Smart Structures and Systems, *Vol. 31, No. 5 (2023) 469-483*
https://doi.org/10.12989/sss.2023.31.5.469

Dynamic characteristics monitoring of wind turbine blades based on improved YOLOv5 deep learning model

W.H. Zhao [1a], W.R. Li*[1,2,3], M.H. Yang [1b], N. Hong [1,2,3c] and Y.F. Du [1,2,3d]

[1] *Institution of Earthquake Protection and Disaster Mitigation, Lanzhou University of Technology, Lanzhou 730050, China*
[2] *International Research Base on Seismic Mitigation and Isolation of GANSU Province, Lanzhou University of Technology, Lanzhou 730050, China*
[3] *Disaster Prevention and Mitigation Engineering Research Center of Western Civil Engineering, Lanzhou University of Technology, Lanzhou 730050, China*

(*Received August 3, 2022, Revised October 29, 2022, Accepted March 19, 2023*)

Abstract. The dynamic characteristics of wind turbine blades are usually monitored by contact sensors with the disadvantages of high cost, difficult installation, easy damage to the structure, and difficult signal transmission. In view of the above problems, based on computer vision technology and the improved YOLOv5 (You Only Look Once v5) deep learning model, a non-contact dynamic characteristic monitoring method for wind turbine blade is proposed. First, the original YOLOv5l model of the CSP (Cross Stage Partial) structure is improved by introducing the CSP2_2 structure, which reduce the number of residual components to better the network training speed. On this basis, combined with the Deep sort algorithm, the accuracy of structural displacement monitoring is mended. Secondly, for the disadvantage that the deep learning sample dataset is difficult to collect, the blender software is used to model the wind turbine structure with conditions, illuminations and other practical engineering similar environments changed. In addition, incorporated with the image expansion technology, a modeling-based dataset augmentation method is proposed. Finally, the feasibility of the proposed algorithm is verified by experiments followed by the analytical procedure about the influence of YOLOv5 models, lighting conditions and angles on the recognition results. The results show that the improved YOLOv5 deep learning model not only perform well compared with many other YOLOv5 models, but also has high accuracy in vibration monitoring in different environments. The method can accurately identify the dynamic characteristics of wind turbine blades, and therefore can provide a reference for evaluating the condition of wind turbine blades.

Keywords: deep learning; dynamic characteristics; improved YOLOv5; structural health monitoring; wind turbine blades

1. Introduction

In the past few decades, people used wind energy instead of fossil energy to reduce carbon dioxide emissions. Wind energy can be seen increasingly wide application because of its clean use, low cost, and low impact on the environment when using wind turbines (Hernandez-Estrada *et al*. 2021). As the demand of wind energy continues to grow, larger wind turbines are required to increase productivity and power generation stability (Kaewniam *et al*. 2022, Sun *et al*. 2022). However, larger blade rotor diameter is indispensable in large wind turbines, which increase the flexibility of the blade but damage easily during operation, such as tower collapse and blade fracture under severe vibration. All these lead to structural safety problems under various environmental factors and operating conditions (Lin *et al*. 2020, Xu *et al*. 2021). As the main components of wind turbine to capture wind energy, blades damaging is accompanied with major safety accidents and serious economic losses. Therefore, accurate monitoring of blade status is of great significance to the operation, maintenance and safety of wind turbines.

Traditional wind turbine blade monitoring is often carried out by using ultrasonic, thermal imaging, telescopes and other equipment (Jasinien *et al*. 2009, Amenabar *et al*. 2011, and Beganovic and Söffker 2016). However, the actual monitoring results are not ideal due to the expensive equipment and limited engineering experience of the inspectors. Contact sensors are mainly used for dynamic characteristic monitoring in civil engineering monitoring, but the use of sensors has the disadvantage of fixing the sensors to the surface of the structure, which not only cause damage to the structure but also substantially modify the structural mode due to the effect of mass loading (Felipe-Sesé and Díaz 2018), resulting to the inaccurate monitoring results. Contact-based wired sensors involve labor-intensive installation processes and seek extensive maintenance, making it difficult to achieve practical dynamic characteristic monitoring. At present, a large number of non-contact sensors are frequently used in structural health monitoring. However, the complexities of data

*Corresponding author, Ph.D., Professor,
E-mail: ce_wrli@lut.edu.cn
[a] Ph.D. Candidate
[b] Master Student
[c] Ph.D.
[d] Ph.D., Professor

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

transmission, time synchronization, and power consumption make the data acquisition for that challengeable, so that non-contact sensors are not suitable for dynamic characteristic monitoring of structures (Sony *et al*. 2019). Hence, there is an urgent need of method for the structural health monitoring of large-scale structures, which is low-cost and high-accuracy.

Due to the application advantages of computer vision technology, such as non-contact, long-distance, fast, low-cost, low labor, and low disturbance to the daily operation of structures, it is promising to consider computer vision-structural health monitoring as a complement to the tranditional one (Onat and Gul 2018, Ehrhardt *et al*. 2017 and Kuddus *et al*. 2019), which has received extensive attention in research and practice field of structural health monitoring. A large number of scholars have applied computer vision monitoring to the health monitoring of actual structures (Spencer *et al*. 2019, Ye *et al*. 2016, Li *et al*. 2019, and Zhou *et al*. 2021). Feng and Feng (2017) used a template matching algorithm to perform a vision-based dense full-field displacement measurement on a simply supported beam, and then verified it in practice of train passing the bridge, whereas the monitoring effect was not good for complex backgrounds. Dong *et al*. (2019) proposed a computer vision-based structural dynamic monitoring method using the optical flow method, which was tested and verified in the large stadium grandstand following by the modal identification, and it turned out that the monitoring results is not accurate under the conditions of illumination changes. Tian *et al*. (2019a) utilized a Gradient Hough Transform (GHT) method to calculate the multi-point dynamic displacement for each setup, identified the mode shapes and compliance matrices of the entire structure, and used a cantilever beam to verify the significance of the method. However, needing to set artificial signs in advance make GHT not suitable for health monitoring of large structures. Zhao *et al*. (2016) developed an APP that can monitor bridge displacement: using a color matching algorithm combined with a smartphone, D-viewer can perform static and dynamic tests on the bridge. However, due to the poor monitoring effect under light conditions and not considering the complex background, it is not conducive to long-term monitoring. Kalaitzakis *et al*. (2019) used digital image correlation (DIC) technology combined with unmanned aerial vehicles to carry out remote structural assessment of railway connections, but the need of spraying artificial markers and constant illumination make it limited in the application of large-scale structural health monitoring. The above vision-based structural health monitoring methods are all carried out by traditional algorithms, which cannot be accurately monitored under actual environmental conditions with complex background and changeable illumination condition. Therefore, a structural health monitoring method suitable for variable actual environmental conditions is required.

In recent years, with the rapid development of computer hardware and remarkable progress in computer vision, deep learning-based structural health monitoring methods have been gradually applied to practical structures. Given the accuracy of target tracking based on deep learning is better than that of traditional algorithms, it is gradually becoming the mainstream in the field of target tracking, which attract a lot of researchers dedicated on it (Lydon *et al*. 2019, Ye *et al*. 2019, and Tian *et al*. 2019b). Dong *et al*. (2020) used deep learning to validate the performance of the method through laboratory experiments and comparative studies on grandstand structures, and field validation on pedestrian bridges. Shao *et al*. (2021) used state-of-the-art deep learning-based key-point detection and matching algorithms to perform experimental tests of target-free measurements on steel cantilevers in the laboratory, and then evaluated the performance of the proposed vision-based 3D vibration displacement measurement method. Song *et al*. (2022) perfectly segmented the background through a deep-learned Fully Convolutional Network (FCN) and Conditional Random Field (CRF), followed by displacement measurements using DIC, and verified the feasibility of the method through experiments under different illuminations conditions. The above deep learning-based structural health monitoring methods not only have high accuracy but also not limited to the monitoring environment. However, it has not been widely used in practical monitoring due to the difficulty of making special datasets, the demanding computer hardware configuration, and the large time cost for training and verification. Hence, the deep learning-based health monitoring of wind turbine structures focuses on the detection of blade surface defects (Xu *et al*. 2019, Yu *et al*. 2020), while more traditional algorithms are used to monitor the dynamic characteristics of blade vibration (Khadka *et al*. 2020, Li *et al*. 2022). Therefore, it is necessary to find a method for monitoring the dynamic characteristics of wind turbine blades based on computer vision, which is suitable for monitoring the dynamic characteristics of blades in different environments.

This paper focuses on monitoring the dynamic characteristics of large wind turbine blades based on computer vision technology and deep learning models. In Section 1, the disadvantages of traditional monitoring methods and the limitations of existing visual algorithms for structural health monitoring are generally described. In Section 2, based on the existing YOLOv5l deep learning model, an improved one is proposed combined with the Deep sort algorithm, which accelerates the speed of network training and can better the accuracy of measure the structural displacement measuring results. In Section 3, the poor dataset of wind turbine blades is expanded by means of modeling method and image expansion technology, among which the actual wind turbine blade is used for testing, and the data augmentation under the modeling is implemented to monitor the actual wind turbine blades. In Section 4, different YOLOv5 models are compared to check their performance, and experimental verification is carried out under the condition of the different lighting and angles during wind turbine monitoring. In Section 5, the edgewise and flap-wise vibration of the wind turbine structure are visually tested, with the results compared with the traditional monitoring laser displacement sensor (LDS) and acceleration sensors, and the dynamic characteristics and modal shape of the wind turbine blade are accordingly

obtained. In Section 6, the main conclusions of this work are summarized.

