

Intelligent Detection of Bolted Connection Damage in Steel Truss Structures Based on LBD-YOLO

Debing Zhuo

Jishou University

Binzhao Guo

`guobinzhao0205@163.com`

Jishou University

Zepu Jiang

Jishou University

Shuowen Li

Jishou University

Zheqian He

Jishou University

Zhongyu Hu

Jishou University

Article

Keywords: bolt connection damage detection, lightweight object detection, YOLO26n, multi-scale feature fusion, steel truss structures

Posted Date: August 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10418932/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Intelligent Detection of Bolted Connection Damage in Steel Truss Structures Based on LBD-YOLO

Debing Zhuo¹, Binzhao Guo^{2,*}, Zepu Jiang¹, Shuowen Li², Zheqian He² & Zhongyu Hu²

¹ College of Intelligent Construction, Jishou University, Zhangjiajie 427000, P.R. China.

² College of Tourism and Urban Planning, Jishou University, Zhangjiajie 427000, P.R. China.

Correspondence: Binzhao Guo (guobinzhao0205@163.com)

ABSTRACT

Bolted connections are critical load-transferring joints in steel truss structures. Damage such as loosening, missing, and rusting directly threatens the service safety of various steel infrastructures, including towers, bridges, and crane systems. However, due to complex structural backgrounds, small target sizes, and weak visual features, existing detection methods struggle to simultaneously achieve satisfactory accuracy, inference efficiency, and edge-deployment capability. To address these challenges, we propose LBD-YOLO, a lightweight bolt damage detection method built upon an improved state-of-the-art object detection framework. We first construct a dedicated dataset of 3,600 images covering four bolt conditions (normal, loosening, missing, and rusted). We then design a collaborative feature enhancement strategy that integrates GSConv and ADown modules to reduce computational redundancy while preserving fine-grained features, embeds a BoTNet module to enhance global context perception, and adopts a BiFPN structure to optimize multi-scale feature fusion. Experimental results show that LBD-YOLO achieves 97.7% mAP@50 and 61.0% mAP@50-95 on the self-built dataset. Compared with the baseline, it reduces parameters by 16.1% and computational cost by 19.2%, while demonstrating strong robustness under various complex scenarios. The proposed method provides an effective technical solution for intelligent high-precision damage detection of bolted connections in steel truss structures.

Keywords: bolt connection damage detection, lightweight object detection, YOLO26n, multi-scale feature fusion, steel truss structures

INTRODUCTION

Steel truss structures are widely utilized in critical infrastructure systems such as bridges, transmission towers, crane systems, and mountain cargo ropeways due to their high load-bearing capacity, light self-weight, and excellent wind resistance¹. Bolted connections constitute the primary assembly method for these structures, involving numerous load-transferring joints. During long-term service, these connections are subjected to combined effects of environmental corrosion, wind-induced vibration, temperature stress, cyclic dynamic loads, and fatigue, making them prone to damage such as loosening, detachment, and corrosion². Failure of bolted connections can lead to redistribution of internal forces, and in severe cases, may cause local instability or even complete collapse, resulting in significant safety incidents and economic losses. Therefore, precise and efficient identification and condition assessment of bolted connection failures are crucial for ensuring the operational safety of steel truss structures throughout their entire lifecycle³.

However, quality inspection of bolted connections in steel truss structures still heavily relies on manual inspections in current engineering practice. In typical scenarios such as cableway towers and large bridges, maintenance personnel must climb the structures to perform visual checks or hammer tests. These methods are not only inefficient and labor-intensive but also pose significant safety risks at heights, with results prone to being influenced by operator experience and subjective judgment—making them inadequate for the frequent maintenance needs of complex facilities with numerous joints⁴⁻⁵. In recent years, structural health monitoring techniques utilizing piezoelectric sensors and acoustic wave detection have offered new approaches for assessing bolt conditions. However, these methods typically require contact-based measurements or extensive sensor deployment within structures, resulting in high costs and significant implementation challenges in remote mountainous areas or complex steel truss environments⁶⁻⁸. In contrast, non-contact inspection methods based on computer vision offer advantages such as high efficiency, low cost, and intuitive results, making them a prominent research focus for identifying bolt failures and addressing pressing practical engineering demands⁹.

With the rapid advancement of deep learning technology, object detection algorithms based on Convolutional Neural Networks (CNNs) have been introduced for bolt damage identification tasks¹⁰. Early studies predominantly employed two-stage detectors such as Faster R-CNN, which, despite offering relatively high detection accuracy, suffered from slow inference speeds¹¹. Single-stage detection algorithms represented by the YOLO (You Only Look Once) series have achieved a better balance between detection speed and accuracy and have been extensively applied in industrial defect

detection¹². Scholars worldwide have conducted substantial research on the application of YOLO models to bolt detection. Wang et al.¹³ constructed a three-stage YOLOv8 detection framework, achieving 94.5% detection accuracy in complex industrial environments. Peng et al.¹⁴ proposed the MS-YOLOv7 model, which improved mAP@50 by 2.52% for transmission line bolt defect detection through the integration of attention mechanisms. Hua et al.¹⁵ performed lightweight modifications to YOLOv8, enhancing computational efficiency while maintaining detection accuracy. Although these studies have achieved notable progress, existing methods still exhibit high miss-detection rates and insufficient robustness to environmental and viewpoint variations when confronted with the complex backgrounds characteristic of truss-type cableway towers, the multi-scale nature of bolt targets, and the subtle deformations caused by early-stage loosening¹⁶.

YOLO26n¹⁷, a recently proposed architecture available as an arXiv preprint, achieves significant reductions in model parameters and computational complexity while maintaining high detection accuracy through the removal of redundant modules and optimization of network structure, demonstrating promising potential for edge deployment. However, its direct application to bolt connection damage identification in steel truss structures still faces the following challenges. First, the lightweight design partially compromises the ability to extract fine-grained features, making it difficult to effectively capture subtle damage characteristics such as corrosion textures and loosening gaps. Second, the pure convolutional architecture exhibits limited capability in modeling global contextual information, which tends to cause false detections in cluttered backgrounds or under occlusion conditions. Third, the original feature fusion strategy yields suboptimal performance when handling dramatic scale variations of bolt targets.

To address these challenges, this paper proposes LBD-YOLO, an improved YOLO26n-based bolt connection damage detection method for steel truss structures. By collaboratively integrating four modules—GSConv, ADown, BoTNet, and BiFPN—this method aims to simultaneously enhance the model's capability for fine-grained damage feature extraction, global semantic perception, and cross-scale feature fusion, while preserving its lightweight characteristics to meet practical edge deployment requirements¹⁸⁻²¹. The main contributions of this work are summarized as follows:

- 1) A dedicated image dataset for bolt connection damage in steel truss structures was constructed, covering four typical states: normal bolts, loosening, missing, and rusted, providing a data foundation for subsequent related research.

- 2) A collaborative feature enhancement network, LBD-YOLO, is proposed, which effectively

addresses the problems of low detection accuracy for small bolt targets and high susceptibility to background interference in complex scenarios.

3) Through ablation studies, comparative experiments, and mobile-edge deployment verification, the comprehensive advantages of the proposed method in terms of detection accuracy, computational efficiency, and on-site robustness are thoroughly demonstrated.

The remainder of this paper is organized as follows. The Methods section introduces the proposed LBD - YOLO network. The Experiments Setup section describes the dataset and experimental configuration. The Results section reports the experimental outcomes, followed by a Discussion. Finally, the Conclusions section summarizes the study and outlines future work.

METHODS

LBD-YOLO model architecture

YOLO26n is a lightweight variant of the YOLO26 series, following an overall three-stage architecture comprising Backbone, Neck, and Head. The Backbone is responsible for progressively extracting multi-scale features from the input image. Its core constituent modules include C3k2, SPPF, and C2PSA. The C3k2 module employs a structure that combines grouped convolutions with residual connections, maintaining efficient feature extraction while reducing computational cost²². The SPPF module applies pooling operations to the feature map using multiple pooling kernels of different sizes connected in series, then concatenates the pooled results with the original feature map for output. This module achieves multi-scale feature fusion while enlarging the receptive field. Compared with the conventional SPP module, SPPF simplifies the computational process by sharing pooling kernels, balancing feature extraction performance with inference speed²³. The C2PSA module is positioned at the junction between the Backbone and the Neck, integrating the lightweight characteristics of the C2 module with the PSA attention mechanism, which effectively enhances the discriminative capability of feature maps and suppresses background interference²⁴.

