Journal of Computers Vol. 34 No. 2, April 2023, pp. 19-27
doi: 10.53106/199115992023043402002

Surface Defect Recognition of Wind Turbine Blades Based on Improved YOLOX-X Model

Changhao Dong[1], Chao Zhang[1*], Jianjun Li[2], Jiaxue Liu[1]

[1] College of Mechanical Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China
Dong812988@gmail.com

[2] College of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China
xidianjj@163.com

Received 27 April 2022; Revised 22 July 2022; Accepted 10 August 2022

Abstract. In order to solve the problem of small data sets and small detected targets in image detection of wind turbine blades. In this paper, we propose an improved YOLOX-X model. Firstly, we use a variety of data set enhancement methods to solve the problem of small data sets. Secondly, an improved Mixup image enhancement method is proposed to enrich the image background. Then, the attention mechanisms of ECAnet and CBAM are introduced to improve the attention of important features. Furthermore, the IOU_LOSS loss function in the original model is replaced with CIOU_LOSS in this paper to improve the positioning accuracy of small target. Last but not least, the overall network uses the Adam optimizer to accelerate network training and recognition. The effectiveness of algorithm is evaluated on a data sets captured by a UAV in a wind farm. Compared with the original YOLOX-X model, our algorithm improves mAP by 4.55%. In addition, compared with other types of YOLO series networks, it is proved that our model is superior to other algorithms.

Keywords: YOLOX-X, deep learning, machine vision, object detection

1 Introduction

Energy is an indispensable material basis for human daily life and industrial production. Wind energy has the characteristics of being renewable and less damaging to the environment. It has attracted the attention of various countries. However, wind energy also has some problems. Wind farm must be built in specific places and it requires large amounts of money for investment and maintenance. The surface of the wind blade will be damaged due to long-term operation, lightning, sandstorm and other harsh natural factors. Due to the long interval between two overhauls, a small damage can cause a big damage if the small defect is not detected and repaired in time. Early detection and resolution of problems can save a lot of money. This makes early detection of defects extremely important. However, traditional manual detection methods are characterised by long time, low efficiency and inaccurate visual recognition. So the machine vision is a good option to solve this problem.

In recent years, many scholars have contributed to the recognition of surface defects of wind turbine blades. Tang et al. [1] proposed a method for surface defect detection of wind turbine blades based on Convolutional Neural Network feature fusion of local binary pattern features and kernel extreme learning machine. Li [2] proposed an adaptive threshold segmentation background method. Meanwhile, the migration learning principle is also applied to an Mask-RCNN network and a method combining traditional image processing with deep learning is proposed to identify the defect of fan blade surface.

In this paper, we introduce a new network YOLOX. This network was proposed by Megvii company [3] in 2021 and has been widely used in various fields due to its excellent competitiveness. Some scholars have used and studied different types of YOLOX networks in different field. Zhang et al. [4] found that there are few studies on the whole fruit count of whole trees. A method for fruit complete yield counting based on YOLOX is proposed. Wang et al. [5] proposed an improved combination network of YOLOX network and Harris algorithm. The combination network could solve the problem of camera shooting chessboard blur, heavy noise and distortion. Zhang et al. [6] find the existing target detection model difficulty for use in complicated fire scenarios. They proposed an improved YOLOX fire scenario detection model. The model detection performance was effectively improved by

* Corresponding Author

introducing a light attention module and channel shuffling techniques. T. Panboonyuen et al. [7] proposed a method of Transformer-Based YOLOX with FPN to detect road asset images. They improved it to 61.5% AP on the Thai highway corpus, outperforming the current best-practice (YOLOv5L) by 2.56% AP on the test-dev data set.

In order to improve the accuracy of small target detection while lacking data set. This paper proposes an improved YOLOX model with YOLOX-X as the basic network framework by a variety of data set enhancement algorithms; an improved Mixup method is proposed and introduced; an attention mechanism is introduced to improve the ability of model to extract small target defect information; the model training loss function is optimized; and a model optimizer is introduced. Excellent detection performance is achieved on the wind turbine blade surface defect data set.

