frontiers | Frontiers in Mechanical Engineering

TYPE Original Research
PUBLISHED 07 October 2025
DOI 10.3389/fmech.2025.1683572

OPEN ACCESS

EDITED BY
Nima Rezazadeh,
Università della Campania Luigi Vanvitelli, Italy

REVIEWED BY
Mohammad Hossein Nejatiamiri,
Birmingham City University, United Kingdom
Shila Fallahy,
Polytechnic of Milan, Italy

*CORRESPONDENCE
Jianli Chen,
jlchan_gdit@hotmail.com

RECEIVED 11 August 2025
ACCEPTED 18 September 2025
PUBLISHED 07 October 2025

CITATION
Chen J, Tong J and Su J (2025) Design of a real-time abnormal detection system for rotating machinery based on YOLOv8.
Front. Mech. Eng. 11:1683572.
doi: 10.3389/fmech.2025.1683572

Design of a real-time abnormal detection system for rotating machinery based on YOLOv8

Jianli Chen*, Jie Tong and Jiang Su

College of Robotics, Guangdong Polytechnic of Science and Technology, Zhuhai, Guangdong, China

To address the issues of low detection accuracy and poor real-time performance in existing methods for detecting minor abnormalities such as cracks, oil leaks, and loose bolts in rotating industrial machinery under dynamic vibration conditions, this paper proposes a lightweight detection system based on YOLOv8 (You Only Look Once version 8) with adaptive feature enhancement. First, this paper employs a temporal motion compensation module based on optical flow to estimate and correct the vibration displacement between adjacent frames. Second, this paper designs a lightweight YOLOv8 network, using depthwise separable convolution instead of traditional convolution. Finally, this paper employs a weighted fusion strategy to improve the accuracy of small object detection in complex backgrounds. This model is deployed on the Jetson AGX Xavier edge computing platform, utilizing FP16 (half-precision floating-point) / INT8 (8-bit integer) quantization and asynchronous pipeline inference to ensure real-time processing capabilities on edge devices. The experimental results show that the method achieves an average detection accuracy of 97.8% (mAP@0.5) and 86.6% (mAP@0.5:0.95), with an average inference speed of 29.5 FPS (frames per second). This demonstrates that the method has reached industrial-grade performance in terms of detection accuracy, real-time performance, and deployment stability, making it highly valuable for practical applications.

KEYWORDS

rotating machinery, YoloV8 model, lightweight network, real-time detection, anomaly detection

1 Introduction

Industrial rotating machinery is the core power unit in key sectors such as energy, manufacturing, and transportation. Its operating status is directly related to production safety and efficiency (Zhang P. et al., 2025; Das et al., 2023; Gawde et al., 2023). With the development of intelligent manufacturing and the Industrial Internet, vision-based anomaly detection technology has gradually become an important means of equipment status monitoring (Jiang, 2022; Cui et al., 2023; Tang et al., 2023). However, complex industrial environments and the high-speed rotation of equipment often cause motion blur in captured images. Existing models struggle to balance accuracy, real-time performance, and the detection of minor anomalies like cracks and oil leaks. This limits the practical development of intelligent operation and maintenance systems (Li et al., 2024a; Ren et al., 2022; Li et al., 2024b). Therefore, it is urgent to build a high-precision, low-latency, and highly robust visual inspection system to achieve real-time and accurate recognition of abnormal conditions in rotating machinery. For abnormal monitoring of industrial equipment, Yadav et al. (2025) introduced an extended adaptive neural fuzzy

inference system method to perform fault diagnosis on rotating mechanical components using infrared thermal imaging; Xiao et al. (2025) proposed a multi-level information fusion fault diagnosis method, which has positive significance for improving multi-sensor data fusion fault diagnosis; Singh and Desai (2023) constructed a defect detection framework of machine vision and convolutional neural network, which can effectively perform image classification and achieve 100% accuracy for good category components; Suo et al. (2022) developed a nuclear fuel rod notch defect detection system based on machine vision to achieve efficient online detection of nuclear fuel rod notches. Natili et al. (2021) used industrial SCADA (Supervisory Control and Data Acquisition) and vibration data to monitor wind turbine bearings at multiple scales. These studies have promoted the development of rotating machinery condition monitoring from different dimensions, but most of them rely on dedicated sensors or offline detection methods, making it difficult to achieve low-cost, all-weather visual real-time monitoring. Visual methods are more flexible.

Among vision-based methods, Yang Y. et al. (2024) improved YOLOv5 for better crack boundary localization. Zhao et al. (2024) built a lightweight model using ShuffleNetv2 (ShuffleNet Version 2) and a coordinate attention mechanism to enhance detection accuracy. Li et al. (2021) proposed an accurate screw detection method that combines Faster R-CNN (Faster Region-based Convolutional Neural Network) and an innovative rotation edge similarity algorithm, achieving a small positioning deviation of 0.094 mm and a classification accuracy of 99.64%. Zhu et al. (2023) proposed a Transformer-based model with excellent feature extraction capabilities that can capture features directly from raw vibration signals; Khan et al. (2023) integrated YOLOv3 and MobileNet single-shot detectors to achieve faster image detection and more accurate positioning. Some of the models in the above methods have large computational loads, making it difficult to implement real-time inference on edge devices. They still have obvious shortcomings in terms of accuracy, speed, or robustness, especially in dynamic vibration scenarios, where it is difficult to balance small target detection capabilities with real-time inference efficiency.

This paper proposes a real-time detection method for abnormal conditions in rotating machinery, achieving collaborative optimization of high precision and low latency through an integrated "perception-compensation-detection-inference" architecture. First, this paper uses industrial cameras to capture video streams and introduces a temporal motion compensation module based on the Farneback optical flow method to achieve pixel-level inter-frame alignment and ensure the input quality of subsequent detection. Second, this paper builds a lightweight network based on YOLOv8, using depthwise separable convolution to reconstruct the backbone, compressing the number of channels to reduce the amount of computation, and embedding an adaptive spatial-channel attention module in the Neck layer to enhance the feature response to minor anomalies. The multi-scale fusion structure of PANet (Path Aggregation Network) was further optimized, and a weighted fusion mechanism with learnable weights was introduced. Finally, the model's FP16/INT8 quantization and asynchronous pipeline inference were implemented on the Jetson AGX Xavier platform through the TensorRT (Tensor Runtime) engine. The system's overall crash rate per frame (CRF) was less than 10^{-6} during continuous 24-h operation. This method innovatively proposes a collaborative mechanism combining optical flow compensation and ASCA (Adaptive Spatial-Channel Attention). It also employs a lightweight network and a learnable weighted fusion strategy to significantly reduce model complexity while maintaining accuracy, enabling efficient, stable, and deployable visual anomaly detection in industrial scenarios.

2 Algorithm design

2.1 Overall system architecture design

This paper constructs an integrated end-to-end visual inspection architecture combining perception, compensation, detection, and reasoning. This architecture uses industrial cameras for visual perception, introduces an optical flow temporal motion compensation module to suppress inter-frame jitter, and employs a lightweight YOLOv8 network combined with the ASCA attention mechanism for anomaly feature extraction. Finally, the TensorRT engine enables low-latency reasoning on the Jetson AGX Xavier platform, forming a closed-loop detection system, as shown in Figure 1.

Figure 1 illustrates the complete technical process of the YOLOv8-based real-time detection system for rotating machinery abnormalities: A high-resolution industrial camera captures RGB (Red, Green, Blue) video streams, which are fed into the temporal motion compensation module. Inter-frame displacement is estimated using optical flow, and back-mapping and bilateral filtering are used for alignment and denoising, resulting in a spatially consistent and clear image sequence as output. The model is then fed into a lightweight YOLOv8 network, with an ASCA module embedded within the Neck structure. A dual-branch spatial and channel-wise attention mechanism is used to enhance feature responses to subtle anomalies, while a weighted fusion strategy is introduced to optimize multi-scale feature transfer. The detection model uses an asynchronous inference pipeline on the Jetson AGX Xavier edge platform, and the final detection results are uploaded to the host computer monitoring interface, establishing an industrial-grade intelligent monitoring system with a closed-loop "perception-compensation-detection-response" process.

2.2 Construction of the temporal motion compensation module

This paper constructs a temporal motion compensation module and uses dense optical flow to estimate and correct the pixel-level motion field between consecutive frames (Pookkuttath et al., 2023; Li et al., 2023). Based on the principles of the Farneback dense optical flow algorithm and the real-time requirements of industrial edge computing, this paper sets the following key parameters to balance estimation accuracy and computational efficiency: the Gaussian pyramid scaling ratio is 0.5, and a total of five pyramid layers are constructed to achieve coarse-to-fine motion estimation

