Smart Structures and Systems, *Vol. 32, No. 1* (*2023*) *23-35*
https://doi.org/10.12989/sss.2023.32.1.023

Corroded and loosened bolt detection of steel bolted joints based on improved you only look once network and line segment detector

Youhao Ni [1,2a], Jianxiao Mao [1,2b], Hao Wang*[1,2], Yuguang Fu [3c] and Zhuo Xi [1,2d]

[1] *Key Laboratory of C&PC Structures of Ministry of Education, Southeast University, Nanjing, 211189, China*
[2] *School of Civil Engineering, Southeast University, Nanjing 211889, China*
[3] *School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore*

(*Received December 4, 2022, Revised April 14, 2023, Accepted June 30, 2023*)

Abstract. Steel bolted joint is an important part of steel structure, and its damage directly affects the bearing capacity and durability of steel structure. Currently, the existing research mainly focuses on the identification of corroded bolts and corroded bolts respectively, and there are few studies on multiple states. A detection framework of corroded and loosened bolts is proposed in this study, and the innovations can be summarized as follows: (i) Vision Transformer (ViT) is introduced to replace the third and fourth C3 module of you-only-look-once version 5s (YOLOv5s) algorithm, which increases the attention weights of feature channels and the feature extraction capability. (ii) Three states of the steel bolts are considered, including corroded bolt, bolt missing and clean bolt. (iii) Line segment detector (LSD) is introduced for bolt rotation angle calculation, which realizes bolt looseness detection. The improved YOLOv5s model was validated on the dataset, and the mean average precision (mAP) was increased from 0.902 to 0.952. In terms of a lab-scale joint, the performance of the LSD algorithm and the Hough transform was compared from different perspective angles. The error value of bolt loosening angle of the LSD algorithm is controlled within 1.09%, less than 8.91% of the Hough transform. Furthermore, the proposed framework was applied to full-scale joints of a steel bridge in China. Synthetic images of loosened bolts were successfully identified and the multiple states were well detected. Therefore, the proposed framework can be alternative of monitoring steel bolted joints for management department.

Keywords: corroded and loosened bolt detection; improved YOLOv5s; linear segment detector; steel bolted joints; vision transformer

1. Introduction

As essential components of steel structures, bolted joints are prone to surface change during long-term operation, such as bolt loosening, bolt corrosion, bolt missing, etc. (Wang *et al*. 2013b, Huynh 2021, Pan and Yang 2021). Subjected to cycling vibration and fluctuating loads, bolt loosening of steel joints frequently occurs (Nikravesh and Goudarzi 2017, Huynh *et al*. 2019). Such surface change would directly lead to the bearing capacity failure of bolted joints, and even cause local failures or overall collapse of the structure (Ta and Kim 2020, Ni *et al*. 2023). Thus, the detection of bolt loosening and bolt corrosion, in an accurate and timely manner, is essential for the safety of infrastructure operation.

During the past two decades, the research and application of structural health monitoring (SHM) have been widely carried out for structure damage detection (Pidaparti 2007, Ye *et al*. 2016, Spencer *et al*. 2019). Generally, the SHM methods adopt contact sensors to monitor the modal properties of components or structures (i.e., frequency, vibration modes, damping etc.) (Sun *et al*. 2020). Damage detection of steel bolted joints based on contact sensors mainly consists of acoustic-based techniques (Chaki and Bourse 2009), wavelet-based method to locate the damage position (Blachowski et al. 2015), the piezoelectric active sensing method (Yang and Chang 2006, Wang *et al*. 2013a), and piezoelectric impedance technique evaluating the tightness degree of bolts (Park *et al*. 2003). However, contact sensors always have limitations, including time-consuming installation and short service life, and are susceptible to environmental changes (temperature, humidity, etc.). Meanwhile, some sensors are consumables with high cost and require periodic inspection and replacement. Large steel structures contain thousands of bolted joints, and it is difficult to rely on contact sensors monitoring one by one.

Recently, compared with the SHM methods, the development of deep learning and computer vision presents alternative solutions to the multiple states detection of bolted joints (Hoskere *et al*. 2018, Cha *et al*. 2018, Forkan *et al*. 2022). The deep-learning-based approaches extract multi-scale and deeper features of bounding objects, so as to improve model generalization ability and achieve high

*Corresponding author, Ph.D., Professor,
E-mail: wanghao1980@seu.edu.cn
[a] Ph.D. Student, E-mail: yhni@seu.edu.cn
[b] Associate Professor, E-mail: jianxiao@seu.edu.cn
[c] Assistant Professor, E-mail: yuguang.fu@ntu.edu.sg
[d] Professor, E-mail: xizhuoxian@seu.edu.cn

ISSN: 1738-1584 (Print), 1738-1991 (Online)

accuracy. For bolt loosening detection, some research of deep-learning-based approaches have emerged with Hough transform. A classifier based on support vector machine and circular Hough transform was employed, and loosening bolts were automatically distinguished from fastening bolts (Cha *et al*. 2016). This method adopted lengths of bolts as damage features index, and the false detection rate is 5% with camera angle limitation. To address the limitation of camera angles, Zhang *et al*. (2020) introduced a bolt loosening detection method based on Faster R-CNN. The object proposals from selective searches was taken as inputs for feature extraction, and the average precision is better than sliding window methods. As small dataset is a limiting factor for model generalization, a bolt looseness detection framework was proposed based on R-CNN model and synthetic images (Pham *et al*. 2020). Data enhancement by synthetic images improved the detection rate from various scenes, while Hough transform is sensitive to environment conditions in line detection. To overcome drawbacks of Hough transform, the tracking algorithm of optical flows was integrated to calculate rotation angle (Pan and Yang 2021). The combination of optical flows and you-only-look-once (YOLO) lightweight network realized reasonable accuracy of continuously bolt rotation monitoring. Moreover, automatic correction of distortion and adaptive bolt angle estimation was aiming at low robustness of Hough transform (Huynh 2021).

Besides bolt loosening detection, some studies have focused on corrosion detection in deep-learning perspective (Gu *et al*. 2018, Xu *et al*. 2020, Ta *et al*. 2022). Based on traditional convolutional neural network (CNN), VGG16 and ZF Net were utilized as input architectures to generate corrosion classifiers (Atha and Jahanshahi 2018). The classification capability is superior to multilayered perceptron network, while sliding window approach leads to severe time consuming. Instead of sliding window approach, Ta *et al.* (2022) proposed a corrosion segment network based on mask region-based convolutional neural network (Mask-RCNN). The segment network adopted Resnet50 as backbone and the mask loss function to regress pixel-wise location of corrosion. Furthermore, ensemble convolutional neural network (ECNN), balancing advantages and eliminating architecture restriction of diverse CNN models, was incorporated to form CorrDetector. Utilizing grid-based and the object annotation with polygonal mask, CorrDetector outperformed the state-of-the-art solution by 4.61% in precision (Forkan *et al*. 2022). In general, the corrosion detection methods based on deep learning, including image classification, object detection and semantic segmentation, basically meet the requirements of steel structure maintenance. Whereas, such network and data were designed for single damage, not applicable to multi-damage detection.

To date, for multi-damage detection algorithms of bolted joints, a limited number of deep-learning-based approaches were conducted (Ta and Kim 2020). Employing a multi-scale CNN, Hoskere *et al* (2018) proposed a six-type damage classification and localization network, including steel corrosion, cracks, etc. One classifier and segmenter were paralleled in architecture, and the filters at multiple scale extracted dimensional features. The aim was to achieve pixel-wise output, while true positives and robustness need further improvement. Faster R-CNN (Ren *et al*. 2015), with region proposal networks sharing multi-scale features, was adopted to form a five-type damage detection method (Cha *et al*. 2018). Both object detector and proposal generator, trained by mini-batch gradient descent end to end with classification and regression loss function, surpass multi-scale CNN in accuracy. In addition, Ta and Kim (2020) integrated R-CNN and Hough transform, where R-CNN is for surface corrosion detection and Hough transform identifies bolt loosening. Nevertheless, the loosening angle by Hough transform depends on the canny detector (Ding and Goshtasby 2001), and the threshold for determining remains still large. Therefore, to improve the capability of CNN-based network still needs to be addressed. Meanwhile, it is essential to further develop the multi-damage detection method for bolted joints, so as to ensure the safety of steel structures.