2. Computer vision-based dynamic characteristics monitoring

The image encodes the entire field of view information in a non-contact manner, potentially addressing the challenges of contact sensor monitoring. The video consists of a series of images, the information of each which is checked in time to obtain the required monitoring information. The monitoring of dynamic characteristics based on computer vision can be divided into four steps: video processing, model building, calculate displacement calculating, and system identification, as shown in Fig. 1.

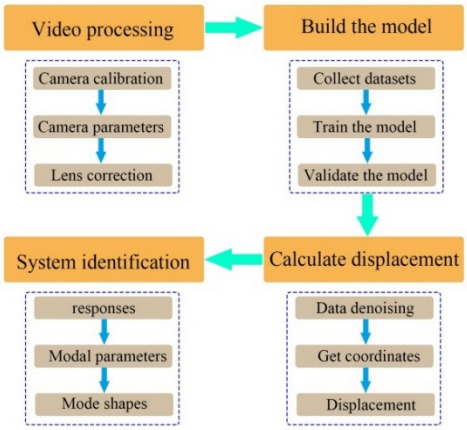


Fig. 1 General steps for monitoring structural dynamic characteristics based on deep learning

2.1 Video processing

The structural dynamic characteristics monitoring based on computer vision mainly uses the camera to capture the video to analyze the structural displacement. However, due to the distortion of the image caused by the distortion of the camera lens when shooting the target, the camera calibration should be performed first to deal with such problems. In order to determine the relationship between the three-dimensional geometric position of a certain point on the surface of a space target and its corresponding point in the image, a geometric model of camera imaging is established, these geometric model parameters of which are camera parameters. The general converting equation from image coordinates to physical coordinates is as follows

$$s\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

The simplified expression is

$$sx = K[R|t]X \tag{2}$$

where s is the scale factor; $(x, y, 1)^T$ is the image coordinate; $K$is the camera internal parameter representing the projection transformation from the three-dimensional real world to the two-dimensional image; $(X, Y, Z, 1)^T$ is the world coordinates. In the intrinsic parameters, f_x and f_y are the focal lengths of the camera in the horizontal and vertical directions, respectively; c_x and c_y are the offsets of the optical axis in the horizontal and vertical directions; γ is the tilt factor; R and t are the camera external parameters representing rigid rotation and translation from

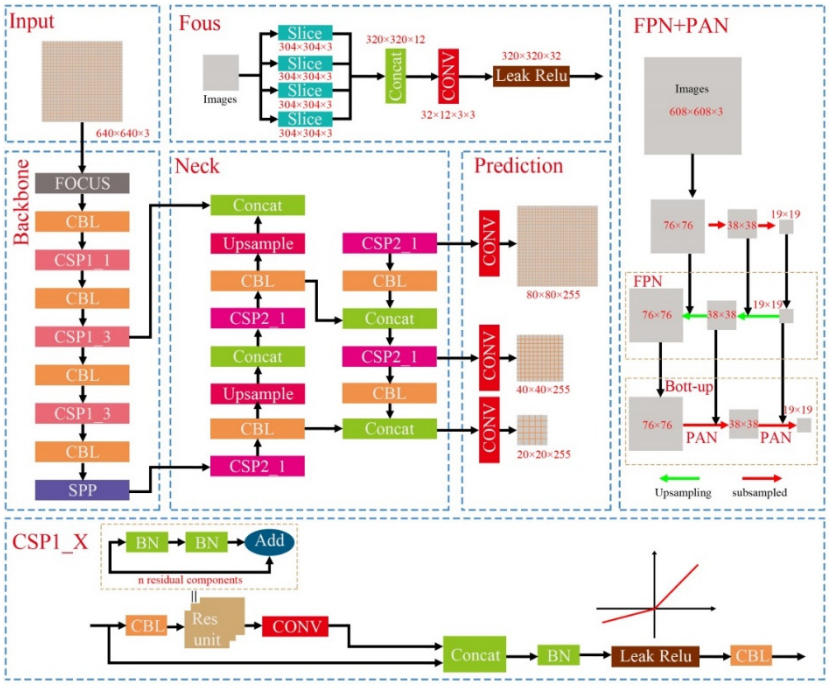


Fig. 2 The overall structure and module diagram of the YOLOv5 model

3D real-world coordinates to 3D camera coordinates; r_{ij} and t_i are elements of R and t, respectively.

The video is calibrated through camera internal parameters, tangential distortion and radial distortion, which can effectively eliminate lens distortion, thereby eliminating image distortion, resulting in more accurate displacement measurement. In this paper, Zhang's calibration method is used to calibrate the camera lens, and then the video calibration is performed (Zhang 2000).

2.2 Target tracking principle based on YOLOv5 deep learning model

Traditional computer vision-based monitoring methods are mainly affected by the external environment (such as changeable illumination conditions, complex backgrounds, etc.), which cannot be accurately monitored, and ultimately result in inaccurate identification result of structural dynamic characteristics. In recent years, with the great success of deep learning, many new target detection methods have been proposed to detect targets from images. In general, these methods can be divided into two categories: two-stage and one-stage, which the two-stage algorithm has higher accuracy but slower detection speed, while the one-stage one has faster detection speed but lower accuracy (Zhu *et al*. 2021). You Only Look Once v5 (YOLOv5) deep learning model has greatly improved in localization accuracy and detection time (Mao *et al*. 2019), compared with the previous YOLO series of algorithms, which is a classic one-stage algorithm (Wu *et al*. 2020, Jia *et al*. 2021).

The YOLOv5 model mainly includes YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x, whose depth and width gradually increase in order. Here, the YOLOv5l model is improved, the network model of which includes four parts: Input, Backbone, Neck, and Prediction. Its main framework and modules are shown in Fig. 2, where the image size is represented by pixels.

(1) Input

Input mainly includes three parts: Mosaic data augmentation, adaptive anchor box calculation and adaptive image scaling. Mosaic data augmentation is to process the image data obtained at the input end by random scaling and random distribution to obtain a richer dataset. For small targets, the overall accuracy is also improved, making the network more robust, and the demand for GPU (Graphic Processing Unit) smaller.

Adaptive anchor boxes compute anchor boxes with deterministic initial settings for actually different datasets. This paper adopts the initial anchor frame set by YOLOv5 with the anchor frame parameters of [116, 90, 156, 198, 373, 326], [30, 61, 62, 45, 59, 119], [10, 13, 16, 30, 33, 23], respectively.

In the commonly used target detection algorithms, the length and width of different images are different. The most common method at present is to uniformly scale the original image to a standard size. The most commonly used image sizes in the YOLO algorithm are 416×416 and 608×608, respectively. After the YOLO algorithm scales and fills the image, there will be black borders of different sizes at the beam end of the image. The excessive filling will lead to information redundancy and affect the overall training speed. Compared with the previous version, YOLOv5 has been optimized in this regard. The original image is adaptively added with the least black borders, which make the black borders at both ends of the image reduced, and the calculation amount is correspondingly reduced, then the target detection speed has also been improved.

(2) Backbone

The Backbone module mainly includes Focus structure and CSP (Cross Stage Partial) structure. The Focus structure is a newly added structure after YOLOv4, which is mainly used for slicing operations. Firstly, the original size of 608×608×3 image is input into the Focus structure for slicing operation to obtain 4 slices of 304×304×3. Secondly, four slices of 304×304×3 are used for depth stitching through the Concat operation, and a feature map of 304×304×12 is obtained. Finally, a new feature map is generated after convolution of 64 convolution kernels (image size is 304×304×64).

The CSP structure in YOLOv5 draws on the CSP Net structure design concept in the YOLOv4 backbone network, the designation of which contains two modules, CSP1_X and CSP2_X. The CSP1_X module is mainly used here, while the CSP2_X is used for the Neck module. The CSP1_X module contains CBL (Conv + BN + Leaky relu) blocks and X residual components so that it can better extract the deep features of the image. This paper uses the same CSP1 structure as YOLOv5l in the Backbone module. The Focus structure and CSP1_X structure are shown in Fig. 1.

(3) Neck

YOLOv5 also applies the CSP structure to the Neck module after YOLOv4, and adopts the structure of CSP2_X. In this paper, the structure of CSP is introduced in detail, main structure of which is similar to that of CSP1. It mainly uses two CBL blocks at the initial end, which increases the convolution layer and makes the recognition of target images more accurate.

Like YOLOv4, the Neck module of YOLOv5 adopts the structure of FPN (Feature Pyramid Networks) combined with PAN (Path Aggregation Network), as shown in Fig. 2. The FPN layer conveys strong semantic features from top to bottom, while the PAN layer is a bottom-up feature pyramid that conveys strong localization features. Both use different backbone layers to aggregate the ability features from different detection layers to further improve feature extraction. The Neck module mainly improves the CSP structure, adopts the CSP2_2 structure, and uses a larger learning rate to accelerate the convergence speed of the model. The subsequent convolution operation reduces the number of residual components and improves the calculation speed of the overall structure.

(4) Prediction

Prediction includes bounding box loss function and NMS (Non Maximum Suppression). GIOU_Loss (Generalized IOU_Loss) function as the loss function of the bounding box are used in YOLOv5 model, which

effectively solves the problem of non-overlapping bounding boxes, and improves the speed and accuracy of the prediction box regression. The weighted NMS is used in the target detection and prediction stage, which enhances the recognition ability for multiple targets and occluded targets, and obtains the optimal target detection frame. GIOU_Loss, as the loss function of bounding box, is actually an upgrade of DIOU (Distance IOU) which the ratio of the square of the distance between the center point of the real box and the predicted one to the square of the diagonal length of the two smallest boxes is used as the measurement standard.