The Neck adopts a PANet structure to fuse the multi-scale features extracted by the Backbone. Through bidirectional top-down and bottom-up pathways, it enhances inter-level information exchange and mitigates the insufficient detection accuracy associated with single-scale features. The Head is responsible for object classification and bounding box regression on the fused features, ultimately outputting the class and location information of bolt targets²⁵.

Although the YOLO26n model offers notable advantages in terms of lightweight design and inference efficiency, its direct application to bolt damage detection presents several limitations. The Backbone relies on standard convolutions as the core feature extraction units, exhibiting limited capability in capturing fine-grained damage features such as rusted textures and loosening gaps. The feature fusion mechanism in the Neck is relatively simple, making it difficult to adequately adapt to the detection requirements for bolts of different sizes in industrial scenarios. Furthermore, the model exhibits weak capability in modeling global semantic information, leading to missed and false detections under complex background conditions or when equipment occlusions are present. Therefore, targeted improvements are necessary to enhance the accuracy of bolt damage detection.

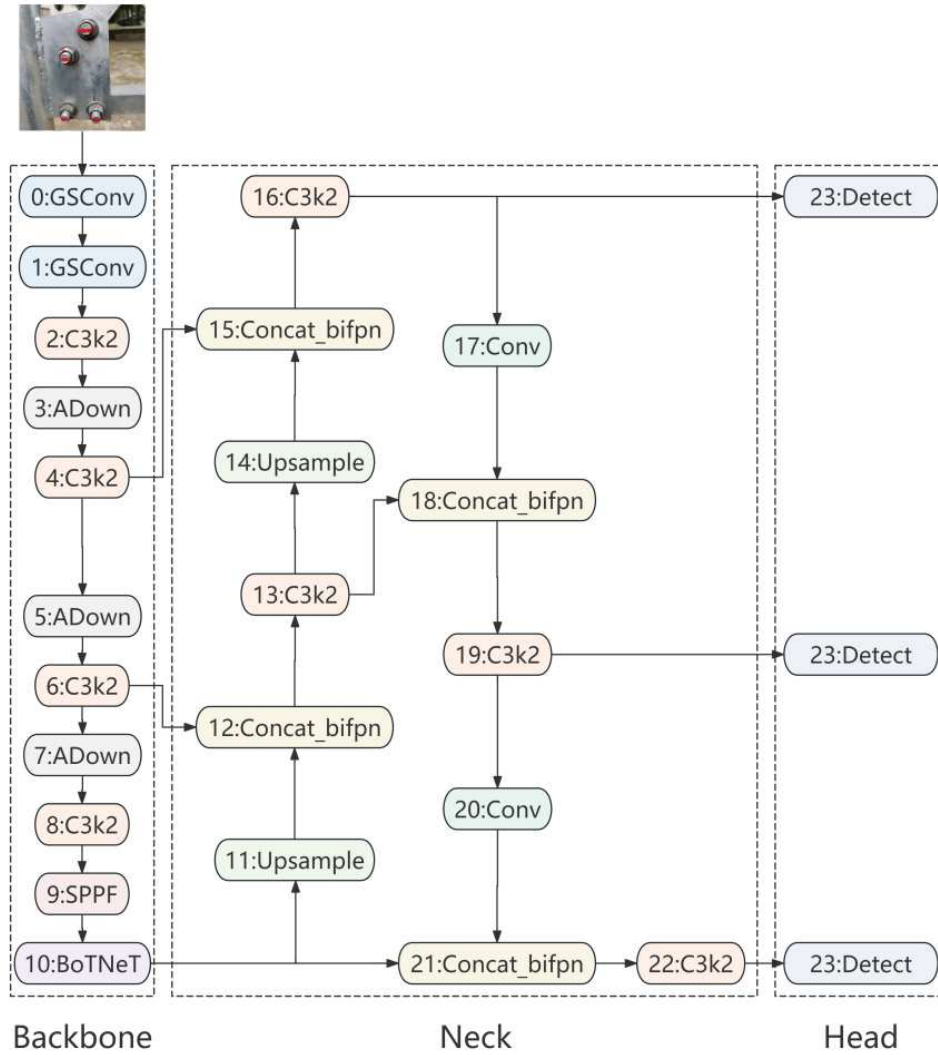


Fig. 1 Network architecture of LBD-YOLO

To address the aforementioned limitations of YOLO26n in bolt damage detection, this paper

proposes an improved lightweight bolt detection model, LBD-YOLO (Lightweight Bolt Detection YOLO). While retaining the three-stage architecture of YOLO26n, the proposed model introduces four key improvement modules: GSConv, ADown, BoTNet, and BiFPN. These modules collaboratively enhance feature extraction efficiency, global semantic modeling capability, and multi-scale feature fusion performance while maintaining the lightweight advantage, thereby better adapting to the real-time detection requirements for bolt damage in industrial scenarios. The overall network architecture of the LBD-YOLO model is illustrated in **Fig. 1**.

The specific improvement strategies are as follows. In the Backbone, GSConv modules are employed to replace the original standard convolutions, enhancing the extraction capability for fine-grained bolt features while reducing computational complexity. Between the Backbone and the Neck, the ADown adaptive downsampling module is embedded to replace the original downsampling operations, reducing the loss of critical information while achieving feature map dimensionality reduction. In the Neck, the BoTNet self-attention module is introduced to leverage multi-head attention mechanisms for enhancing the model's ability to capture global semantic information, effectively alleviating missed and false detections caused by target occlusion or background interference in complex scenes. In the Head, the BiFPN bidirectional weighted feature pyramid structure is integrated to replace the original simple feature fusion approach, optimizing multi-level feature interaction through a cross-scale weighted fusion mechanism and improving the model's detection adaptability for bolts of different sizes.

GSConv lightweight convolution module

GSConv (Group Shuffle Convolution) is an efficient convolution module designed for lightweight object detection tasks, first proposed by Li et al.¹⁸ Its core design principle is to reduce computational complexity while maintaining or even improving feature extraction capability by hybridizing the output features of standard convolutions and depthwise separable convolutions. The operation of GSConv can be formally described as:

$$F_{out} = \text{Shuffle}(\text{Concat}(F_{SC}, F_{DSC})) \quad (1)$$

where F_{SC} denotes the feature map extracted by the standard convolution; F_{DSC} denotes the feature map extracted by the depthwise separable convolution; $\text{Concat}(\cdot, \cdot)$ denotes the concatenation operation along the channel dimension; and $\text{Shuffle}(\cdot)$ denotes the channel shuffling operation. The channel

shuffling mechanism effectively promotes information interaction among different convolution groups and enhances the diversity and representational capacity of features.

Compared with standard convolutions, GSConv reduces the computational cost (FLOPs) by approximately 50%-60%, while achieving feature extraction performance comparable to or even slightly better than that of standard convolutions. This module maintains lightweight characteristics while effectively capturing fine-grained texture information in images, making it particularly suitable for extracting subtle damage features such as bolt rusted textures and loosening gaps.

ADown adaptive downsampling module

ADown (Adaptive Downsampling) is derived from the YOLOv9 model¹⁹, and serves as an efficient adaptive downsampling structure designed for lightweight networks. This module adopts a dual-path parallel design that maximizes the preservation of critical feature information while reducing the spatial dimensions of the feature map. Its operation can be expressed as:

$$F_{\text{out}} = \text{Concat} \left(\text{Conv}_{3 \times 3, s=2}(F_{\text{in}}), \text{Conv}_{1 \times 1}(\text{AvgPool}(F_{\text{in}})) \right) \quad (2)$$

where F_{in} is the input feature map. The first path employs a 3×3 convolution with a stride of 2 to perform direct downsampling, extracting discriminative features through a learning-based approach during dimensionality reduction. The second path first applies average pooling for downsampling, followed by a 1×1 convolution to adjust the channel dimensions, thereby preserving more of the original statistical information. The outputs from both paths are concatenated along the channel dimension to form the fused downsampled feature map.