2 The Brief of YOLOX-X Model

The YOLOX model is based on the YOLOv3 model [8] to add an effective tricks for improvement. The paper use the YOLOX-X model in this paper is an x-large version of the original model, with a huge number of network layers and the slowest in training time, but also the most accurate network model. The YOLOX-X model consists of four parts. Input, Back-bone, Neck and Prediction. As is shown in Fig. 1.

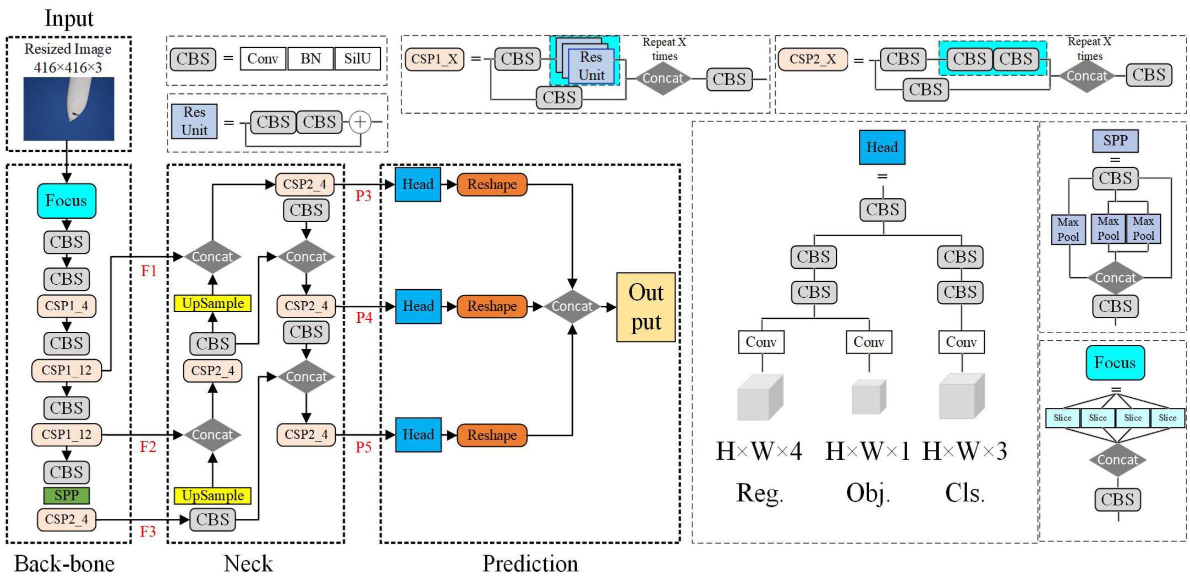

Fig. 1. YOLOX-X model network structure diagram

In the Input part, the original model input image of 640×640×3 was changed to 416×416×3 to unify the input image sizes of different models for a more accurate comparison with different types of YOLO models.

In Back-bone, the Focus layer [9] consists of the slice operation and the CBS, which can guarantee the low parameter and images will be fully trained. The CBS is a combination of Convolution, Batch Normalization and SilU activation function. The SilU activation function has the characteristics of no upper bound and lower bound, smooth and non-monotonic. The Modified CSPnet [10] is used in the YOLOX model to extract deep semantic information from input images. The SPP layer consists of a maximum pooling operation with pooling kernel sizes of 5×5, 9×9, 13×13 and the base convolution. It can effectively increase the receptive field.

In Neck, FPN [11] and PAN methods are used to merge shallow features information with deep features information. The feature layers from shallow to deep are F1, F2, F3 and the output feature maps are P3, P4, P5.

In Prediction, the YOLOX head is called Decoupled head. It is consists of a shared CBS, two additional CBSs for each branch and a separate convolution for each task. The Reg. is used to predict the offset of object. The Obj. is used to determine whether the feature point has a corresponding object. The Cls. is used to determine the type of object contained in each feature point. In the last, stacking the three predictions results and send these to the Output.

For the overall model, the YOLOX uses the Mosaic enhancement method. When the model is trained, the images will be rotated, zoomed and gamut converted. The four images then combined into a single image. The purpose of this is: firstly, it can effectively enrich the background of the detected object. Also it can make the detection result more accurate. Secondly, when the Batch Normalization works. The Mosaic method can efficiently improve the training speed. As is shown in Fig. 2.