The specific calculation equation of GIOU_Loss is as follows

$$GIOU\_Loss = 1 - GIOU = 1 - (\frac{|S_A|}{|S_B|} - \frac{|S_C - S_B|}{|S_C|}) \quad (3)$$

where S_A is the area of the intersection of box1 and box2, S_B is the area of the union of box1 and box2, and S_C is the smallest area of the outer rectangle between rectangles 1 and 2.

2.3 Improved YOLOv5 model

YOLOv5l is a standard model with the increasing number of convolution kernels included and improved detection accuracy, while the detection speed has decreased compared with the first two. YOLOv5x is an extended model with the largest number of convolution kernels, the highest detection accuracy among the four models, and the slowest detection speed. These four models correspond to different network depths. At the same time, each model has the defects of long training speed and low detection accuracy, so its YOLOv5 depth learning model needs to be improved.

The improved YOLOv5 deep learning model is mainly mended base on the YOLOv5l one. First, the Mosaic augmentation method is performed on the collected images in the input end as the same as that in YOLOv51 where 4 images are used for Mosaic data augmentation. On this basis, the original dataset was enhanced and Mosaic data augmentation was carried out using 9 images which are randomly cropped, scaled and then randomly arranged before stitching together to form a image. By this operation, the numbers of overall small target sample are increased followed by diversifying the dataset, which finally improves the network training speed. Using the slice operation in the Focus structure of the Backbone module, the input image dataset (three-channel image with size of 608×608×3) is divided into four slices (slice size is 304×304×3). Secondly, the four slices are deeply concatenated using the Concat operation to generate feature maps (image size is 304 × 304 ×12). Then, referring to the YOLOv5l model, the convolution layer consisting of 64 convolution kernels performs convolution operations to generate a new feature map (image size is 304×304×64). After passing through the Focus structure, it is passed to the CBL module, where the image is output through a convolutional layer of 128 convolution kernels and normalization and activation functions. The rest of the CSP structure and SPP structure belonging to the Backbone module are exactly the same as the YOLOv5l model. The Neck module mainly improves the CSP structure by adopting the CSP2_2 structure and replacing the Resunit with 2n CBLs. One branch of Resunit passes through 1 CBL layer first, and then performs the Concat operation after 2n CBLs, whereas the other branch passes through 1 CBL for Concat, and in the end the last two branches are merged and then go through a CBL to increase the model's nonlinear regression ability by using the Leaky Relu function. The improved YOLOv5 model promoted the calculation speed of the overall structure while ensuring the detection accuracy.

In order to achieve the effect of high-precision and clearly classified image tracking, the improved YOLOv5 depth learning model is combined with the Deep sort algorithm, the idea of which is to input the IOU (Intersection over Union) of the detection frame obtained by the target detection algorithm and the predicted tracking frame into the Hungarian algorithm to achieve maximum matching within an acceptable loss cost. The YOLOv5 model performs linear assignment to associate targets between frames and then classifies these targets and assigns IDs (Identifications). The Deep sort algorithm adds the occurrence information of the target to the calculation of the matching between frames. When the target is occluded but later reappears, the ID can be correctly matched, thus reducing the ID loss. The Deep sort algorithm as a whole includes two parts: state estimation and trajectory processing. The overall process of YOLOv5 combined with the Deep sort algorithm is shown in Fig. 3.

2.4 Displacement calculation

According to YOLOv5 and Deep sort algorithm, the coordinate $P_i(x_i, y_i)$ of each frame in the video stream and the target coordinate $P_0(x_0, y_0)$ of the first frame are obtained. The displacement in the coordinate of the structure image can be calculated by the following equation (unit: pixel)

$$d_I = P_i - P_0 \quad (4)$$

However, for the actual structural health monitoring, the displacement d_S (unit: mm) in the physical coordinates of the structure needs to be output, so the target scale *SF* needs to be calculated by using the scale factor method, and

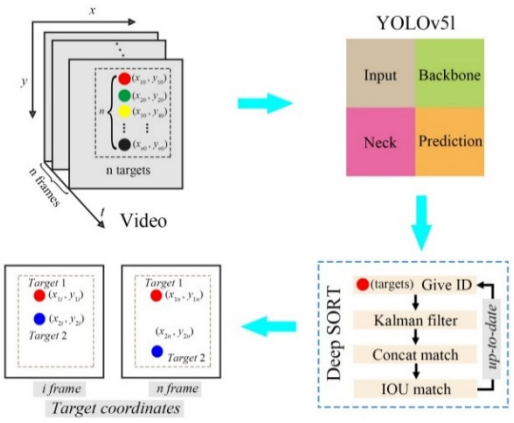


Fig. 3 YOLOv5l combined with Deep sort algorithm

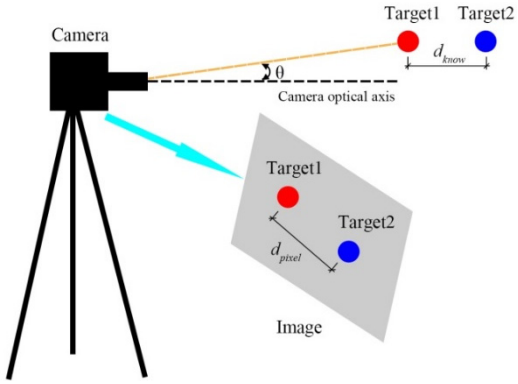


Fig. 4 Case of non-perpendicula camera optical lens axis

then converted into the displacement in physical coordinates by the Eq. (5) (Feng *et al*. 2015). The camera vibration may be caused by the ground vibration when using the camera to monitor the structure. The camera vibration can be eliminated by using the background fixed point and high pass filtering methods (Perry and Guo 2021 and Hoskere *et al*. 2019).

In general, the *SF* can be obtained from the actual physical size and pixel size. In the actual monitoring process, the optical axis of the camera cannot be perpendicular to the moving direction of the target due to various external factors. At this time, the actual displacement of the measured target and the camera image will form an angle, as shown in Fig. 4.

At this time, the calculation equation of the scale factor *SF* is

$$SF = \frac{d_{physical}}{d_{pixel}\, cos^2\, \theta} \tag{5}$$

where $d_{physical}$ is the actual physical size of the target; and d_{pixel} is the pixel size of the target in the image; $\theta$is the angle between the optical axis of the camera and the measured target.

Therefore, the displacement d_s in physical coordinates of the structural vibration (unit: mm) is as followed

$$d_S = SF \times d_I \tag{6}$$

2.5 Dynamic characteristic parameter identification

After obtaining the dynamic displacement of each position of the structure by the above method, the frequency domain information of the structure is got by means of fast Fourier transform and power spectral density. Finally, the dynamic characteristic parameters of the structure are acquired through the correlation calculation on the obtained natural vibration frequency of the structure.

3. Modeling-based approach to dataset augmentation

The premise of efficient application of deep learning algorithms is to have a large number of datasets, the collection of which has always been a major obstacle to the broad development of deep learning (Liang 2019, Narazaki *et al*. 2021). In common recognition fields such as pedestrian detection and vehicle recognition, a large number of scholars and companies have collected public datasets. However, in other niche fields, the collection of datasets is more difficult. Because wind farms are located in harsh environments and wind turbines continue to grow in size, it is difficult for researchers to collect datasets by taking images of blades in practice. Therefore, it is urgently necessary to develop a simple and efficient data collection method for wind turbine blades.

3.1 Wind turbine blade dataset collection based on PBGMs

In view of the lack of deep learning datasets, people gradually use virtual images to expand the datasets. Narazaki *et al*. (2019) propose an approach to augment the dataset with physics-based graphics models (PBGMs), where damage hotspots identified by finite element analysis are used to achieve photorealistic synthetic damage scenarios. Then, Hoskere *et al*. (2022) used PBGMs to generate images and associated ground truth annotations for semantic segmentation of structural parts and damage, and monocular depth estimation for structural part localization. Therefore, using PBGMs to generate images can also establish a dataset of wind turbine blade structure, and then perform structural displacement monitoring and dynamic characteristics identification.

Blender software is a free and open source 3D graphics and image software developed by the Blender Foundation, providing a series of animation short film production solutions from modeling, animation, material, rendering, to audio processing, video editing and so on. Blender software has built-in advanced film and television solutions such as green screen keying, camera reverse tracking, mask processing, and post-node synthesis, which can create models that are very similar to the real world. Therefore, it is especially suitable for the production of datasets for deep learning.

Generally, the general steps of creating through blender software can be divided into wind turbine structure making, material texture filling, correlation of related scenes, light background adjustment and rendering. Therefore, the same wind turbine structure can be combined with different scenes for image expansion. In this paper, blender software is employed to make the wind turbine structure, and the photos of the wind turbine blade training dataset are made in combination with the common scenes of the actual wind turbine structure such as different environment, varied types of wind turbines, and various lightning, as shown in Fig. 5. Different from other damage identification based on PBGMs, this paper does not pay attention to the surface defects of blades, so it is only necessary to model and expand the intact blades through blender software, and finally identify them through the depth learning model.