Compared with conventional strided convolution or pooling-based downsampling, the adaptive dual-path mechanism of the ADown module effectively reduces information loss of small objects during the downsampling process, which is particularly beneficial for preserving the edge contours and detail features of small-sized targets such as bolts.

BoTNet self-attention module

BoTNet (Bottleneck Transformer Network) was first proposed by Srinivas et al.²⁰ at CVPR 2021. It is a hybrid architecture that integrates convolutional neural networks with the Transformer self-attention mechanism²⁹. Its core innovation lies in replacing the 3×3 spatial convolution in the ResNet bottleneck block³⁰ with a multi-head self-attention layer, thereby significantly enhancing the model's global context

modeling capability while maintaining relatively low computational overhead.

The basic building unit of the BoTNet module is a bottleneck block incorporating multi-head self-attention. The computation of multi-head self-attention follows the standard scaled dot-product attention mechanism:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W_O \quad (4)$$



Fig. 2 Architecture of the BoTNet module.

The structure of the BoTNet module is illustrated in **Fig. 2**. Q , K , and V denote the Query, Key, and Value matrices, respectively; d_k is the dimension of the key vectors; h is the number of attention heads; and W_O is the output projection matrix. The multi-head self-attention mechanism captures long-range dependencies among features from different representation subspaces, effectively compensating for the limited global receptive field inherent in pure convolutional networks.

BiFPN bidirectional weighted feature pyramid

BiFPN (Bidirectional Feature Pyramid Network) was first proposed by Tan et al.²¹ in the EfficientDet model. It is an efficient bidirectional weighted feature pyramid structure. Compared with conventional FPN³¹ and PANet²⁵, BiFPN introduces two core improvements: cross-scale bidirectional connections and a weighted feature fusion mechanism.

In terms of connectivity, BiFPN removes intermediate nodes with only a single input edge to simplify the network topology, while simultaneously adding skip connections between input and output nodes at the same level, thereby fusing more original feature information without significantly increasing computational cost. In terms of the fusion mechanism, BiFPN adopts a fast normalized fusion strategy that assigns learnable weights to input features at different scales:

$$F_{\text{out}} = \sum_i \frac{w_i}{\epsilon + \sum_j w_j} \cdot I_i \quad (5)$$

where w_i denotes the learnable weight assigned to the i -th input feature branch; ϵ represents a tiny constant introduced to avoid division by zero in the normalization operation; and I_i stands for the feature map of the i -th input branch. Benefiting from this weighted fusion design, the model can dynamically balance the importance of multi-scale features in the fusion stage, which effectively boosts the detection performance of objects with various sizes.

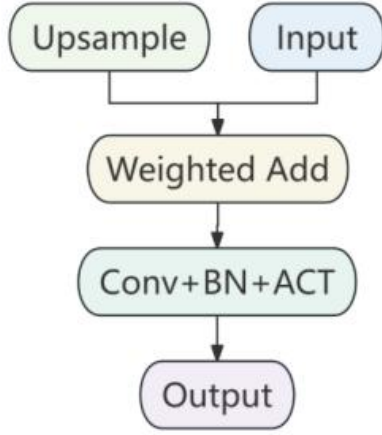


Fig. 3 Architecture of the original BiFPN module.

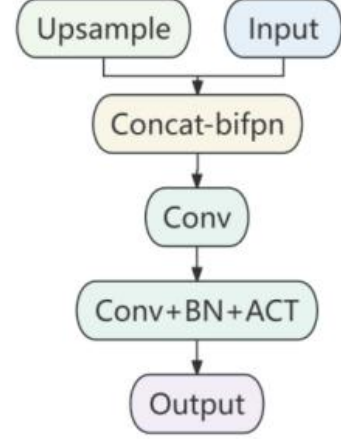


Fig. 4 Architecture of the Concat_bifpn module.

The structure of the original BiFPN module is shown in **Fig. 3**. In LBD-YOLO, to further enhance feature interaction between the Backbone and the Head, the Concat_bifpn operation³² is employed to implement feature concatenation fusion, the structure of which is illustrated in **Fig. 4**.

EXPERIMENTS SETUP

Data collection and marking principle

The dataset comprises four bolt condition categories: normal bolt, loosening bolt, missing bolt, and rusted bolt, as summarized in **Table 1**. During dataset construction, industrial anti-loosening marker pens were employed to paint anti-loosening marks on bolt surfaces. These marks exhibit strong adhesion, excellent weather resistance, chemical corrosion resistance, rapid drying, and are difficult to erase. The anti-loosening mark serves as a reliable visual management tool that can intuitively reflect whether a bolt has loosened or been disassembled. The specific procedure is as follows: after the bolt is fully tightened, a red marker paint is applied transversely from the end face of the bolt head, with a continuous stroke traversing both the bolt top surface and the outer wall of the external nut to form a complete and continuous mark line. During routine inspections, any fracture, misalignment, or offset of

the red mark line indicates nut rotation and bolt loosening, necessitating prompt re-tightening.

Condition	Feature	Illustration
Normal Bolt	Anti-loosening mark aligned	
Loosening Bolt	Anti-loosening mark misaligned	
Missing Bolt	Bolt completely missing, only hole remains	
Rusted Bolt	Obvious reddish-brown rust on bolt surface, blurred texture	

Table 1 Bolt connection damage conditions.

The bolt image data used in this study were acquired from two sources.

1)**Laboratory truss-type cableway tower test model.** As shown in **Fig. 5(a)**, the tower has a total height of 9 m and consists of three stories, each with a height of 3 m. The bottom plan dimensions are 4 m × 4 m, the top plan dimensions are 1 m × 1 m, and the plan dimensions of the two middle stories are 3 m × 3 m and 2 m × 2 m, respectively. The model is fabricated from Q235 carbon structural steel. The main members are made of angle steel L125×8, L100×8, and L90×6; the horizontal members and diagonal braces are made of angle steel L75×5, L63×5, and L50×5, respectively. The ends of the members are connected to the gusset plates by 1 to 2 M16 bolts, and the four main columns are anchored to the ground using anchor bolts.



(a) Experimental model of a truss-type cableway tower



(b) Experimental model of a hexagonal-section steel tubular truss tower

Fig. 5 Photographs of the test models.

2) Hexagonal-section steel tubular truss tower experimental model. This model is a regular hexagonal steel-pipe tower structure situated on the outdoor ground level, as shown in **Fig. 5(b)**. The model has a side length of 1.5 m and a total height of 6 m, comprising four levels. The bottom level is 0.5 m above the ground, while the heights of the other three levels are 1.8 m each. The tower columns are fabricated from Q235 seamless steel pipes with a diameter of 16 cm. All horizontal members and diagonal braces are made of equal-leg angle steel L40×4. The ends of the members are connected to the column gusset plates by two M20 bolts per joint, and each column is anchored to the ground using four anchor bolts.

Prior to data collection, all bolts were first tightened using a torque wrench, and anti-loosening marks were applied. Subsequently, to simulate various damage conditions, a subset of bolts was randomly selected and manually loosened to create loosening conditions, some bolts were randomly removed to create missing conditions, and some bolts were replaced with corroded bolts of the same specification to create corrosion conditions.

The image acquisition equipment employed in this study was a first-person-view (FPV) drone (DJI Avata 2) in **Fig. 6**. This drone offers high maneuverability and flexibility, equipped with a high-definition camera and a real-time image transmission system, and supports binocular visual positioning.

Even in indoor environments with weak GPS signals or within steel truss structures, the drone can achieve precise hovering and stable flight through visual positioning. Consequently, this equipment is suitable not only for rapid inspection of the exterior of large structures but also for capturing images within confined interior spaces, making it well-suited for damage detection of engineering structures in complex environments.



(a) FPV drone



(b) Acquisition site

Fig. 6 Image acquisition equipment.