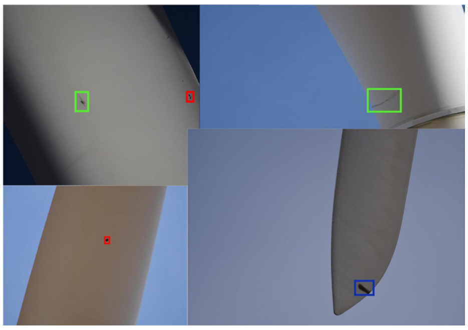

Fig. 2. Mosaic data augmentation

Although, the YOLOX-X model performs well in normal images detection, it still has the following problems on the defect of wind turbine blade surface images detection.

(1) Due to the special factor of wind farms. The number of wind farms is small and most of the surface defects of the wind blade can be removed by the staff in time, the defect images are not saved at scale. These factors lead to the lack of surface defect images of wind turbine blades, which brings difficulties to deep learning.

(2) Although the YOLOX-X model has a deep network. The overall network model has a relatively small number of channels in a single layer, which is not enough to retain the information in the small target defect images. Especially after successive convolution, important detail information is seriously lost. It can affect the judgment of small targets.

(3) Although there is no need to consider the issue of FPS, the overall network training time and recognition time of the whole network are too long, which will still cause problems in detection.

3 The Improved YOLOX-X Model

In order to solve the problem of the lack number of data sets and the difficulty of detecting small defects, while considering the training time and recognition speed factors. The following improvements are made to the YOLOX-X model:

Firstly, the total number of data set is enlarged using multiple data set augmentation methods. The Mosaic method from the original YOLOX model is applied to the expanded data set. A modified Mixup image enhancement method is used in the YOLOX model to enrich the image background and improve the detection accuracy.

Secondly, without changing the model size and image channels, the ECAnet attention mechanism, the CBAM (Convolutional Block Attention Module), which combines channels and space into one, is introduced to pay more attention to the effective information of the target and improve the detection accuracy of small targets.

Then, the loss function of network training is optimized. The CIOU loss function replaces the bounding box coordinate prediction loss term in the original network IOU loss function. It can improve the accuracy and convergence speed of the model for defect localization.

Finally, the Adam optimizer is used for the overall network to obtain the optimal network parameters while accelerating the network training speed and recognition speed.

3.1 Data Set Enhancement

The number of original images is 198, including three types of images, stained, scratched and holes. Using the random rotation, zooming, flipping, brightness and contrast enhancement methods to enlarge the number of images to 500. This makes the model more random and improves the generalization ability.

Meanwhile, an improved image enhancement method Mixup [12] is introduced into the YOLOX-X model. The principle of the Mixup method is a combination of two images overlapping to reorganize the image. In common, the Mosaic and Mixup methods should be used separately when the model is trained. This can increase the training time. The improved Mixup method can reduce the training time and enrich the image background. The principle of the improved Mixup method is to use Mixup for processing after the Mosaic method has been completed. As is shown in Fig. 3.

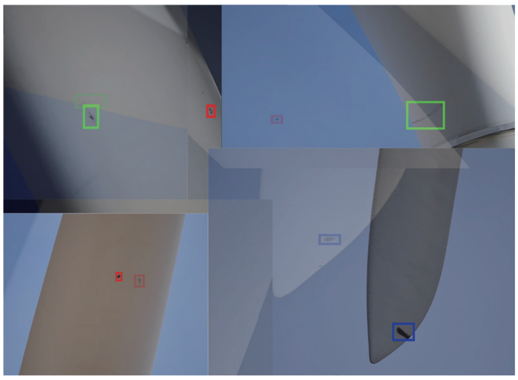

Fig. 3. An improved Mixup method

3.2 Attention Mechanism

Since the image pixels are 416×416. The pixels of the F1-F3 feature maps are 13×13, 26×26, 52×52, respectively, which can effectively identify the corresponding pixels of 32× 32, 16×16 and 8×8. But it is difficult to recognize pixels below 8×8. By introducing ECAnet block between Back-bone and Neck, F1, F2, F3 positions, and introducing CBAM block between Neck and Head, P5, P4, P3 positions to improve the detection ability of small targets [13]. The red marked part is the improved network structure, as is shown in the Fig. 4.