In this study, a framework for corroded and loosened bolts detection of steel bolted joints is proposed. The main idea is to identify the multiple surface damage and loosening bolts based on improved deep learning network and image processing techniques. The main contributions of this study are as follows. 1) Vision Transformer (ViT) is introduced to improve the feature extraction capability of you-only-look-once version 5s (YOLOv5s) network. The accuracy improvement by ViT is quantified, and three types of bolt states are detected, namely, corroded bolt, bolt missing and clean bolt. 2) Based on boundaries by canny edge detector, line segment detector (LSD) algorithm is designed to recognize nut edges for estimation of bolt looseness angle. 3) The proposed framework realizes corroded and loosened bolts detection of bolted joints, providing reference for the maintenance of steel structures. The feasibility and reliability of the framework were validated in laboratory testing and in a steel truss arch bridge, respectively.

The rest of this study is organized as follows. Section 2 illustrates the overview of methodology, including YOLOv5s, perspective transformation, and the LSD algorithm. The proposed detection framework of bolted steel joints are shown in Section 3. Section 4 presents improvement of YOLOv5s by ViT and laboratory test of bolt damage detection. Section 5 determines the bolt looseness angle and compares the LSD algorithm with the Hough transform. Field test is implemented on a real bridge. Finally, Section 6 concludes this study.

2. Overview of methodology

2.1 You only look once v5

In deep learning-based object detection algorithms, the emergence of YOLO network was epoch-making (Redmon *et al*. 2016, Redmon and Farhadi 2017, 2018, Bochkovskiy *et al*. 2020, Ge *et al*. 2021). YOLO adopts global image detection to reduce the background detection error, and directly outputs the positions and categories of the bounding

boxes. As a one-stage object detection network, the inference speed is superior to and the inference accuracy is competitive with that of the two-stage network, respectively. To date, the researchers have proposed subsequent versions of YOLO, including YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOX, etc. (Redmon *et al*. 2016, Redmon and Farhadi 2017, 2018, Bochkovskiy *et al*. 2020, Ge *et al*. 2021).

As a stable version, YOLOv5 is in application most widely for industrial scenes. The feature pyramid network (FPN) and perceptual adversarial network (PAN) are adopted in the neck of YOLOv5. FPN is a top-down structure, transmitting semantic features of the high level without the positioning information. PAN adds a bottom-up pyramid following FPN and transfers positioning features of the lower layer.

Focus structure is the core of the image slice operation, which reduces network parameters and floating-point computations. The CSP structure consists of convolutional layers, batch normalized layers, leaky relu layers, and residual layers. YOLOv4 adopts one CSP structure, while YOLOv5 adopts two CSP structures both in the backbone and neck. Residual layers enhances the gradient value during backpropagation between layers, effectively preventing the gradient disappearance. Generalized intersection over union (GIOU), binary cross entropy, and logits loss function are the loss function of bounding box loss, class probability loss and target score, respectively.

2.2 Perspective transformation

Due to the change of camera vision angle, perspective distortion of images occurs. Perspective transformation, known as projection mapping, is the process of projecting the target image onto a new view plane and corrects the distortion caused by perspective angle (Mezirow 1978, Muthalagu *et al*. 2020). Fig. 1 shows the schematic diagram of perspective transformation between image planes. Based on perspective transformation matrix and detected corner coordinates, the object perspective is transformed from the world coordinate plane to the camera projection plane.

The expression for perspective transformation is as follows

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = T \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} T_1 & T_2 \\ T_3 & a_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{22} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \quad (1)$$

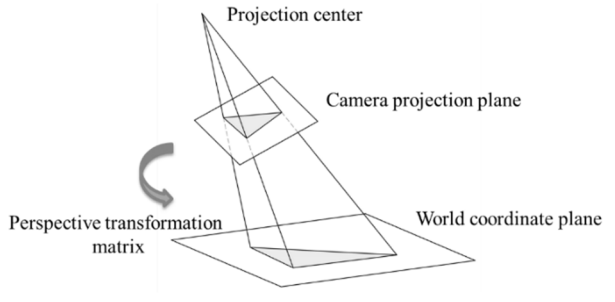


Fig. 1 Perspective transformation diagram between image planes

$$T_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad (2)$$

$$T_2 = [a_{13} \quad a_{23}]^T \quad (3)$$

$$T_3 = [a_{31} \quad a_{32}] \quad (4)$$

where T represents the perspective transformation matrix, T_1 is the linear transformation matrix, including scale change, rotation, shear, etc., T_2 is the perspective matrix, T_3 is the translation matrix. $[u, v, w]^T$ and $[x, y, z]^T$ are the pixel homogeneous coordinates of the original image and the image after perspective correction, respectively.

Fig. 2 presents the generation process of perspective-corrected steel bolted joint image. The bolt states are detected and the center point coordinates of the bolts are calculated from bounding boxes. Afterwards, the outermost coordinates of four bolts are extracted accordingly. Given the joint drawing, perspective transformation matrix can be calculated to obtain perspective-corrected steel joint. The joint shape is arbitrarily quadrilateral, and the corrected image is adopted for the subsequent bolt edge detection.

2.3 You only look once v5

Line segment detector (LSD) is a line detection and segmentation algorithm in linear processing time (Von Gioi *et al*. 2008). Compared with the Hough transform, LSD possesses the advantages of sub-pixel accuracy, controllable false detection rates, and no need to adjust parameters. On the basis of image edge detection, LSD is adopted to extract hexagon edges of nuts.

Contours of gray images are special areas, where the gray values vary significantly in the integer range. To reduce quantization artifacts and aliasing in line detection, image scaling is used to preprocess images in LSD, avoiding frequent breakpoints in detected lines. Subsequently, the global image gradients in two directions and gradient angles are computed, and the gradients are sorted into 1024 bins according to magnitude. The concept of level-line is proposed in LSD, that is, each pixel owns boundary line perpendicular to the gradient direction. If the level-line of each pixel is traversed, the level-line field of the image would be generated. In the level-line field, pixels with the largest gradient magnitude are successively taken as the seeds in the storage bins. The pixel regions, with the same level-line angle within certain limits of tolerance, are referred to as line support regions. Areas with small change in pixel values are usually slow gradient variation areas, which are distinguished from boundaries. Pixels with gradient magnitude less than the threshold are rejected for the line support regions construction. Fig. 3 shows a schematic diagram of angle error estimation caused by quantization noise.

In the generation process of line support regions, assuming an ideal image i and a quantization noise n, as follows

$$\tilde{\imath} = i + n \quad (5)$$

$$\nabla\tilde{\imath} = \nabla i + \nabla n \quad (6)$$

(a) Joint drawing (b) Detected bolts (c) Perspective-corrected steel joint

Fig. 2 Generation process of perspective-corrected steel bolted joint image

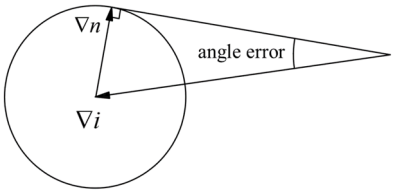


Fig. 3 Angle error estimation caused by quantization noise

$$|E_{\mathrm{AGL}}| \leq arcsin\left(\frac{q}{|\nabla i|}\right) \tag{7}$$

$$\rho = \frac{q}{sin\,\tau} \tag{8}$$

where ∇n is the gradient operator of the quantization noise, ∇i is the gradient operator of the image, $\nabla \tilde{i}$ is the real gradient operator of the image, q is the boundary on ∇i, and E_{AGL} is the angular error of level-line. The gradient threshold is determined by ρ and τ. q is the possible error caused by quantization artifacts in the gradient value calculation, and τ is the angle tolerance value used in the region growing process.

Generally, the pixel values are in the integer range [0,255], and the pixel gradient of error is mostly 1. To obtain a conservative boundary, q is set to 2. Pixels less than the threshold are eliminated from for line-support regions construction. Based on experimental analysis, the angle tolerance value τ is set to $\pi/8$ optimally.

Fig. 4 presents the schematic diagram of region connectivity in gradient directions. Assume that pixel regions are solid units, and that the gradient magnitude of each pixel is the "mass" of the pixel. The center of the enclosing rectangle formed by line-support regions is the centroid of pixel regions. The principal direction of the enclosing rectangle is the first inertia axis of the region, and the width and length are set to the minimum value that the enclosing rectangle covers the line-support regions.

