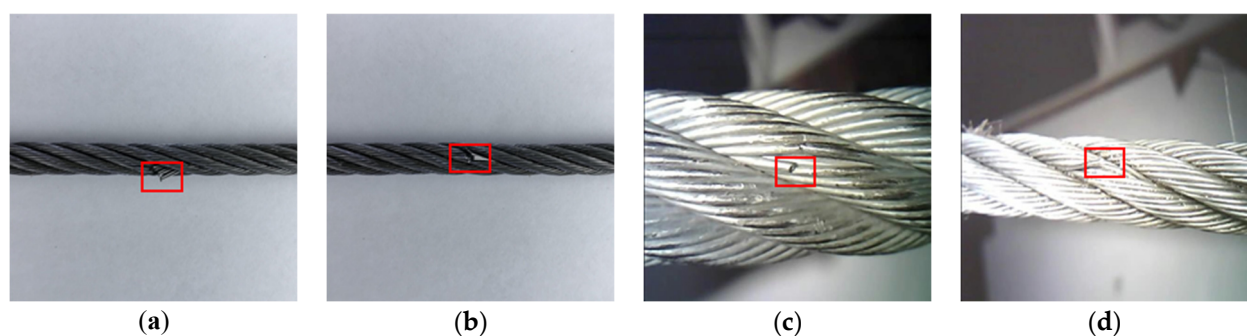


Figure 1. Representative surface-defect images from the focal wire-rope dataset: (a,b) break defects and (c,d) thunderbolt defects. Red boxes indicate the annotated defect regions.

The dataset was divided at the image level into a training set of 1054 images, a validation set of 131 images, and a test set of 133 images. These subsets contain 1214, 150, and 148 annotated defect instances, respectively. Of the 1512 instances in the complete dataset, 822 are labelled as break and 690 as thunderbolt. Among the 1318 images, 1160 contain a

single annotated defect, 157 contain multiple defects, and one image has an empty annotation file and was retained as a background sample. Thirty-four images contain both defect categories, including 26 training images, three validation images, and five test images. The maximum number of annotated defects in one image is eight.

The training set was used for parameter optimization, the validation set for checkpoint selection, and the test set exclusively for final evaluation. The image-level split was completed before construction of the blurred test subsets, preventing any derived test image from entering training or model selection. All annotations were stored in YOLO format. Detailed rope specifications were not consistently recorded during data acquisition. Accordingly, the present results demonstrate performance on the collected dataset but should not be interpreted as cross-specification validation.

In online inspection scenarios, the wire rope usually moves continuously along its axial direction. Affected by rope speed, camera exposure time, and illumination conditions, the captured images are prone to motion blur along the rope length direction. A conventional clear-image test set cannot fully reflect the reliability of the algorithm under this working condition. To evaluate the robustness of detectors under motion blur, this study applied axial motion blur with different intensities to the 133 original test images while keeping their labels unchanged, thereby constructing three blurred test subsets corresponding to mild, moderate, and severe blur. Each blurred subset contains 133 images, and a total of 399 axial motion-blurred test images were generated.

Figure 2 qualitatively illustrates the appearance of the same wire-rope image under progressively increasing axial motion blur. Kernel lengths of 7, 15, and 25 pixels were applied along the rope axis to provide a clear visual comparison of different degradation levels. As the kernel length increases, the strand texture and defect boundaries become progressively stretched and weakened along the axial direction. These images are presented only to illustrate the visual characteristics of axial motion blur.

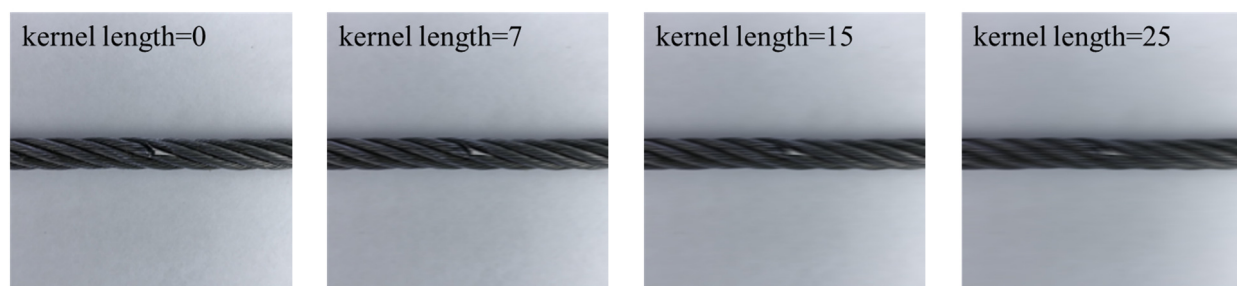


Figure 2. Axial motion-blur degradation at different kernel lengths.

3. Network Structure and Improved Method

This study adopts YOLOv8n [21] as the baseline and develops RMP-YOLOv8n for wire-rope surface-defect detection under axial motion blur. As shown in Figure 3, the proposed network consists of a backbone, an RMP-HFEM module, a multi-scale feature-fusion neck, and a four-scale detection head. Compared with standard YOLOv8n, RMP-YOLOv8n introduces a P2 detection branch for small-defect localization and embeds RMP-HFEM into the shallow feature-extraction stage for motion-prior-guided feature enhancement.

Wire-rope defects generally occupy small image regions and exhibit weak boundaries against periodic strand textures. Standard YOLOv8n performs detection using the P3, P4, and P5 feature maps, whose corresponding strides are 8, 16, and 32. For early broken wires, local scars, and bright-spot defects, the P3 resolution may be insufficient to preserve weak edges and local texture discontinuities after repeated down-sampling. A P2 branch with a

stride of 4 is therefore added to retain finer spatial and geometric information. Together with the original P3, P4, and P5 heads, it forms a four-scale detection structure.