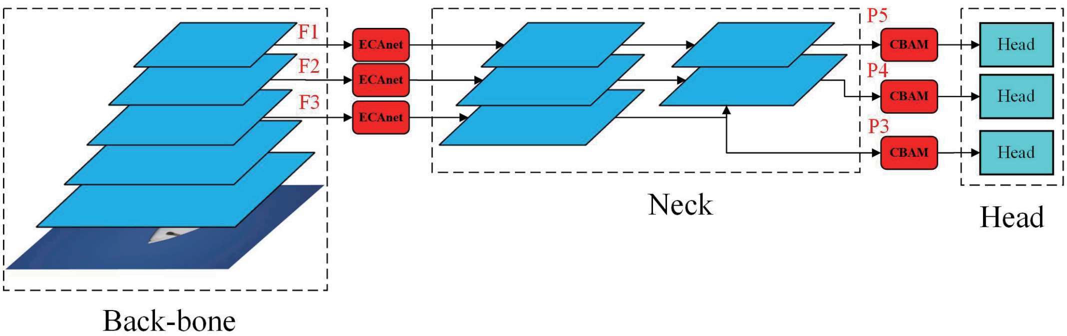

Fig. 4. An improved YOLOX-X method

First, adding ECAnet [14] between the Back-bone and Neck. After the image has undergone rough extraction of Back-bone features, the features are not very obvious. Adding ECAnet can effectively improve the feature extraction performance. As a channel attention mechanism module, the ECAnet module can explore the relevance of existing data and highlight important features. The performance can be greatly improved when only a small number of parameters are added and no need to change the number of model channels.

As a channel attention mechanism, ECAnet improves the defect of the previous channel attention mechanism SEnet. It proposes a local cross-channel interaction strategy without dimensional reduction and a method of adaptively selecting the size of one-dimensional convolution kernel to achieve performance optimization. As is shown in Fig. 5.

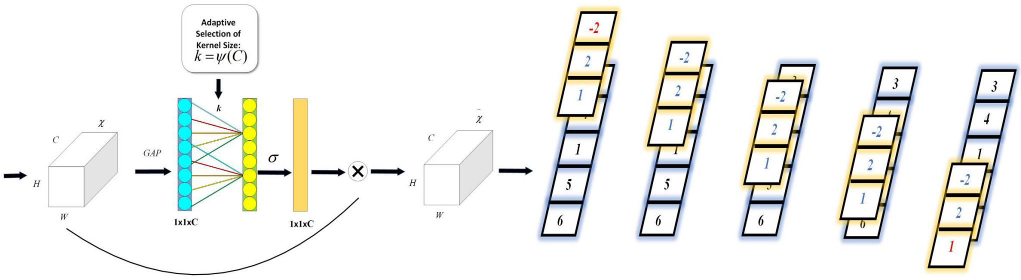

Fig. 5. ECAnet attention mechanism

Secondly, a lightweight attention module, the Convolution Block Attention Module (CBAM) [15] is added before the Neck structure after the Head structure. The CBAM is an attention mechanism module that combines spatial and channel. CBAM contains two independent sub-modules, the Channel Attention Module (CAM) and the Spatial Attention Module (SAM), which perform channel and spatial attention enhancement respectively. After the images has undergone feature extraction and feature merge of backbone and neck, the features are relatively obvious. Adding CBAM block can make the network parameters more accurate.

The entire CBAM process is divided into two parts, CAM and SAM. In CAM, the input feature map F(H×W×C) is subjected to global max pooling and global average pooling based on width and height to obtain two 1×1×C feature maps. Then, the feature maps are sent to a two-layer neural network MLP is implemented to realize feature transformation and feature extraction and finally generate channel attention feature. In SAM, the feature map F’ output by the channel attention module is used as the input feature map of the module. First, a channel-based global max pooling and global average pooling to get two H×W×1 feature maps. Then, the SAM performs the contact operation on the two feature maps based on the channels. After a 7×7 convolution operation, the dimensionality is reduced to 1 channel. Final, the CBAM get the final generated features. As is shown in Fig. 6.