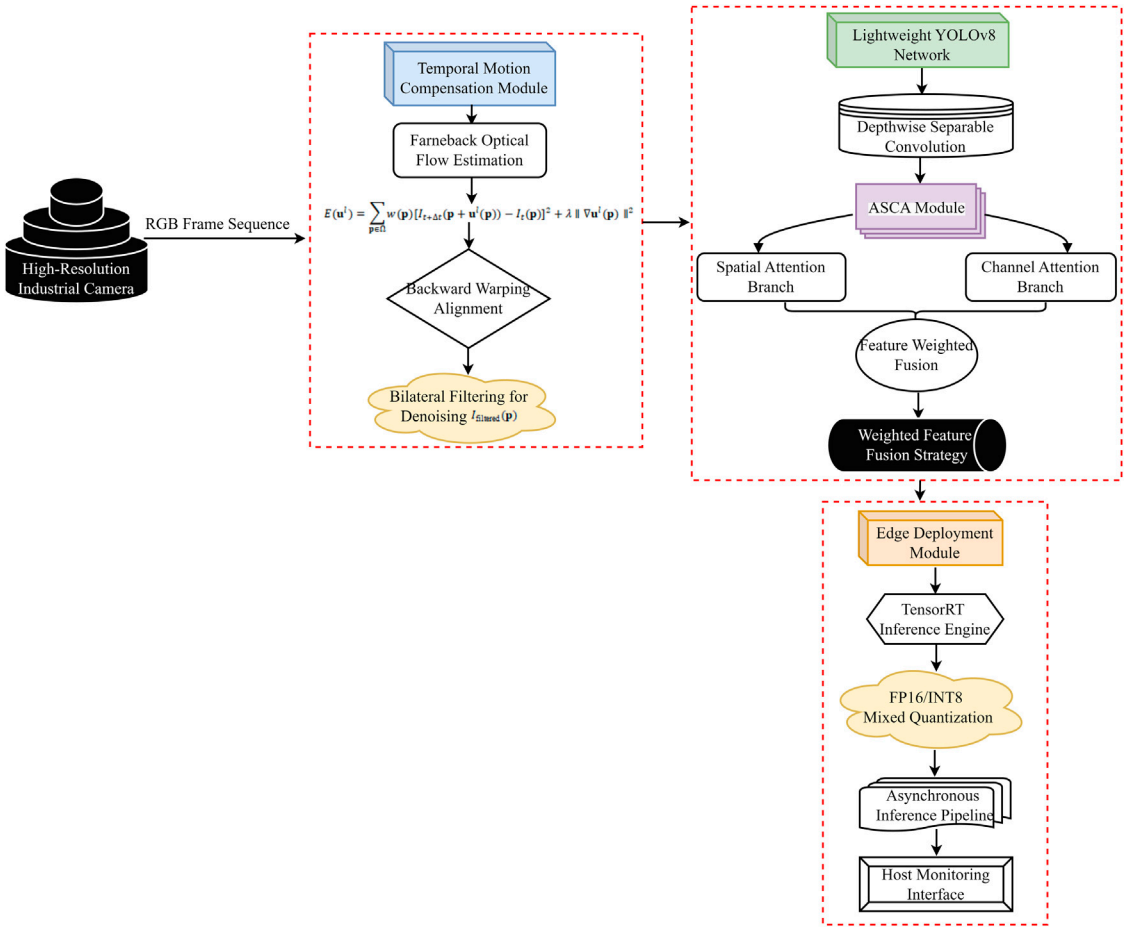

FIGURE 1
Overall architecture.

and effectively handle large displacements; the spatial search window size is set to 15 × 15 pixels, which is sufficient to capture local motion patterns while avoiding excessive computational effort; the number of optical flow field optimization iterations is set to 3, which can ensure convergence while controlling latency; the neighborhood size of the polynomial expansion is set to 5 to fit local image brightness changes; and the standard deviation of the Gaussian kernel used in the fitting process is set to 1.2 to ensure sub-pixel estimation accuracy. Assume that the three consecutive frames in the input video stream are $I_{t\text{-}1}$, I_t, I_{t+1}, respectively. The middle frame I_t is used as the reference frame, and motion estimation is performed on the previous and next frames $I_{t\text{-}1}$ and I_{t+1} based on Farneback optical flow. For the pixel $\mathbf{p} = (x, y)$, the image intensity function $I(\mathbf{p})$ in its neighborhood can be expressed as:

$$I(\mathbf{p}) = \mathbf{p}^T A\mathbf{p} + \mathbf{b}^T\mathbf{p} + c \tag{1}$$

In Equation 1, A is a quadratic coefficient matrix, $\mathbf{b}$ is a linear term vector, and c is a constant term. Let the displacement field of the previous and next frames relative to the reference frame be $\mathbf{u}_t(\mathbf{p}) = (u_x(\mathbf{p}), u_y(\mathbf{p}))$, and the solution is based on the grayscale invariance assumption and the multi-scale pyramid strategy. At the scale layer l, the optical flow field $\mathbf{u}^l$ satisfies the following minimization objective function:

$$E(\mathbf{u}^l) = \sum_{\mathbf{p}\in\Omega} w(\mathbf{p})\left[I_{t+\Delta t}(\mathbf{p} + \mathbf{u}^l(\mathbf{p})) - I_t(\mathbf{p})\right]^2 + \lambda \| \nabla \mathbf{u}^l(\mathbf{p}) \|^2 \tag{2}$$

In Equation 2, Ω is the local spatial window, $w(\mathbf{p})$ is the Gaussian weighted kernel, and λ is the regularization coefficient used to suppress the violent fluctuations of the optical flow field. The dense displacement fields $\mathbf{u}_{t\text{-}1\to t}$ and $\mathbf{u}_{t+1\to t}$ are obtained by solving the Euler-Lagrange equation through iterative optimization. The reverse mapping of the previous frame $I_{t\text{-}1}$ and the next frame I_{t+1} is performed to achieve image alignment, as shown in Equation 3:

$$I'_{t\text{-}1}(\mathbf{p}) = I_{t\text{-}1}(\mathbf{p} + \mathbf{u}_{t\text{-}1\to t}(\mathbf{p})), I'_{t+1}(\mathbf{p}) = I_{t+1}(\mathbf{p} - \mathbf{u}_{t+1\to t}(\mathbf{p})) \tag{3}$$

The interpolation process uses bilinear interpolation to ensure sub-pixel accuracy:

$$I(\mathbf{p} + \mathbf{u}) = \sum_{i,j\in\{0,1\}} (1\text{-}i')(1\text{-}j')I(x + i, y + j) \tag{4}$$

In Equation 4, $i' = u_x\text{-}\lfloor u_x \rfloor$ and $j' = u_y\text{-}\lfloor u_y \rfloor$. To suppress the edge blur and noise accumulation introduced in the optical flow

interpolation process, the aligned image $I_t^{'}$ is post-processed by applying bilateral filtering:

$$I_{\text{filtered}}(\mathbf{p}) = \frac{1}{W}\sum_{\mathbf{q}\in\Omega} G_s(\|\mathbf{p}\text{-}\mathbf{q}\|) G_r\left(\left|I'(\mathbf{p})\text{-}I'(\mathbf{q})\right|\right) I'(\mathbf{q}) \quad (5)$$

In Equation 5, G_s is the spatial Gaussian kernel, G_r is the grayscale similarity Gaussian kernel, and W is the normalization coefficient. The three aligned output images $I_{t\text{-}1}^{'}$, I_t, $I_{t+1}^{'}$ serve as the input of the subsequent detection network to ensure that feature extraction is performed on a spatially consistent image sequence.

2.3 Lightweight YOLOv8 network reconstruction

This paper constructs a lightweight YOLOv8 network based on YOLOv8, achieving efficient model reconstruction through structural reparameterization and channel compression strategies (Liu et al., 2025; Liu et al., 2024). Taking the standard convolution layer as an example, the calculation process of the input feature map $\mathbf{X} \in \mathbb{R}^{H\times W\times C_{\text{in}}}$ and the standard 3 × 3 convolution kernel $\mathbf{K} \in \mathbb{R}^{3\times 3\times C_{\text{in}}\times C_{\text{out}}}$ can be expressed as Equation 6:

$$Y_{i,j,k} = \sum_{a=1}^{3}\sum_{b=1}^{3}\sum_{c=1}^{C_{\text{in}}} X_{i+a-2,j+b-2,c}\cdot K_{a,b,c,k} + b_k, k = 1,\ldots,C_{\text{out}} \quad (6)$$

We replace all standard convolutions in the backbone (except the first layer) with depthwise separable convolutions. This splits the operation into two steps: a depthwise convolution (spatial filtering per channel) and a pointwise convolution (1x1 channel fusion) (Qin et al., 2025; Zhang et al., 2025a). First, the depthwise convolution performs spatial filtering on each input channel independently:

$$X_{\text{dw}}^{(c)} = X^{(c)} * K_{\text{dw}}^{(c)}, c = 1,\ldots,C_{\text{in}} \quad (7)$$

In Equation 7, $K_{\text{dw}}^{(c)} \in \mathbb{R}^{3\times 3}$ is the 3 × 3 depth kernel of the cth channel, and * represents the two-dimensional convolution operation. Point-by-point convolution achieves information fusion between channels through 1 × 1 convolution:

$$Y = X_{\text{dw}} * K_{\text{pw}}, K_{\text{pw}} \in \mathbb{R}^{1\times 1\times C_{\text{in}}\times C_{\text{out}}} \quad (8)$$

In Equation 8, the first convolution layer retains the standard 3 × 3 convolution to maintain sensitivity to the underlying texture and edge features. In the Neck and Head modules, the number of output channels of each layer is further compressed uniformly. Let the original number of channels be C, and after compression, it is shown in Equation 9:

$$C' = \lfloor C\cdot(1-\alpha)\rfloor, \alpha = 0.15 \quad (9)$$

The compressed feature fusion layer and detection head were redesigned based on reduced channels to ensure consistent overall network width. The standard bottleneck blocks in all C2f modules were replaced with lightweight versions. Their internal convolutional layers also adopted a depthwise separable structure, and cross-layer connections remained unchanged to preserve gradient paths. The resulting lightweight YOLOv8 network, while maintaining the native YOLOv8 detection head structure and loss function, achieves efficient reconstruction of backbone feature extraction, providing lightweight and effective feature input for subsequent attention enhancement and multi-scale fusion.