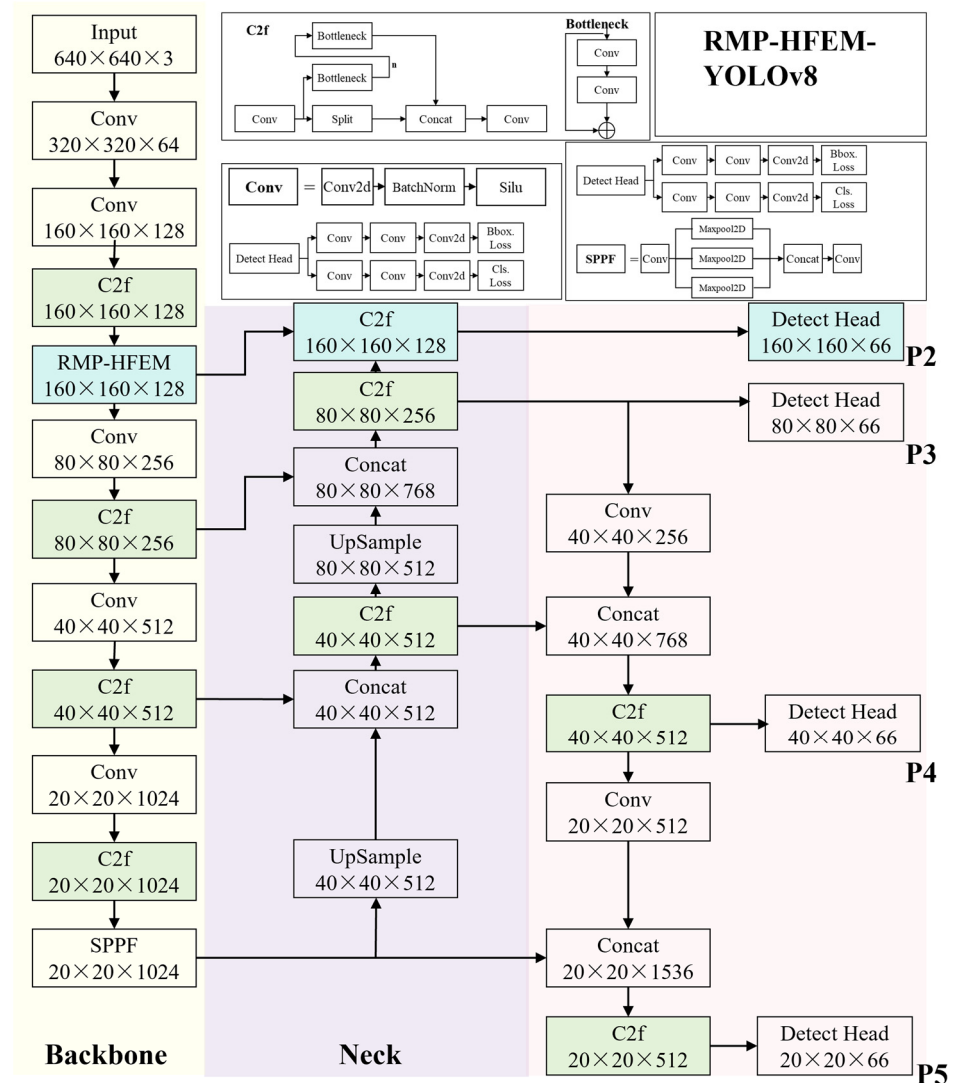


Figure 3. Architecture of RMP-YOLOv8n.

As previously stated, during online inspection, the wire rope moves continuously along its axial direction. In the exposure stage, defect boundaries and local high-frequency textures are stretched along the rope length direction, further weakening the difference between defects and the strand-texture background. The motion direction of the wire rope is determined by the equipment structure, camera installation position, and running path, and can therefore be regarded as a known engineering prior. Based on this characteristic, RMP-HFEM is designed to introduce the axial motion direction into the shallow feature enhancement process. This module is not an image deblurring network and does not reconstruct a clear image before detection. Instead, it enhances defect residual information that remains discriminative under blurred conditions at the feature level.

The internal structure of RMP-HFEM is shown in Figure 4. The module consists of a wire-rope axial edge response, a wire-rope normal edge response, a Laplacian high-frequency response, an axial motion high-pass response, and a learnable local convolution response. The axial edge response is used to characterize texture variations distributed along the rope length direction, while the normal edge response captures local abrupt changes perpendicular to the rope direction. The Laplacian high-frequency response

enhances direction-independent local details, and the axial motion high-pass response highlights residual signals along the motion-blur direction. The learnable local convolution response is used to preserve the network's own data-driven representation capability. These responses are subsequently fused to regulate the enhancement of the learnable feature, as formulated below.

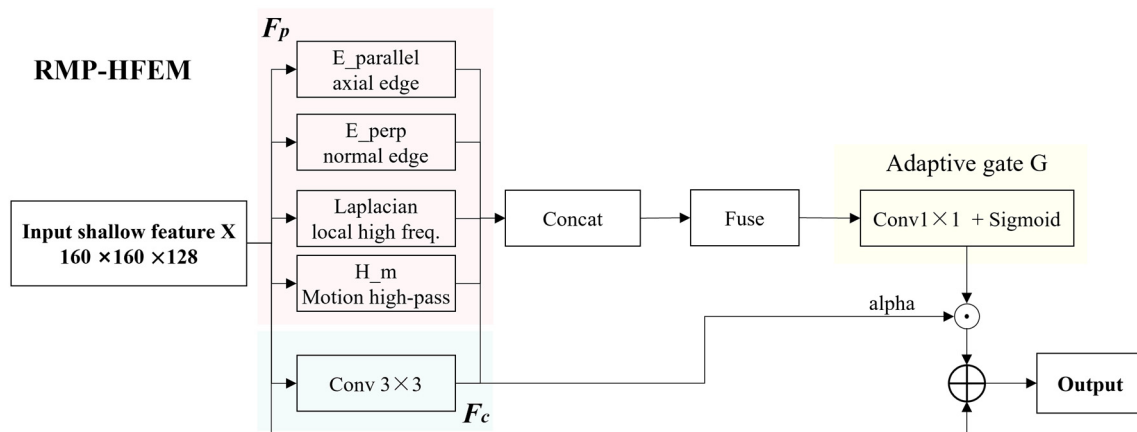


Figure 4. Architecture of RMP-HFEM.

Let X denote the input shallow feature. In the fixed camera configuration, the rope axial direction is determined by the installation geometry. Its angle relative to the horizontal image axis is denoted by θ . In the present implementation, θ is set to -20° after image alignment according to the dominant rope direction. Two Sobel derivative kernels are first defined as:

$$K_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, K_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

The axial and normal directional kernels are obtained by projecting the two-dimensional gradient onto the rope axial direction and its normal direction:

$$K_{\parallel} = \cos \theta K_x + \sin \theta K_y, \\ K_{\perp} = -\sin \theta K_x + \cos \theta K_y.$$

The corresponding feature responses are calculated using channel-wise depthwise convolution:

$$E_{\parallel}(X) = X_{*dw} K_{\parallel}, E_{\perp}(X) = X_{*dw} K_{\perp}$$

where $*dw$ denotes channel-wise depthwise convolution. The axial response measures local intensity variation along the rope direction, whereas the normal response captures complementary transverse discontinuities. This directional construction follows classical gradient-based edge analysis and orientation-aware filtering [24,25].

To preserve local high-frequency variations independent of the rope direction, a fixed Laplacian kernel is adopted:

$$K_L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}, L(X) = X_{*dw} K_L$$

This branch highlights local intensity changes associated with small bright spots, scars, and weak texture discontinuities. Unlike the directional derivatives, the Laplacian provides isotropic high-frequency evidence [25].

The motion high-pass branch is designed to emphasize local residual variations along the known axial motion direction. Let $\delta \in \mathbb{R}^{5 \times 5}$ denote a discrete unit-impulse kernel whose central element is one and whose remaining elements are zero. Let $B_\theta \in \mathbb{R}^{5 \times 5}$ denote a normalized two-point axial averaging kernel aligned with the rope-axis direction θ . Its two sampling positions are symmetrically located at $c \pm 2(\cos \theta, \sin \theta)$, where c denotes the kernel center. Bilinear interpolation is used when the sampling coordinates are non-integer, and the weights of B_θ are normalized to sum to one. The motion high-pass kernel is obtained by mean-centering and L_1 -normalized the difference between the central impulse and the axial averaging kernel:

$$\tilde{K}_m = \delta - B_\theta,$$

$$K_m = \frac{\tilde{K}_m - \text{mean}(\tilde{K}_m)}{\|\tilde{K}_m - \text{mean}(\tilde{K}_m)\|_1},$$

where $\tilde{K}_m$ denotes the unnormalized motion high-pass kernel and $\text{mean}(\cdot)$ calculates the arithmetic mean of all kernel elements. Mean subtraction enforces a zero-DC response, while L_1 normalization constrains the response magnitude and improves numerical consistency across different axial directions.

$$H_m(X) = X_{*dw} K_m,$$

The motion high-pass response compares the central feature with neighboring responses along the rope axis. It is not a deblurring operation; it supplies a directional residual cue that may remain informative after axial averaging has weakened the original boundary [25].