3.2 Dataset augmentation

Although blender software can produce a large number of wind turbine blade images under different conditions, the

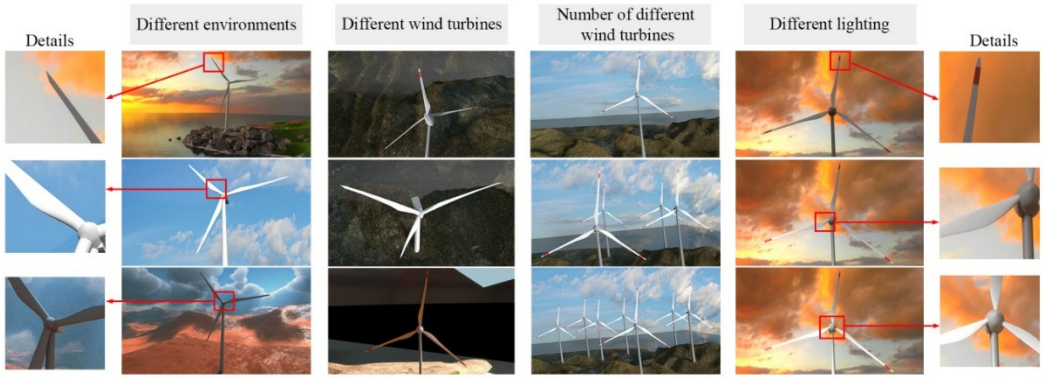


Fig. 5 Images of wind turbine blades under different conditions using blender software

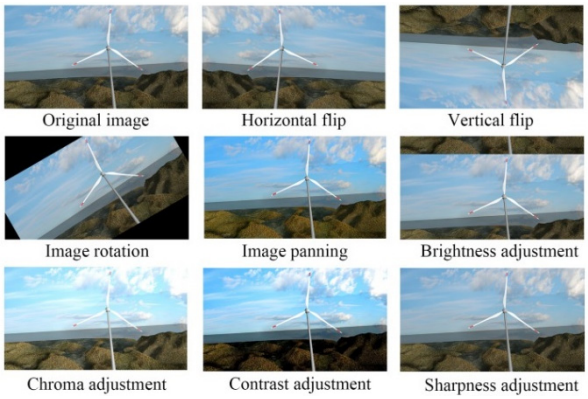


Fig. 6 Image data expansion

number of images required for deep learning datasets is relatively large, and the use of blender software alone is not enough to meet the dataset production requirements. In addition, blender not only takes a lot of time to produce images, but also requires high computer configuration, so more data expansion methods need to be developed. Horizontal flip, Vertical flip, Image rotation, Image panning, Brightness adjustment, Chroma adjustment, Contrast adjustment, Sharpness adjustment and other methods are adopted in this paper for expansion. Fig. 6 is an image of a wind turbine blade image expansion in different ways.

Although there are not many forms of expanded datasets, parameters such as angle, brightness, chromaticity, etc. can be adjusted for data expansion, from a small sample dataset to a large dataset.

3.3 Dataset augmentation verification

The dataset can be supplemented by blender and image augmentation, but it is only used as a real dataset in practical identification. In this paper, 85 images are created using the blender, expanding which to 2210 images as a dataset, followed by testing using the YOLOv5 deep learning model. The monitoring sites were blade tip, maximum chord and blade root. The test dataset is divided into training set and validation set according to the ratio of 8:2, and part of the divided validation set is shown in Fig. 7.

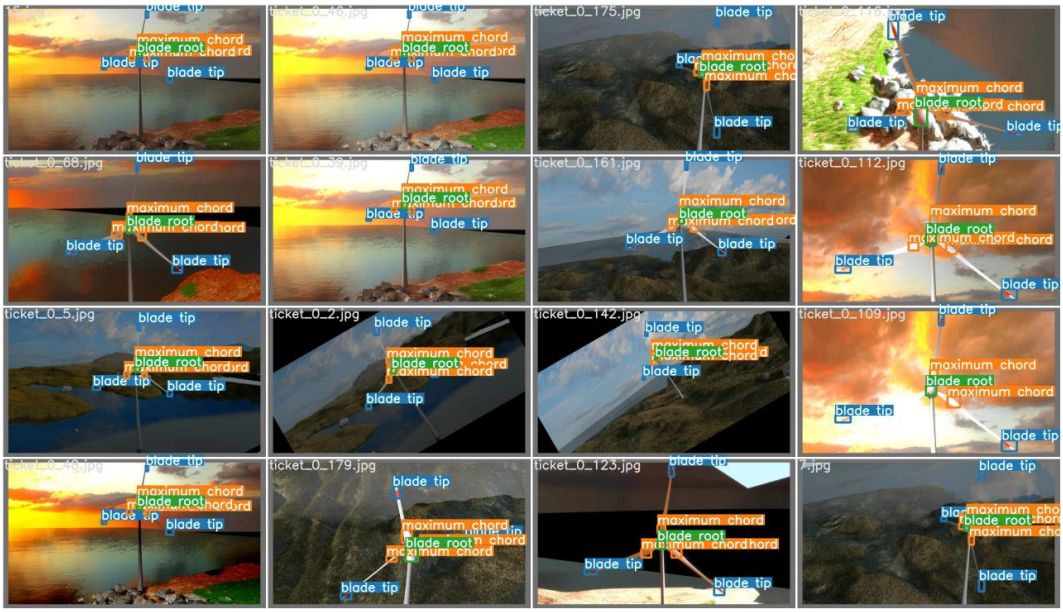


Fig. 7 Validation dataset

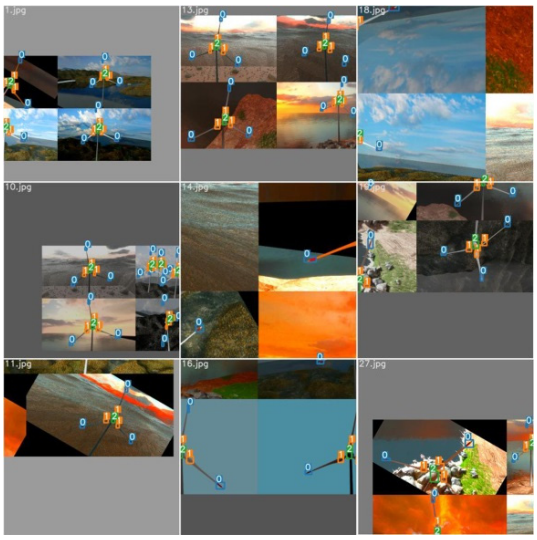


Fig. 8 Mosaic data augmentation

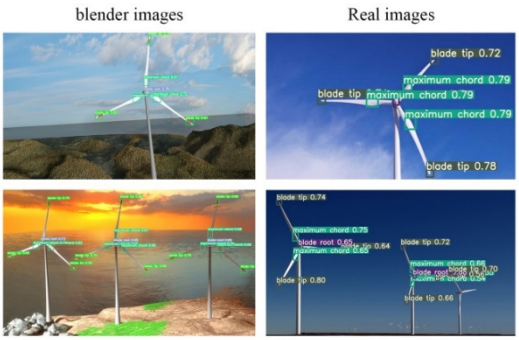


Fig. 9 Practical effect test

Then, the Mosaic data augmentation described in Section 2.2 is performed on the dataset followed by processing the image data in two ways: random scaling and random distribution to enrich the dataset, as shown in Fig. 8.

Finally, the weights are obtained after training for actual effect testing. In this paper, blender images and actual images that are not in the dataset are adopted for testing. The test effect is shown in Fig. 9.

It can be seen from Fig. 9 that both the modeled images and the actual images can accurately identify the preset type, and the confidence level is greater than 0.72, showing a good detection effect. Therefore, modeling-based data augmentation methods can be extended to the production of deep learning small-sample datasets so that health monitoring in more fields can be expanded.

This paper uses the improved YOLOv5 depth learning model combined with the Deep sort algorithm to monitor. Therefore, each training target is given a different ID in the first frame of video processing. Although it is the same object during training, different IDs are assigned to this object category during program running. For example, three identical blades will be given different IDs, which will ensure that the target tracking coordinates of multiple blades will not be confused.

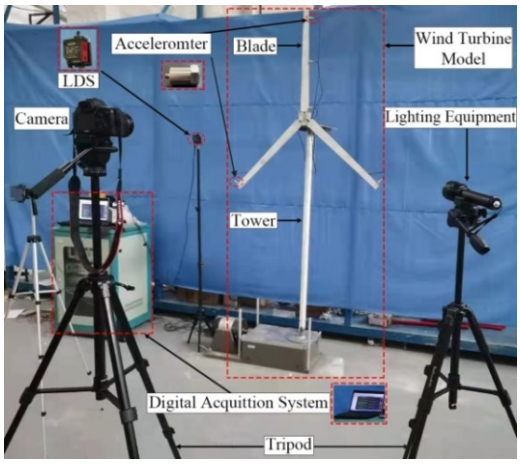


Fig. 10 Experimental equipment

4. Experimental verification

4.1 Experimental device

The quantification of structural vibration displacement is an important means to evaluate the dynamic performance of a structure under various external dynamic loads, the error of which is particularly important, accordingly. The structure of the large-scale wind turbine is scaled by 1:60 with the height of 1.63 m and the blade length of 0.73 m. The diameter of the bottom and top tower is 65 mm and 42 mm, respectively. The wall thickness of the bottom tower is 1.1mm, while the one of the top tower is 0.33 mm. In this test, the vibration test is carried out on the scaled wind turbine model under the simulated shutdown state.

The video image data adopts Canon 5D4 camera equipped with a lens of 24 mm ~ 105 mm zoom. The video resolution is set to 1080 P, and the frame rate 50 fps to simulate the scaled wind turbine model, light source, tripod, etc. In order to verify the accuracy of computer vision monitoring results of structural displacement, the laser displacement sensor of Banner250U with a sampling frequency of 50 Hz and the INV9812 accelerator with a sampling frequency of 50 Hz is used. The test equipment is shown in Fig. 10.