2.4 Adaptive spatial-channel attention module embedding

This paper embeds an adaptive spatial-channel attention module after each C2f module in the Neck layer of a lightweight YOLOv8 network to achieve dual-dimensional joint feature enhancement (Ding et al., 2024; An and Shi, 2024). Given an input feature map $\mathbf{F} \in \mathbb{R}^{H\times W\times C}$, the ASCA module generates a joint weight map using parallel spatial and channel attention branches. The spatial attention branch first performs max pooling and average pooling on the input feature map along the channel dimension, generating two spatial descriptors, as shown in Equation 10:

$$M_{max}(x,y) = \max_{c\in C} F(x,y,c), A_{\text{avg}}(x,y) = \frac{1}{C}\sum_{c=1}^{C} F(x,y,c) \quad (10)$$

After concatenating $M_{max}(x,y)$ and $A_{\text{avg}}(x,y)$ along the channel, they are input into a 7 × 7 depthwise separable convolution layer to capture large receptive field spatial context information. Let the convolution kernel be $K_s \in \mathbb{R}^{7\times 7\times 1}$, and the output is activated by the Sigmoid function to generate the spatial attention weight map $\mathbf{w}_s \in [0,1]^{H\times W\times 1}$:

$$\mathbf{W}_s = \sigma\left(\text{Conv}_{7\times 7}^{\text{dw}}\left(\text{Concat}\left(M_{max}, A_{\text{avg}}\right)\right)\right) \quad (11)$$

In Equation 11, $\sigma(\cdot)$ is a Sigmoid function. The channel attention branch adopts an improved Squeeze-and-Excitation (SE) structure. First, global average pooling is performed on the input F to compress the spatial dimension, as shown in Equation 12:

$$z_c = \frac{1}{H\cdot W}\sum_{i=1}^{H}\sum_{j=1}^{W} F(\text{i},\text{j},\text{c}), \text{c} = 1,\ldots,\text{C} \quad (12)$$

Then a nonlinear transformation is performed through a two-layer fully connected network, introducing a dimensionality reduction ratio, as shown in Equations 13, 14:

$$\mathbf{q} = \text{ReLU}(W_1\mathbf{z}), W_1 \in \mathbb{R}^{C/r\times \text{C}} \quad (13)$$

$$\mathbf{e} = \sigma(W_2\mathbf{q}), W_2 \in \mathbb{R}^{\text{C}\times C/r} \quad (14)$$

In Equation 15, the output channel weight vector forms the channel attention matrix $\mathbf{W}_c \in \mathbb{R}^{1\times 1\times C}$. The ASCA module adopts an element-by-element multiplication fusion strategy to jointly apply spatial and channel weights to the original feature map:

$$F_{\text{out}}(x,y,c) = F(x,y,c)\cdot \mathbf{W}_s(x,y,1)\cdot \mathbf{W}_c(1,1,c) \quad (15)$$

that is:

$$F_{\text{out}} = F \otimes (\mathbf{W}_s \odot \mathbf{W}_c) \quad (16)$$

In Equation 16, $\otimes$ represents element-by-element multiplication, and $\odot$ is the outer product expansion operation to ensure the alignment of weight dimensions. This structure achieves a synergistic enhancement of spatial positioning sensitivity and channel semantic selectivity without introducing additional

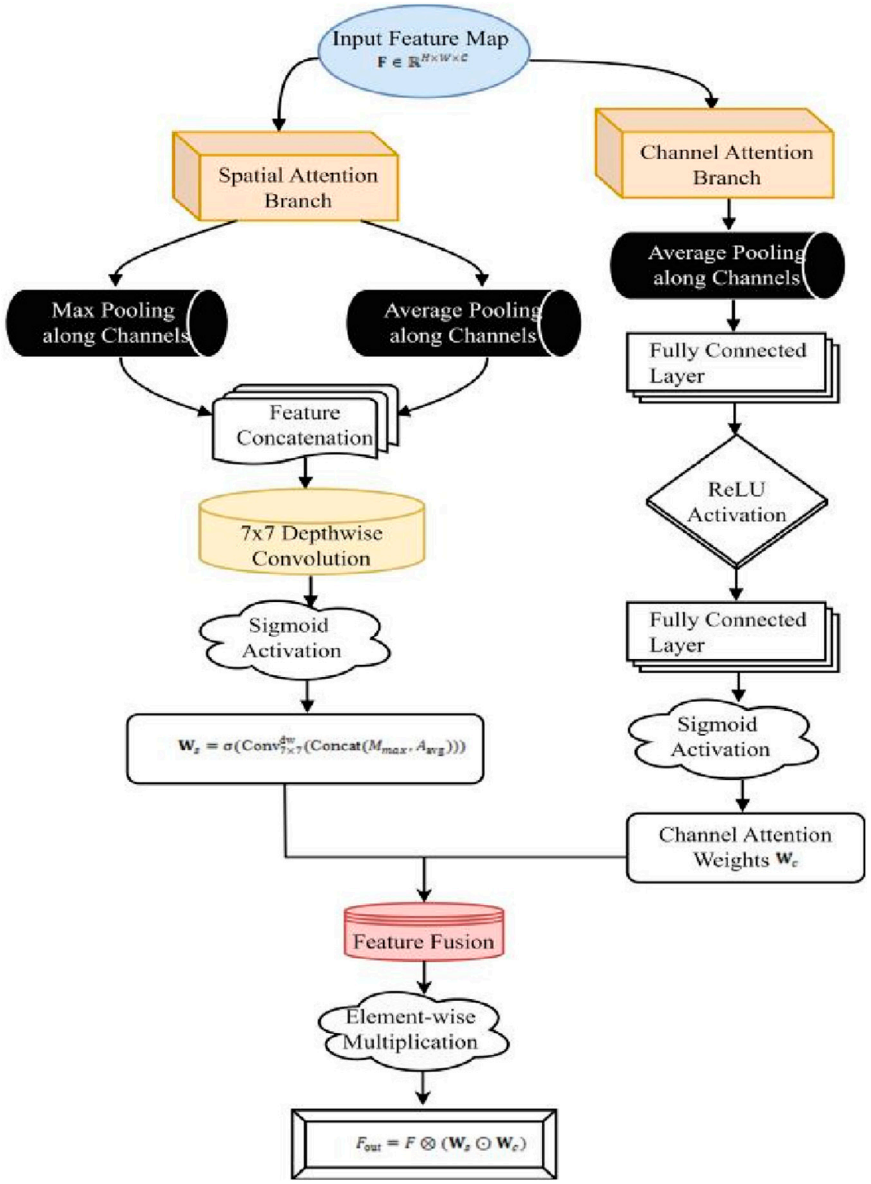

FIGURE 2
ASCA module architecture.

branches or complex gating mechanisms, strengthening the response intensity to low-contrast, small-area abnormal areas, and providing more discriminative feature representations for subsequent multi-scale fusion and target detection. The ASCA module structure is shown in Figure 2.

Figure 2 illustrates the ASCA module's structural flow: The input feature map is fed simultaneously into the spatial attention branch and the channel attention branch. The spatial branch extracts spatial saliency information through channel-wise max and average pooling. After concatenation, it undergoes a 7 × 7 depthwise separable convolution to capture the large receptive field context. Sigmoid activation is then used to generate spatial attention weights. The channel branch performs global average pooling on the input, passes it through a two-layer fully connected network and ReLU (Rectified Linear Unit) activation, and outputs channel attention weights. Finally, the spatial weights and channel weights are combined and applied to the original feature map through element-by-element multiplication, achieving dual enhancement of spatial localization and channel semantics. This significantly improves the model's responsiveness to low-contrast and small-area anomalies, providing a more discriminative feature representation for subsequent detection heads.

2.5 Multi-scale feature fusion optimization

This paper optimizes the original multi-scale feature fusion structure PANet of YOLOv8 and proposes a weighted feature fusion mechanism guided by learnable weights (Xu et al., 2024;

Yang L. et al., 2024). This mechanism is deployed at the key fusion nodes of the top-down (up-sampling) and bottom-up (down-sampling) paths, replacing traditional splicing or simple addition operations to achieve efficient alignment and fusion of high-level semantic information and low-level detail features. Assume that in a certain fusion layer, the semantic feature map $F_{\text{high}} \in \mathbb{R}^{H\times W\times C}$ from the high layer is up-sampled and fused with the feature map $F_{\text{low}} \in \mathbb{R}^{H\times W\times C}$ from the backbone of the same layer. Traditional PANet uses channel splicing, as shown in Equation 17:

$$F_{\text{cat}} = \text{Concat}\left(F_{\text{high}}, F_{\text{low}}\right) \in \mathbb{R}^{H\times W\times 2C} \tag{17}$$

Then the convolution layer is used to reduce the dimension. This paper proposes a weighted fusion strategy and defines the fusion output as:

$$F_{\text{out}} = \alpha \cdot F_{\text{high}} + \beta \cdot F_{\text{low}} \tag{18}$$

In Equation 18, α and β are learnable scalar weights, which are automatically optimized through back propagation to ensure that the network adaptively adjusts the contribution ratio of high- and low-level features according to the input content. To enhance stability, a normalization constraint is introduced, and the Softmax normalization form is adopted:

$$\tilde{a} = \frac{e^{w_1}}{e^{w_1} + e^{w_2}} \tag{19}$$

$$\tilde{b} = \frac{e^{w_2}}{e^{w_1} + e^{w_2}} \tag{20}$$

In Equations 19, 20, are trainable parameters, initialized to 0 to ensure balanced fusion in the initial training phase. The fusion process can be rewritten as:

$$F_{\text{out}} = \tilde{\alpha} \cdot U\left(F_{\text{high}}\right) + \tilde{\beta} \cdot F_{\text{low}} \tag{21}$$

In Equation 21, $U(\cdot)$ represents the upsampling operation. This weighted fusion module is embedded in each P3, P4, and P5 fusion node of the Neck part. Since small anomalies have higher spatial resolution in low-level features but are semantically ambiguous, this mechanism automatically enlarges the $\tilde{\beta}$ when small targets exist through training, thereby enhancing the ability to retain details. In downsampling fusion, the downsampled feature $F_{\text{down}} \in \mathbb{R}^{H/2\times W/2\times C}$ is fused with the same-layer feature F_{mid}, and the weighted strategy is also adopted, as shown in Equation 22:

$$F_{\text{out}} = \gamma \cdot F_{\text{down}} + \delta \cdot F_{\text{mid}}, \tilde{\gamma}, \tilde{\delta} = \text{Softmax}\left(w_3, w_4\right) \tag{22}$$

All weight parameters w_i are embedded in the network as independent learnable variables and participate in end-to-end backpropagation optimization. The gradient of the loss function can be expressed as Equation 23:

$$\frac{\partial \mathcal{L}}{\partial w_i} = \frac{\partial \mathcal{L}}{\partial F_{\text{out}}} \cdot \frac{\partial F_{\text{out}}}{\partial \tilde{\alpha}} \cdot \frac{\partial \tilde{\alpha}}{\partial w_i} \tag{23}$$

During the inference phase, all weights are solidified, eliminating the need for additional computational overhead. This fusion mechanism improves the semantic consistency and spatial sensitivity of multi-scale features without significantly increasing the number of parameters.