The four fixed-prior responses are concatenated as

$$F_p = \text{Cat}[E_{\parallel}(X), E_{\perp}(X), L(X), H_m(X)].$$

In parallel, a learnable local branch extracts data-driven features:

$$F_c = \text{Conv}_{3 \times 3}(X).$$

The fixed prior responses do not directly form the final enhanced feature. They are fused with F_c to estimate an adaptive gate:

$$F = \text{Conv}_{1 \times 1}(\text{Cat}[F_p, F_c]),$$

$$G = \sigma(\text{Conv}_{1 \times 1}(F)),$$

where σ denotes the Sigmoid function. The module output is:

$$Y = X + \alpha(G \odot F_c),$$

where element-wise multiplication modulates the learnable local feature and the residual scaling coefficient is initialized to 0.35 and optimized during training. The fixed kernels are non-trainable depth-wise filters; the local 3×3 convolution, fusion layer, and gate-generating 1×1 convolution are learnable.

The residual connection preserves the original feature representation, while the adaptive gate regulates the contribution of the enhanced feature according to the combined directional and high-frequency responses. When these responses contain defect-related residual information, the corresponding feature components can be strengthened; in periodic strand-texture regions, unnecessary enhancement can be suppressed. The module

output is therefore not determined solely by the fixed kernels, and the data-driven representation capacity of the YOLOv8 backbone is retained.

The individual operations used in RMP-HFEM are based on established techniques. Depth-wise convolution enables efficient channel-wise spatial filtering [26], fixed directional kernels encode orientation priors [24], SE and CBAM perform adaptive feature recalibration [27,28], and residual learning and scaling facilitate the optimization of additional feature transformations [29,30]. These operations are not claimed as separate methodological innovations. The distinction of RMP-HFEM lies in using equipment-derived axial, normal, Laplacian, and motion high-pass responses to condition feature enhancement, rather than relying solely on generic feature statistics.

4. Experimental Settings

Two complementary experiment groups were designed. First, YOLOv8n, YOLOv8n + P2, YOLOv8n-RMP-HFEM, and RMP-YOLOv8n were trained under the same clear-image protocol and evaluated on the original clear test set to isolate structural effects. Here, RMP-YOLOv8n denotes the final structural model that integrates both the P2 small-object detection branch and the RMP-HFEM module. Second, the same clear-trained structural variants were evaluated on paired axial motion-blur subsets generated from the held-out test images. This design separates clear-image structural ablation from robustness evaluation under previously unseen degradation.

The dataset was divided at the image level before any blur generation. No synthetically blurred image was used for training, validation, checkpoint selection, or hyperparameter adjustment. After model training and selection had been completed, each of the 133 held-out test images was processed using axial motion-blur kernels of 7, 15, and 25 pixels to construct the mild, moderate, and severe subsets, respectively. The blur direction was aligned with the test-set rope direction of -19.79° , and the original annotations were retained because the degradation altered image appearance without changing defect locations. Thus, the test images were held out independently from model development, whereas the three degradation subsets were paired transformations of the same fixed test sample rather than mutually independent draws.

In the intended fixed-camera inspection system, the axial motion direction is determined by the wire-rope running path and camera installation geometry rather than estimated independently from each image. The prior angle can therefore be calibrated during system installation and remains approximately constant under normal guided operation. To evaluate tolerance to small calibration or operating deviations, an additional direction-mismatch experiment was conducted using the trained checkpoints without retraining. The RMP-HFEM prior angle was fixed at -20° , while the actual blur direction was offset by -10° , -5° , 0° , $+5^\circ$, and $+10^\circ$ from the nominal test-set direction of -19.79° . The kernel length was fixed at 15 pixels so that only the angular mismatch was varied.

All experiments were conducted on a Windows workstation equipped with an NVIDIA GeForce RTX 3060 GPU with 12 GB of memory. The software environment comprised Python 3.10.20, PyTorch 2.5.1 with CUDA 12.4, and Ultralytics 8.2.103. A consistent training configuration was adopted for all compared models, including an input resolution of 640×640 pixels, 100 training epochs, and a batch size of 16. Automatic mixed precision was disabled throughout the experiments.

The standard training-time augmentation settings provided by Ultralytics were applied to the clear training images. HSV perturbation, random translation, random scaling, horizontal flipping, and Mosaic augmentation were enabled, while rotation, shear, vertical flipping, MixUp, and Copy-Paste were disabled. No task-specific motion-blur augmentation was introduced. The detector configuration and augmentation operations followed

the Ultralytics implementation [21]. Training was completed for the full 100 epochs without early stopping, and the checkpoint with the highest validation fitness, calculated as $0.1 \times \text{mAP@50} + 0.9 \times \text{mAP@50:95}$, was used for final evaluation.

Each of the four YOLOv8n-based structural variants was trained three times using random seeds 0, 1, and 2. The dataset split, training settings, and checkpoint-selection method were identical in all runs. Each trained model was evaluated on the same clear, mild, moderate, and severe test sets. The final results are reported as the mean $\pm$ standard deviation of the three runs. The mild, moderate, and severe test sets were generated from the same 133 held-out test images, so that only the blur intensity changed during comparison. Because only three independent training runs were conducted, the reported standard deviations should be interpreted as a preliminary indication of reproducibility across the examined random seeds rather than as a precise estimate of the underlying training variance or a confidence interval.

5. Experimental Results

Experiments were conducted in three aspects: clear-image detection performance, structural ablation, and axial motion-blur robustness. The clear-image detection experiment was used to evaluate the basic recognition capability of different detectors on the original test set. The structural ablation experiment was used to analyze the effects of the P2 small-object detection branch and the RMP-HFEM module on detection performance. The axial motion-blur robustness experiment was used to evaluate the ability of the models to maintain detection performance under online imaging degradation.

5.1. Performance Evaluation of Baseline Networks

Table 1 presents the performance of the YOLO-series detectors on the clear test set. YOLOv8n achieved 94.88% mAP@50 and 46.58% mAP@50:95 with 3.01 M parameters and 8.1 GFLOPs, providing a strong balance between accuracy and computational cost. YOLOv8s reached 95.08% mAP@50 and 49.51% mAP@50:95, the highest clear-image accuracy among the compared conventional models, but its parameter count and computational cost increased to 11.13 M and 28.4 GFLOPs. YOLOv5n achieved 91.46% recall with an inference latency of 2.57 ms. By comparison, YOLOv10n and YOLOv10s achieved lower mAP@50 values of 82.87% and 90.41%, respectively.

Table 1. Detection results of different baseline detection models on the clear test set.

Model	Params/M	GFLOPs	Precision/%	Recall/%	mAP@50/%	mAP@50:95/%	Inference/ms
YOLOv5n	2.50	7.1	91.38	91.46	93.58	44.42	2.57
YOLOv8n	3.01	8.1	92.77	90.64	94.88	46.58	3.61
YOLOv8s	11.13	28.4	94.68	91.06	95.08	49.51	4.41
YOLOv10n	2.70	8.2	74.93	76.53	82.87	40.66	3.69
YOLOv10s	8.04	24.4	80.98	86.22	90.41	44.50	4.66

The clear-image results show that YOLOv8n has favorable basic detection performance and lightweight characteristics, making it suitable as the baseline for subsequent structural improvement and robustness enhancement. The following experiments in this study do not take the highest accuracy on clear images as the only objective, but instead focus on analyzing the ability of the model to maintain detection performance under axial motion blur.

5.2. Convergence Behavior and Accuracy of the Improved Algorithms

Four structural variants were trained for the ablation study: YOLOv8n, YOLOv8n + P2, YOLOv8n-RMP-HFEM, and RMP-YOLOv8n. Figure 5 presents their loss trajectories. Training and validation losses decreased and stabilized for all models without evident divergence. The curves confirm comparable optimization behavior, but do not by themselves demonstrate robustness to motion blur.