The principal direction of the enclosing rectangle is set to the eigenvector angle corresponding to the minimum eigenvalue of the matrix, as follows

$$M = \begin{pmatrix} m^{xx} & m^{xy} \\ m^{xy} & m^{yy} \end{pmatrix} \tag{9}$$

$$m^{xx} = \frac{\sum_{j \in Re\,gion} G(j) \cdot (x(j) - c_x)^2}{\sum_{j \in Re\,gion} G(j)} \tag{10}$$

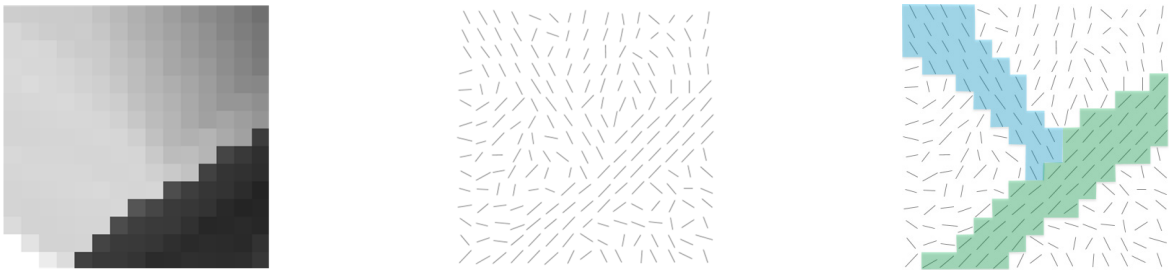
$$m^{yy} = \frac{\sum_{j \in Re\,gion} G(j) \cdot \left(y(j) - c_y\right)^2}{\sum_{j \in Re\,gion} G(j)} \tag{11}$$

$$m^{xy} = \frac{\sum_{j \in Re\,gion} G(j) \cdot (x(j) - c_x)\left(y(j) - c_y\right)}{\sum_{j \in Re\,gion} G(j)} \tag{12}$$

where $G(j)$ is the gradient magnitude of the pixel j, $x(j)$ and $y(j)$ are the coordinates of the pixel, and c_x and c_y are the center coordinates of the enclosing rectangle, respectively.

3. Proposed detection framework

Fig. 5 presents the proposed framework for corroded and loosened bolts detection of steel bolted joints, including the following four steps. Firstly, bolted joint images are input into the improved YOLOv5s network model, including the backbone, the bottleneck and detectors.



(a) Image (b) Level-line field (c) Line support regions

Fig. 4 Schematic diagram of region connectivity in gradient directions

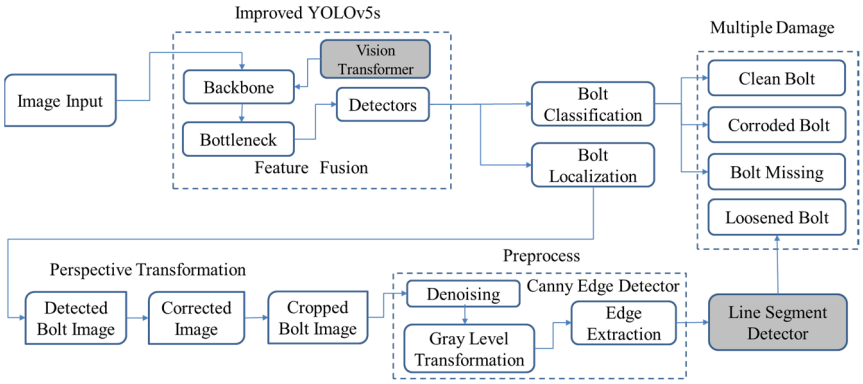
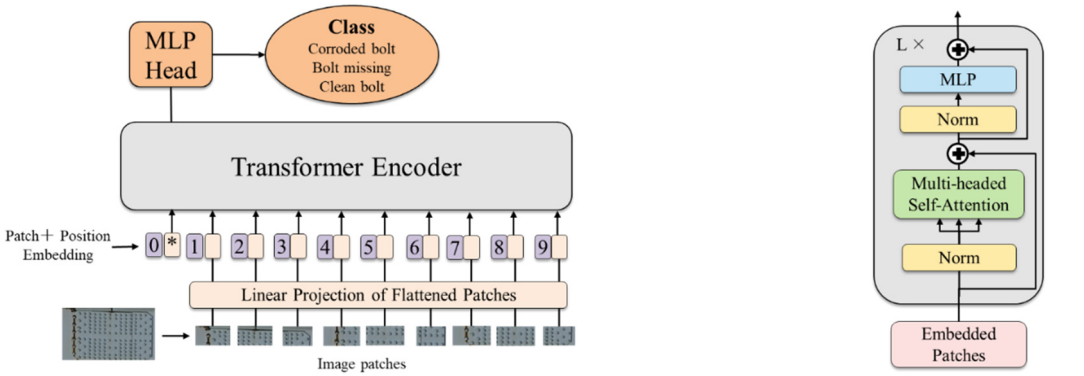


Fig. 5 The proposed framework for corroded and loosened bolts detection



(a) Vision transformer (b) Transformer encodery

Fig. 6 Vision transformer structure (ViT)

Secondly, through bolt classification, bolt states are identified as clean bolt, corroded bolt or bolt missing. Meanwhile, through bolt localization, bolt pixel coordinates are determined. Thirdly, the perspective-corrected images are obtained through perspective transformation matrix, and cropped bolt images are sub-images of individual bolts. Afterwards, through denoising and grey-level transformation, the Canny edge detector is to extract the edges of individual nuts. Finally, the LSD algorithm is to detect the straight line of nut edges, and the bolt rotation angle is calculated to determine loosened bolt.

4. Improved YOLOv5s algorithms

4.1 Vision transformer

For natural language processing (NLP) tasks, a model framework named Transformer was presented by the Google Brain team based on self-attention mechanism. (Vaswani *et al*. 2017). The framework mainly adopts multi-attention mechanism, residual network structure and position encoding. Based on Transformer, Dosovitskiy *et al*. (2020) proposed Vision Transformer (ViT) in computer vision (CV). Combining image patch decomposition and position encoding, sequence to sequence output prediction was successfully implemented in image classification tasks.

ViT mainly consists of four parts, namely, Patch embedding, Position embedding, Transformer encoder and Multi-layer perceptron (MLP) head. Among them, the Transformer encoder is the core component.

In the part of Patch embedding, the image is uniformly divided into multiple patches to a sequence. The image $x \in \Re^{H\times W\times C}$ is converted into two-dimensional patches $x_P \in \Re^{N\times(P^2\times C)}$, where $H \times W$ is the original image resolution, C is the number of channels, P^2 is the resolution of embedded patches, $D = P^2 \times C$ is the dimension of image patches and $N = \frac{H\times W}{P^2}$ is the sequence length. The output of the linear projection layer is referred as patch embedding, whose dimension remains $D \times N$.

According to class token in BERT (Devlin *et al.* 2018), a learnable embedding $z_0^0 = x_{\text{class}}$ is prepended to the embedded patch sequence, and the final state is z_L^0. The output of z_L^0 from Transformer encoder is y in Eq. (16). Position embedding are added to the embedded patch sequence with the original position information and the dimension is $N + 1$. The patch embedding and the position embedding together serve as the input vector of Transformer encoder. Transformer encoder consists of Multi-headed self-attention layer (MSA), Layernorm (LN) and Multi-layer perceptron (MLP) head (Vaswani *et al.* 2017). MSA is the extension of self-attention layer with multiple self-attention operations, seen in Eq. (14). LN is a normalized layer combined with the residual layer (Wang *et al*. 2019, Baevski and Auli 2018). MLP consists of two

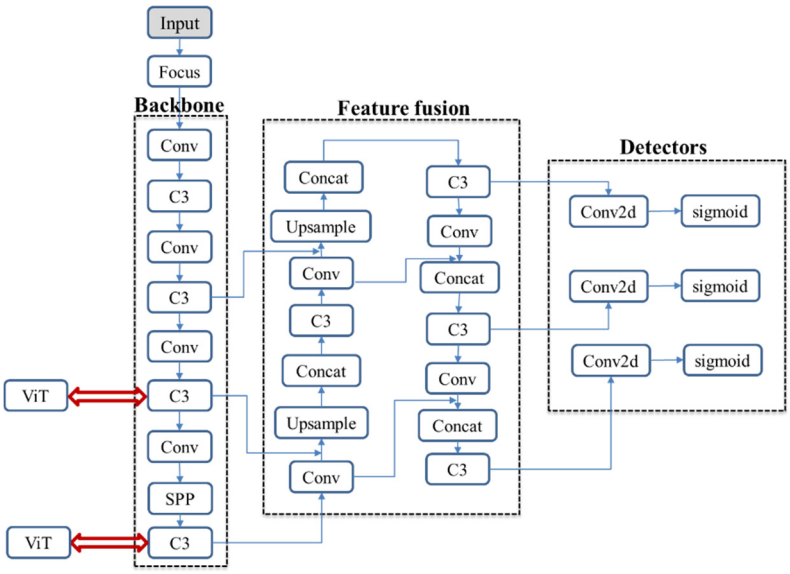


Fig. 7 The network architecture of improved YOLOv5s algorithms

layers with the nonlinear GELU and the module is seen in Eq. (15). The residual connection is adopted, including $z_{\ell-1}$ in Eq. (14) and $z_{\ell}^{'}$ in Eq. (15).