2.6 Model quantization and edge deployment

This paper deploys the optimized lightweight YOLOv8 model on the Jetson AGX Xavier edge computing platform, employing a collaborative model compression and hardware acceleration strategy to achieve efficient migration from the training domain to the inference domain (Elhanashi et al., 2024; Ling et al., 2023). First, export the weight model trained in PyTorch to the ONNX (Open Neural Network Exchange) intermediate representation format, preserving the complete computational graph structure. Using the NVIDIA TensorRT engine, the model is quantized using a mixture of FP16 (half-precision floating point) and INT8 (8-bit integer). The TensorRT engine deeply optimizes the computational graph, including layer fusion (combining convolution, batch normalization, and activation functions into a single computing unit), memory reuse, and kernel automatic tuning, to maximize the computing power of edge devices. During the deployment phase, the model is loaded into the GPU (Graphics Processing Unit) memory of the Jetson AGX Xavier, and the CUDA (Compute Unified Device Architecture) stream mechanism is enabled to implement multi-frame asynchronous processing (Bai et al., 2024; Zhang et al., 2025b). The final system uses the GStreamer framework to achieve video stream acquisition and result visualization. The detection results are encapsulated in JSON (JavaScript Object Notation) format, including bounding box coordinates, category labels, and confidence levels. They are uploaded to the industrial monitoring center in real time via UDP (User Datagram Protocol), completing end-to-end closed-loop deployment (Balogh and Vidács, 2022; Cobanoglu et al., 2025).

3 Experiment and verification

3.1 Experimental design

The experimental platform was equipped with an NVIDIA Jetson AGX Xavier running Ubuntu 20.04, and image acquisition was performed using an industrial camera. The experimental data was derived from field operating data of rotating machinery and included RGB images of the rotating machinery under four operating conditions: rated load, overload, high temperature, and high vibration. Each anomaly category (The Kappa values of cracks, oil leaks, loose bolts, damaged protective covers, and broken belts are 0.87, 0.85, 0.91, 0.90, and 0.88, respectively.) was independently annotated by three engineers with bounding boxes. If there are significant differences in the annotation results of the three engineers for the same image, an expert review meeting will be initiated. The three engineers will jointly review the image, discuss and compare the device history and reference standard samples, and ultimately reach a consensus and determine a unique “gold standard” label. The dataset was split into training, validation, and test sets with an 8: 1:1 ratio. The dataset is not open to the public. The experimental setup is shown in Table 1.

TABLE 1 Experimental setup.

Item	Description
Hardware Platform	NVIDIA Jetson AGX Xavier
Camera Model	Basler acA 2000-50gc
Operating System	Ubuntu 20.04
Image Resolution	1920 × 1080
Operating Conditions	Nominal Load, Overload, Misalignment, High Temperature, Dusty Environment
Anomaly Classes	Crack, Oil Leak, Bolt Loosening, Guard Damage, Belt Breakage
Data Split	Training: (80%); Validation: (10%); Test: (10%)

TABLE 2 Precision and recall rates for each category.

Anomaly class	Precision	Recall
Crack	97.00%	97.60%
Oil Leak	97.60%	97.80%
Bolt Loosening	96.80%	97.10%
Guard Damage	98.30%	98.50%
Belt Breakage	98.20%	98.40%

3.2 Detection accuracy: mAP@0.5 evaluation

As shown in Table 2, on the test set, a confidence score threshold of 0.5 is used to evaluate the precision and recall of each anomaly category. According to Table 3, calculating AP@0.5 (Average Precision @0.5) and summarizing mAP@0.5 (mean Average Precision @0.5) to comprehensively measure detection accuracy under different conditions. AP is calculated using the 11-point interpolation method, as shown in Equation 24:

$$\mathrm{AP} = \frac{1}{11} \sum_{r \in \{0, 0.1, \ldots, 1\}} \underset{\tilde{r} \geq r}{max} P(\tilde{r}) \tag{24}$$

Figure 3 shows the AP@0.5 performance of Faster R-CNN, YOLOv8s, and proposed method for five types of rotating machinery anomalies under four industrial operating conditions (rated load, overload, high temperature, and strong vibration). The overall mAP@0.5 for each condition is also plotted. The horizontal axis represents the anomaly category, and the vertical axis represents the average detection accuracy. The paper's method significantly outperforms all operating conditions and categories, achieving an average mAP@0.5 of 97.8% and 98.5% under rated load, far exceeding Faster R-CNN's 91.7% and YOLOv8s's 95.7%. Under strong vibration conditions, proposed method maintains a mAP@0.5 of 97.1%, while Faster R-CNN drops to 85.0% and YOLOv8s to 91.7%. For minor anomalies such as "cracks" and "oil leaks," the paper's method achieves 97.3% and 97.9% AP@0.5, respectively, under high-temperature conditions, while Faster R-CNN's AP drops to 87.6% and 85.4%, demonstrating greater environmental adaptability.

This performance advantage comes from our method's multi-layer optimization. The temporal motion compensation module suppresses vibration-induced blur. This allows feature extraction on aligned frames, reducing positioning drift. The ASCA attention mechanism uses joint spatial and channel weighting to enhance the response of weak features when contrast decreases due to high temperatures. The lightweight YOLOv8 network, combined with a weighted feature fusion strategy, reduces computational overhead while maintaining high sensitivity to small objects, thus avoiding sudden drops in accuracy caused by resource scheduling fluctuations. Faster R-CNN is sensitive to noise due to its two-stage structure, and YOLOv8s lacks a dedicated enhancement mechanism, resulting in significant performance degradation under complex working conditions.

TABLE 3 Ablation experiment.

Experiment ID	Model Configuration	Temporal motion compensation	ASCA attention	Weighted fusion	Depthwise separable convolution	mAP@0.5 (%)	mAP@0.5:0.95 (%)
(a)	Baseline	—	—	—	—	93.7	80
(b)	Baseline + Motion Compensation	√	—	—	—	94.8	81.6
(c)	Baseline + ASCA Attention	—	√	—	—	94.5	81.3
(d)	Baseline + Weighted Fusion	—	—	√	—	94.1	80.9
(e)	Baseline + Depthwise Separable Convolution	—	—	—	√	93.9	80.5
(f)	(b) + (c)	√	√	—	—	95.3	82.2
(g)	(f) + (d)	√	√	√	—	96.7	83.9
(h)	(e) + (c) + (d)	—	√	√	√	95.9	82.7
(i)	Proposed Method	√	√	√	√	97.8	87

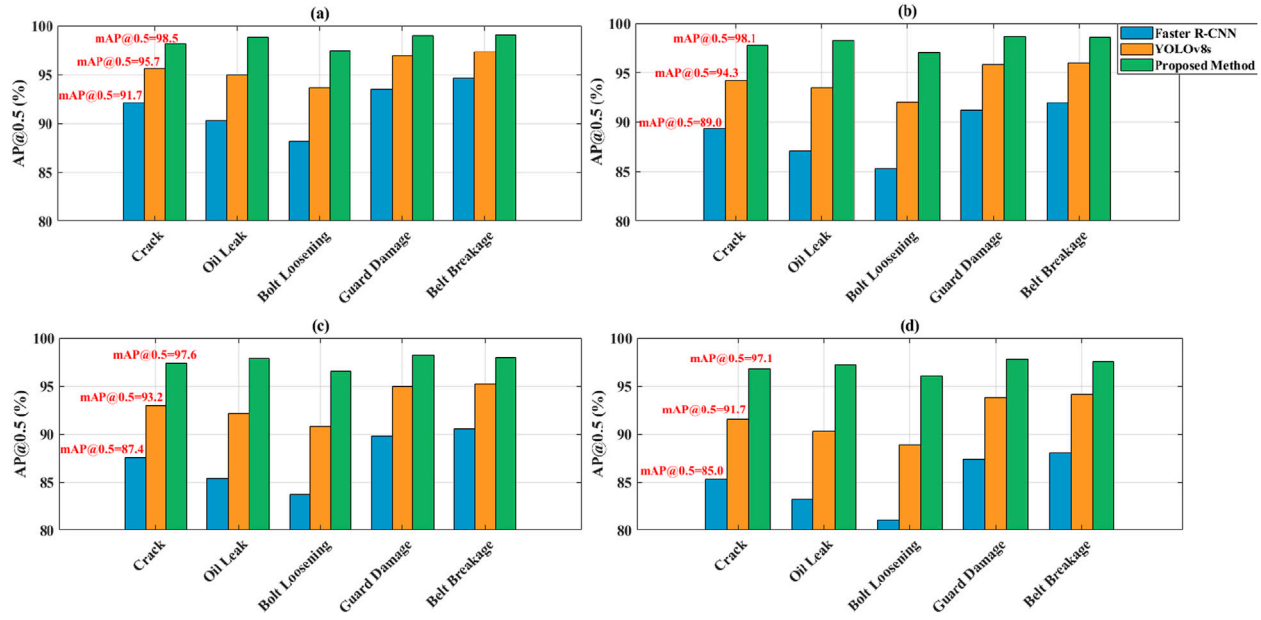

FIGURE 3
Comparison of mAP@0.5 of different methods under four industrial conditions. **(a)** Rated Load. **(b)** Overload. **(c)** High Temperature. **(d)** Strong Vibration.