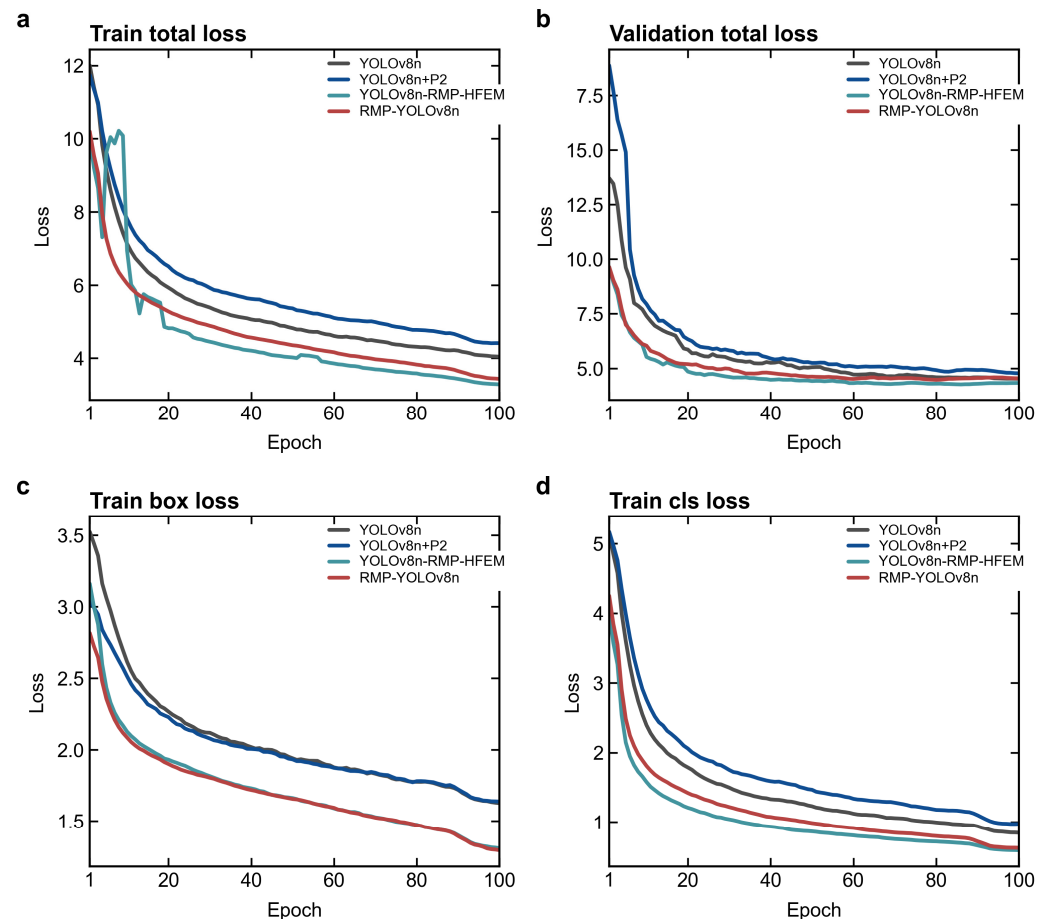


Figure 5. Training and validation loss curves: (a) training total loss; (b) validation total loss; (c) training box loss; and (d) training classification loss.

Figure 6 shows the mAP@50 and mAP@50:95 curves of the main models on the validation set. YOLOv8n-RMP-HFEM and RMP-YOLOv8n exhibit faster mAP improvement in the early training stage and enter a stable range after approximately 30–40 epochs. Compared with YOLOv8n and YOLOv8n + P2, the models incorporating motion-prior high-frequency enhancement achieve generally higher mAP@50:95 values, indicating more stable bounding-box localization quality. In the later training stage, RMP-YOLOv8n maintains high mAP@50 and mAP@50:95 values, suggesting that the proposed structure does not disrupt the normal convergence process of YOLOv8n and can further improve feature learning capability.

Importantly, the training curves mainly reflect the optimization process of the models and the accuracy variation on the validation set, and cannot alone prove reliability under online motion-blur conditions. To further evaluate the ability of the models to maintain performance under degraded imaging conditions, axial motion-blur test sets were constructed and robustness experiments were conducted.

Table 2 presents the three-run structural ablation results. Adding P2 increases mAP@50 from $94.87 \pm 0.12\%$ to $95.41 \pm 0.10\%$, while mAP@50:95 changes only from $46.58 \pm 0.11\%$

to $46.60 \pm 0.10\%$. P2 therefore provides a modest gain in coarse-threshold detection, but does not materially change stricter localization accuracy.

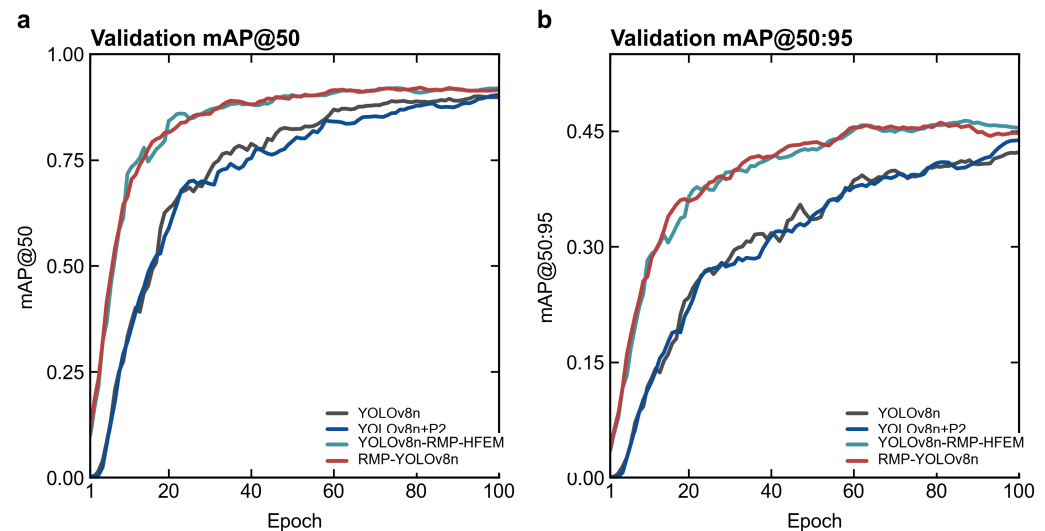


Figure 6. Validation mAP curves of the compared models: (a) mAP@50 and (b) mAP@50–95.

Table 2. Structural ablation results on the clear test set.

Model	Precision/%	Recall/%	mAP@50/%	mAP@50:95/%	Params/M	GFLOPs	Inference/ms
YOLOv8n	92.75 ± 0.07	90.64 ± 0.07	94.87 ± 0.12	46.58 ± 0.11	3.006	8.1	3.61
YOLOv8n + P2	92.84 ± 0.09	90.66 ± 0.07	95.41 ± 0.10	46.60 ± 0.10	3.199	12.8	4.06
YOLOv8n-RMP-HFEM	91.84 ± 0.09	94.55 ± 0.07	93.38 ± 0.09	42.25 ± 0.08	3.010	8.9	3.66
RMP-YOLOv8n	90.92 ± 0.10	94.61 ± 0.09	93.21 ± 0.08	40.87 ± 0.10	3.203	13.6	4.15

RMP-HFEM alone raises recall from $90.64 \pm 0.07\%$ to $94.55 \pm 0.07\%$, while mAP@50 decreases to $93.38 \pm 0.09\%$. The complete RMP-YOLOv8n reaches $94.61 \pm 0.09\%$ recall and $93.21 \pm 0.08\%$ mAP@50. These results indicate that the motion-prior module improves defect sensitivity but does not provide an accuracy advantage on clear images. The simultaneous increase in recall and decrease in precision suggests that high-frequency enhancement may also strengthen some defect-like responses in the periodic strand texture. The practical value of this trade-off is examined further through the axial motion-blur robustness experiments.