After introducing the algorithm principles discussed above, the next essential step for achieving the YOLOv5 deep learning model is training on image data. In this paper, the wind turbine blade vibration video dataset is used for training, and the image of the Open CV deep learning platform is built for training. Specific test environment setting as follows: the graphics card is NVIDIA GeForce Rtx 3060, the video memory is 16 GB, Pytorch uses version 1.7.1, CUDA version 10.1, the selected compilation language is Python 3.7, the overall operating environment is performed in Pycharm, and the batch size is set to 16. The dataset mainly uses the collected video images, the labelimg software to label the image targets, eventually a total of 100 epochs are trained.

Table 1 Quantitative analysis of four models

Class	mAP (%)	Training time (h)
YOLOv5s	97.16	3.823
YOLOv5m	97.88	4.766
YOLOv5l	98.14	5.899
Improved YOLOv5	98.96	4.473

4.2 Improved YOLOv5 model vibration test validation

In order to verify the improved YOLOv5 deep learning model, the YOLOv5s, YOLOv5m, YOLOv5l and the improved YOLOv5 are chosen to train image data. On this basis, quantitative analysis of these four models is carried out comparing the mAP (Mean Average Precision) and training time respectively, as shown in Table 1.

It can be seen from Table 1 that the improved YOLOv5 model not only ensures the accuracy, but also improves the overall training speed. Through comparative analysis, it can be seen that YOLOv5s has the shortest training time among the four models, the corresponding accuracy rate of which is the lowest. Compared with the improved YOLOv5, the Mosaic method is improved at the input end and the CSP structure is improved in the Neck module, and the detection effect is very obvious. Compared with the YOLOv5m and YOLOv5l structures, the improved YOLOv5 also has 1.08% and 0.82% improvement in accuracy. The most obvious is that the training time is significantly shortened with the using time of only 4.473h, which achieved the goal of lightweight, rapid model of machine learning and computer vision.

The scaled wind turbine model is applied with initial displacement in edgewise and flap-wise, and then is released to make it vibrate freely. Different deep learning models are used for training, and finally the blade tip displacement is obtained through the Deep sort algorithm. The vibration time domain image is shown in Fig. 11.

It can be seen from Fig. 11 that the YOLOv5s and YOLOv5m models do not perform well in displacement measurement due to the shallow network depth, which is mainly manifested in the identification error at the displacement peak. YOLOv5l and improved YOLOv5 have the best recognition performance and are closest to the time-history curve of LDS. While, improved YOLOv5 is more suitable for vibration displacement measurement as it saves more than an hour in training time.

In order to quantify the accuracy of the model, the Root Mean Square Error (RMSE), correlation coefficient (ρ) and coefficient of determination (R^2) are introduced for error analysis. Their equations are:

$$\mathrm{RMSE} = \sqrt{\sum_i (x_v(i) - x_s(i))^2 / n} \tag{7}$$

$$\rho = \frac{|\sum_i (x_s(i) - \mu_s) \times (x_v(i) - \mu_v)|}{\sqrt{\sum_i (x_s(i) - \mu_s)^2} \sqrt{\sum_i (x_v(i) - \mu_v)^2}} \tag{8}$$

$$R^2 = 1 - \frac{\sum_i (x_v(i) - x_s(i))^2}{\sum_i (x_s(i) - \mu_s)} \tag{9}$$

RMSE is calculated using Eq. (7), where n is the total number of monitoring, x_v and x_s are displacement data from vision monitoring and laser displacement sensors, respectively, which measure the deviation between the measured value and the reference one. ρ is calculated by Eq. (8), where μ_v and μ_s are the average values of the two displacement trajectories, ρ varies from 0 to 1, which ρ = 1 represents complete correlation, and ρ = 0 means that the two recording tracks have no correlation. The calculation formula of R^2 is Eq. (9), which is used to determine the matching degree of the two recorded tracks. R^2 also belongs to [0, 1] with the unit representing the similarity of the two recorded tracks. The comparison of vibration monitoring data errors of the four deep learning models is shown in Table 2.

It can be seen from Table 2 that improved YOLOv5 performs the best among the three error indicators, especially for and showing that vibration time-history data can be accurately measured by improved YOLOv5. Although YOLOv5l and improved YOLOv5 have similar errors, the improved YOLOv5, which is referred to as

Table 2 Error analysis of four deep learning models

Class	RMSE	ρ	R^2
YOLOv5s	1.155	0.956	0.903
YOLOv5m	1.131	0.957	0.907
YOLOv5l	0.139	0.989	0.998
Improved YOLOv5	0.104	0.993	0.999

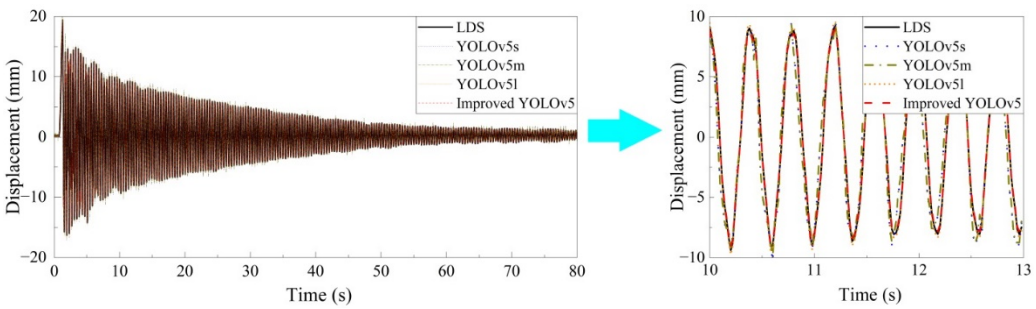


Fig. 11 Comparison of vibration displacement recognition by different deep learning models

YOLOv5 below, is chosen due to its better training duration and mAP accuracy.

4.3 Validation of experiments under changing illumination conditions

Limited by changeable illumination conditions, the displacement measuring by traditional algorithms is usually inaccurate. In response to this problem, the YOLOv5 dataset for illumination changes are applied for training followed by a comparison between the former and a color matching algorithm here. In order to simulate the illumination change, a light source is used to illuminate the wind turbine blade structure with strong light. After the video acquisition system is built, the vibration of the tip of the wind turbine blade is captured by the camera. During the collection process, the blade tips were irradiated with lights of different intensities at different time periods through the light source, as shown in Fig. 12. To highlight the monitoring performance of YOLOv5 depth learning model, light sensitive color matching algorithm is used for comparison, and color matching algorithm requires colored artificial markers. Therefore, the color artificial markers in Fig. 12 are mainly used for displacement monitoring of color matching algorithm. For YOLOv5 deep learning model, artificial markers are not required for auxiliary monitoring. Besides this test, other tests do not use artificial markers for auxiliary monitoring. The vibration time-history curves measured by YOLOv5 and the color matching algorithm are compared in Fig. 13.

As can be seen from Fig. 13, the color matching algorithm cannot accurately capture the measurement points of the illumination change, resulting in a non-negligible error in the displacement measurement. The YOLOv5 model has been trained using the illumination change dataset. It still maintains good tracking performance in the process of illumination change, and the time history curve is consistent with the LDS display, showing that YOLOv5 has strong robustness to the displacement measurement under changing illumination conditions, overcoming the shortcomings of traditional algorithms for displacement measurement under changing illumination conditions. Therefore, the YOLOv5 deep learning model is suggested more widely used in structural health monitoring under more environmental changes.

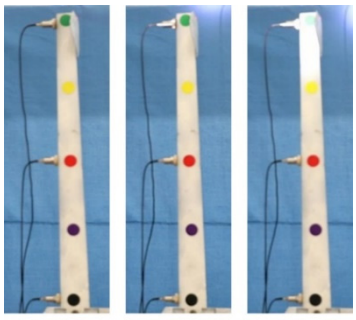


Fig. 12 Image acquisition under the illumination of different light sources

4.4 Experimental verification from different angles

In the case of actual engineering monitoring, it is difficult for computer vision-based displacement monitoring to ensure that the optical axis of the camera is always perpendicular to the surface of the target point, which cannot completely ignore the error caused by the angle to the visual measurement. In this experiment, the vibration displacement of the wind turbine blade structure was extracted using the improved YOLOv5 depth vision model at the shooting angles of 0°, 5° and 10°. The displacement calculation of the generated angle adopts Eq. (6). Through the vibration test of 80 s, the displacement measurement results at different angles are shown in Fig. 14.

It can be seen from Fig. 14 that the visual measurement results from different angles are consistent with the overall displacement time history of the LDS. However, when the displacement is less than 3 mm, the measuring result is no longer accurate due to the influence of the angle. In order to quantify the measurement error, PDF (Probability Density Function) was adopted and the absolute error also was compared accordingly, which can be seen in Fig. 15.