This method, through a closed-loop "perception-compensation-enhancement" design, achieves stable detection accuracy in dynamic industrial environments, demonstrating strong engineering practicality.

3.3 High accuracy and robustness analysis

Using the mAP@0.5:0.95 metric, the paper calculated and averaged the AP at 10 levels with an IoU (Intersection over Union) threshold ranging from 0.5 to 0.95 and a step size of 0.05 to comprehensively evaluate the model's overall performance and robustness under high localization accuracy requirements, as shown in Equation 25:

$$\mathrm{mAP}^{(s)} = \frac{1}{10}\sum_{i=1}^{10}\mathrm{AP@0.5} + 0.05i \tag{25}$$

Figure 4 shows the AP@0.5:0.9 performance of Faster R-CNN, YOLOv8s, and proposed method for five types of rotating machinery anomalies under four operating conditions. The vertical axis represents the anomaly category, and the horizontal axis represents the detection accuracy. The paper's method significantly outperforms all categories and working conditions. In "crack" detection, the AP reaches 87.6% under rated load conditions, far exceeding Faster R-CNN's 78.1%. For the low-contrast anomaly "oil leakage", the paper's method maintains 84.2% under strong vibration conditions, while Faster R-CNN's performance drops to 67.5%. The overall mAP@0.5:0.95 is 88.5% under rated load, and even under strong vibration conditions, it still reaches 84.5%. Faster R-CNN achieves 77.7% and 69.6% respectively, and YOLOv8s achieves 83.4% and 76.7% respectively, showing significant performance degradation. The average mAP@0.5:0.95 of the paper's method reaches 86.6%, indicating that the paper's method is more robust to minor defects in complex industrial environments.

This advantage lies in the refined collaborative optimization achieved by this method between model structure and industrial scenario adaptability. By deeply integrating a lightweight YOLOv8 network backbone with the ASCA module, this approach enables the network to focus on local texture variations and structural anomalies, rather than relying solely on geometric outlines. This allows the network to activate responses to minor defects even under fuzzy conditions. Furthermore, a weighted feature fusion mechanism assigns higher propagation weights to underlying high-resolution features, achieving enhanced spatial fidelity while maintaining semantic richness, effectively alleviating the feature sparsity problem caused by downsampling. Hybrid FP16/INT8 quantization maintains accuracy stability at the edge, preventing low-precision inference from further weakening its sensitivity to minor anomalies. Faster R-CNN is limited by its region proposal mechanism's preference for large objects, and YOLOv8s struggles to maintain high discrimination for small objects under resource-constrained conditions without targeted optimization. This method, through end-to-end reconstruction tailored to industrial defect characteristics, achieves consistently reliable detection output in complex and dynamic environments.

3.4 Inference speed measurement

The model was run continuously for 10 min on a Jetson AGX Xavier, recording the end-to-end processing time (T_{total}) for 1800 frames. This includes the entire process from image

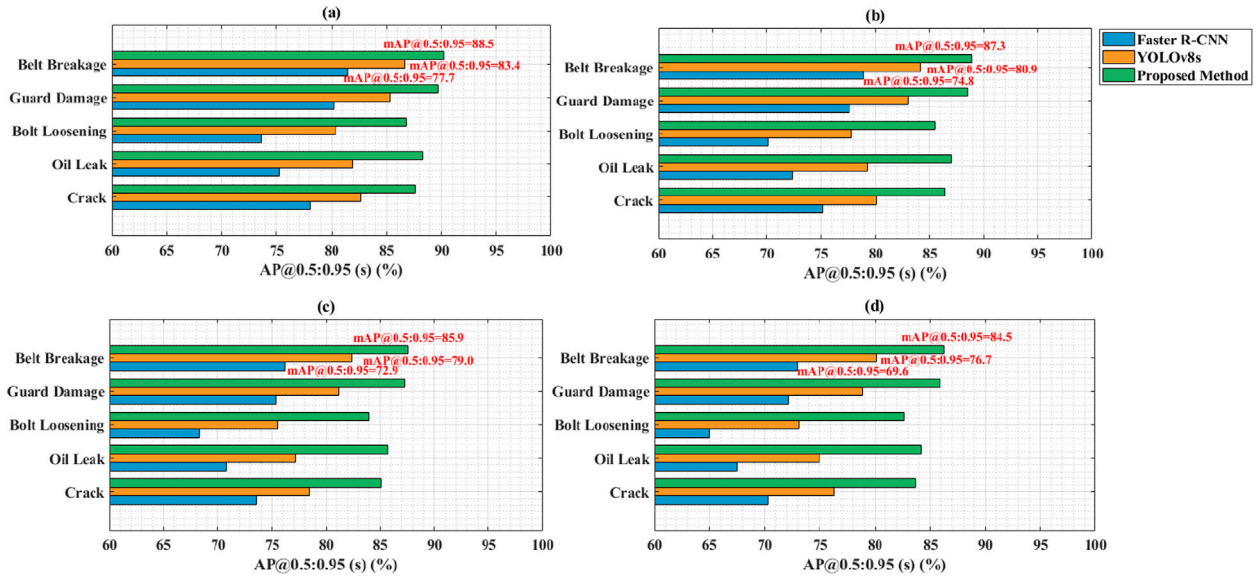

FIGURE 4
Comparison of mAP@0.5:0.95 of different methods under four industrial conditions. **(a)** Rated Load. **(b)** Overload. **(c)** High Temperature. **(d)** Strong Vibration.

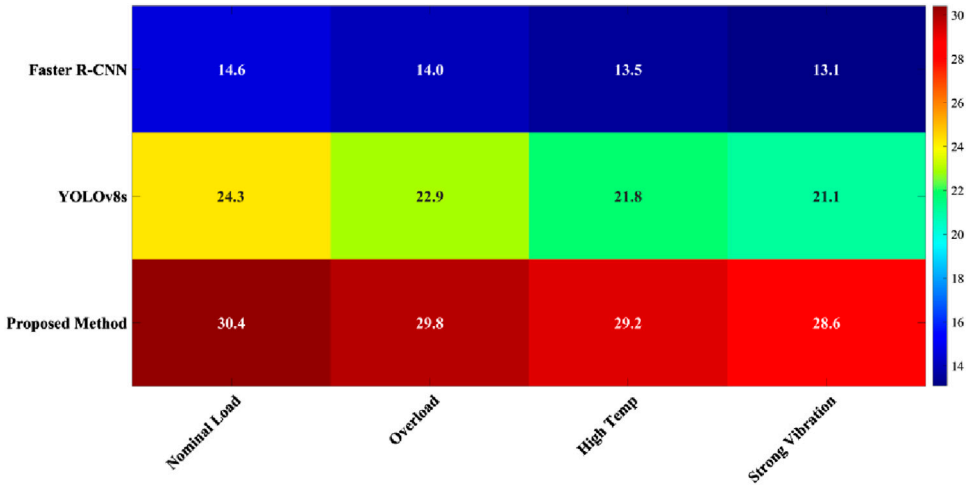

FIGURE 5
Comparison of end-to-end inference speed of different methods under four industrial conditions.

acquisition, preprocessing, inference, and post-processing. Inference speed is calculated as Equation 26:

$$\mathrm{FPS} = \frac{1800}{T_{\mathrm{total}}} \tag{26}$$

Figure 5 shows the inference speed performance of Faster R-CNN, YOLOv8s, and proposed method under four typical industrial conditions in the form of a heat map. The horizontal axis represents the four conditions (rated load, overload, high temperature, and strong vibration), and the vertical axis represents the detection method. The color depth reflects the frame rate. As can be seen from the figure, proposed method exhibits the brightest color under all conditions, with an inference speed that remains stable between 28.6 and 30.4 FPS, averaging 29.5 FPS. This is significantly higher than the comparison methods, achieving a stable output of over 30 FPS under rated load, demonstrating its potential to meet the needs of industrial real-time detection. Faster R-CNN achieved only 13.1–14.6 FPS, while YOLOv8s achieved 21.1–24.3 FPS. Under strong vibration and high temperature conditions, the proposed method achieved frame rates of 28.6 and 29.2 FPS, significantly higher than Faster R-CNN's 13.1 and 13.5 FPS, and YOLOv8s's 21.1 and 21.8 FPS, respectively. This demonstrates greater environmental adaptability and resource scheduling stability, validating its

feasibility in achieving real-time response in complex field scenarios.

This inference efficiency advantage stems from the system-level collaborative optimization implemented in this paper, from model structure to deployment strategy. The lightweight YOLOv8 network significantly reduces computational density through depthwise separable convolutions, reducing redundant parameters while maintaining feature extraction capabilities, significantly reducing the amount of computation per frame. FP16/INT8 hybrid quantization fully leverages the Tensor Core hardware acceleration capabilities of the Jetson AGX Xavier to improve floating-point throughput. The TensorRT engine performs layer fusion, memory reuse, and automatic kernel tuning on the computational graph, maximizing the computing power of edge devices. The lightweight design reduces the risk of thermal throttling under high temperatures and loads, preventing sudden frame rate drops due to CPU (Central Processing Unit)/GPU throttling, and maintaining stable output under dynamic conditions. Faster R-CNN is difficult to compress due to its two-stage redundant computations, and YOLOv8, while fast, is still limited by its unoptimized inference path. This approach, through a three-pronged strategy of "lightweighting + quantization + engine acceleration," achieves efficient and reliable operation of edge intelligence in harsh industrial environments.