A total of 600 images were collected from the two tower models. During the acquisition process, variations in lighting conditions, camera angles, and background complexity were fully considered to ensure the diversity and representativeness of the dataset.

Data preprocessing and annotation

After preparing the raw image dataset, the LabellImg annotation tool was used for manual labeling. Each bolt was precisely localized and classified using a bounding box. Four annotation categories were defined with the following label names: normal bolt (“normal_bolt”), loosening bolt (“loosening_bolt”), missing bolt (“missing_bolt”), and rusted bolt (“rust_bolt”). To ensure annotation quality, all images were independently labeled by two researchers and subsequently cross-checked and corrected by a senior expert.

Dataset partition and augmentation

The 600 collected bolt images were partitioned into 500 training images and 100 validation images. To expand the training samples and enhance the model's generalization capability, offline data augmentation (including rotation, scaling, flipping, and brightness/contrast adjustments) was applied to the training set of 500 images. These transformations simulate drone imaging under different angles,

distances, and lighting conditions in **Fig. 7**. After augmentation, the training set was expanded to 3,500 images. In contrast, the validation set remained as the original 100 images and did not undergo any augmentation, ensuring the objectivity of the evaluation. Finally, a complete dataset comprising 3,600 images (3,500 for training, 100 for validation) was constructed. The entire dataset contains a total of 19,719 annotated bolt instances. The distribution of instances across categories is presented in **Table 2**: 8,960 normal bolts, 6,919 loosening bolts, 1,509 missing bolts, and 2,331 rusted bolts, which satisfies the sample diversity requirements for model training and evaluation.

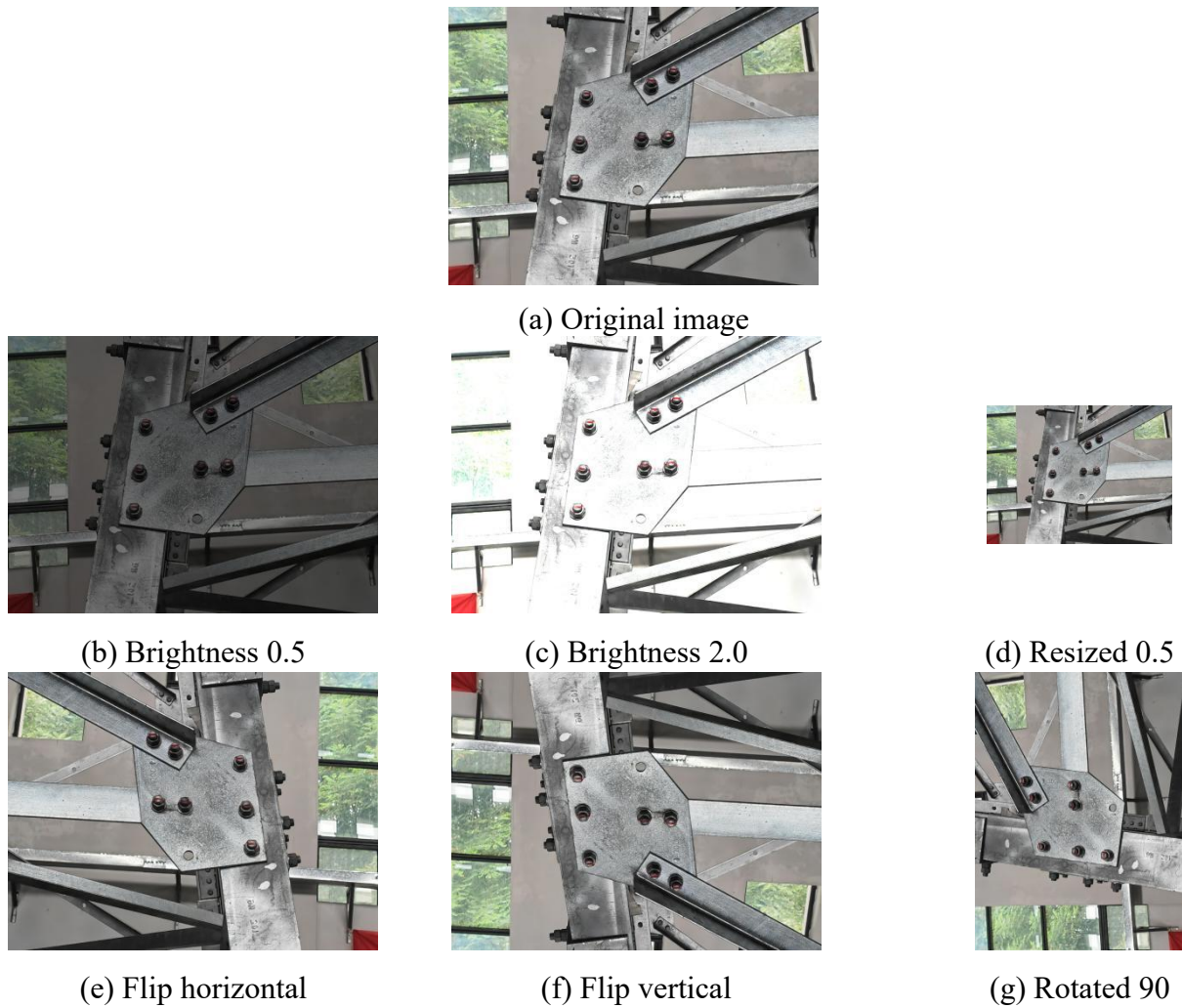


Fig. 7 Examples of data augmentation

Classification	Label	Instance Count
Normal Bolt	normal_bolt	8,960

Loosening Bolt	loosening_bolt	6,919
Missing Bolt	missing_bolt	1,509
Rusted Bolt	rust_bolt	2,331
Total		19,719

Table 2 Statistics of instance counts for each category in the dataset.

Training configuration

The proposed lightweight bolt detection model, LBD-YOLO, was trained using the stochastic gradient descent (SGD) optimizer with an initial learning rate of 0.01, a momentum factor of 0.937, and a weight decay coefficient of 0.0005. The learning rate decay followed a cosine annealing schedule. The total training period was set to 100 epochs, with the first 3 epochs dedicated to warmup. The batch size was set to 16, and all input images were uniformly resized to 640×640 pixels. The computational platform and programming environment for the LBD-YOLO model are listed in **Table 3**.

Software/Hardware	Specification
Operating System	Windows 11 64-bit
CPU	Intel(R) Core(TM) i9-14900K
GPU	NVIDIA GeForce RTX 3090 24G
RAM	64 GB
Environment	CUDA 11.8, Python 3.9.19, PyTorch 2.3.1, NumPy 1.24.2, Pandas 2.2.3, Matplotlib 3.9.2

Table 3 Computational platform and environment configuration.

Evaluation metrics

To comprehensively evaluate the detection performance of the LBD-YOLO model, this study adopted the following commonly used metrics in the object detection field: precision (P), recall (R), mean average precision at an Intersection over Union (IoU) threshold of 0.5 (mAP@50), and mean average precision averaged across multiple IoU thresholds from 0.5 to 0.95 (mAP@50-95). Additionally, to assess the lightweight characteristics and inference efficiency of the model, the number of parameters (Parameters), computational cost in giga floating-point operations (GFLOPs), and model file size were employed as supplementary metrics. To evaluate the model's performance on mobile deployment, the Android application package (APK) size, real-time inference speed (FPS), average confidence score,

and memory footprint were also adopted as evaluation indicators.