The computational cost of the proposed method was evaluated at an input resolution of 640×640 pixels. After standard convolution–batch-normalization fusion for inference, RMP-YOLOv8n contains 3.203 M parameters and requires 13.6 GFLOPs, compared with 3.006 M parameters and 8.1 GFLOPs for YOLOv8n. The parameter count therefore increases by approximately 6.55%, while the computational cost increases by approximately 67.90%. Most of the additional computation is introduced by the high-resolution P2 detection branch, as YOLOv8n + P2 already contains 3.199 M parameters and requires 12.8 GFLOPs. Relative to YOLOv8n + P2, RMP-HFEM adds only approximately 0.004 M parameters and 0.8 GFLOPs. Thus, RMP-HFEM introduces limited parameter overhead, whereas the main computational increase results from feature processing at the 160×160 P2 scale. Encouragingly, despite the additional computational overhead introduced by the modified structures, all improved models maintain satisfactory real-time performance.

5.3. Performance on the Blurred Dataset

Table 3 reports the fixed-seed performance of models trained on clear images and evaluated on the axial motion-blur test sets. All models degraded markedly under severe blur. YOLOv8n decreased from 94.88% mAP@50 on clear images to 48.27% under severe

blur, a reduction of 46.61 percentage points. Recall decreased from 90.64% to 43.79%, a reduction of 46.85 percentage points. Although YOLOv8s achieved the highest clear-image mAP@50 in this fixed-seed comparison, its severe-blur mAP@50 fell to 25.53%, corresponding to a 69.55-percentage-point decrease. These fixed-seed results indicate that clear-image accuracy alone does not predict robustness to axial motion blur.

Table 3. Fixed-seed robustness results of clear-trained models on the axial motion-blur test sets.

Model	Clear mAP@50/%	Mild/%	Moderate/%	Severe/%	mAP@50 Drop	Clear Recall/%	Severe Recall/%	Recall Drop
YOLOv5n	93.58	93.68	79.99	47.82	45.76	91.46	44.70	46.76
YOLOv8n	94.88	95.85	80.35	48.27	46.61	90.64	43.79	46.85
YOLOv8s	95.08	94.58	75.51	25.53	69.55	91.06	35.01	56.05
YOLOv10n	82.87	82.14	57.09	31.04	51.83	76.53	22.16	54.37
YOLOv10s	90.41	92.09	75.82	42.92	47.49	86.22	35.15	51.07

Table 4 reports the three-run robustness results of YOLOv8n and its structural variants on the paired axial motion-blur subsets generated from the held-out test set. Under severe blur, YOLOv8n achieves $48.31 \pm 0.11\%$ mAP@50 and $43.76 \pm 0.14\%$ recall, with a clear-to-severe mAP@50 reduction of 46.56 percentage points. The three runs produced similar severe-blur results, indicating that the previously reported fixed-seed value of 48.27% was not an isolated low result. Given the limited number of runs, this observation is interpreted only as preliminary evidence of reproducibility across the examined seeds. YOLOv8n-RMP-HFEM retains $89.97 \pm 0.09\%$ mAP@50 and $86.56 \pm 0.07\%$ recall under severe blur, while YOLOv8n + P2 achieves $88.12 \pm 0.14\%$ mAP@50 and $82.79 \pm 0.11\%$ recall. These results indicate that preserving shallow spatial details through the P2 branch improves robustness but does not fully account for the gain provided by motion-prior-guided feature enhancement.

Table 4. Robustness of the clear-trained structural variants on paired axial motion-blur subsets generated from the held-out test set (mean $\pm$ standard deviation over three seeds).

Model	Clear mAP@50/%	Mild/%	Moderate/%	Severe/%	Clear Recall/%	Severe Recall/%
YOLOv8n	94.87 ± 0.12	95.88 ± 0.08	80.32 ± 0.16	48.31 ± 0.11	90.64 ± 0.07	43.76 ± 0.14
YOLOv8n-RMP-HFEM	93.38 ± 0.09	92.53 ± 0.06	90.92 ± 0.13	89.97 ± 0.09	94.55 ± 0.07	86.56 ± 0.07
YOLOv8n + P2	95.41 ± 0.10	91.17 ± 0.12	93.42 ± 0.07	88.12 ± 0.14	90.66 ± 0.07	82.79 ± 0.11
RMP-YOLOv8n	93.21 ± 0.08	92.52 ± 0.09	90.63 ± 0.05	90.62 ± 0.08	94.61 ± 0.09	87.28 ± 0.15

Two local non-monotonic variations are observed in Table 4. The mAP@50 of YOLOv8n increases from $94.87 \pm 0.12\%$ on clear images to $95.88 \pm 0.08\%$ under mild blur, while that of YOLOv8n + P2 increases from $91.17 \pm 0.12\%$ under mild blur to $93.42 \pm 0.07\%$ under moderate blur. These results should not be interpreted as evidence that motion blur generally improves detection. Mild or moderate directional smoothing may suppress part of the repetitive strand texture while retaining the principal defect structure, allowing some borderline detections to exceed the IoU threshold of 0.5. Moreover, the three degradation subsets are derived from the same limited set of 133 test images and 148 defect instances. These local reversals are therefore regarded as dataset-specific responses rather than a general performance trend. The marked decreases under severe blur confirm the overall adverse effect of increasing motion degradation.

RMP-YOLOv8n achieves the highest severe-blur mAP@50 of $90.62 \pm 0.08\%$ and recall of $87.28 \pm 0.15\%$. Compared with the three-seed YOLOv8n baseline, it improves severe-

blur mAP@50 and recall by 42.31 and 43.52 percentage points, respectively. Its mAP@50 decreases by only 2.59 percentage points from clear to severe blur, compared with the 46.56-percentage-point reduction in YOLOv8n. Relative to YOLOv8n-RMP-HFEM, the complete model provides additional improvements of 0.65 percentage points in mAP@50 and 0.72 percentage points in recall under severe blur. Although this additional gain is modest, the results indicate that combining motion-prior-guided enhancement with the P2 branch provides the most stable overall performance under severe axial motion blur.

Although the preceding results demonstrate the robustness of the proposed method when the motion blur is aligned with the calibrated axial direction, slight directional disturbances may still occur during wire-rope operation because of vibration, lateral displacement, or small variations in the running trajectory. Consequently, the actual blur direction may deviate from the predefined directional prior, and the tolerance of the proposed method to such deviations requires further evaluation.

Under moderate motion blur (kernel length $L = 15$ pixels), the actual blur direction was varied relative to the calibrated axial prior, and the corresponding detection results are reported in Table 5. Under the matched condition, RMP-YOLOv8n achieved $90.63 \pm 0.05\%$ mAP@50, $89.13 \pm 0.09\%$ recall, and $51.32 \pm 0.08\%$ mAP@50:95. At direction offsets of -5° and $+5^\circ$, mAP@50 decreased by 3.22 and 1.89 percentage points, respectively. When the offsets increased to -10° and $+10^\circ$, the corresponding decreases reached 5.37 and 4.25 percentage points. Similar variations were observed in recall and mAP@50:95. These results indicate that RMP-YOLOv8n retains a certain tolerance to small deviations from the calibrated axial direction, whereas larger mismatches cause a more evident performance decrease. The experiment does not treat the motion direction as an unknown variable; instead, it quantifies the allowable deviation around the direction predetermined by the fixed camera–wire-rope geometry. In practical deployment, the directional prior should be recalibrated when the camera position or wire-rope running path changes.