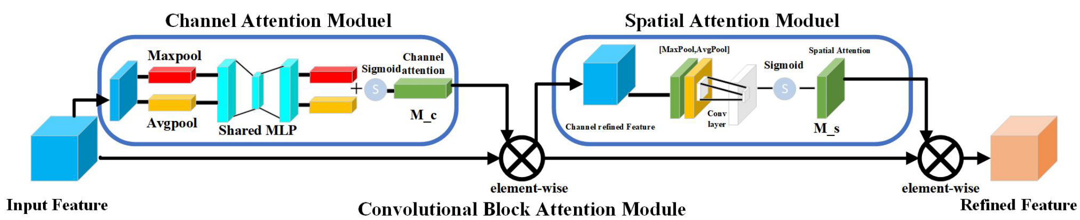

Fig. 6. The CBAM attention mechanism

3.3 Optimize Loss Function

In the YOLOX-X model. During the object detection training, it is found that the target localization loss converges slowly, especially for small targets in the images. Because the target is small, the ground-truth box and the predicted box are often overlapped and contained. This seriously affects the small target detection accuracy. The CIOU [16] loss function is an improved version of the IOU loss function. It introduce center position and aspect ratio error calculations, which solve the problem of overlapping with the ground-truth box and the predicted box. It makes the loss function to converge faster. The loss function is shown in formula (1).

$$CIoU = 1 - IoU + \frac{\rho^2(A,B)}{c^2} + \alpha v \tag{1}$$

The IoU (Intersection-over-Union) is the overlap rate between the ground-truth box and the predicted box. ρ denotes the shortest linear distance between two points. A and B are centroids of the two boxes. c denotes the diagonal distance of the smallest circumscribed rectangle between the ground-truth box and the predicted box.α denotes the coordination scale parameter, v is used to measure the aspect ratio consistency parameter of the box.

3.4 Adam Optimizer

The Adam (adaptive moment estimation) [17] is a popular algorithm in deep learning because it achieves good results. The Adam algorithm is differs from the traditional stochastic gradient descent. The SGD keeps a single learning rate updating all weights. The learning rate does not change during training. However, the Adam can design independent adaptive learning rates for different parameters. It can effectively improve the model recognition accuracy and reduce the training time and recognition time.

4 Experiment and Result Analysis

Model training and performance evaluation are done on GPU servers. The server hardware is configured as: Intel Core i7-7700HQ CPU(2.80GHz), RAM 16GB, the Graphic card is NVDIA GeForce GTX 1070. The soft environment: Python3.7.0, tensorflow_gpu-2.0.0, CUDA10.0.

In the experiment, an unpublished data set of surface defects of wind turbine blades photographed by UAV in the wind farm is selected to evaluate and test the performance of the algorithm. The original data set contains 198 images, which were expanded to 500. There are three categories, stains, scratches and holes. The detected targets of this data set are all small targets. The images are labeled with labelImg software and converted to VOC format for storage. After inputting the data set into the model, 90% of the expanded data set is used as the training set+validation set and the remaining 10% as the test set. When the 100 Epoch training is completed, the optimal model weights and loss function images are obtained. The accuracy of the model improves significantly in the initial stage, with large fluctuations in the loss function, and after 50 epochs, the accuracy and loss function levelled off. In the 90 epoch, the accuracy reaches its highest. In addition, with the continuous improvement of the model, the convergence speed and recognition accuracy are significantly improved. The loss curve image is shown in Fig. 7.