As shown in Fig. 15(a), with the increscent of the angle, the measured displacement error increases accordingly. The main reason is that when the shooting target has an angle, the measured physical displacement is smaller than the actual one, indicating that the existing structure produces plane vibration, which will lead to an increase in error. Fig. 15(b) shows the absolute error under different angle measurements. When the optical axis of the camera is perpendicular to the surface of the target point (that is, at 0°), the error is less than 0.5 mm, and the maximum error is only 1.324 mm at 10°. Although the error will increase with

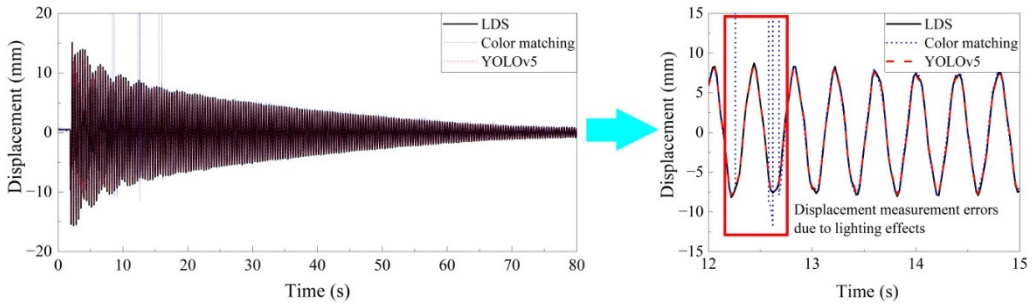


Fig. 13 Comparison of time history curves measured by YOLOv5 and color matching algorithms

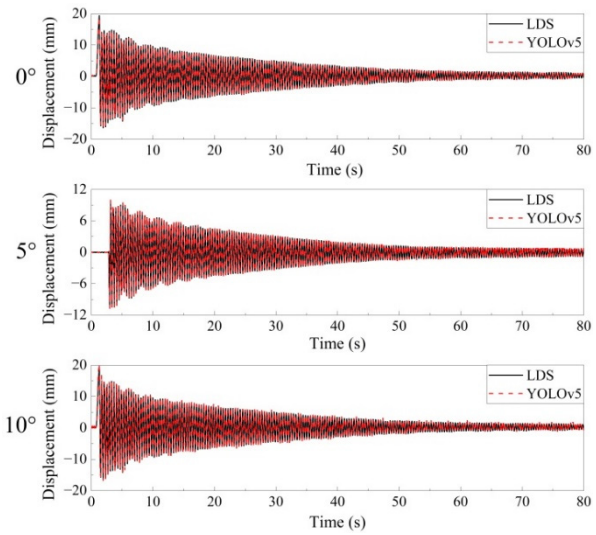
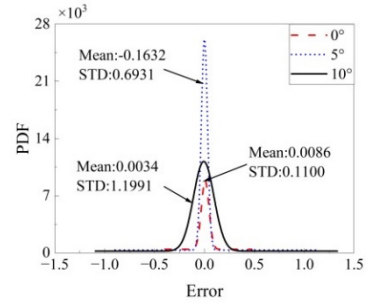
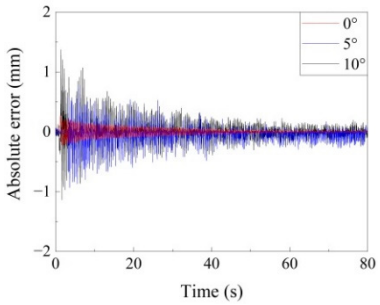


Fig. 14 Comparison of displacement time history at different angles



(a) PDF distribution



(b) Absolute error

Fig. 15 Comparison of displacement errors at different angles

the increase of the angle, the error is within the controllable range of 0.5 mm. The YOLOv5 model still meets the engineering requirements when measuring the structural error when an angle exists. Therefore, the YOLOv5 model can meet the accuracy requirements for the dynamic characteristic identification of wind turbine blades.

In conclusion, the improved YOLOv5 deep learning model is higher than other YOLOv5 deep learning models in training speed and mAP value, and can also adapt to the actual engineering application environment of illumination change and angles. Therefore, the improved YOLOv5 depth learning model can be applied to the displacement response

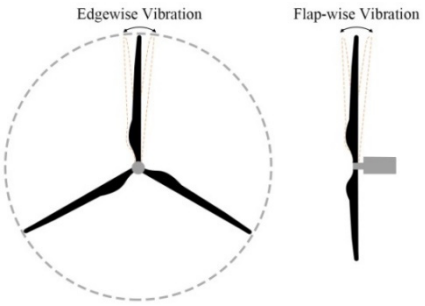


Fig. 16 Edgewise and flap-wise vibration of wind turbine

monitoring of wind turbine blades in common environments.

5. Wind turbine blade dynamic characteristics test

5.1 Wind turbine blade natural vibration frequency test

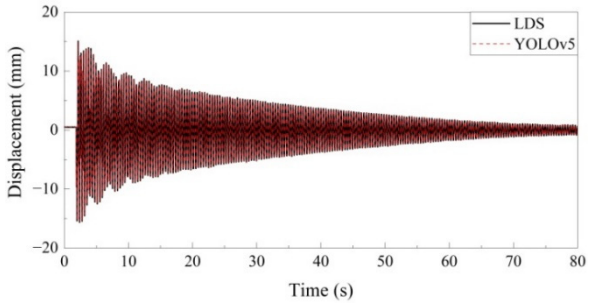
This paper mainly explores the wind turbine blades in edgewise and flap-wise vibration. Although the dynamic characteristic parameters along the edgewise and flap-wise direction are related, but this paper mainly uses monocular vision measurement. Therefore, the camera is used twice in this test in the edgewise and flap-wise directions. The vibration forms are shown in Fig. 16. The experimental device adopts the scaled wind turbine blade described in Section 4.1, the YOLOv5 deep learning model for displacement measurement under the condition of given excitation, and the LDS and acceleration sensor for verification. The time-domain and frequency-domain images of the measured edgewise and flap-wise vibration of the wind turbine blades are shown in Fig. 17.

It can be seen from Fig. 17 that the displacement time history tracked using the YOLOv5 deep learning model is consistent with the displacement time history measured by LDS as a whole, and the natural vibration frequency domain converted to the frequency domain by FFT (Fast Fourier Transform) is also consistent with LDS. The natural vibration frequencies of the wind turbine blades measured by vibration are shown in Table 3.

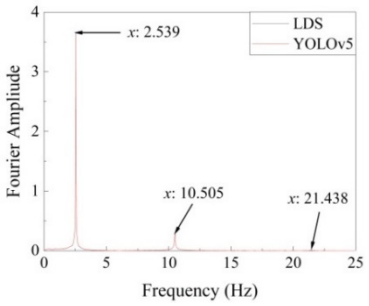
It can be seen that the measuring results of wind turbine blades using the YOLOv5 deep learning model in the sway and waving directions are completely consistent with that by the LDS, indicating that it meets the wind turbine natural vibration frequency test requirements. Combined with the frequency domain information in Fig. 17, it can be found that the recognition effect of the third-order natural vibration frequency of the wind turbine blade in the edgewise direction is not good, which requires higher-precision equipment, such as a high-speed camera for photographing and recognition.

5.2 Comparison of vision sensor and traditional sensors performance

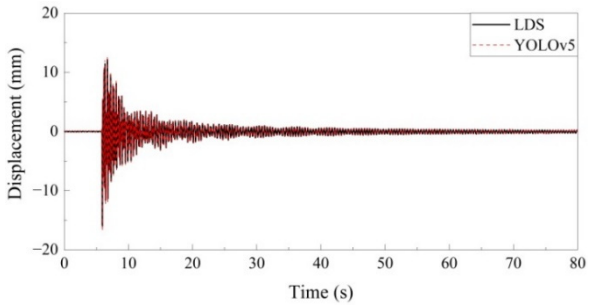
Accelerometers have wide application in structural health monitoring for their high sensitivity and wide



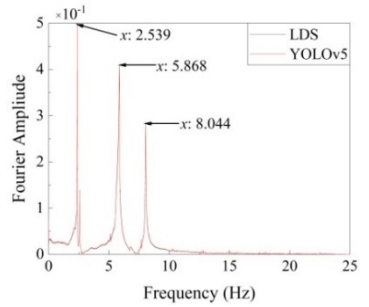
(a) Time domain comparison in the edgewise vibration direction



(b) Frequency domain comparison in the edgewise vibration direction



(c) Time domain comparison in the flap-wise vibration direction



(d) Frequency domain comparison in the flap-wise vibration direction

Fig. 17 Wind turbine vibration time domain and frequency domain monitoring comparison

Table 3 Error analysis of four deep learning models

Vibration direction		LDS	YOLOv5	Error (%)
Edgewise vibration	1st	2.539	2.539	0.00
	2st	10.505	10.505	0.00
	3st	21.438	21.438	0.00
Flap-wise vibration	1st	2.539	2.539	0.00
	2st	5.8868	5.8868	0.00
	3st	8.044	8.044	0.00

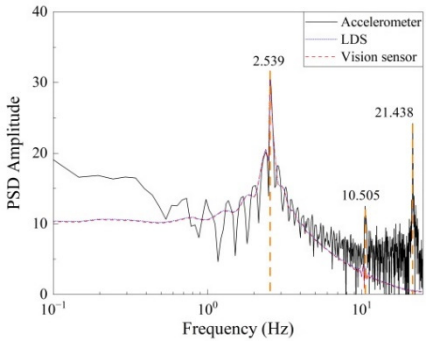


Fig. 18 PSD comparison of different sensors

frequency domain measurement range. However, the accelerometer must be mounted on the structure, which is not only time-consuming but also labor-intensive with unstable signal transmission. Therefore, more convenient, non-contact and damage-free vision sensors are suggested strongly.

In order to verify the accuracy of the accelerometer sensor and the vision sensor, the vision sensors are adopted in this experiment for displacement tracking, and the LDS and the acceleration sensor are also chosen to compare and verify. Fig. 18 illustrate the PSD (Power Spectral Density) results measured by the three sensors.

It can be seen from Fig. 18 that the PSD results obtained from displacement measured by the LDS and the vision sensor are consistent. While the measured higher-order natural frequencies for accelerometers are more accurate due to its greater sensitivity with the almost same PSD results with the former two. Therefore, it is possible to use a vision sensor to measure the low-order natural frequency of the structure. If higher order natural frequencies need to be measured, accelerometers or high-speed cameras can be adopted. Most civil engineering structures have low natural vibration frequencies, so general cameras can also meet the measurement needs.