Table 5. Sensitivity of RMP-YOLOv8n to motion-direction mismatch.

Direction Offset/ $^\circ$	Actual Blur Angle/ $^\circ$	mAP@50/%	Recall/%	mAP@50:95/%
−10	−29.79	85.26 ± 0.15	83.72 ± 0.17	46.15 ± 0.14
−5	−24.79	87.41 ± 0.11	86.05 ± 0.12	48.26 ± 0.10
0	−19.79	90.63 ± 0.05	89.13 ± 0.09	51.32 ± 0.08
+5	−14.79	88.74 ± 0.08	87.66 ± 0.11	49.57 ± 0.09
+10	−9.79	86.38 ± 0.13	85.10 ± 0.15	47.03 ± 0.12

5.4. Visualization Analysis of Detection Results

Figure 7 compares the detection results of the four structural variants on representative clear images. All models successfully localize the selected wire-rope defects, indicating that the introduction of the P2 branch and RMP-HFEM does not impair basic defect recognition. Differences are mainly observed in confidence stability and the alignment of predicted boxes with weak defect boundaries. These qualitative observations are consistent with the clear-image ablation results in Table 2, where the four models exhibit relatively small performance differences and RMP-YOLOv8n does not show an absolute accuracy advantage under clear conditions.

Figure 8 compares the detection results under progressively increasing axial motion blur. All displayed models were trained exclusively on clear images. Each column represents a blur level applied to the same original sample, while each row corresponds to a different detector. As the kernel length increases, defect boundaries and local texture discontinuities are progressively stretched and weakened along the rope axis. YOLOv8n exhibits missed detections and localization deviations under moderate and severe blur.

The P2 branch retains finer spatial information, but YOLOv8n + P2 still shows unstable responses when the defect boundary is strongly degraded.

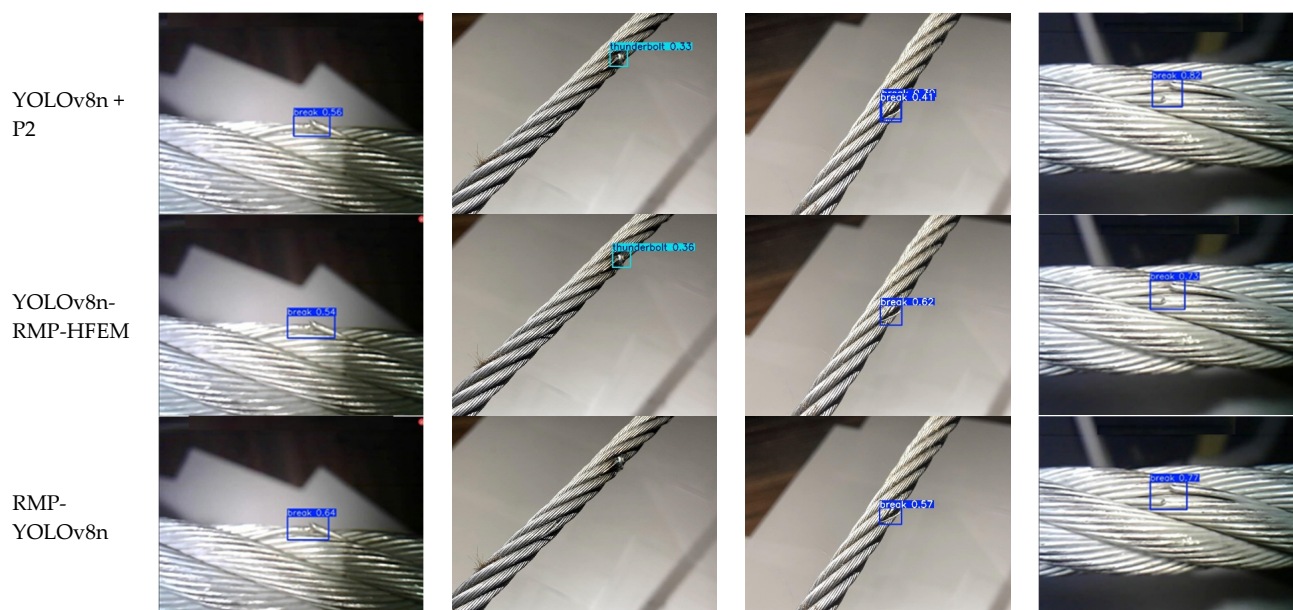


Figure 7. The detection visualization results of different models on clear sample.

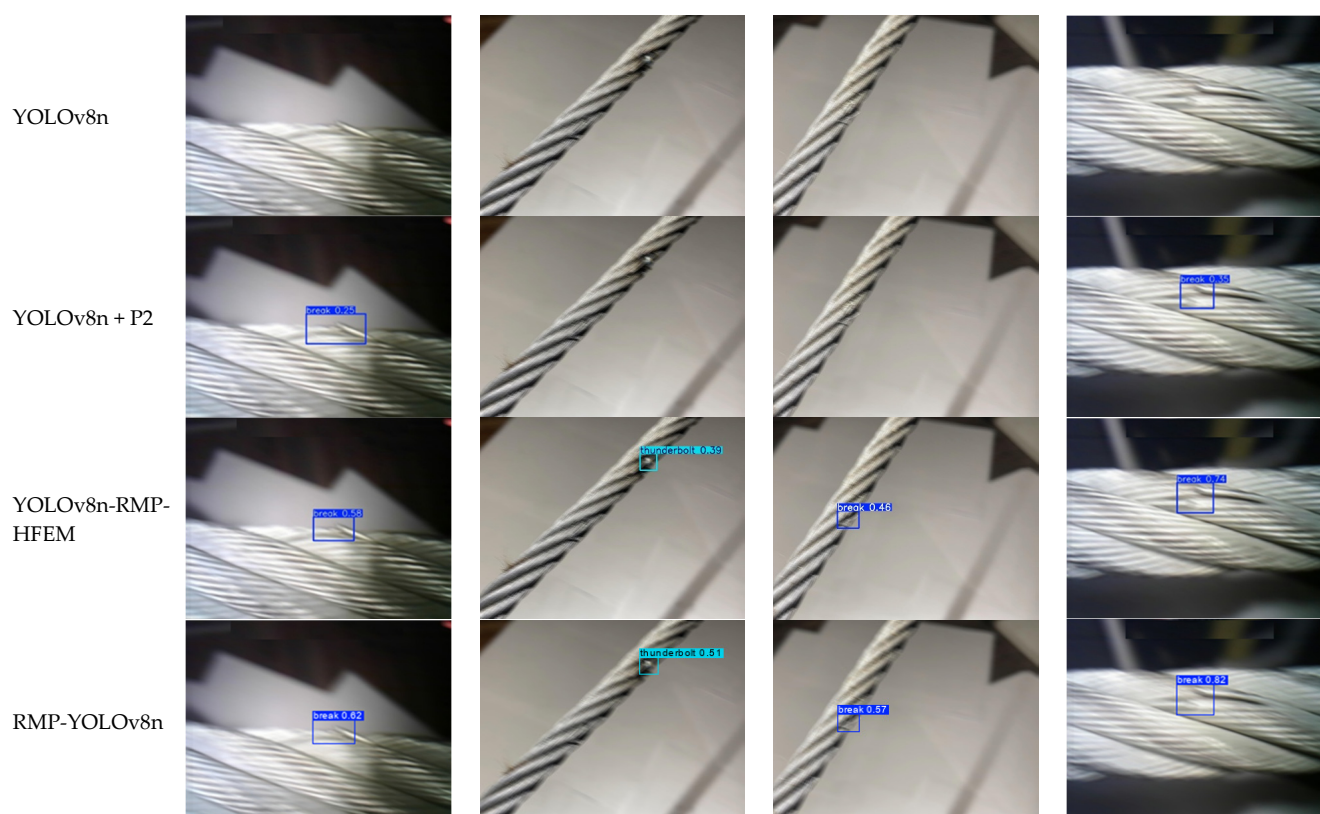


Figure 8. Detection results under axial motion blur. Rows denote models and columns denote test samples; all models within a column use the same input image, with kernel length indicated above the column.