$$z_0 = \left[x_{\text{class}};\ x_p^1 E;\ \ x_p^2 E;\cdots;\ \ x_p^N E\right] + E_{\text{pos}},\quad E \in \Re^{(P^2 \bullet C)\times D},\quad E_{\text{pos}} \in \Re^{(N+1)\times D} \tag{13}$$

$$z_{\ell}^{'} = MSA\left(LN(z_{\ell-1})\right) + z_{\ell-1},\quad \ell = 1 \ldots L \tag{14}$$

$$z_{\ell} = MLP\left(LN\left(z_{\ell}^{'}\right)\right) + z_{\ell}^{'},\quad \ell = 1 \ldots L \tag{15}$$

$$y = LN(z_L^0) \tag{16}$$

where z_0 is the image patches sequence, E is the patch embedding of z_0, E_{pos} is the position embedding of z_0, LN is the *LN* module, MSA is the MSA module, MLP is the *MLP* module, and ℓ is the cycle number of MSA and *MLP* module.

4.2 Precision improvement by vision transformer

For small object detection tasks of bolts, it is essential to improve the recognition ability of spatial distributed features. YOLOv5s is a small model of the YOLOv5 series. In the backbone of YOLOv5s, the C3 module consists of the convolution layer, the normalized layer, the LeakyRelu activation layer, and multiple residual structure layers. The C3 module adopts convolution by sliding window operation without considering the weighting of small targets, and does not suppress useless information in different feature channels. Therefore, ViT is adopted to replace the C3 module, and the MSA mechanism could enhance the extraction of spatial distributed features.

Fig. 7 shows the network architecture of improved YOLOv5s algorithms. In the backbone network, the third and fourth C3 module are replaced by ViT.

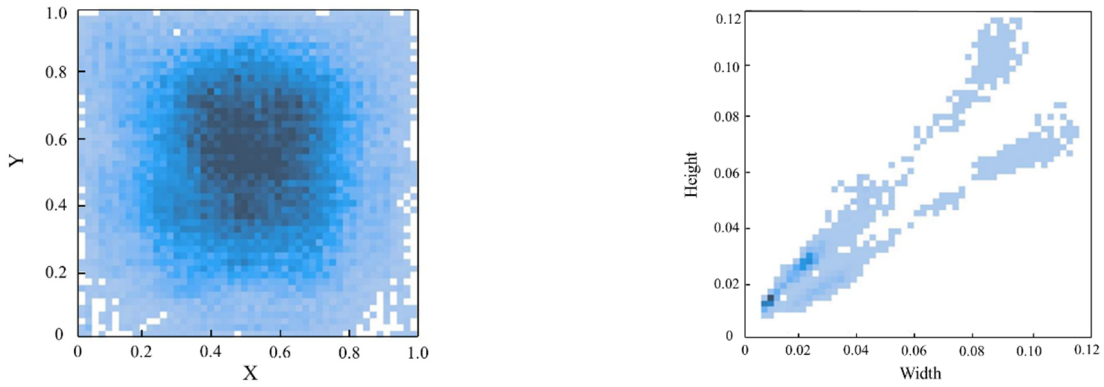
4.3 Dataset preparation and training

To train the network, this study established a labeled dataset of bolt damage. The dataset was collected from fabricated steel bolted joints for laboratory test and joints of bridges by unmanned aerial vehicle (UAV) photography. Image resolutions are 4608 × 3456, 5184 × 3888, respectively. The dataset was labeled for three types of bolt states, namely, clean bolt, corroded bolt and bolt missing. To expand the dataset, data augmentation methods were adopted, including image rotation, shifting, flipping, mirror, zoom, Gaussian blur, etc. In general, the dataset contains 830 images of bolted joints, with 61045 bounding boxes of targeted bolts labeled. The dataset was randomly divided into the training, validating and testing dataset with the proportions of 80%, 10% and 10%, respectively.

Transfer learning transmits the knowledge of similar task, which saves training time and realize high precision with fine tuning for new task (Weiss *et al*. 2016). YOLOv5s models are pre-trained on the COCO dataset, which is open source (Caesar *et al*. 2018). Not consisting of images with bolts, the COCO dataset contains 200,000 images with 80 object classes and 500,000 labeled bounding boxes. The pre-trained weight of YOLOv5s was initially transferred for improved YOLOv5s algorithm.

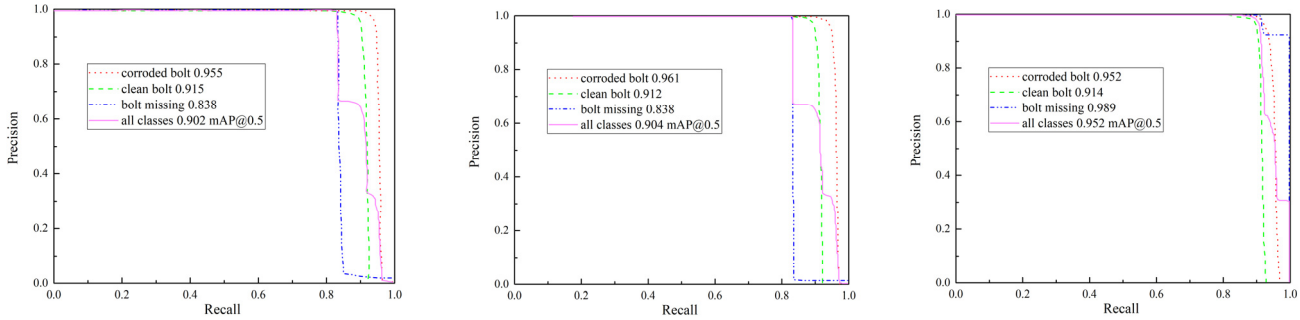
The optimizer of stochastic gradient descent with momentum (SGDM) was adopted, and the momentum was set to 0.937 with an initial learning rate of 0.01. To prevent overfitting of model, the weight decay of regularization was set to 0.0005. The training process was proceeded on an NVIDIA GeForce RTX 3090Ti GPU with 24 GB memory, and the computing power reached 35.6 TFLOPS.

Fig. 8 illustrates the distribution of dataset labels. Fig. 9(a) is the dimensionless label location distribution, and the bounding boxes of the bolts are mainly located near the image center. Fig. 9(b) is the dimensionless boundingbox size distribution. As seen in the figure, the ratio of width to height of the bounding boxes is near the slope line of 45 degree. The perspective angle of the camera to steel bolted



(a) Dimensionless label location distribution (b) Dimensionless boundingbox size distribution

Fig. 8 The distribution of dataset labels



(a) Precision-recall curve of YOLOv5s (b) Precision-recall curve of YOLOv5s with one ViT (c) Precision-recall curve of YOLOv5s with two ViTs