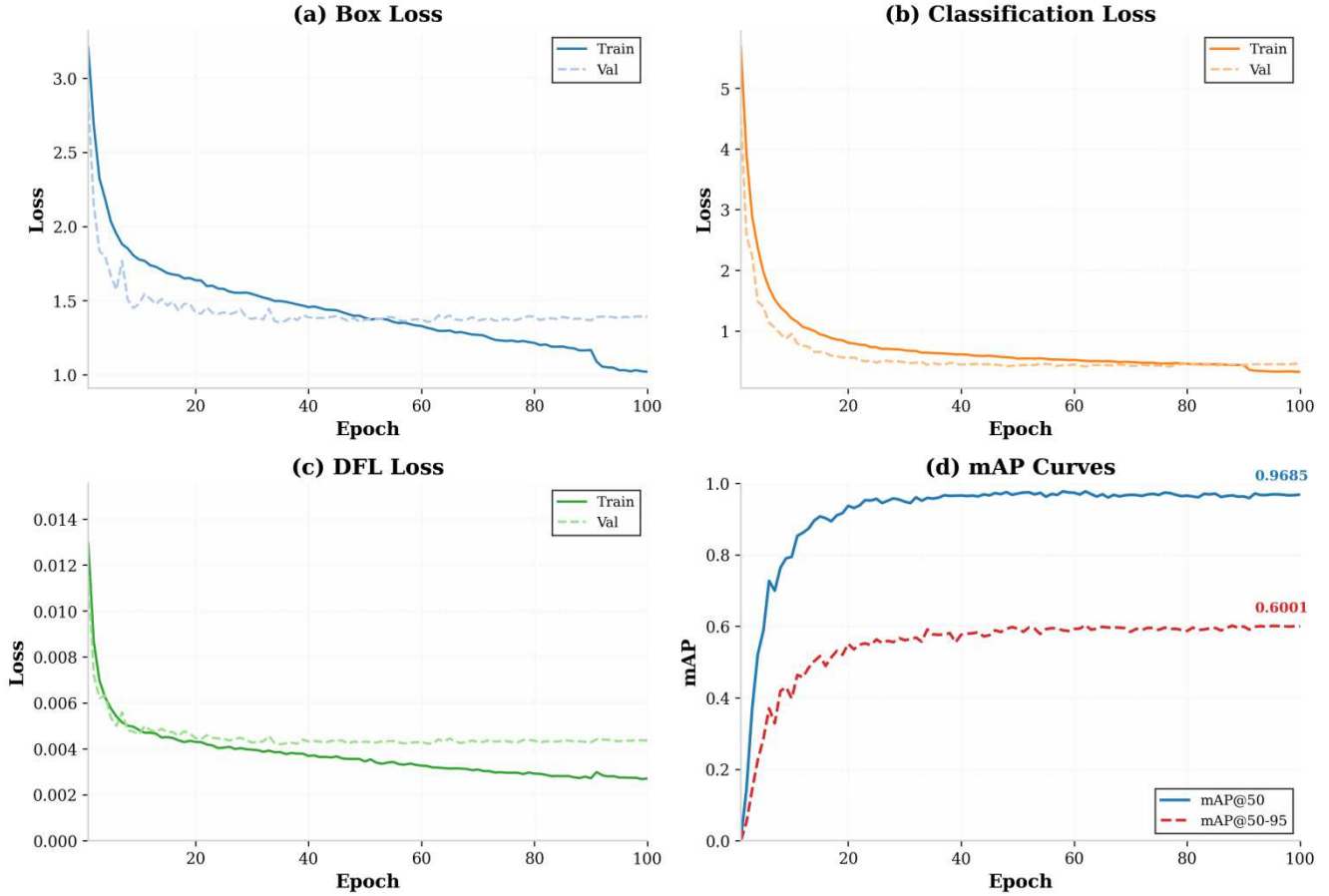


Fig. 8 Training process curves of LBD-YOLO.

Fig. 8 presents the training curves of LBD-YOLO on the self-built bolt dataset, including box loss, classification loss, DFL loss, and mAP curves. As shown in panels (a) – (c), all three losses decline rapidly during the early stage: the box loss drops from 3.21 at epoch 1 to 1.54 at epoch 30, the classification loss from 5.68 to 0.68, and the DFL loss from 0.013 to 0.004. After epoch 60, the decreasing trends gradually level off, and by the end of 100 epochs, the three losses converge to approximately 1.02, 0.33, and 0.0027, respectively. Throughout the entire training process, the training and validation loss curves remain closely aligned with consistent trends, with no obvious divergence or rebound in the validation losses, indicating that the model generalizes well without overfitting. Panel (d) displays the mAP@50 and mAP@50-95 curves, both of which rise steadily as training proceeds. The mAP@50 increases rapidly from near zero to 94.9% within the first 30 epochs and eventually reaches 97.7%, while the more stringent mAP@50-95 also exhibits a sustained upward trend, finally converging to 61.0%. Both precision curves stabilize in the later epochs with minimal fluctuation, further

confirming that the proposed collaborative feature enhancement strategy not only improves detection accuracy but also maintains training stability. Overall, the training curves demonstrate that LBD-YOLO achieves fast convergence, high accuracy, and strong generalisation capability, making it well suited for reliable deployment in real-world engineering scenarios.

RESULTS

Ablation study

To systematically evaluate the contribution of each improved module to the model performance, a series of ablation experiments were conducted under identical training settings using YOLO26n as the baseline. The results are summarized in **Table 4**.

GSConv	ADown	BoTNet	BiFPN	P	R	mAP@50	mAP@50-95	Parameters	GFLOPs	Time (h)
×	×	×	×	91.8%	93.2%	97.0%	60.4%	2,375,616	5.2	3.965
√	×	×	×	95.4%	91.6%	96.4%	58.5%	2,373,696	5.1	3.786
×	√	×	×	95.4%	92.6%	97.2%	60.1%	2,029,952	4.3	3.715
×	×	√	×	95.9%	90.5%	97.1%	60.2%	2,341,440	5.2	3.997
×	×	×	√	95.6%	89.5%	96.9%	60.4%	2,375,628	5.2	3.555
√	√	√	√	95.9%	93.2%	97.7%	61.0%	1,993,868	4.2	3.893

Table 4 Results of the ablation study.

The standalone introduction of the GSConv module increased the precision from 91.8% to 95.4%, while marginally reducing the computational complexity, indicating that efficient convolution operations can effectively enhance fine-grained feature extraction capability. The ADown module improved mAP@50 from 97.0% to 97.2% and simultaneously reduced model complexity, validating its effectiveness in preserving discriminative information during the downsampling process. The incorporation of the BoTNet module raised the precision to 95.9%, demonstrating that global context modeling can effectively suppress false detections against the complex background of cableway towers. The BiFPN module, through enhanced cross-scale feature interaction, improved multi-scale detection performance.

When all four modules were integrated, LBD-YOLO achieved the best overall performance: mAP@50 reached 97.7%, mAP@50-95 reached 61.0%, while the number of parameters and GFLOPs were reduced by 16.1% and 19.2%, respectively. These results indicate that a significant complementary

effect exists among the four modules, enabling simultaneous improvements in detection accuracy and computational efficiency.

Comparison with state-of-the-art methods

To further evaluate the effectiveness of LBD-YOLO, comparative experiments were conducted against current mainstream detection models. As shown in **Table 5**, LBD-YOLO achieves a favorable trade-off between detection accuracy and computational cost.

Model	mAP@50	mAP@50-95	R	Parameters	GFLOPs	Model Memory Usage (MB)
YOLO26n	97.0%	60.4%	93.2%	2,375,616	5.2	5.12
YOLO12n	97.7%	61.2%	92.2%	2,557,508	6.3	5.24
YOLOv9t	97.6%	60.8%	92.8%	1,972,156	7.6	4.40
YOLOv8n	97.7%	60.4%	93.4%	3,006,428	8.1	5.94
YOLOv6	96.8%	59.9%	92.2%	4,234,140	11.8	8.37
YOLOv3	96.6%	59.8%	91.2%	12,129,720	18.9	23.20
LBD-YOLO	97.7%	61.0%	93.2%	1,993,868	4.2	4.39

Table 5 Results of the comparison experiment.

LBD-YOLO achieved an mAP@50 of 97.7%, outperforming YOLOv9t¹⁹, YOLOv6²⁶, and YOLOv3²⁷, while matching YOLOv8n²⁸ and YOLO12n³³, with only 1,993,868 parameters and 4.2 GFLOPs of computational resources. Compared with YOLO12n, LBD-YOLO reduced the number of parameters and computational complexity by 22.0% and 33.3%, respectively, while maintaining comparable detection accuracy. In contrast to YOLOv9t, LBD-YOLO reduced the GFLOPs by 44.7% and achieved slightly higher detection accuracy.