YOLOv8n-RMP-HFEM and RMP-YOLOv8n retain more stable defect localization under severe blur, with fewer missed detections and better alignment between the predicted boxes and visible defect regions. This behavior is consistent with the three-seed

results in Table 4: under severe blur, YOLOv8n decreases to $48.31 \pm 0.11\%$ mAP@50 and $43.76 \pm 0.14\%$ recall, whereas RMP-YOLOv8n retains $90.62 \pm 0.08\%$ mAP@50 and $87.28 \pm 0.15\%$ recall. The selected examples do not independently establish robustness, but visually illustrate the performance trends obtained from the complete test set.

Taken together, the clear-image and blurred-image results show that clear-image accuracy alone is insufficient to characterize the reliability of online wire-rope inspection. The main benefit of RMP-HFEM is not a higher clear-image mAP, but improved retention of defect-related responses under axial motion degradation. Incorporating the known rope direction into shallow feature enhancement provides a more stable detection response to previously unseen axial blur without using blurred images during training.

6. Discussion

Surface-defect detection of hoist wire ropes should be considered together with motion degradation during online industrial imaging, rather than treated solely as a conventional object detection task under clear-image conditions. In a fixed inspection device, the dominant rope direction is constrained by the equipment structure, camera installation, and running path. This known axial direction provides an engineering prior for constructing controlled robustness tests and guiding shallow feature enhancement.

RMP-HFEM introduces axial edge and motion high-pass information into shallow feature processing. The clear-image ablation results show that high-frequency enhancement does not inherently improve detection accuracy. Periodic strand textures can generate high-frequency responses similar to those produced by defect boundaries, causing background structures to be enhanced together with true defects. In RMP-HFEM, the fixed responses do not replace data-driven features; they provide directional evidence for regulating the learnable enhancement feature. The robustness improvement is associated with the alignment between this directional evidence and the expected axial degradation.

The clear-image results also reveal a measurable trade-off. Compared with YOLOv8n, RMP-YOLOv8n decreases mAP@50 from 94.87% to 93.21%, while recall increases from 90.64% to 94.61% and precision decreases from 92.75% to 90.92%. This pattern suggests that the proposed structure preserves more weak defect responses but may also strengthen some defect-like patterns in the periodic strand texture. A 1.66-percentage-point reduction in clear-image mAP@50 may be acceptable in safety-oriented inspection scenarios where missed defects carry greater consequences than a limited increase in false alarms. In applications dominated by clear images and requiring strict false-alarm control, the detection threshold and operating strategy should be selected according to the actual risk balance. The present study does not establish an optimal switching strategy between clear and degraded operating conditions.

The gating operation itself is not claimed as the principal innovation. Depth-wise convolution enables efficient channel-wise spatial filtering. SE and CBAM provide general feature-recalibration mechanisms without assuming a predefined degradation direction. RMP-HFEM differs by conditioning feature enhancement on axial, normal, Laplacian, and motion high-pass responses derived from the known rope direction. Lei et al. [14] addressed motion-blurred defects in hoisting steel belts using dynamic up-sampling and adaptive feature reconstruction. Their method modifies feature sampling to recover spatial detail, whereas RMP-HFEM retains the original sampling process and introduces the equipment-derived motion direction into shallow feature modulation. The two approaches address a similar degradation problem through different technical routes.

The prior-guided design may also be applicable to fixed online inspection of belt conveyors, elevator hoisting belts, steel strips, and cables, where the dominant motion direction is constrained by the mechanical layout. Its scope is limited to systems with a

known and relatively stable motion direction and should not be generalized to arbitrary camera motion or unconstrained moving objects. The three blur levels were constructed as paired transformations of the same 133 held-out test images, allowing degradation severity to be compared while preserving image content and annotations. The reported standard deviations describe the variation observed across the three examined training seeds, but three runs are insufficient for a precise estimate of training variability. They are therefore interpreted only as preliminary reproducibility evidence and do not represent confidence intervals or test-sample uncertainty. The strict separation of the test set prevents data leakage, although broader generalization requires independently collected field images.

Several limitations remain. First, the motion-blurred samples are synthetic images with controlled directions, whereas actual hoist environments may also contain vibration blur, defocus, exposure variation, dust, water stains, illumination flicker, and non-uniform motion. Second, the dataset contains only two defect categories and does not support cross-specification validation across different rope diameters, constructions, lay directions, imaging devices, and operating speeds. Although normal rope-texture regions within the annotated images provide background supervision, the absence of independent defect-free sequences prevents evaluation of long-term false-alarm behavior under continuous operation. The reported results support defect localization and axial-blur robustness at the image level, rather than complete validation of an online alarm system. Third, the current detector processes individual frames and does not exploit temporal continuity. Optical flow, tracking, or temporal feature aggregation may improve prediction stability in video streams, but these methods also introduce additional computation and sensitivity to vibration and illumination changes. Finally, the fixed prior angle (-20°) is closely aligned with the synthetic test-blur direction (-19.79°). The direction-mismatch experiment indicates that the model retains limited tolerance within $\pm 5^\circ$, while larger deviations lead to a more evident performance decrease. Nevertheless, this evaluation remains limited to controlled synthetic blur within $\pm 10^\circ$, and robustness under real non-uniform directional disturbances requires further validation. In addition, blur robustness was evaluated primarily using mAP@50 and Recall because the principal engineering concern is the preservation of defect sensitivity under degraded imaging conditions. The stricter mAP@50:95 metric was not included and may provide complementary evidence regarding localization uncertainty under motion blur.

7. Conclusions

This paper proposes RMP-YOLOv8n, an axial motion-prior-guided detector for surface defects in hoist wire ropes. The method combines a P2 small-object detection branch with RMP-HFEM, using the known axial motion direction to guide shallow feature modulation. All models were trained exclusively on clear images, while robustness was evaluated on paired axial motion-blur subsets generated from the held-out test images.

The results show that clear-image accuracy alone does not reliably predict robustness to motion degradation. Across three training seeds, YOLOv8n decreases from $94.87 \pm 0.12\%$ mAP@50 on clear images to $48.31 \pm 0.11\%$ under severe blur, whereas RMP-YOLOv8n decreases from $93.21 \pm 0.08\%$ to $90.62 \pm 0.08\%$ and retains $87.28 \pm 0.15\%$ recall. These results indicate that motion-prior-guided feature enhancement improves resistance to controlled axial degradation without requiring blurred images during training.

The evidence remains limited to the present two-class dataset and synthetically generated axial motion blur. Within this scope, the study provides a technically feasible approach for incorporating predictable equipment motion into robust defect detection in fixed online inspection systems. Validation using real motion degradation, defect-free

operating sequences, and additional wire-rope specifications remains necessary before full industrial deployment.