Fig. 9 Precision improvement comparison among YOLOv5s, one ViT and two ViTs

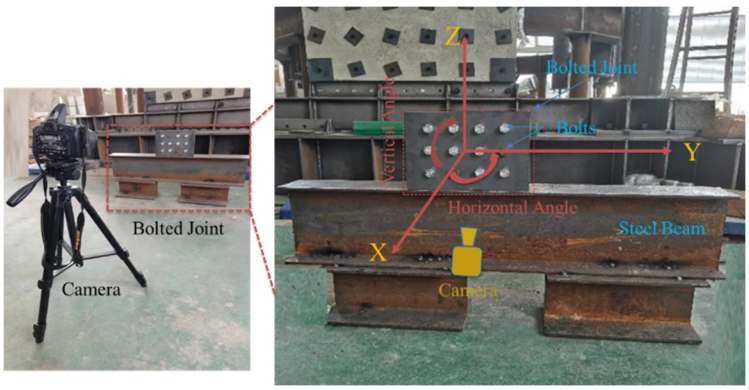


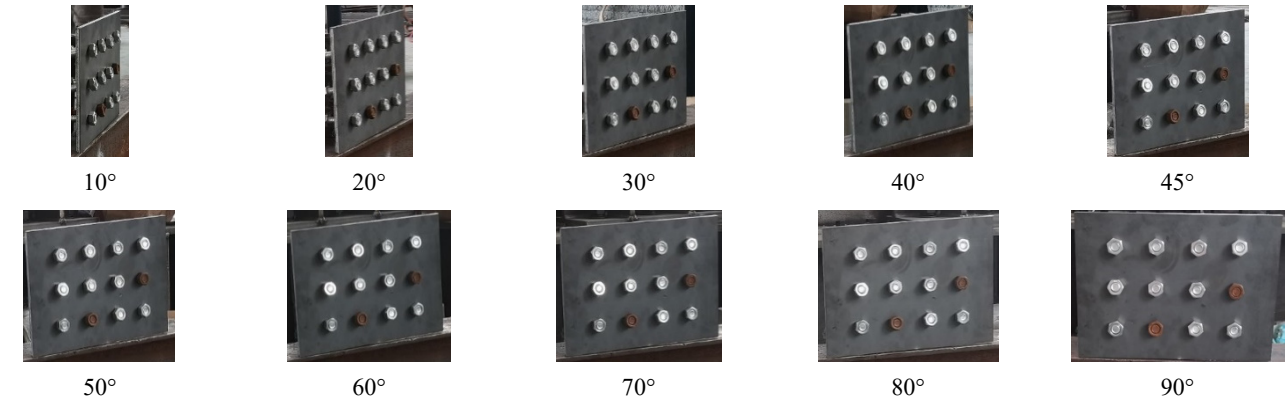
Fig. 10 Schematic diagram of steel bolted joint test in laboratory

joints has an effect on the ratio.

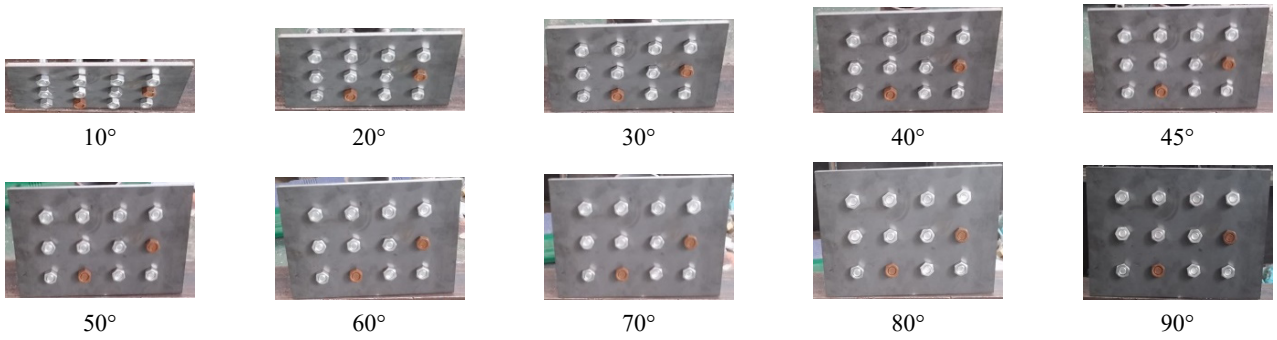
Fig. 9 presents the precision improvement comparison among YOLOv5s, YOLOv5s with one ViT and YOLOv5s with two ViTs. Fig. 10(a)~10(c) are precision-recall curve of the above models, respectively. For YOLOv5s with one ViT, the fourth C3 module is replaced by ViT. The confidence threshold of detection was set to 0.5. As for corroded bolt in the figure, the AP of YOLOv5s was increased from 0.955 to 0.961 by one ViT, while the AP of the two ViTs decreased slightly to 0.952. The AP difference is within 2 percent, which is regarded as the random distribution of validating and testing dataset. In terms of bolt missing, the AP of two ViTs has been significantly improved, from 0.838 to 0.989. In the descending section of the curve in the figure, precision is inversely proportional to recall. The larger the area covered by the curve, the larger the AP of the model. Generally, one ViT improves mean average precision (mAP) of YOLOv5s slightly, while two ViTs improves mAP from 0.902 to 0.952 by 5.0 percent, evidencing the feature extraction capability of two ViTs.

4.4 Laboratory testing

In this study, a steel bolted joint was fabricated and



(a) Horizontal perspective



(b) Vertical perspective

Fig. 11 Steel bolted joint from different perspective with the distance of 1.5 meters

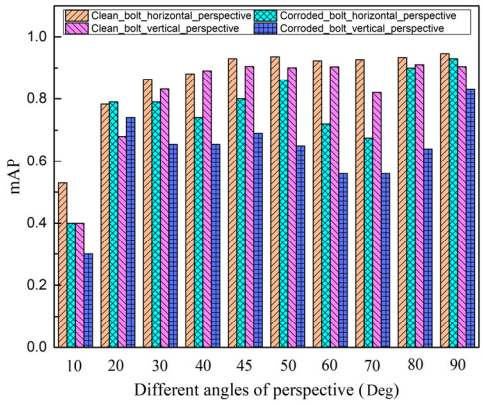


Fig. 12 The mAP of improved YOLOv5s from different perspective angles

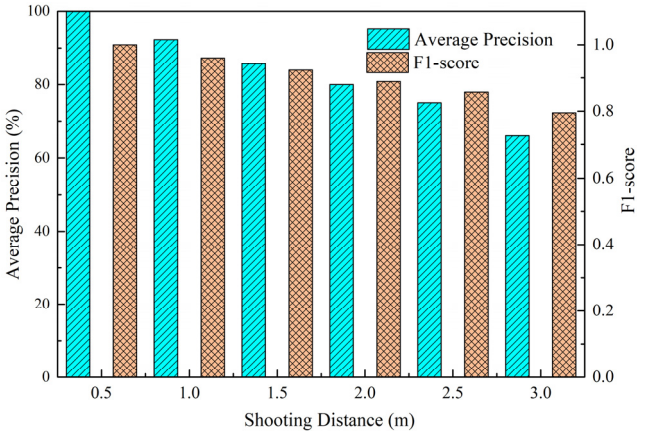


Fig. 13 Steel bolted joint from different distance

tested in laboratory to evaluate the applicability of the proposed framework. Fig. 10 presents the schematic diagram of steel bolted joint test in laboratory. The plate size of bolted joint is 250 mm width × 200 mm height × 8 mm thickness, containing 12 fastened bolts (20 mm diameter) at the spacing of 50 mm. Figs. 11(a) and (b) illustrate steel bolted joint from horizontal and vertical perspective with the distance of 1.5 meters, respectively. The perspective angle of bolted joints ranges from 0° to 90°.

The damage detection accuracy of improved YOLOv5s was compared in different scenes, including horizontal and vertical perspective, and the shooting distance. Fig. 12 shows the mAP of improved YOLOv5s from different perspective angles with the shooting distance of 1.5 meters. Two states of the bolts were detected, including clean bolt and corroded bolt. As observed in Fig. 12, in general, with the increase of perspective angles, the mAP of the two states showed an upward trend. The state of clean bolt maintained the mAP over 90% under a perspective angle of 45° to 90°. Considering the calculation precision of bolt loosening angle, it is recommended to maintain the horizontal and vertical perspective angle between 45° and 90°.

Fig. 13 shows the average precision and F1-score of improved YOLOv5s at different shooting distances under a

horizontal and vertical perspective angle of 90°. The shooting distance ranged from 0.5 m to 3.0 m, and the step size was 0.5 m. As depicted in Fig. 13, with the increase of the shooting distance, the average precision and F1-score value decreased steadily. With the shooting distance between 0.5 meter and 2 meters, the average precision was above 80%. The F1-score takes into account both the precision and the recall rate, exceeding 0.8 with the distance between 0 meter and 2.5 meters.

5. Bolt looseness angle detection

5.1 Determination of bolt looseness angle

Most research on bolt loosening mechanics are based on the bolt loosening test device diagram proposed by Junker (1969). Jiang *et al*. (2003) conducted the relationship between the pre-tightening torque and the rotation angle of the loosened bolts. As the pre-tightening torque of the bolt remains or decreases to near 85%, the bolt rotation angle is within 2 degree. As the pre-tightening torque decreases from near 85% to near 25%, the bolt rotation angle increases from 2 degree to 20 degree. Afterwards, the pre-tightening torque decreases sharply to near zero with the increase of rotation angle. Therefore, based on deep learning and computer vision, it is feasible to diagnose bolt loosening by detecting the rotation angle of the nut.