3.5 Model lightweight evaluation

The total number of model parameters was calculated using PyTorch's torchsummary tool, as shown in Equation 27:

$$\text{Params} = \sum_{l\in\{\text{Conv,Linear}\}} \left(k_h \cdot k_w \cdot C_{\text{in}}^{l} \cdot C_{\text{out}}^{l}\right) \tag{27}$$

Floating Point Operations (FLOPs) are calculated using an approximate MACs × 2 (two FLOPs per multiplication and addition), and only the convolutional and fully connected layers in the forward propagation are counted, as shown in Equation 28:

$$\text{FLOPS}_l = 2 \cdot H_l \cdot W_l \cdot C_{\text{in}}^{l} \cdot C_{\text{out}}^{l} \cdot k_h \cdot k_w \tag{28}$$

The total FLOPs is the sum of all layers, which evaluates the computational complexity of the model.

Figure 6 compares the model complexity of Faster R-CNN, YOLOv8s, and proposed method, focusing on two key lightweight metrics: parameter count (params) and floating-point operations (FLOPs). As can be seen, Faster R-CNN has a high parameter count of 41.2 million and 158.6 gigabytes of FLOPs, resulting in significant computational overhead and difficulty meeting the resource constraints of edge devices. While YOLOv8s has been optimized to 3.2 million parameters and 8.7 gigabytes of FLOPs, demonstrating some potential for deployment, it still outperforms proposed method. The lightweight YOLOv8 network architecture proposed in this paper, incorporating the ASCA module and a weighted fusion strategy, further reduces the number of parameters to 2.26M and the number of FLOPs to 6.2G, representing reductions of 29.4% and 28.7%, respectively, compared to YOLOv8s. This significantly reduces the computational burden while maintaining high detection accuracy, demonstrating its enhanced lightweight

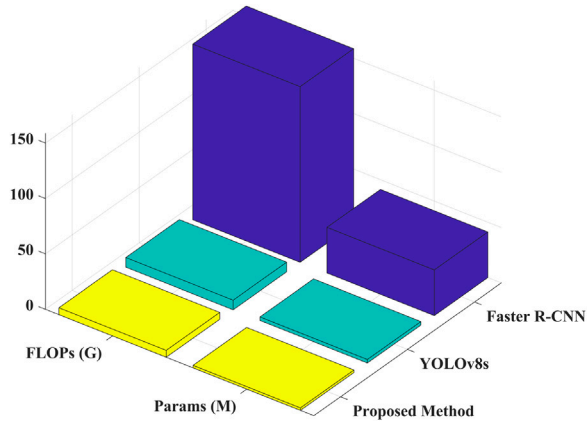

FIGURE 6
Comparison of model complexity of different methods: number of parameters and floating-point operations.

advantages and providing the structural foundation for real-time inference on edge platforms such as the Jetson AGX Xavier.

This improvement results from a refined network structure and module optimization. We replace standard convolutions with depthwise separable ones, which greatly reduces parameter redundancy. This allows the backbone network to maintain its receptive field while significantly reducing computational density. The ASCA attention module is designed as a lightweight structure, effectively enhancing the representation of key features and avoiding the computational explosion caused by traditional attention mechanisms. The weighted feature fusion mechanism optimizes information flow through learnable weight scalars without adding learnable parameters, balancing performance and efficiency. Faster R-CNN incurs a significant amount of redundant computation due to its region proposal network and RoI (Region of Interest) pooling. While compact, YOLOv8s still retains many standard convolutional modules and lacks specialized compression for industrial edge scenarios. This method achieves high-precision anomaly detection with minimal resource consumption through a collaborative design approach combining streamlined architecture and enhanced functionality, truly enabling efficient and sustainable intelligent monitoring of rotating machinery.

3.6 MBR testing

Two test subsets were constructed: one containing clear images (no blur), and the other simulating motion blur using a Gaussian convolution kernel. MBR is defined as Equation 29:

$$\text{MBR} = \frac{\text{mAP}_{\text{clear}}}{\text{mAP}_{\text{blur}}} \tag{29}$$

Figure 7 shows the mAP@0.5 performance of Faster R-CNN, YOLOv8s, and the paper's method for clear and blurred images under four typical working conditions. The Motion Blur Robustness (MBR) ratio is also plotted for each working condition. The horizontal axis represents the four working conditions (rated load, overload, high temperature, and strong vibration), and the

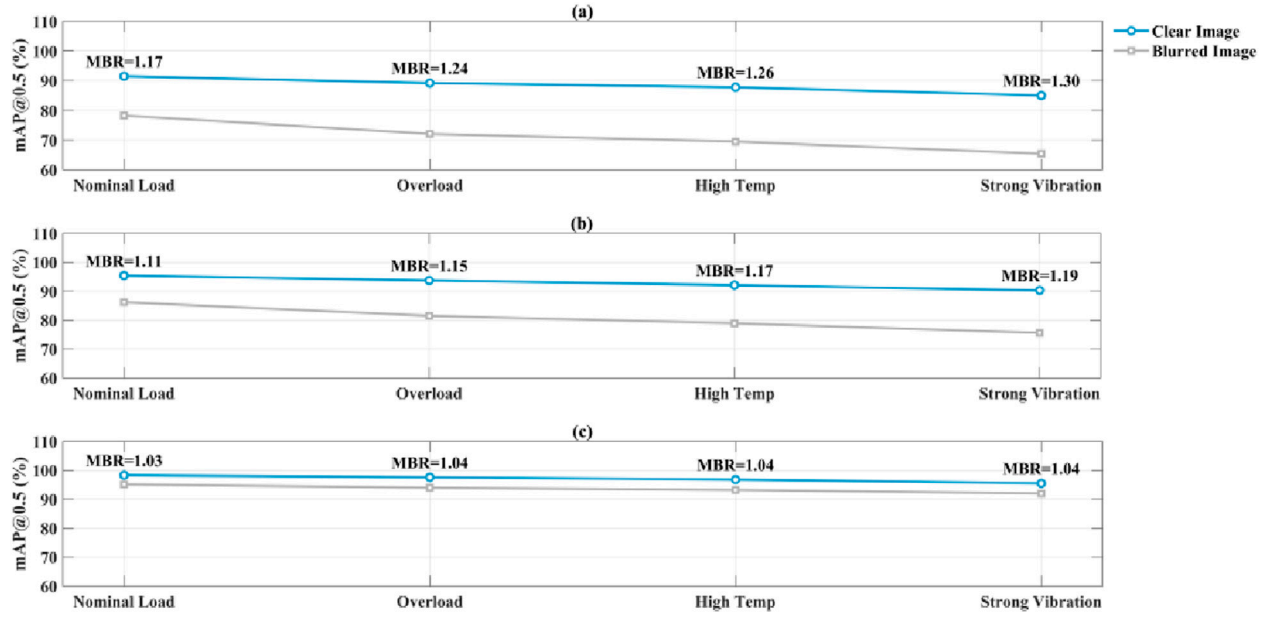

FIGURE 7
MBR. **(a)** Faster R-CNN. **(b)** YOLOv8s. **(c)** Proposed method.

vertical axis represents detection accuracy. As shown in the figure, Faster R-CNN's mAP@0.5 dropped from 85.0% to 65.4% under strong vibration, with an MBR of 1.30, indicating significant performance degradation. YOLOv8s performed even better, but still saw a drop from 90.3% to 75.6% (MBR = 1.19). However, proposed method only saw a drop from 95.5% to 92.0% under strong vibration, with an MBR of 1.04, significantly lower than the other methods. Across all conditions, proposed method's MBR remained consistently between 1.03 and 1.05, and the blurred image curve was nearly parallel to the clear image, demonstrating its strong ability to suppress vibration-induced motion blur.

This exceptional robustness stems from the proposed system's deep optimization of modeling the consistency of spatiotemporal features. The optical flow temporal motion compensation module introduced in this paper not only achieves pixel-level alignment between frames but also reconstructs stable feature inputs through dense displacement fields, effectively mitigating feature drift and blurring caused by vibration. The ASCA attention mechanism uses spatial weights to focus on structural changes in abnormal regions rather than edge strength, enabling the network to activate key features such as cracks and oil leaks even in low-resolution conditions. The weighted feature fusion strategy enhances the cross-layer transfer of underlying details, preventing the suppression of small object features caused by blur in deep networks. Faster R-CNN relies on fixed anchor frames and is sensitive to clear contours. YOLOv8s, while fast, lacks a dynamic compensation mechanism. This proposed method, through a closed-loop design of "motion alignment - feature enhancement - fusion optimization," maintains high-fidelity detection performance in complex industrial vibration environments, truly meeting the stability requirements for real-time monitoring of abnormal conditions in rotating machinery.

3.7 Deployment stability evaluation

The system runs continuously for 24 h at the industrial site, automatically recording its status every 5 min. Failures are defined as: process crash, memory overflow, or missed detection of three or more consecutive frames (manually verified). The total number of detected frames and the number of failures are used to evaluate the long-term operational reliability of the system, as shown in Equation 30.

$$\mathrm{CRF} = \frac{N_{\mathrm{fail}}}{N_{\mathrm{total}}} \tag{30}$$

Figure 8 compares the CRFs of Faster R-CNN, YOLOv8s, and proposed method under four typical industrial operating conditions. The horizontal axis represents the four operating conditions (rated load, overload, high temperature, and strong vibration), and the vertical axis represents the CRF. Faster R-CNN's CRF increases from 1.16×10^{-6} under rated load to 6.56×10^{-6} under strong vibration, indicating poor stability in complex environments. YOLOv8s performs better, with its CRF increasing from 7.72×10^{-7} to 2.31×10^{-6}, but still showing a clear upward trend. The proposed method achieves a CRF as low as 3.86×10^{-7} under rated load and overload conditions, and only rises to 7.72×10^{-7} and 1.16×10^{-6} under high temperature and strong vibration conditions, respectively. These results consistently remain significantly lower than those of other methods, demonstrating the strongest adaptability to these conditions. These results demonstrate the high reliability of proposed method during continuous operation, meeting the core requirement for long-term stable detection in industrial settings.