8. Future Work

Future work will focus on validating the proposed method using real motion-blurred videos and defect-free operating sequences collected under field conditions. Rope diameter, construction, lay direction, operating speed, and imaging conditions will be recorded systematically to support cross-condition evaluation, followed by deployment and long-term testing on industrial edge devices. Adaptive and multi-directional priors will also be investigated to accommodate lateral vibration, rope rotation, and changes in camera viewpoint. For multimodal assessment, visual detections will be aligned with magnetic-flux-leakage, acoustic-emission, and tension measurements according to acquisition time and rope position. Modality-specific features or prediction scores will then be integrated using signal-quality-aware weighting, allowing unreliable measurements from individual sensors to contribute less to the final defect assessment.

Author Contributions: Y.W.: Writing—original draft; B.S.: Algorithm development; writing—original draft. C.L.: Data curation; algorithm development; W.S.: Data curation; algorithm development. X.Z.: Dataset annotation; results statistics; S.S.: Algorithm development and testing; N.J.: Algorithm development. M.Z.: Conceptualization; methodology; writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Science Foundation for Young Scientists of China (Grant No. 52504165) and the Open Fund of the State Key Laboratory of Intelligent Coal Mining and Strata Control (No. SKLIS202507).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: Authors Wenbin Sun, Xueyi Zhao and Mengchao Zhang were employed by the company Libo Heavy Industries Science and Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Xia, L.; Li, G.; Huang, K.; Peng, Y.; Tang, Y.; Zhou, Z.; Deng, R.; Chang, X. A Review of Tribological Behavior of Wire Ropes: Generation, Characteristics, Effects, and Protection. *Lubricants* **2026**, *14*, 62. [\[CrossRef\]](#)
2. Peng, H.; Zhang, Y.; Shangguan, L.; Zhao, M.; Li, B.; Yang, L.; Liu, Y. Review of Failure Mechanisms of Steel Wire Ropes Under Heavy-Load Conditions and the Anti-Friction Effects of Gel-like Grease. *Gels* **2025**, *11*, 900. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Olszyna, G.; Gašić, V.; Tytko, A. Some aspects on quantification of the wear at steel wire ropes. *Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.* **2022**, *236*, 2032–2041.
4. Li, Y.; Liu, J.; Tang, X.; Pan, J.; Liu, W.; Huang, Y.; Li, Z. Fault diagnosis methods for electromechanical special equipment: Review and prospects. *Meas. Sci. Technol.* **2025**, *36*, 076115. [\[CrossRef\]](#)
5. Xue, J.; Hu, E.; Sun, Y. A wire rope defect signal enhancement method based on spatial domain array sensors combined with multi-process adaptive signal processing. *Struct. Health Monit.* **2025**, 14759217251360644. [\[CrossRef\]](#)
6. Liu, S.; Sun, Y.; Jiang, X.; Kang, Y. A review of wire rope detection methods, sensors and signal processing techniques. *J. Nondestruct. Eval.* **2020**, *39*, 85. [\[CrossRef\]](#)
7. Mazurek, P. A comprehensive review of steel wire rope degradation mechanisms and recent damage detection methods. *Sustainability* **2023**, *15*, 5441. [\[CrossRef\]](#)
8. Zhou, P.; Zhou, G.; Zhu, Z.; He, Z.; Ding, X.; Tang, C. A review of non-destructive damage detection methods for steel wire ropes. *Appl. Sci.* **2019**, *9*, 2771. [\[CrossRef\]](#)

9. Wang, Y.; Chen, P.; Wei, Q.; Qi, J.; He, C.; Zhou, C. Multi-channel fusion scale transformed signals with magnetic leakage for damage detection in steel wire ropes. *Nondestruct. Test. Eval.* **2025**, *41*, 4884–4909. [\[CrossRef\]](#)
10. Rossteutscher, I.; Blaschke, O.; Dötzer, F.; Uphues, T.; Drese, K.S. Improved EMAT sensor design for enhanced ultrasonic signal detection in steel wire ropes. *Sensors* **2024**, *24*, 7114. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Xu, F.; Jiang, Z.; Jiang, Q.; Wang, X.S. Damage detection and assessment of broken wires in cables of a bridge based on magnetic flux leakage. *Exp. Tech.* **2023**, *47*, 907–920. [\[CrossRef\]](#)
12. Zejli, H.; Gaillet, L.; Laksimi, A.; Benmedakhene, S. Detection of the presence of broken wires in cables by acoustic emission inspection. *J. Bridge Eng.* **2012**, *17*, 921–927. [\[CrossRef\]](#)
13. Wu, H.; Liu, S.; Xu, Z.; Ji, Z.; Qian, M.; Yuan, X.; Wang, Y. An Interpretability Method for Broken Wire Detection. *Sensors* **2025**, *25*, 4002. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Lei, D.; Wei, Y.; Niu, Y.; Pan, J. Vision-based surface damage recognition method for elevator hoisting steel belts considering motion blur. *Eng. Res. Express* **2026**, *8*, 065503. [\[CrossRef\]](#)
15. Qian, M.; Wang, Y.; Liu, S.; Xu, Z.; Ji, Z.; Chen, M.; Wu, H.; Zhang, Z. Real time wire rope detection method based on Rockchip RK3588. *Sci. Rep.* **2025**, *15*, 30625. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Hong, J.; Ye, K.; Qiu, S. Study on lightweight strategies for L-YOLO algorithm in road object detection. *Sci. Rep.* **2025**, *15*, 7649. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Zhang, M.; Yue, Y.; Jiang, K.; Li, M.; Zhang, Y.; Zhou, M. Hybrid compression optimization based rapid detection method for non-coal conveying foreign objects. *Micromachines* **2022**, *13*, 2085. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Quan, Y.; Zhang, D.; Zhang, L.; Tang, J. Centralized feature pyramid for object detection. *IEEE Trans. Image Process.* **2023**, *32*, 4341–4354. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Xue, M.; Chen, M.; Peng, D.; Guo, Y.; Chen, H. One spatio-temporal sharpening attention mechanism for light-weight YOLO models based on sharpening spatial attention. *Sensors* **2021**, *21*, 7949. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Yang, Z.; Guan, Q.; Zhao, K.; Yang, J.; Xu, X.; Long, H.; Tang, Y. Multi-branch auxiliary fusion yolo with re-parameterization heterogeneous convolutional for accurate object detection. In *Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*; Springer Nature: Singapore, 2024; pp. 492–505.
21. Varghese, R.; Sambath, M. Yolov8: A novel object detection algorithm with enhanced performance and robustness. In *Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*; IEEE: New York, NY, USA, 2024; pp. 1–6.
22. Hendrycks, D.; Dietterich, T. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv* **2019**, arXiv:1903.12261.
23. Michaelis, C.; Mitzkus, B.; Geirhos, R.; Rusak, E.; Bringmann, O.; Ecker, A.S.; Bethge, M.; Brendel, W. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv* **2019**, arXiv:1907.07484.
24. Luan, S.; Chen, C.; Zhang, B.; Han, J.; Liu, J. Gabor convolutional networks. *IEEE Trans. Image Process.* **2018**, *27*, 4357–4366. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Woods, R.E.; Gonzalez, R.C. *Digital Image Processing*, 3rd ed.; Prentice Hall: Hoboken, NJ, USA, 2021.
26. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
27. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
28. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, 8–14 September 2018; pp. 3–19.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
30. Touvron, H.; Cord, M.; Sablayrolles, A.; Synnaeve, G.; Jégou, H. Going deeper with image transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, QC, Canada, 11–17 October 2021; pp. 32–42.

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.