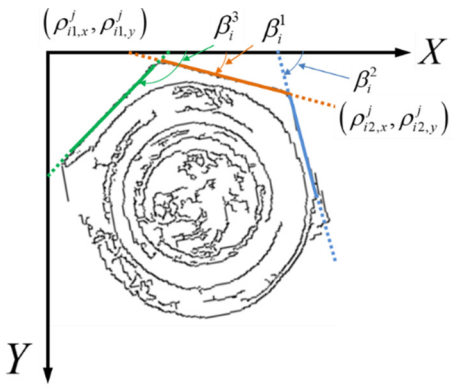


Fig. 14 Schematic diagram of bolt rotation angle calculation

To quantify the rotation angle, the images of steel bolted joints in the initial state are selected as the reference image. Firstly, the improved YOLOv5s is adopted to detect the bolt state. The perspective correction of the joint image is carried out by perspective transformation matrix, and the images of individual bolts are cropped out respectively. After denoising and gray level transformation, edges of individual nuts are obtained through Canny edge detector. Afterwards, the LSD algorithm is used to detect the straight line of the nut edges for rotation angle calculation.

Fig. 14 presents schematic diagram of bolt rotation angle calculation. Assume that the horizontal X axis is the reference axis and that the nut rotates clockwise as a positive direction. The LSD algorithm is used to obtain the endpoint coordinates of the detected nuts edges. Combined with the reference axis, the line segment slope is calculated to obtain the rotation angle.

$$\beta_i^j = arctan\left(\frac{\rho^j_{i1,y} - \rho^j_{i2,y}}{\rho^j_{i1,x} - \rho^j_{i2,x}}\right) \tag{17}$$

$$\alpha_i = \frac{1}{n}\sum_{j=1}^{n}\alpha_i^j = \frac{1}{n}\sum_{j=1}^{n} mod\left(\frac{\beta_i^j}{\pi/3}\right) \tag{18}$$

where β_i^j denotes the calculation angle of j^{th} edge for the i^{th} bolt, and $(\rho^j_{i1,x}, \rho^j_{i1,y})$ and $(\rho^j_{i2,x}, \rho^j_{i2,y})$ denote the terminal coordinates of the line segment, respectively. α_i denotes the average rotation angle of the i^{th} bolt, α_i^j denotes the rotation angle of j^{th} edge for the i^{th} bolt. $mod(\cdot)$ is the remainder operation of β_i^j, n denotes the number of line segments identified for single nut.

The rotation angle α_i generally follows a normal distribution. Assume that the mean and standard deviation of α_i are μ_{α_i} and σ_{α_i}, respectively.

According to the three-sigma criterion, the probability of α_i falling in the interval $(\mu_{\alpha_i} - 3\sigma_{\alpha_i}, \mu_{\alpha_i} + 3\sigma_{\alpha_i})$ is 0.997 (Pukelsheim 1994). If Eq. (19) is satisfied, the bolt is determined as loosened.

$$|\alpha_i| > \mu_{\alpha_i} + 3\sigma_{\alpha_i} \tag{19}$$

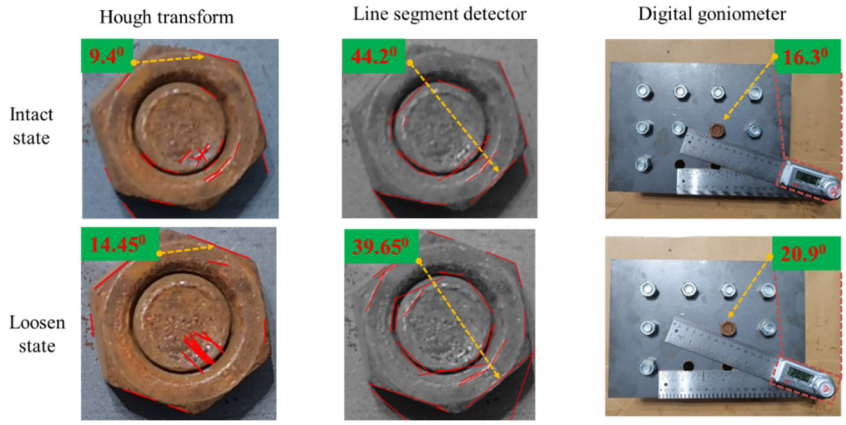


Fig. 15 Measuring results of rotation angles

Table 1 Performance of algorithms compared with digital goniometer

Case	Hough transform	Line segment detector	Digital goniometer
Intact state angle	9.4°	44.2°	16.3°
Loosen state angle	14.45°	39.65°	20.9°
Rotation angle	5.05°	4.550	4.6°
Calculation Error	8.91%	1.09%	/

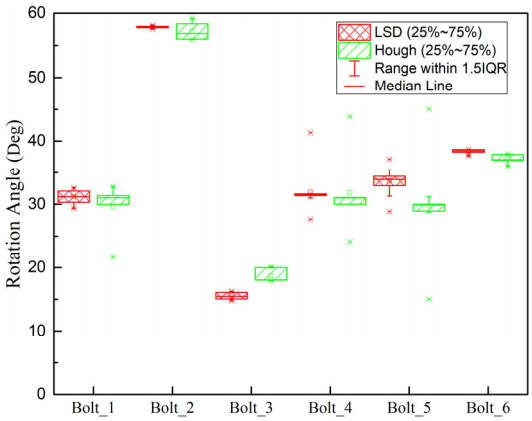


Fig. 16 Box plots of bolt rotation angles distribution

5.2 Comparison with Hough transform algorithm

To verify the accuracy of the LSD algorithm, the image of the bolts with intact and loosen state were captured from the steel bolted joint. A digital goniometer was used to measure the rotation angle in two states, as shown in Fig. 15. Meanwhile, the LSD algorithm and the Hough transform were adopted to detect the nut edge, and the line segment with highest confidence value was taken for rotation angle calculation.

Table 1 shows the performance of algorithms compared with digital goniometer. The rotation angle measured by the digital goniometer was 4.60°, which was considered as ground truth. The rotation angle calculated by the LSD algorithm was 4.55°, and the Hough transform was 5.05°. The calculation error of the LSD algorithm was 1.09%, much lower than that of the Hough transform 8.91%. Therefore, the threshold of the LSD algorithm to determine bolt loosening is smaller than that of the Hough transform, and LSD algorithm is more accurate for loosening detection. In the aspect of rotation angle calculation value, the fluctuation range of Hough transform is higher than that of LSD algorithm.

To further verify the applicability of the LSD algorithm, six bolts of the joint were selected to analyze the error distribution of rotation angles. Fig. 16 shows box plots of bolt rotation angles distribution. As seen in the figure, interquartile range (IQR) and the range within 1.5 IQR of bolt rotation angles by the LSD algorithm were smaller than that of the Hough transform. There existed data outliers in the figure, and the fluctuation range could be found. Both algorithms were affected by various lighting condition and other environmental factors. Basically, the result revealed that the LSD algorithm is more robust than the Hough transform in the rotation angle calculation. Further, the bolt loosening angle threshold could be determined according to the three-sigma criterion.

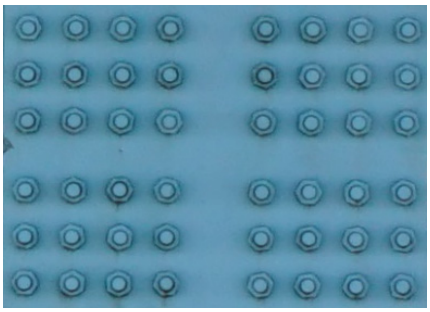
5.3 Field test on a real bridge

For engineering practice, Nanjing Dashengguan Yangtze River Bridge, the control project of Beijing-Shanghai high-speed railway in China, was selected in this study. The bridge is a steel truss arch bridge built in 2012. According to design drawings, the bridge consists of nearly 10 million high-strength bolts. To ensure the structure safety and service life, the inspection of steel bolted joint is essential for routine maintenance. Fig. 17 shows Dashengguan Yangtze River Bridge. DJI M300 UAV was used to conduct flight inspection and obtain in-situ images, and the UAV was about 30 meters away from the bridge. The truss arch foot joint was selected, where the detected area was labeled by the red dashed box (see Fig. 17(b)).