5.3 Identification of dynamic characteristic parameters of wind turbine blades

Through the above measurement of the natural vibration frequency of the wind turbine blade, the accuracy of the visual sensor and the traditional sensor is compared by PSD herein. The response output of the blade monitoring point is the indispensable condition for analyzing the dynamic characteristics of the blade. Hence, in this test, 5 monitoring points from P0 to P4 were set at the blade from bottom to top. By applying excitation to the blade tip, the above 5 points are monitored by the visual sensor, while the three points of P0, P2 and P4 are measured using 3 acceleration

Table 4 Error analysis of four deep learning models

Monitoring points	Edgewise vibration (mm)	Flap-wise vibration (mm)
P4	16.098	17.184
P3	14.431	14.268
P2	11.773	11.326
P1	9.539	8.763
P0	7.804	6.231

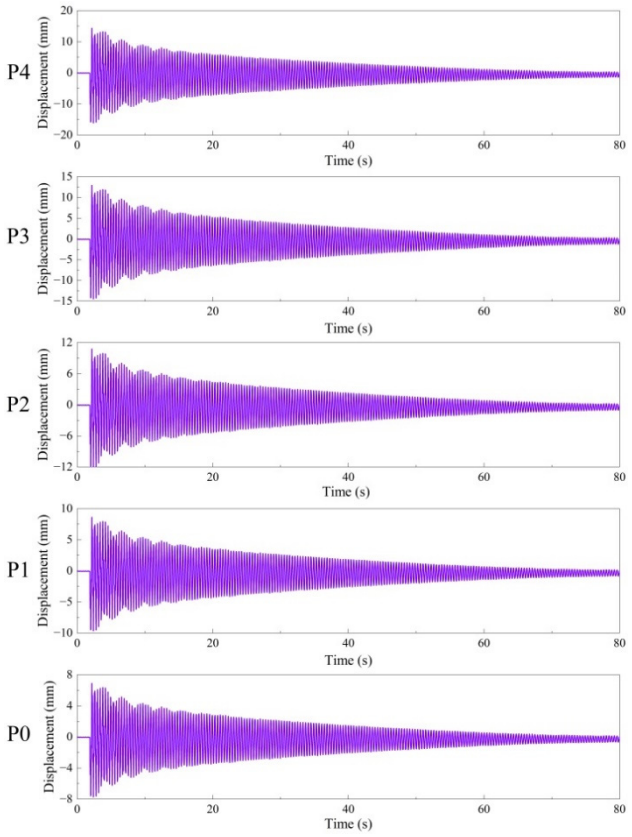


Fig. 19 Displacement response of five monitoring points of the blade

sensors. In order to make this paper concise, the displacement of the five monitoring points measured by the test is only plotted in the direction of the edgewise vibration, as shown in Fig. 19.

It can be seen from Fig. 19 that the overall vibration laws of the displacement responses of the five monitoring points of the blade in the edgewise direction are consistent. In order to draw accurate mode shapes, the displacement response peaks in the edgewise and flap-wise vibration directions are listed in Table 4.

It can be seen from Table 4 that the peaks of the displacement response in the edgewise and flap-wise directions decrease in turn from the blade tip to the blade root, which conforms to the vibration rule. Because the blade vibration in the flap-wise direction belongs to the out-of-plane vibration, the peak reduction is faster than that in the edgewise direction. From the above displacement

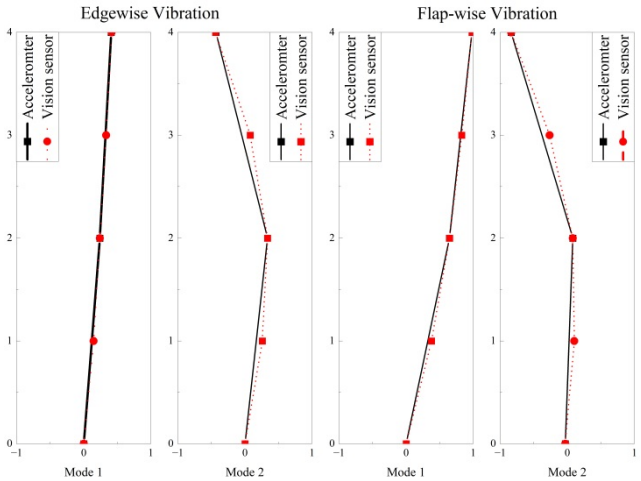


Fig. 20 Mode shapes of blade edgewise and flap-wise vibration direction

response, the mode shapes of the blade edgewise and flap-wise vibration direction are obtained as shown in Fig. 20.

As can be seen, the resulting mode shape obtained from vision sensor is smoother than that from the accelerometer, due to its advantage of monitoring more points, which are beneficial to the analysis of structural dynamic characteristics. Therefore, the vision sensors are more conducive to identify the dynamic characteristics of wind turbine blades compared with traditional accelerometers sensors.

6. Conclusions

In this study, a remote structural dynamic characteristics monitoring method was developed based on the improved YOLOv5 deep learning model combined with a modeling-based dataset expansion method. Then comprehensive experiments were carried out, including comparative experiments among different influential factors and wind turbine test experiments, to study its performance. The following conclusions can be drawn:

(1) Introducing the CSP2_2 structure to mend the YOLOv5 deep learning model make the residual component reduced and the network training speed improved. On this basis, the Deep sort algorithm is added to better the accuracy of structural displacement monitoring. Compared with other deep learning models, the training time of the improved YOLOv5 model is greatly shortened, with the highest mAP value.

(2) A wind turbine model which is similar to the actual environment through blender software is established based on the modeling-based dataset expansion method. Combined with image expansion technology, the original dataset for scarce wind turbine blade is expanded, which provides better support for training deep learning models.

(3) The improved YOLOv5 deep learning model has better performance in monitoring vibration

displacement than other YOLOv5 models, and can achieve relatively accurate monitor displacement result under changeable illumination conditions. The maximum error in displacement monitoring with an angular deviation of 10° is only 1.324 mm, which meet actual monitoring needs.

(4) Using the improved YOLOv5 visual model to test the dynamic characteristics of the wind turbine in edgewise and flap-wise directions, the results show that the model can accurately identify the natural vibration frequency of the object structure. With the aid of vision sensors, multiple points can be monitored simultaneously, so that the mode shape of visual monitoring is smoother than that from accelerometer, which is more conducive to the analysis of structural dynamic characteristics. However, the vision sensor is not sensitive to high-order, and cannot accurately capture the high-order natural frequency compared with the accelerometer. To overcome this problem, a high-speed camera was suggested for visual monitoring.

Acknowledgments

The work described in this paper was jointly supported by the National Science Foundation of China (Grant Nos. 52068049 and 51908266), the Science Fund for Distinguished Young Scholars of Gansu Province (No. 21JR7RA267), and Hongliu Outstanding Young Talents Program of Lanzhou University of Technology.

References

Amenabar, I., Mendikute, A., López-Arraiza, A., Lizaranzu, M. and Aurrekoetxea, J. (2011), "Comparison and analysis of non-destructive testing techniques suitable for delamination inspection in wind turbine blades", *Compos. B. Eng.*, **42**(5), 1298-1305. https://doi.org/10.1016/j.compositesb.2011.01.025

Beganovic, N. and Söffker, D. (2016), "Structural health management utilization for lifetime prognosis and advanced control strategy deployment of wind turbines: An overview and outlook concerning actual methods, tools, and obtained results", *Renew. Sustain. Energy Rev.*, **64**, 68-83. https://doi.org/10.1016/j.rser.2016.05.083

Dong, C.Z., Celik, O. and Catbas, F.N. (2019), "Marker-free monitoring of the grandstand structures and modal identification using computer vision methods", *Struct. Health Monit.*, **18**(5-6), 1491-1509. https://doi.org/10.1177/1475921718806895

Dong, C.Z., Celik, O., Catbas, F.N., O'Brien, E.J. and Taylor, S. (2020), "Structural displacement monitoring using deep learning-based full field optical flow methods", *Struct. Infrastruct. Eng.*, **16**(1), 51-71. https://doi.org/10.1080/15732479.2019.1650078

Du, Y., Zhou, S., Jing, X., Peng, Y., Wu, H. and Kwok, N. (2020), "Damage detection techniques for wind turbine blades: A review", *Mech. Syst. Signal Process.*, **141**, 106445. https://doi.org/10.1016/j.ymssp.2019.106445

Ehrhardt, D.A., Allen, M.S., Yang, S. and Beberniss, T.J. (2017), "Full-field linear and nonlinear measurements using continuous-scan laser doppler vibrometry and high speed three-dimensional digital image correlation", *Mech. Syst. Signal Process.*, **86**, 82-97. https://doi.org/10.1016/j.ymssp.2015.12.003

Felipe-Sesé, L. and Díaz, F.A. (2018), "Damage methodology approach on a composite panel based on a combination of Fringe Projection and 2D Digital Image Correlation", *Mech. Syst. Signal Process.*, **101**, 467-479. https://doi.org/10.1016/j.ymssp.2017.09.002

Feng, D. and Feng, M.Q. (2017), "Experimental validation of cost-effective vision-based structural health monitoring", *Mech. Syst. Signal Process.*, **88**, 199-211. https://doi.org/10.1016/j.ymssp.2016.11.021

Feng, D., Feng, M.Q., Ozer, E. and Fukuda, Y. (2015), "A vision-based sensor for noncontact structural displacement measurement", *Sensors*, **15**(7), 16557-16575. https://doi.org/10.3390/s150716557