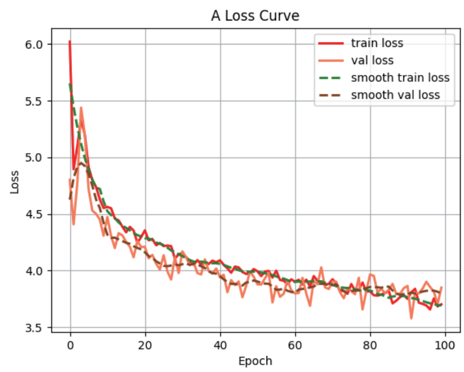

Fig. 7. The improved YOLOX-X loss curve image

The following six figures is shown in Fig. 8.
(1)YOLOX-X+ECAnet
(2)YOLOX-L+Mixup+CIOU+sgd
(3)YOLOX-X+Mixup+CIOU+sgd
(4)YOLOX-X+Mixup+CBAM+CIOU+Adam
(5)YOLOX-X+Mixup+ECAnet+CBAM+CIOU+Adam

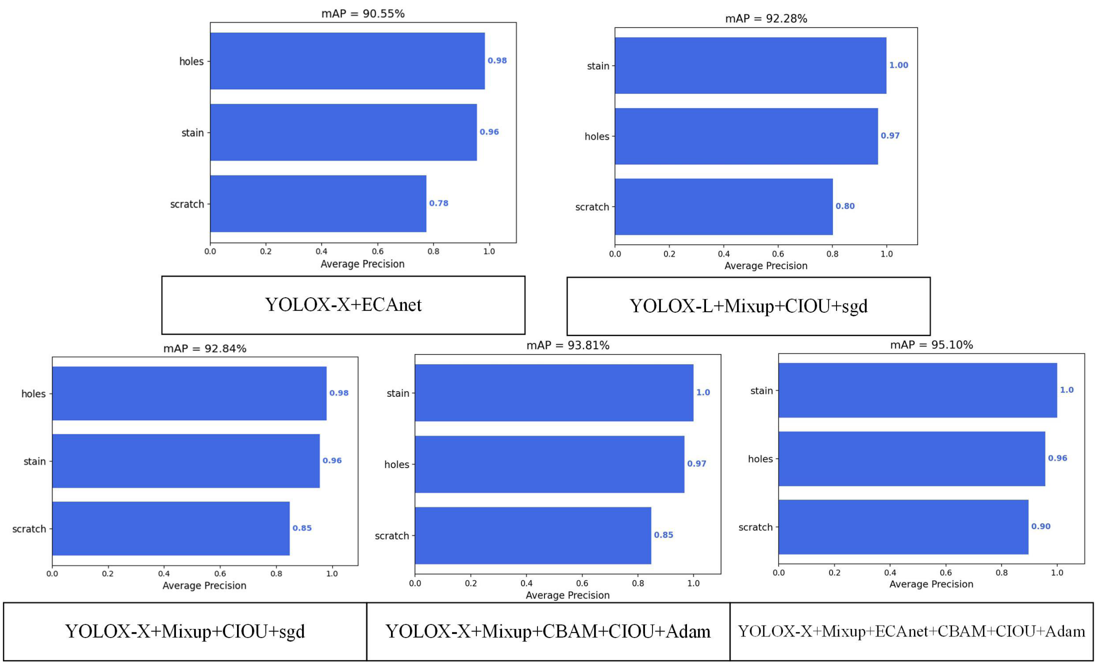

Fig. 8. Comparison of progressively improved models

According the Fig. 8, when the attention mechanism is added to the original model, the detection of small targets, especially the average precision of scratch images has a slight increase. After adding all the tricks, the overall mAP reaches the highest.

This paper uses the improved YOLOX-X model. Compared with other types of deep learning networks, this model is more competitive in image target recognition. It adds an attention mechanism to improve the recognition rate of small targets and focuses more on the recognition of small targets. When five tricks are added to the original YOLOX-X model, it can be seen that the recognition rate has improved compared to other models. After the 100 Epoch training, the model is fully trained and is especially competitive for small target recognition. Scratch defects are characterised by inhomogeneity and different sizes, and their AP fluctuates. Compared with other YOLO model, the improved YOLOX-X model has the highest recognition rate in scratch defect. The comparison for different models is shown in Table 1.