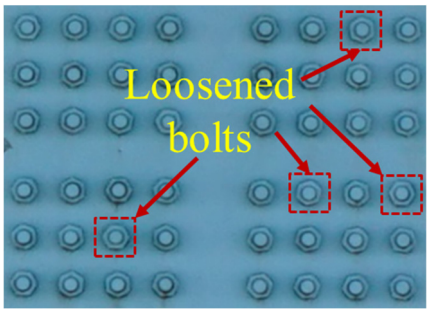
(a) UAV inspection

(b) Location of steel bolted joint detected

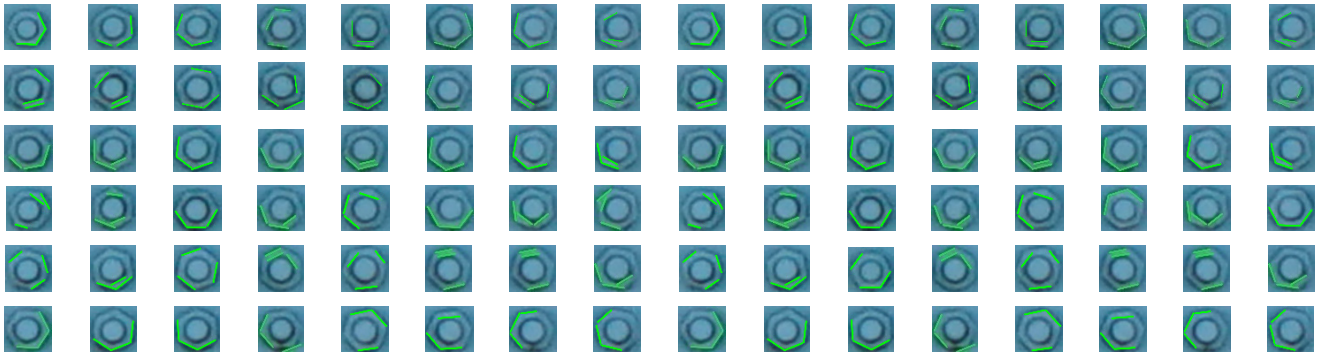
Fig. 17 Dashengguan Yangtze River Bridge



(a) The initial joint image for reference



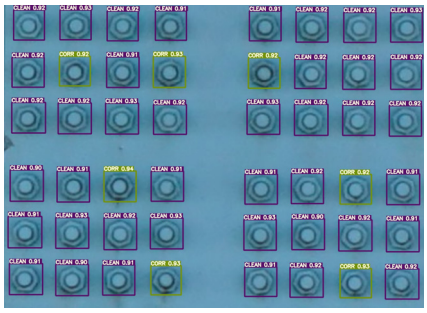
(b) The synthetic image for comparison



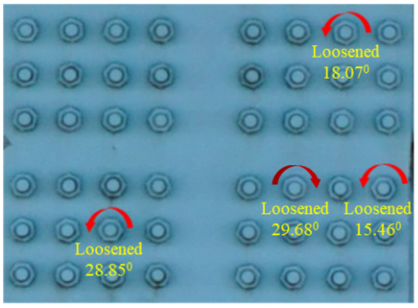
(c) Line detection of the initial joint image

(d) Line detection of the synthetic image

Fig. 18 Test result of the nut edge lines



(a) Corroded bolt detection results



(b) Bolt looseness detection results

Fig. 19 Detection result of bolts including bolt looseness

Generally, the bolt looseness monitoring is a long-term process and the bolt looseness event is a random phenomenon to capture. Thus, synthetic images were used as an auxiliary means for testing. As seen in Fig. 18(a), based on the initial joint image for reference, four bolts are chosen to simulate with a certain rotation angle in the detected area. As shown in Fig. 18(b), the synthetic image for comparison consists of 4 loosened bolts and the other 44 intact bolts. The synthetic image was adopted for calculating the bolt rotation angle. Figs. 18(c) and (d) present the line detection result of the initial joint image and the synthetic image, respectively. For line detection of nut edges, three edges with the highest confidence value were generally selected, and most of the lines detected are in good agreement with nut edges.

Fig. 19 shows the detection result of bolts including bolt looseness. In Fig. 20(a), state of clean bolt and corroded bolt were corrected detected with high confidence. According to the three-sigma criterion (Pukelsheim 1994) and bolt rotation angle (Jiang *et al.* 2003), the angle threshold for loosening determination was 3.59°. Four synthetic loosen bolts were accurately identified, and the rotation angles were 18.07°, 28.85°, 29.68° and 15.46°, respectively. Thus, the rotation angles exceeded the angle threshold and the four bolts could be determined as loosened bolts.

6. Conclusions

In this study, a framework based on improved YOLOv5s network and the LSD algorithm was proposed for corroded and loosened bolts detection of steel bolted joints. Firstly, ViT is introduced to replace the third and fourth C3 module of YOLOv5s network. Secondly, through improved YOLOv5s network, bolt state including corroded bolt, bolt

missing and clean bolt are identified, and the bolts for checking looseness are cropped as individual images respectively. Afterwards, perspective transformation was designed to correct the camera vision angle. The edges of single bolt were extracted by canny detector, and the LSD algorithm was introduced to detect the straight lines of bolt edges. Sequentially, the edge rotation angles of bolts were calculated and the criterion for judging bolt looseness was determined. Finally, the proposed framework was verified on the steel bolted joint fabricated in the laboratory and applied to the cross-river steel arch bridge. The proposed method leads to the following conclusions:

- ViT increases the attention weights of feature channels and the feature extraction capability. The mAP of detection model by YOLOv5s with ViTs was increased from 0.902 to 0.952.
- For the bolt loosening angle calculation, the LSD algorithm is more robust than the Hough transform. The test error of the LSD algorithm is 1.09%, less than 8.91% of the Hough transform.
- The recommended camera shooting angle should be between 45 and 90 degrees both horizontally and vertically.
- The proposed framework can realize corroded bolt detection and bolt loosening angle calculation in good performance.

Nonetheless, the proposed detection framework still have some limitations worth studying in the future. The angle threshold for bolt loosening determination at different camera resolutions needs to be re-calibrated, which reduces the generality of the method. To distinguish paint discoloration or improper surface coating from the corrosion of the steel structure material is worthy of further study in the future. In addition, the real-time performance of the proposed framework for real bridge application is worthy of optimization.

Acknowledgments

The authors greatly acknowledge the financial support from the National Natural Science Foundation of China (51978155, 52108274) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant number: SJCX21_0056). The authors also would like to thank Postdoctoral Fellow Yiming Zhang of The Hong Kong Polytechnic University, doctoral candidate Hui Gao, doctoral candidate Ruijun Liang and doctoral candidate Zhijie Yuan of Southeast University for paper writing and computing. Finally, contributions by the anonymous reviewers are also highly appreciated.

References

Atha, D.J. and Jahanshahi, M.R. (2018), "Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection", *Struct. Health Monitor*., **17**(5), 1110-1128. https://doi.org/10.1177/1475921717737051

Baevski, A. and Auli, M. (2018), "Adaptive input representations for neural language modeling", arXiv preprint arXiv: 1809.10853. https://doi.org/10.48550/arXiv.1809.10853

Blachowski, B., Swiercz, A. and Pnevmatikos, N. (2015), "Experimental verification of damage location techniques for frame structures assembled using bolted connections", *Proceedings of the 5th International Conference on Computational Methods in Structural Dynamics and Earthquake Engineering*, pp. 05-25.

Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020), "Yolov4: Optimal speed and accuracy of object detection", arXiv preprint arXiv:2004.10934. https://doi.org/10.48550/arXiv.2004.10934

Caesar, H., Uijlings, J. and Ferrari, V. (2018), "Coco-stuff: Thing and stuff classes in context", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1209-1218.