Hernandez-Estrada, E., Lastres-Danguillecourt, O., Robles-Ocampo, J.B., Lopez-Lopez, A., Sevilla-Camacho, P.Y., Perez-Sariñana, B.Y. and Dorrego-Portela, J.R. (2021), "Considerations for the structural analysis and design of wind turbine towers: A review", *Renew. Sust. Energ. Rev.*, **137**, 110447. https://doi.org/10.1016/j.rser.2020.110447

Hoskere, V., Park, J.W., Yoon, H. and Spencer Jr, B.F. (2019), "Vision-based modal survey of civil infrastructure using unmanned aerial vehicles", *J. Struct. Eng.*, **145**(7), 04019062. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002321

Hoskere, V., Narazaki, Y. and Spencer Jr, B.F. (2022), "Physics-based graphics models in 3D synthetic environments as autonomous vision-based inspection testbeds", *Sensors*, **22**(2), 532. https://doi.org/10.3390/s22020532

Jasinien, E., Raiutis, R., Voleiis, A., Vladiauskas, A., Mitchard, D. and Amos, M. (2009), "NDT of wind turbine blades using adapted ultrasonic and radiographic techniques", *INSIGHT*, **51**(9), 477-483. https://doi.org/10.1016/j.ymssp.2019.106445

Jia, W., Xu, S., Liang, Z., Zhao, Y., Min, H., Li, S. and Yu, Y. (2021), "Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector", *IET Image Process*, **15**(14), 3623-3637. https://doi.org/10.1049/ipr2.12295

Kalaitzakis, M., Kattil, S.R., Vitzilaios, N., Rizos, D. and Sutton, M. (2019), "Dynamic structural health monitoring using a DIC-enabled drone", *Proceedings of 2019 International Conference on Unmanned Aircraft Systems* (*ICUAS*), Atlanta, GA, USA, June.

Kaewniam, P., Cao, M., Alkayem, N.F., Li, D. and Manoach, E. (2022), "Recent advances in damage detection of wind turbine blades: A state-of-the-art review", *Renew. Sust. Energ. Rev.*, **167**, 112723. https://doi.org/10.1016/j.rser.2022.112723

Khadka, A., Fick, B., Afshar, A., Tavakoli, M. and Baqersad, J. (2020), "Non-contact vibration monitoring of rotating wind turbines using a semi-autonomous UAV", *Mech. Syst. Signal Process.*, **138**, 106446. https://doi.org/10.1016/j.ymssp.2019.106446

Kuddus, M.A., Li, J., Hao, H., Li, C. and Bi, K. (2019), "Target-free vision-based technique for vibration measurements of structures subjected to out-of-plane movements", *Eng. Struct.*, **190**, 210-222. https://doi.org/10.1016/j.engstruct.2019.04.019

Li, S., Guo, Y., Xu, Y. and Li, Z. (2019), "Real-time geometry identification of moving ships by computer vision techniques in bridge area", *Smart Struct. Syst., Int. J.*, **23**(4), 359-371. https://doi.org/10.12989/sss.2019.23.4.359

Li, W., Zhao, W., Gu, J., Fan, B. and Du, Y. (2022), "Dynamic characteristics monitoring of large wind turbine blades based on target-free DSST vision algorithm and UAV", *Remote Sens.*, **14**(13), 3113. https://doi.org/10.3390/rs14133113

Liang, X. (2019), "Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization", *Comput.-Aided Civil Inf.*, **34**(5), 415-430. https://doi.org/10.1111/mice.12425

Lin, Z., Cevasco, D. and Collu, M. (2020), "A methodology to develop reduced-order models to support the operation and

maintenance of offshore wind turbines", *Appl. Energy*, **259**, 114228. https://doi.org/10.1016/j.apenergy.2019.114228

Lydon, D., Taylor, S.E., Lydon, M., del Rincon, J.M. and Hester, D. (2019), "Development and testing of a composite system for bridge health monitoring utilising computer vision and deep learning", *Smart Struct. Syst.*, *Int. J.*, **24**(6), 723-732. https://doi.org/10.12989/sss.2019.24.6.723

Mao, Q.C., Sun, H.M., Liu, Y.B. and Jia, R.S. (2019), "Mini-YOLOv3: real-time object detector for embedded applications", *IEEE Access*, **7**, 133529-133538. 10.1109/ACCESS.2019.2941547

Narazaki, Y., Hoskere, V., Eick, B.A., Smith, M.D. and Spencer Jr., B.F. (2019), "Vision-based dense displacement and strain estimation of miter gates with the performance evaluation using physics-based graphics models", *Smart Struct. Syst.*, *Int. J.*, **24**(6), 709-721. https://doi.org/10.12989/sss.2019.24.6.709

Narazaki, Y., Hoskere, V., Yoshida, K., Spencer Jr., B.F. and Fujino, Y. (2021), "Synthetic environments for vision-based structural condition assessment of Japanese high-speed railway viaducts", *Mech. Syst. Signal Process.*, **160**, 107850. https://doi.org/10.1016/j.ymssp.2021.107850

Onat, O. and Gul, M. (2018), "Application of artificial neural networks to the prediction of out-of-plane response of infill walls subjected to shake table", *Smart Struct. Syst.*, *Int. J.*, **21**(4), 521-535. https://doi.org/10.12989/sss.2018.21.4.521

Perry, B.J. and Guo, Y. (2021), "A portable three-component displacement measurement technique using an unmanned aerial vehicle (UAV) and computer vision: a proof of concept", *Measurement*, **176**, 109222. https://doi.org/10.1016/j.measurement.2021.109222

Shao, Y., Li, L., Li, J., An, S. and Hao, H. (2021), "Computer vision based target-free 3D vibration displacement measurement of structures", *Eng. Struct.*, **246**, 113040. https://doi.org/10.1016/j.engstruct.2021.113040

Song, Q., Wu, J., Wang, H., An, Y. and Tang, G. (2022), "Computer vision-based illumination-robust and multi-point simultaneous structural displacement measuring method", *Mech. Syst. Signal Process.*, **170**, 108822. https://doi.org/10.1016/j.ymssp.2022.108822

Sony, S., Laventure, S. and Sadhu, A. (2019), "A literature review of next-generation smart sensing technology in structural health monitoring", *Struct. Control Health Monit.*, **26**(3), e2321. https://doi.org/10.1002/stc.2321

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, S., Wang, T., Yang, H. and Chu, F. (2022), "Damage identification of wind turbine blades using an adaptive method for compressive beamforming based on the generalized minimax-concave penalty function", *Renew. Energ.*, **181**, 59-70. https://doi.org/10.1016/j.renene.2021.09.024

Tian, Y., Zhang, J. and Yu, S. (2019a), "Vision-based structural scaling factor and flexibility identification through mobile impact testing", *Mech. Syst. Signal Process.*, **122**, 387-402. https://doi.org/10.1016/j.ymssp.2018.12.029

Tian, Y., Zhang, J. and Yu, S. (2019b), "Rapid impact testing and system identification of footbridges using particle image velocimetry", *Comput.-Aided Civil Inf.*, **34**(2), 130-145. https://doi.org/10.1111/mice.12390

Wu, D., Lv, S., Jiang, M. and Song, H. (2020), "Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments", *Comput. Electron. Agric.*, **178**, 105742. https://doi.org/10.1016/j.compag.2020.105742

Xu, D., Wen, C. and Liu, J. (2019), "Wind turbine blade surface inspection based on deep learning and UAV-taken images", *J. Renew. Sustain. Energy*, **11**(5), 053305. https://doi.org/10.1063/1.5113532

Xu, D., Liu, P.F. and Chen, Z.P. (2021), "Damage mode identification and singular signal detection of composite wind turbine blade using acoustic emission", *Compos. Struct.*, **255**, 112954. https://doi.org/10.1016/j.compstruct.2020.112954

Ye, X.W., Dong, C.Z. and Liu, T. (2016), "Image-based structural dynamic displacement measurement using different multi-object tracking algorithms", *Smart Struct. Syst.*, *Int. J.*, **17**(6), 935-956. https://doi.org/10.12989/sss.2016.17.6.935

Ye, X.W., Jin, T. and Yun, C.B. (2019), "A review on deep learning-based structural health monitoring of civil infrastructures", *Smart Struct. Syst.*, *Int. J.*, **24**(5), 567-585. 10.12989/sss.2019.24.5.567

Yu, Y., Cao, H., Yan, X., Wang, T. and Ge, S.S. (2020), "Defect identification of wind turbine blades based on defect semantic features with transfer feature extractor", *Neurocomputing*, **376**, 1-9. https://doi.org/10.1016/j.neucom.2019.09.071

Zhang, Z. (2000), "A flexible new technique for camera calibration", *IEEE Trans. Pattern Anal. Machine Intell.*, **22**(11), 1330-1334. DOI: 10.1109/34.888718

Zhao, X., Liu, H., Yu, Y., Zhu, Q., Hu, W., Li, M. and Ou, J. (2016), "Displacement monitoring technique using a smartphone based on the laser projection-sensing method", *Sens. Actuator A Phys.*, **246**, 35-47. https://doi.org/10.1016/j.sna.2016.05.012

Zhou, Y., Zhou, S., Hao, G. and Zhang, J. (2021), "Bridge influence line identification based on big data and interval analysis with affine arithmetic", *Measurement*, **183**, 109807. https://doi.org/10.1016/j.measurement.2021.109807

Zhu, L., Geng, X., Li, Z. and Liu, C. (2021), "Improving yolov5 with attention mechanism for detecting boulders from planetary images", *Remote Sens.*, **13**(18), 3776. https://doi.org/10.3390/rs13183776

BS