Table 1. Comparison of progressively improved models

Model	Stain/%	Holes/%	Scratch/%	mAP/%
YOLOX-X+ECAnet	96.08	98.04	77.53	90.55
YOLOX-L+Mixup+CIOU+sgd	99.92	96.77	80.15	92.28
YOLOX-X+Mixup+CIOU+sgd	95.91	97.88	84.73	92.84
YOLOX-X+Mixup+CBAM+CIOU+Adam	99.63	96.91	84.89	93.81
YOLOX-X+Mixup+ECAnet+CBAM+CIOU+Adam	99.57	95.89	89.84	95.10

The final recognition images is shown in Fig. 9:

Fig. 9. Image of surface defects of wind turbine blades identified

Experiments have shown that holes, scratches, stains and other problems on the surface of fan blades can be effectively and accurately identified. Even if there are more than two different defects on the same blade, they can be identified separately, which provides effective solutions for real-life practical problems.

5 Conclusion

Aiming at the problems of lack of data set and small detected targets in the detection of wind turbine blade surface defects, an improved YOLOX-X model is proposed. A dataset enhancement method is used to introduce the ECAnet and CBAM dual attention mechanism, optimizing the training loss function, and introducing the Adam optimizer. On the basis of not significantly enhancing the model volume, strengthen the ability of the model to extract features, improve recognition accuracy. In data set testing, compared with different YOLOX series model, the improved YOLOX-X network model achieves the highest accuracy of 95.10%. It lays a foundation for the task of identifying surface defects of wind turbine blades, and will continue to improve the overall network and improve the identification accuracy in the future.

6 Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 51965052); National Natural Science Foundation of China (Grant No. 51865045); Scientific Research Project of Higher Education Institutions of Inner Mongolia Autonomous Region (NJZY22114).

References

[1] Z. Tang, D. Sun, Y. Li, P. Lu, Defect detection model of wind turbine blade based on feature fusion, Electronic measurement technology 45(11)(2022) 161-166.

[2] M. Li, Detection of surface defects of wind turbine blades based on deep learning, [dissertation] Beijing: North China Electric Power University, 2021.

[3] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, YOLOX: Exceeding YOLO Series in 2021. <https://arxiv.org/abs/2107.08430>, 2021 (accessed 06.08.2021).

[4] Y. Zhang, W. Zhang, J. Yu, L. He, J. Chen, Y. He, Complete and accurate holly fruits counting using YOLOX object detection, Computers and Electronics in Agriculture (198)(2022) 107062.

[5] G. Wang, H. Zheng, X. Zhang, A robust Checkerboard Corner Detection Method for Camera Calibration Based on Improved YOLOX, Frontiers in Physics (9)(2021) 819019.

[6] J. Zhang, S. Ke, Improved YOLOX Fire Scenario Detection Method, Wireless Communications and Mobile Computing 2022(2022) 9666265.

[7] T. Panboonyuen, S. Thongbai, W. Wongweeranimit, P. Santitamnont, K. Suphan, C. Charoenphon, Object detection of road assets using transformer-based yolox with feature pyramid decoder on thai highway panorama, Information (Switzerland) 13(1)(2022) 5.

[8] J. Redmon, A. Farhadi, Yolov3: An Incremental Improvement. <https://arxiv.org/abs/1804.02767>, 2018 (accessed 08.04.2018).

[9] H. Zhang, Y. Ban, L. Guo, Y. Jin, L. Chen, Detection method of remote sensing image ship based on YOLOv5, Electronic Measurement Technology 44(8)(2021) 87-92.

[10] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, I.-H. Yeh, CSPNET: A New Backbone that Can Enhance Learning Capability of CNN, in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020.

[11] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proc. Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[12] H. Zhang, M. Cisse, Y. Dauphin, D. Lopez-Paz, Mixup: Beyond Empirical Risk Minimization, in: Proc. The 6th International Conference on Learning Representations (ICLR), 2018.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need. <https://arxiv.org/abs/1706.03762>, 2017 (accessed 06.12.2017).

[14] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zou, Q. Hu, ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks, in: Proc. Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[15] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: Convolutional Block Attention Module, <https://arxiv.org/abs/1807.06521>, 2018 (accessed 18.07.2018).

[16] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in: Proc. of the AAAI Conference on Artificial Intelligence (AAAI), 2020.

[17] D.P. Kingma, J. Ba, Adam: A method for Stochastic Optimization, in: Proc. The 3rd International Conference on Learning Representations (ICLR), 2015.