This exceptional stability comes from our system-level robust design. The lightweight network reduces computational overhead,

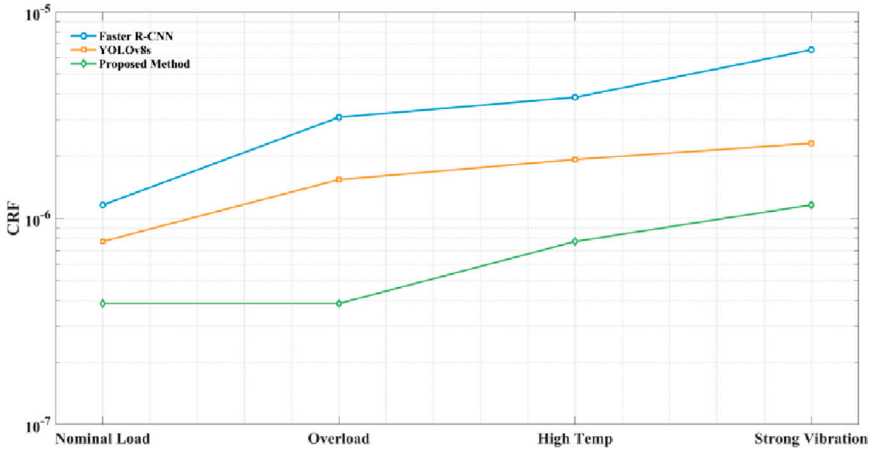

FIGURE 8
Deployment stability evaluation.

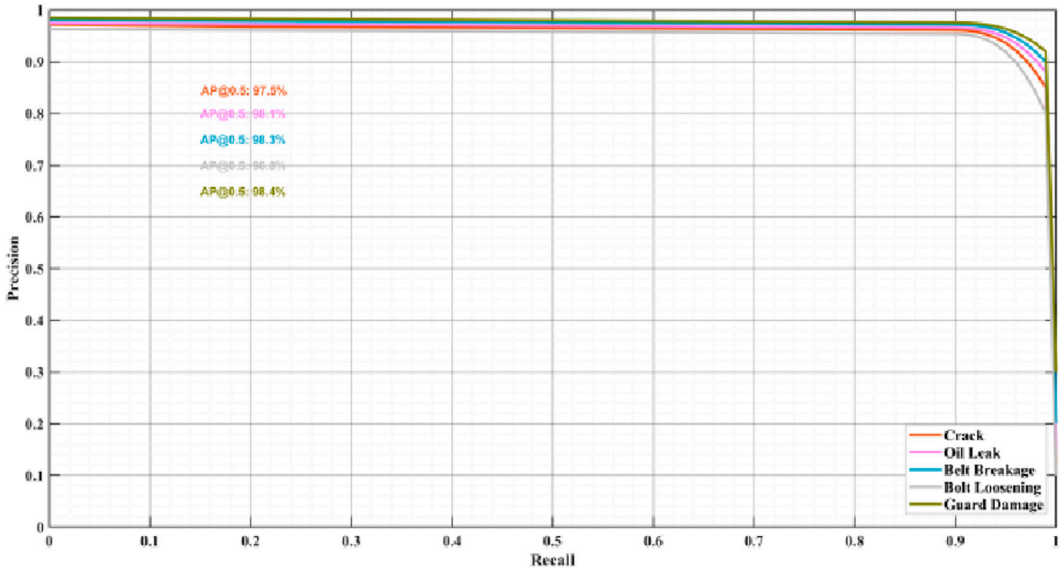

FIGURE 9
PR curve.

preventing thermal throttling and memory overflow on edge devices. The TensorRT engine's efficient memory management mechanism reduces the risk of resource leakage, ensuring long-term process operation without crashes. Temporal motion compensation effectively suppresses continuous missed detections caused by vibration and avoids the accumulation of misjudgments caused by inter-frame misalignment. Model quantization and asynchronous inference mechanisms balance computing resource scheduling, preventing I/O (Input/Output) blocking or inference backlogs. The system's asynchronous CUDA stream mechanism enables parallel pipeline execution of data acquisition, preprocessing, and model inference, effectively avoiding task blocking and resource idling, further improving operational smoothness and stability. Faster R-CNN, due to its high resource consumption, easily reaches hardware limits, and YOLOv8s, while lightweight, lacks dynamic adaptation mechanisms. However, this method, through its "lightweight structure + edge optimization + motion robustness" design, achieves extreme failure rate reduction in complex industrial environments, truly achieving deployable and reliable intelligent monitoring.

3.8 Fine-grained performance analysis: precision, recall, PR curve, and confusion matrix

To further evaluate the fine-grained detection performance of our proposed method for various rotating machinery anomalies, this

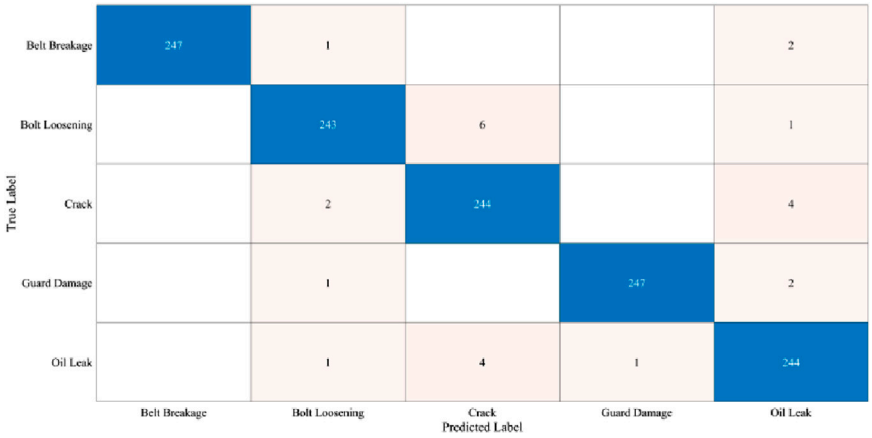

FIGURE 10
Confusion matrix.

section provides additional analysis of the precision, recall, PR curve, and confusion matrix for each category.

3.8.1 Precision and recall analysis

Proposed method achieved precision and recall exceeding 96.5% for all five anomaly categories, with an average of 97.6%. The "broken belt" and "broken guard" categories achieved the highest precision, reaching 98.2% and 98.3%, respectively, and the highest recall, reaching 98.4% and 98.5%, respectively, demonstrating that the model is able to effectively capture these potentially serious faults.

3.8.2 PR curve

As shown in Figure 9, the PR curves of all categories are close to the upper right corner of the coordinate system, and the area under the curve (AUC) is close to 1, indicating that the model has excellent comprehensive performance for all types of anomalies.

3.8.3 Confusion matrix

As can be seen from the Figure 10, the diagonal elements of the confusion matrix (that is, the number of correctly classified samples) are much higher than the off-diagonal elements, indicating that the model has a very strong classification ability.

3.9 Ablation experiment

To deeply analyze the independent contributions of each proposed module to the final performance, this paper designed a systematic ablation experiment. Using YOLOv8s as the base model, the experiment gradually added the key components proposed in this paper to evaluate their impact on detection accuracy.

This study validated the effectiveness and synergistic gains of the core components of the proposed "perception-compensation-detection-inference" integrated architecture through systematic ablation experiments. Using the standard YOLOv8s with a mAP@0.5 of 93.7% and a mAP@0.5:0.95 of 80% as a baseline, the experiments showed that introducing temporal motion compensation, ASCA attention, weighted fusion, or depthwise separable convolutions individually all improved performance. Motion compensation increased mAP@0.5%–94.8%, demonstrating its key role in suppressing dynamic vibration blur. ASCA attention and weighted fusion increased mAP@0.5%–94.5% and 94.1%, respectively, validating their ability to enhance subtle anomaly features. Finally, lightweight depthwise separable convolutions increased mAP@0.5%–93.9%, laying the foundation for subsequent deployment. Further combined experiments showed that combining motion compensation with ASCA attention achieved a mAP@0.5 of 95.3%. When the first three enhanced modules were integrated into the standard network, mAP@0.5 and mAP@0.5:0.95 jumped to 96.7% and 83.9%, respectively, fully demonstrating the synergistic effect between the modules. Ultimately, when all innovative modules were fully integrated into this method, system performance reached its peak, with mAP@0.5 and mAP@0.5:0.95 reaching 97.8% and 87%, respectively. This not only far exceeded the baseline but also significantly outperformed all intermediate combinations, strongly demonstrating the comprehensive superiority of this integrated design in solving the problem of real-time, high-precision anomaly detection in complex operating conditions of rotating machinery.