These results demonstrate that LBD-YOLO achieves an excellent balance between detection performance and model complexity, providing technical feasibility for its deployment on resource-constrained industrial edge devices.

Per-class detection performance analysis

To further assess the robustness of LBD-YOLO across different damage types, the detection performance for four bolt conditions (normal, loosening, missing, and rusted) was analyzed. The results

are presented in **Table 6**.

Classification	P	R	mAP@50	mAP@50-95
All	95.9%	93.2%	97.7%	61.0%
Normal Bolts	95.2%	92.7%	98.3%	66.1%
Loosening Bolts	92.2%	94.5%	97.7%	61.7%
Missing Bolts	99.4%	94.3%	98.9%	59.4%
Rusted Bolts	96.7%	91.3%	96.7%	56.1%

Table 6 Per-class detection results.

Overall, LBD-YOLO exhibited stable performance across all categories, with an overall precision of 95.9%, a recall of 93.2%, an mAP@50 of 97.7%, and an mAP@50-95 of 61.0%. This consistently strong performance indicates that the proposed collaborative optimization strategy enhances the feature representation capability for each category while maintaining favorable generalization ability in complex scenarios.

Among all categories, the missing bolt detection task achieved the highest recognition accuracy, with a precision of 99.4% and an mAP@50 of 98.9%. This performance is primarily attributed to the global context modeling capability introduced by BoTNet, which enables the model to distinguish missing bolt regions from complex backgrounds by leveraging surrounding structural information. The low false detection rate is particularly critical for bolt connection safety inspection of cableway towers, as missed detections may lead to severe safety hazards.

Visualization of Detection Results

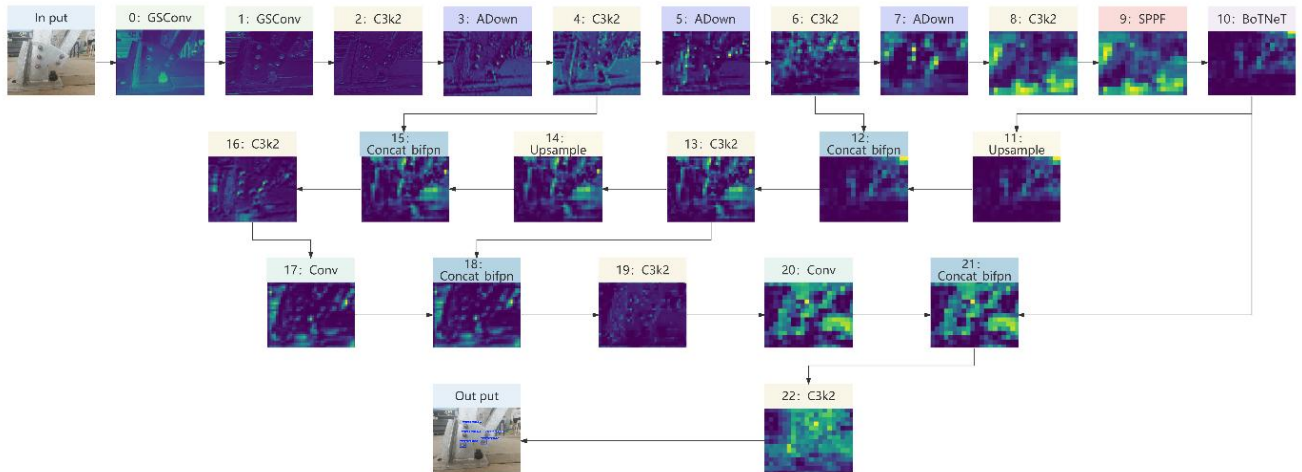


Fig. 9 Visualization of feature extraction and detection results.

Fig. 9 illustrates the feature extraction visualization and bolt damage detection results of the LBD-YOLO model, validating its robustness in feature representation, multi-scale detection, and complex scenarios. The GSConv module enhances fine-grained feature extraction capability, enabling clear presentation of subtle damage textures while suppressing background interference. The ADown module preserves edge contours and fine details of small bolts during downsampling. The BoTNet module strengthens global semantic modeling, accurately delineating bolt contours under occlusion or cluttered backgrounds. The BiFPN module unifies detection features across different scales, mitigating the insufficient activation of tiny targets. These visualization results demonstrate that the synergistic integration of the proposed improvement modules effectively enhances the feature expression capability and overall detection performance of LBD-YOLO for industrial bolt damage detection.

Mobile deployment validation

To evaluate the practical applicability of the LBD-YOLO model, mobile deployment experiments were conducted on an OPPO Find X8 smartphone equipped with a MediaTek Dimensity 9400 processor, 16 GB of RAM, and 512 GB of storage. The trained model was converted to the NCNN mobile inference framework format and integrated into a mobile application developed for real-time bolt detection.

The deployment results are summarized in **Table 7**. The application requires only 11.4 MB of storage space and 424 MB of runtime memory, achieving an inference speed of 20.2 FPS. These results remain practically acceptable for on-site bolt inspection of cableway towers and validate the feasibility of deploying the LBD-YOLO model on mobile devices.

APP Size (MB)	FPS	Average Confidence	Memory Usage (MB)
11.4	20.2	0.85	424

Table 7 Mobile deployment metrics of LBD-YOLO.

This deployment performance primarily stems from the lightweight architecture design of LBD-YOLO. The integration of GSConv and ADown effectively reduces computational redundancy, while the collaborative optimization strategy maintains high detection accuracy without significantly increasing computational overhead, thereby achieving a balance between detection performance and deployment efficiency.

The successful deployment on Android devices demonstrates that LBD-YOLO possesses practical value in the field of intelligent inspection and is applicable to mobile detection, unmanned aerial

vehicles, and edge computing scenarios.

Field test validation on a transmission tower

To further evaluate the generalization capability of the proposed LBD-YOLO model in real-world, unconstrained outdoor environments, an additional field test was conducted on an in-service steel transmission tower located in Zhangjiajie, Hunan, China. This tower is a typical lattice steel structure assembled with numerous bolted gusset-plate connections, and it differs markedly from the laboratory cableway tower models used for training in terms of structural form, member cross-sections, connection details, and surrounding background complexity. The outdoor setting introduces multiple challenging factors: cluttered vegetation and sky backgrounds, frequent and drastic illumination changes, cross-occlusion among structural members, and continuously varying shooting distances and viewing angles. Consequently, this test serves as a rigorous cross-structure assessment of the model's transferability under zero-shot conditions.

For this field test, the same FPV drone (DJI Avata 2) described in the Data collection section was employed to acquire an independent set of 100 images. All images were manually annotated following the same protocol as the main dataset to provide reliable ground-truth labels. The model weights trained exclusively on the laboratory dataset were directly applied to this new scenario without any fine-tuning or retraining, corresponding to a strict zero-shot inference setting. The quantitative detection performance is summarized in **Table 8**, and representative detection results are visualized in **Fig. 10**.

Model	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50-95 (%)
LBD-YOLO	91.6	79.8	91.2	54.5

Table 8 Field test performance of LBD-YOLO on the transmission tower.