Cha, Y.J., You, K. and Choi, W. (2016), "Vision-based detection of loosened bolts using the Hough transform and support vector machines", *Automat. Constr*., **71**, 181-188. https://doi.org/10.1016/j.autcon.2016.06.008

Cha, Y.J., Choi, W., Suh, G., Mahmoudkhani, S. and Büyüköztürk, O. (2018), "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types", *Comput-Aided. Civil Infrastr., Eng*., **33**(9), 731-747. https://doi.org/10.1111/mice.12334

Chaki, S. and Bourse, G. (2009), "Stress level measurement in prestressed steel strands using acoustoelastic effect", *Exp. Mech*., **49**(5), 673-681. https://doi.org/10.1007/s11340-008-9174-9

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018), "Bert: Pre-training of deep bidirectional transformers for language understanding", arXiv preprint arXiv:1810.04805. https://doi.org/10.48550/arXiv.1810.04805

Ding, L. and Goshtasby, A. (2001), "On the Canny edge detector", *Pattern Recognit*., **34**(3), 721-725. https://doi.org/10.1016/S0031-3203(00)00023-6

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G. and Gelly, S. (2020), "An image is worth 16x16 words: Transformers for image recognition at scale", arXiv preprint arXiv:2010.11929. https://doi.org/10.48550/arXiv.2010.11929

Forkan, A.R.M., Kang, Y.B., Jayaraman, P.P., Liao, K., Kaul, R., Morgan, G., Ranjan, R. and Sinha, S. (2022), "CorrDetector: A framework for structural corrosion detection from drone images using ensemble deep learning", *Expert Syst. Appl*., **193**, 116461. https://doi.org/10.1016/j.eswa.2021.116461

Ge, Z., Liu, S., Wang, F., Li, Z. and Sun, J. (2021), "Yolox: Exceeding yolo series in 2021", arXiv preprint arXiv:2107.08430. https://doi.org/10.48550/arXiv.2107.08430

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, T. (2018), "Recent advances in convolutional neural networks", *Pattern Recogn*., **77**, 354-377. https://doi.org/10.1016/j.patcog.2017.10.013

Hoskere, V., Narazaki, Y., Hoang, T. and Spencer Jr, B. (2018), "Vision-based structural inspection using multiscale deep convolutional neural networks", arXiv preprint arXiv:1805.01055. https://doi.org/10.48550/arXiv.1805.01055

Huynh, T.C. (2021), "Vision-based autonomous bolt-looseness detection method for splice connections: Design, lab-scale evaluation, and field application", *Automat. Constr*., **124**, 103591. https://doi.org/10.1016/j.autcon.2021.103591

Huynh, T.C., Park, J.H., Jung, H.-J. and Kim, J.T. (2019), "Quasi-autonomous bolt-loosening detection method using vision-based deep learning and image processing", *Automat. Constr*., **105**, 102844. https://doi.org/10.1016/j.autcon.2019.102844

Jiang, Y., Zhang, M. and Lee, C.-H. (2003), "A study of early stage self-loosening of bolted joints", *J. Mech. Des*., **125**(3), 518-526. https://doi.org/10.1115/1.1586936

Junker, G.H. (1969), "New criteria for self-loosening of fasteners under vibration", *Sae Transact*., **78**, 314-335. https://www.jstor.org/stable/44563013

Mezirow, J. (1978), "Perspective transformation", Adult education, **28**(2), 100-110. https://doi.org/10.1177/074171367802800202

Muthalagu, R., Bolimera, A. and Kalaichelvi, V. (2020), "Lane detection technique based on perspective transformation and histogram analysis for self-driving cars", *Comput. Electr. Eng*., **85**, 106653. https://doi.org/10.1016/j.compeleceng.2020.106653

Ni, Y.H., Mao, J.X., Wang, H., Xi, Z. and Xu, Y.F. (2023), "Toward high-precision crack detection in concrete bridges using deep learning", *J. Perform. Constr. Facil*., **37**(3), 04023017. https://doi.org/10.1061/JPCFEV.CFENG-4275

Nikravesh, S.M.Y. and Goudarzi, M. (2017), "A review paper on looseness detection methods in bolted structures", *Lat. Am. J. Solids Struct*., **14**, 2153-2176. https://doi.org/10.1590/1679-78254231

Pan, X. and Yang, T. (2021), "Image-based monitoring of bolt loosening through deep-learning-based integrated detection and tracking", *Comput-Aided. Civil Infrastr. Eng*., **37**(10), 1207-1222. https://doi.org/10.1111/mice.12797

Park, G., Sohn, H., Farrar, C.R. and Inman, D.J. (2003), "Overview of piezoelectric impedance-based health monitoring and path forward", *Shock Vib. Digest*, **35**(6), 451-464. https://doi.org/10.48550/arXiv.1805.01055

Pham, H.C., Ta, Q.B., Kim, J.T., Ho, D.D., Tran, X.L. and Huynh, T.C. (2020), "Bolt-loosening monitoring framework using an image-based deep learning and graphical model", *Sensors*, **20**(12), 3382. https://doi.org/ 10.3390/s20123382

Pidaparti, R.M. (2007), "Structural corrosion health assessment using computational intelligence methods", *Struct. Health Monitor*., **6**(3), 245-259. https://doi.org/10.1177/1475921707081975

Pukelsheim, F. (1994), "The three sigma rule", *Am. Stat*., **48**(2), 88-91. https://doi.org/10.1080/00031305.1994.10476030

Redmon, J. and Farhadi, A. (2017), "YOLO9000: better, faster, stronger", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263-7271.

Redmon, J. and Farhadi, A. (2018), "Yolov3: An incremental improvement", arXiv preprint arXiv:1804.02767. https://doi.org/10.48550/arXiv.1804.02767

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016), "You only look once: Unified, real-time object detection", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788.

Ren, S., He, K., Girshick, R. and Sun, J. (2015), "Faster R-CNN: Towards real-time object detection with region proposal networks", *IEEE Trans. Pattern Anal. Mach. Intell*., **39**(6), 1137-1149. https://doi.org/ 10.1109/TPAMI.2016.2577031

Spencer Jr, B.F., Hoskere, V. and Narazaki, Y. (2019), "Advances in computer vision-based civil infrastructure inspection and monitoring", *Eng.*, **5**(2), 199-222. https://doi.org/10.1016/j.eng.2018.11.030

Sun, L., Shang, Z., Xia, Y., Bhowmick, S. and Nagarajaiah, S. (2020), "Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection", *J. Struct. Eng*., **146**(5), 04020073. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002535

Ta, Q.B. and Kim, J.T. (2020), "Monitoring of corroded and loosened bolts in steel structures via deep learning and Hough transforms", *Sensors*, **20**(23), 6888. https://doi.org/10.3390/s20236888

Ta, Q.-B., Huynh, T.C., Pham, Q.Q. and Kim, J.T. (2022), "Corroded bolt identification using mask region-based deep learning trained on synthesized data", *Sensors*, **22**(9), 3340. https://doi.org/10.3390/s22093340

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017), "Attention is all you need", *Adv. Neural Inform. Process*., **30**.

Von Gioi, R.G., Jakubowicz, J., Morel, J.M. and Randall, G. (2008), "LSD: A fast line segment detector with a false detection control", *IEEE Trans. Pattern Anal. Mach. Intell*., **32**(4), 722-732. https://doi.org/10.1109/TPAMI.2008.300

Wang, T., Song, G., Wang, Z. and Li, Y. (2013a), "Proof-of-concept study of monitoring bolt connection status using a piezoelectric based active sensing method", *Smart Mater. Struct*., **22**(8), 087001. https://doi.org/10.1088/0964-1726/22/8/087001

Wang, T., Song, G., Liu, S., Li, Y. and Xiao, H. (2013b), "Review of bolted connection monitoring", *Int. J. Distrib. Sens. Netw*., **9**(12), 871213. https://doi.org/10.1155/2013/871213

Wang, C., Wang, N., Ho, S.C., Chen, X. and Song, G. (2019), "Design of a new vision-based method for the bolts looseness detection in flange connections", *IEEE Trans. Ind. Electron*., **67**(2), 1366-1375. https://doi.org/10.1109/TIE.2019.2899555

Weiss, K., Khoshgoftaar, T.M. and Wang, D. (2016), "A survey of transfer learning", *J. Big Data*, **3**(1), 1-40. https://doi.org/10.1186/s40537-016-0043-6

Xu, J., Gui, C. and Han, Q. (2020), "Recognition of rust grade and rust ratio of steel structures based on ensembled convolutional neural network", *Comput-Aided. Civil Infrastr. Eng*., **35**(10), 1160-1174. https://doi.org/10.1111/mice.12563

Yang, J. and Chang, F.K. (2006), "Detection of bolt loosening in C-C composite thermal protection panels: I. Diagnostic principle", *Smart Mater. Struct*., **15**(2), 581. https://doi.org/10.1088/0964-1726/15/2/041

Ye, X.W., Dong, C.Z. and Liu, T. (2016), "A review of machine vision-based structural health monitoring: methodologies and applications", *J. Sensors*, 2016. https://doi.org/10.1155/2016/7103039

Zhang, Y., Sun, X., Loh, K.J., Su, W., Xue, Z. and Zhao, X. (2020), "Autonomous bolt loosening detection using deep learning", *Struct. Health Monitor*., **19**(1), 105-122. https://doi.org/10.1177/1475921719837509

BS