4 Conclusion

This paper proposes a real-time detection method for abnormal conditions in rotating machinery, building an integrated "perception-compensation-detection-inference" architecture. It uses optical flow to compensate for temporal motion and suppress vibration-induced motion blur. A lightweight YOLOv8 network is designed, embedding an adaptive spatial-channel attention module to enhance the response to subtle anomaly features. Learnable weighted fusion is then used to optimize multi-scale feature transfer. This approach, combined with the TensorRT engine, implements FP16/INT8 quantization and asynchronous inference on the Jetson AGX Xavier platform. Experiments show that this method achieves an average mAP@0.5 of 97.8% and a mAP@0.5:0.95 of 86.6% under complex operating conditions, with a stable inference speed of 28.6–30.4 FPS. This approach combines high precision, strong

robustness, and real-time performance, providing a reliable technical solution for the intelligent operation and maintenance of industrial rotating equipment.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JC: Writing – review and editing, Writing – original draft, Methodology. JT: Formal Analysis, Resources, Project administration, Writing – review and editing. JS: Visualization, Investigation, Writing – review and editing, Formal Analysis.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by Guangdong University innovation team project (2023KCXTD062).

Conflict of interest

The 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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

An, J., and Shi, Z. (2024). YOLOv8n-enhanced PCB defect detection: a lightweight method integrating spatial–channel reconstruction and adaptive feature selection. *Appl. Sci.* 14 (17), 7686. doi:10.3390/app14177686

Bai, T., Duan, J., Wang, Y., Fu, H., and Zong, H. (2024). Fasteners quantitative detection and lightweight deployment based on improved YOLOv8. *Rev. Sci. Instrum.* 95 (10), 105108. doi:10.1063/5.0214188

Balogh, M., and Vidács, A. (2022). Optimizing camera stream transport in cloud-based industrial robotic systems. *Infocommunications J.* 14 (1), 36–42. doi:10.36244/icj.2022.1.5

Cobanoglu, H. C., Ay, B., Bulut, F., and Samli, R. (2025). Towards efficient video stream analysis: a distributed deep learning framework: the DiVA approach. *Trait. Du. Signal* 42 (3), 1541–1552. doi:10.18280/ts.420326

Cui, Y., Liu, Z., and Lian, S. (2023). A survey on unsupervised anomaly detection algorithms for industrial images. *IEEE Access* 11, 55297–55315. doi:10.1109/access.2023.3282993

Das, O., Das, D. B., and Birant, D. (2023). Machine learning for fault analysis in rotating machinery: a comprehensive review. *Heliyon* 9 (6), e17584. doi:10.1016/j.heliyon.2023.e17584

Ding, P., Zhan, H., Yu, J., and Wang, R. (2024). A bearing surface defect detection method based on multi-attention mechanism Yolov8. *Meas. Sci. Technol.* 35 (8), 086003. doi:10.1088/1361-6501/ad4386

Elhanashi, A., Dini, P., Saponara, S., and Zheng, Q. (2024). TeleStroke: real-time stroke detection with federated learning and YOLOv8 on edge devices. *J. Real-Time Image Process.* 21 (4), 121. doi:10.1007/s11554-024-01500-1

Gawde, S., Patil, S., Kumar, S., and Kotecha, K. (2023). A scoping review on multi-fault diagnosis of industrial rotating machines using multi-sensor data fusion. *Artif. Intell. Rev.* 56 (5), 4711–4764. doi:10.1007/s10462-022-10243-z

Jiang, W. (2022). A machine vision anomaly detection system to industry 4.0 based on variational fuzzy autoencoder. *Comput. Intell. Neurosci.* 2022 (1), 1–10. doi:10.1155/2022/1945507

Khan, D., Waqas, M., Tahir, M., Islam, S., Amin, M., jan, L., et al. (2023). Revolutionizing real-time object detection: YOLO and MobileNet SSD integration. *J. Comput. and Biomed. Inf.* 6 (01), 41–49. doi:10.56979/601/2023

Li, X., Li, M., Wu, Y., Zhou, D., Liu, T., Hao, F., et al. (2021). Accurate screw detection method based on faster R-CNN and rotation edge similarity for automatic screw disassembly. *Int. J. Comput. Integr. Manuf.* 34 (11), 1177–1195. doi:10.1080/0951192x.2021.1963476

Li, X., Yu, S., Lei, Y., and Yang, B. (2023). Intelligent machinery fault diagnosis with event-based camera. *IEEE Trans. Industrial Inf.* 20 (1), 380–389. doi:10.1109/tii.2023.3262854

Li, Y., Zhou, Y., and Liu, H. (2024a). Research on the influence of image motion blur on the effectiveness of machine vision-based metal scraps separation system. *J. Material Cycles Waste Manag.* 26 (4), 2509–2517. doi:10.1007/s10163-024-01989-5

Li, Y., Wu, Y., Gao, K., and Yang, H. (2024b). Early bolt loosening detection method based on digital image correlation. *Sensors* 24 (16), 5397. doi:10.3390/s24165397

Ling, Q., Isa, N. A. M., and Asaari, M. S. M. (2023). Precise detection for dense PCB components based on modified YOLOv8. *IEEE Access* 11, 116545–116560. doi:10.1109/access.2023.3325885

Liu, M., Zhang, M., Chen, X., Zheng, C., and Wang, H. (2024). YOLOv8-LMG: an improved bearing defect detection algorithm based on YOLOv8. *Processes* 12 (5), 930. doi:10.3390/pr12050930

Liu, S., Tohti, G., Geni, M., He, H., Wu, Z., and Liao, C. (2025). Micro vibration detection algorithm for rotating machinery based on visual target detection. *Signal, Image Video Process.* 19 (4), 275. doi:10.1007/s11760-025-03863-9

Natili, F., Daga, A. P., Castellani, F., and Garibaldi, L. (2021). Multi-scale wind turbine bearings supervision techniques using industrial SCADA and vibration data. *Appl. Sci.* 11 (15), 6785. doi:10.3390/app11156785

Pookkuttath, S., Gomez, B. F., Elara, M. R., and Thejus, P. (2023). An optical flow-based method for condition-based maintenance and operational safety in autonomous cleaning robots. *Expert Syst. Appl.* 222, 119802. doi:10.1016/j.eswa.2023.119802

Qin, G., Zou, Q., Li, M., Deng, Y., Mi, P., Zhu, Y., et al. (2025). Surface defect detection on industrial drum rollers: using enhanced YOLOv8n and structured light for accurate inspection. *PloS one* 20 (2), e0316569. doi:10.1371/journal.pone.0316569

Ren, Z., Fang, F., Yan, N., and Wu, Y. (2022). State of the art in defect detection based on machine vision. *Int. J. Precis. Eng. Manufacturing-Green Technol.* 9 (2), 661–691. doi:10.1007/s40684-021-00343-6

Singh, S. A., and Desai, K. A. (2023). Automated surface defect detection framework using machine vision and convolutional neural networks. *J. Intelligent Manuf.* 34 (4), 1995–2011. doi:10.1007/s10845-021-01878-w

Suo, X., Liu, J., Dong, L., Shengfeng, C., Enhui, L., and Ning, C. (2022). A machine vision-based defect detection system for nuclear-fuel rod groove. *J. Intelligent Manuf.* 33 (6), 1649–1663. doi:10.1007/s10845-021-01746-7

Tang, B., Chen, L., Sun, W., and Lin, Z. (2023). Review of surface defect detection of steel products based on machine vision. *IET Image Process.* 17 (2), 303–322. doi:10.1049/ipr2.12647

Xiao, X., Li, C., He, H., Huang, J., and Yu, T. (2025). Rotating machinery fault diagnosis method based on multi-level fusion framework of multi-sensor information. *Inf. Fusion* 113, 102621. doi:10.1016/j.inffus.2024.102621

Xu, J., Zhu, X., Shi, L., Li, J., and Guo, Z. (2024). Squeeze-and-excitation attention and bi-directional feature pyramid network for filter screens surface detection. *J. Electron. Imaging* 33 (4), 043044. doi:10.1117/1.jei.33.4.043044

Yadav, E., Chawla, V. K., Angra, S., and Yadav, S. (2025). The Fault diagnosis of different rotating machine elements by using infrared thermography images and extended adaptive neuro-fuzzy inference system: an experimental evaluation. *MAPAN*, 1–19. doi:10.1007/s12647-025-00838-6

Yang, Y., Zuo, J., Li, L., Wang, X., Yin, Z., and Ding, X. (2024). Crack identification method for magnetic particle inspection of bearing rings based on improved Yolov5. *Meas. Sci. Technol.* 35 (6), 065405. doi:10.1088/1361-6501/ad3181

Yang, L., Chen, G., Liu, J., and Guo, J. (2024). Wear state detection of conveyor belt in underground mine based on retinex-YOLOv8-EfficientNet-NAM. *IEEE Access* 12, 25309–25324. doi:10.1109/access.2024.3363834

Zhang, P., Chen, R., Yang, L., Zou, Y., and Gao, L. (2025). Recent progress in digital twin-driven fault diagnosis of rotating machinery: a comprehensive review. *Neurocomputing* 634, 129914. doi:10.1016/j.neucom.2025.129914

Zhang, Y., Liang, S., Li, J., and Pan, H. (2025a). Yolov8s-DDC: a deep neural network for surface defect detection of bearing ring. *Electronics* 14 (6), 1079. doi:10.3390/electronics14061079

Zhang, Y., Wang, S., Wang, J., Zhao, Y., and Chen, Z. (2025b). DHNet: a surface defect detection model utilizing multi-scale convolutional kernels. *J. Real-Time Image Process.* 22 (1), 45. doi:10.1007/s11554-025-01623-z

Zhao, H., Gao, Y., and Deng, W. (2024). Defect detection using shuffle Net-CA-SSD lightweight network for turbine blades in IoT. *IEEE Internet Things J.* 11 (20), 32804–32812. doi:10.1109/jiot.2024.3409823

Zhu, S., Liao, B., Hua, Y., Zhang, C., Wan, F., and Qing, X. (2023). A transformer model with enhanced feature learning and its application in rotating machinery diagnosis. *ISA Trans.* 133, 1–12. doi:10.1016/j.isatra.2022.07.016