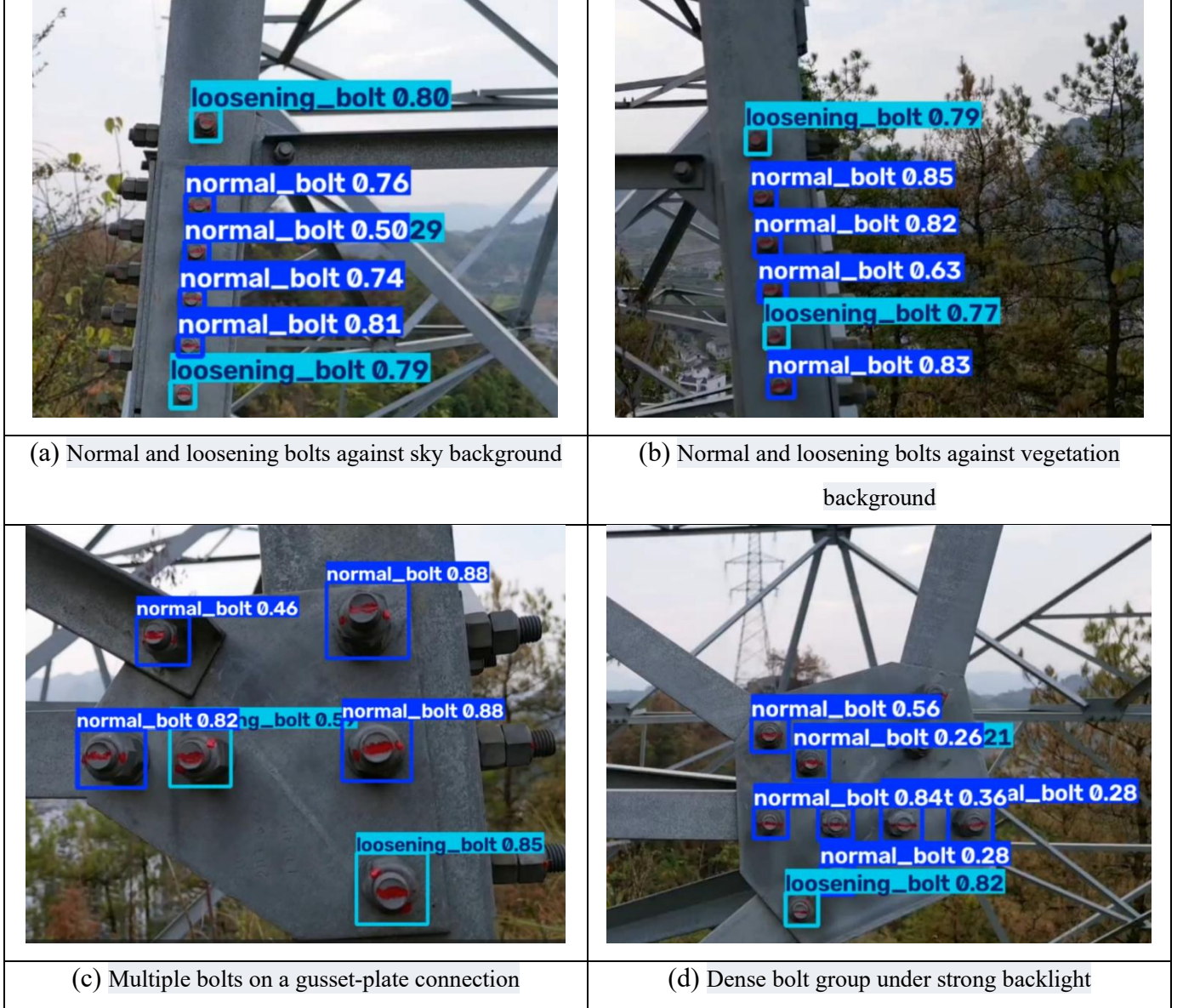


Fig. 10 Representative bolt detection results of LBD-YOLO in the transmission tower field test.

Under this challenging zero-shot test, LBD-YOLO achieved a precision of 91.6%, a recall of 79.8%, an mAP@50 of 91.2%, and an mAP@50-95 of 54.5%. Compared with the results on the laboratory test set (95.9% precision, 93.2% recall, 97.7% mAP@50, and 61.0% mAP@50-95), the four metrics decreased by 4.3, 13.4, 6.5, and 6.5 percentage points, respectively. The relatively larger drop in recall indicates that the model faces increased missed-detection challenges when transferred to an unseen real structure, which can be attributed to the cumulative adverse factors inherent in outdoor transmission tower inspection, including extreme lighting, severe occlusions, and the differences in structural geometry and bolt appearance between the two tower types.

The visualization results in **Fig. 10** qualitatively confirm these observations. LBD-YOLO

accurately localized and classified bolts against cluttered sky and vegetation backgrounds (**Fig. 10(a)** and **(b)**) as well as on multi-bolt gusset-plate connections (**Fig. 10(c)**). Even for a dense bolt group imaged under strong backlight (**Fig. 10(d)**), all bolts were successfully detected; however, partially occluded and small-sized bolts received relatively low confidence scores (0.26–0.56), suggesting that extreme illumination, occlusion, and insufficient image quality remain the primary factors limiting on-site detection performance. The performance gap between laboratory and field tests is therefore largely explainable by these real-world degradation factors, rather than a fundamental failure of the model architecture.

The promising transferability, despite the measurable drop, can be attributed to the collaborative architectural design of LBD-YOLO. Specifically, the BoTNet global self-attention module suppresses interference from complex foliage and sky backgrounds by modeling the contextual relationships between bolts and surrounding members; the BiFPN weighted multi-scale fusion mechanism effectively captures bolts of varying sizes caused by different drone shooting distances and angles; and the GSConv and ADown modules preserve the fine-grained edge and texture features of small bolt targets during lightweight feature extraction and downsampling. Overall, this field test demonstrates that the proposed method possesses good generalization capability and can be extended from cableway towers to other lattice steel structures, although further improvements—such as augmenting the training set with simulations of mark fading, blur, and complex backgrounds, or employing domain-adaptive strategies—are desirable to bridge the remaining performance gap.

DISCUSSION

The proposed LBD-YOLO model achieves 97.7% mAP@50 and 61.0% mAP@50-95 on the self-constructed bolt damage dataset, representing gains of 0.7 and 0.6 percentage points over the baseline YOLO26n, while reducing parameters and FLOPs by 16.1% and 19.2%, respectively. This simultaneous improvement in accuracy and efficiency is attributed to the complementary roles of the four integrated modules.

GSConv reduces computational redundancy via hybrid standard/depthwise separable convolutions with channel shuffling, preserving fine-grained edge and texture features of bolts. ADown adopts a dual-path downsampling strategy that mitigates information loss for small-sized targets. The multi-head self-attention in BoTNet expands the receptive field and enables contextual reasoning between bolts and

surrounding members, which is particularly effective for detecting missing bolts (99.4% precision). BiFPN adaptively balances multi-scale contributions through learnable weights, alleviating the large scale variation among tower bolt targets. Collectively, these modules form a collaborative enhancement chain from local detail to global semantics and from single-scale to multi-scale representation.

Compared with mainstream YOLO variants, LBD-YOLO outperforms YOLOv9t, YOLOv6 and YOLOv3 in accuracy, and matches YOLOv8n and YOLO12n while using substantially fewer parameters and lower computational cost. This advantage stems from the high-quality self-built dataset covering four typical damage states, the effective compensation for fine-grained feature loss through GSConv and ADown, and the background-suppression capability provided by BoTNet in cluttered tower scenes. However, the $mAP@50-95$ for rusted bolts (56.1%) remains notably lower than other categories, likely due to visual confusion between rust textures and background stains, as well as the gradual nature and blurred boundaries of corrosion, which hinder precise localization under high IoU thresholds.

The model also exhibits limitations in detecting severely rusted bolts or under extreme occlusion, and misclassification may occur when bolt holes are covered by oil or snow. Although deployment on a smartphone has been validated, adaptation to more resource-constrained embedded devices requires further investigation. Future work will focus on addressing these practical challenges through targeted data expansion and model optimisation.

CONCLUSIONS

This paper presents LBD-YOLO, a lightweight bolt damage detection model built upon YOLO26n, specifically designed for intelligent inspection of bolted connections in steel truss structures. By synergistically integrating GSConv, ADown, BoTNet, and BiFPN, the model enhances fine-grained feature extraction, global semantic perception, and multi-scale fusion capabilities while preserving a compact architecture suitable for edge deployment.

On the self-constructed dataset covering normal, loosening, missing, and rusted bolts, LBD-YOLO achieves 97.7% $mAP@50$ and 61.0% $mAP@50-95$, with 95.9% precision and 93.2% recall. Compared to the baseline, it reduces parameters by 16.1% and computational cost by 19.2%, while delivering accuracy superior or comparable to several mainstream YOLO models with significant lightweight advantages. Real-time inference at 20.2 FPS on an OPPO Find X8 smartphone confirms its engineering

practicality. Furthermore, zero-shot field tests on an in-service transmission tower demonstrate promising generalization to unseen real-world structures, achieving 91.6% precision and 91.2% mAP@50, which highlights its potential for broad deployment in power grid and infrastructure inspection.

The proposed method offers an effective technical solution for automated, high-precision condition assessment of bolted connections in steel truss infrastructures. Ongoing efforts will further enhance its robustness under adverse environmental conditions and facilitate low-power embedded implementation for wider engineering applications.

ACKNOWLEDGEMENTS

This research was funded by National Natural Science Foundation of China (Grant No. 52268049) and 2025 Open Fund for Bidding Projects of Provincial-Level Scientific Research Platform, Hunan Base of Natural and Cultural Heritage (Grant No. ZRYC2504).

FUNDING DECLARATION

This research was funded by the National Natural Science Foundation of China (Grant No. 52268049) and the 2025 Open Fund for Bidding Projects of Provincial-Level Scientific Research Platform, Hunan Base of Natural and Cultural Heritage (Grant No. ZRYC2504).

AUTHOR CONTRIBUTIONS

D.Z. and B.G. conceived the experiments. D.Z. conducted the experiments, analyzed the results and wrote the main manuscript text. Z.J., S.L., Z.H. and Z.Hu. assisted with data collection. B.G. supervised the work and acquired funding. All authors reviewed the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY

The datasets generated or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Wang, Y. et al. Review of health monitoring and intelligent fault diagnosis for high-strength bolts: failure mechanisms, multi-modal sensing, and data-driven approaches. *Buildings* 16, 691 (2026). <https://doi.org/10.3390/buildings16040691>
2. Yang, L. et al. Fatigue life prediction method for bolted joints based on equivalent structural stress under tensile-compressive loading. *Int. J. Fatigue* 190, 108625 (2025). <https://doi.org/10.1016/j.ijfatigue.2024.108625>
3. Li, Q. et al. Evaluating bolt connecting faults in ring-type structures with nonlinear fault-induced loads and transmissibility functions. *Mech. Syst. Signal Process.* 191, 110212 (2023). <https://doi.org/10.1016/j.ymssp.2023.110212>
4. Qin, Y. & Yuan, L. Intelligent bolt loosening detection in transmission towers using acoustic signature analysis and machine learning. *Processes* 13, 1111 (2025). <https://doi.org/10.3390/pr13041111>
5. Jiang, S. et al. Automatic inspection of bridge bolts using unmanned aerial vision and adaptive scale unification-based deep learning. *Remote Sens.* 15, 328 (2023). <https://doi.org/10.3390/rs15020328>
6. Na, W. S. Bolt loosening detection using impedance based non-destructive method and probabilistic neural network technique with minimal training data. *Eng. Struct.* 226, 111228 (2021). <https://doi.org/10.1016/j.engstruct.2020.111228>
7. Wu, C., Xiao, Q. & Zhu, D. Piezoelectric ceramic sensor-based structural health monitoring of diagonally braced H-shaped steel structures. *Sci. Rep.* 13, 14964 (2023). <https://doi.org/10.1038/s41598-023-40758-6>
8. Zhao, L. et al. Acoustoelastic theory and mode analysis of bolted structures under preload. *Machines* 12, 822 (2024). <https://doi.org/10.3390/machines12110822>
9. Malik, Z. et al. A review on vision-based deep learning techniques for damage detection in bolted joints. *Asian J. Civ. Eng.* 25, 5697-5707 (2024). <https://doi.org/10.1007/s42107-024-01139-0>
10. Luo, J. et al. A novel anti-loosening bolt looseness diagnosis of bolt connections using a vision-based technique. *Sci. Rep.* 14, 11441 (2024). <https://doi.org/10.1038/s41598-024-62560-8>
11. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137-1149 (2017). <https://doi.org/10.1109/TPAMI.2016.2577031>
12. Redmon, J. et al. You only look once: unified, real-time object detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 779-788 (2016). <https://doi.org/10.48550/arXiv.1506.02640>
13. Wang, Y. et al. Robust three-stage deep learning and image processing framework for automated loose bolt detection in complex environments. *Autom. Constr.* 180, 106531 (2025). <https://doi.org/10.1016/j.autcon.2025.106531>
14. Peng, L. et al. Detection of bolt defects on transmission lines based on multi-scale YOLOv7. *IEEE Access* 12, 156639-156650 (2024). <https://doi.org/10.1109/ACCESS.2024.3485965>
15. Hua, G. et al. An enhanced YOLOv8-based bolt detection algorithm for transmission line. *IET Gener. Transm. Distrib.* 18, 4065-4077 (2024). <https://doi.org/10.1049/gtd2.13330>
16. Pan, X. & Yang, T. Y. Image-based monitoring of bolt loosening through deep-learning-based integrated detection and tr

acking. *Comput.-Aided Civ. Infrastruct. Eng.* 37, 1207-1222 (2022). <https://doi.org/10.1111/mice.12797>

17. Sapkota, R., Cheppally, R. H., Sharda, A. & Karkee, M. YOLO26: key architectural enhancements and performance benchmarking for real-time object detection. *arXiv preprint arXiv:2509.25164* (2025). <https://doi.org/10.48550/arXiv.2509.25164>
18. Li, H. et al. Slim-neck by GSConv: a lightweight-design for real-time detector architectures. *J. Real-Time Image Process.* 21, 62 (2024). <https://doi.org/10.1007/s11554-024-01436-6>
19. Wang, C.-Y., Yeh, I.-H. & Liao, H.-Y. M. YOLOv9: learning what you want to learn using programmable gradient information. In *Computer Vision – ECCV 2024, Lecture Notes in Computer Science*, vol. 15089 (Springer, Cham, 2025). https://doi.org/10.1007/978-3-031-72751-1_1
20. Srinivas, A. et al. Bottleneck transformers for visual recognition. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 16519-16529 (2021). <https://doi.org/10.48550/arXiv.2101.11605>
21. Tan, M., Pang, R. & Le, Q. V. EfficientDet: scalable and efficient object detection. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 10781 – 10790 (2020). <https://doi.org/10.1109/cvpr42600.2020.01079>
22. Zhou, K. & Jiang, S. Forest fire detection algorithm based on improved YOLOv11n. *Sensors* 25, 2989 (2025). <https://doi.org/10.3390/s25102989>
23. He, K., Zhang, X., Ren, S. & Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1904-1916 (2015). <https://doi.org/10.1109/TPAMI.2015.2389824>
24. Khanam, R. & Hussain, M. YOLOv11: an overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725* (2024). <https://doi.org/10.48550/arXiv.2410.17725>
25. Liu, S. et al. Path aggregation network for instance segmentation. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 8759-8768 (2018). <https://doi.org/10.1109/CVPR.2018.00913>
26. Li, C. et al. YOLOv6: a single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976* (2022). <https://doi.org/10.48550/arXiv.2209.02976>
27. Redmon, J. & Farhadi, A. YOLOv3: an incremental improvement. *arXiv preprint arXiv:1804.02767* (2018). <https://doi.org/10.48550/arXiv.1804.02767>
28. Cai, Z. et al. YOLOv8n-FAWL: object detection for autonomous driving using YOLOv8 network on edge devices. *IEEE Access* 12, 158376-158387 (2024). <https://doi.org/10.1109/ACCESS.2024.3480976>
29. Vaswani, A. et al. Attention is all you need. In *Proc. Advances in Neural Information Processing Systems* 30, 5998-6008 (2017). <https://doi.org/10.48550/arXiv.1706.03762>
30. He, K. et al. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 770-778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
31. Lin, T. et al. Feature pyramid networks for object detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2117-2125 (2017). <https://doi.org/10.1109/CVPR.2017.106>
32. Gao, J. et al. Augmented weighted bidirectional feature pyramid network for marine object detection. *Expert Syst. Appl.* 237, 121688 (2024). <https://doi.org/10.1016/j.eswa.2023.121688>
33. Tian, Y., Ye, Q. & Doermann, D. YOLOv12: Attention-Centric Real-Time Object Detectors. *arXiv preprint arXiv:2502.12524* (2025). <https://doi.org/10.48550/arXiv.2502.12524>