Research On Surface Loss Detection Algorithm Of Wind Turbine Blade Based On FRE-DETR Network

Jianwei Yang[1*], Xiaocong Chen[2], Shengxian Cao[3], Bo Zhao[3], Zhenhao Tang[3], Gong Wang[3], Xingyu Li[3], and Han Gao[3]

[1] *Zhongdian Huachuang Electric Power Technology Research Co., Ltd., 215123, Suzhou, China.*
[2] *Hubei Zhongdian Chunyangshan wind power Co., LTD., 430040, Wuhan, China.*
[3] *Northeast Electric Power University, 132012, Jilin, China.*
[*] *Corresponding author. E-mail: aaayjw@163.com*

Received: Dec. 30, 2024; Accepted: Feb. 25, 2025

Wind turbine generators operate in harsh areas for a long time, resulting in frequent problems such as blade breakage, and traditional blade defect detection methods have low detection accuracy. In this paper, an end-to-end target detection algorithm FRE-DETR based on wind turbine blade defects is designed, and the detection speed and detection accuracy of the end-to-end detection model are further improved by redesigning the feature extraction location in the backbone network and proposing a feature selection and fusion module. FRE-DETR is tested on a wind turbine blade defect dataset, and the results show that the model improves the detection accuracy by 2% compared with RTDETR-R18. The inference speed is already higher than RTDETR-R18 when the step size is larger than 2. The Gflops of the model is only 66.8% of that of RTDETR-R18, which also greatly reduces the computational requirements of the hardware when deployed. FRE-DETR meets the requirements of real-time detection.

Keywords: Integrated energy system; End-to-end algorithm; Wind Turbine Blade Defect; Target Detect; Computer Vision

http://dx.doi.org/10.6180/jase.202512_28(12).0002

1. Introduction

In recent years, with the continuous development of the new energy industry, wind energy, solar energy, nuclear energy, and other new energy industries have significantly contributed to promoting sustainable economic development. Wind energy has attracted people's attention with its clean and renewable characteristics and has become a mainstream new energy industry. Wind turbine blade is the most indispensable component in wind power generation equipment [1]. When the blades of a wind turbine are defective, it impacts the turbine's power generation efficiency and, if left unrepaired, can lead to costly downtime. In such cases, the only solution may be to replace the blades instead of repairing them. This will cause significant losses to the economic benefits of the wind turbine. The operation of the wind turbine causes great safety hazards [2]. Therefore, the preliminary inspection and repair of wind turbine blades is essential.

Visual inspection in non-destructive wind turbine blade defect detection testing has greater advantages in terms of real-time, high efficiency and safety [3]. Visual inspection of wind turbine blade defects is divided into infrared light state and visible light state detection. Focusing on a single category restricts infrared detection capabilities and may obscure visible cracks and other defects in wind turbine blades, as external flaws can often be misleading. Hence, its development speed is relatively slow compared to the visible light state defect detection.

Visual inspection of blade defects in the visible light state has more detection categories and high detection ef-

ficiency, and defect types are more intuitive and clearer. Since the emergence of machine learning, many scholars have focused on wind turbine blade defects under the visual detection of wind turbine crack defects described by Haar features by training traditional cascade classifiers [4]. However, similar classifiers that classify by cascade often only recognize and classify one class, which makes it difficult to meet the detection requirements at this stage. Subsequently, with the emergence of deep learning, various neural networks represented by ResNet and mask rcnn are proposed to detect blade defects [5, 6], achieving excellent detection results. Still, the depth of the ResNet and mask rcnn networks is too deep, resulting in a model volume that is too large. The hardware requirements are also high when detecting, making it difficult to deploy to the UAV for real-time inspection functions.

With the continuous development of target detection, faster and more accurate YOLO series of two-stage target detection algorithms based on the CNN framework have been proposed. Many scholars have made targeted improvements to the detection algorithms of the YOLO series for the characteristics of each wind turbine blade defect type. Among them, Liu's research introduces a model tailored for detecting defects in wind turbine blades, addressing issues such as paint removal and small target leakage. By incorporating an attention mechanism, Liu enhances the feature pyramid structure, effectively minimizing the risk of overlooking small targets during the feature extraction phase. This innovation maximizes the use of small target information, resulting in a notable improvement in detection accuracy [7–11]. Furthermore, Liu's work demonstrates the implementation of a lightweight model to replace the original large-volume network module, leading to a more efficient YOLO detection model [12–15]. Additionally, redundant channels are removed through channel pruning of the original network, significantly reducing the model's size. By applying a distillation process, the larger model informs the learning of the pruned model, allowing for much higher detection accuracy and reduced processing speed. High detection accuracy and speed reduction, great edge equipment in the UAV on the wind turbine blade defects detection speed. YOLO algorithm, however, realizes the real-time detection task of the wind turbine blade defects. The types of fan blade defects are numerous and varied. When the YOLO detection algorithm is tasked with classifying a wide range of datasets and categories, it struggles with the long-term memory of the dataset. This limitation can lead to fatigue, negatively impacting detection and classification accuracy.

With the continuous development of neural networks, the network model represented by Transformer, which has long-term memory and is used to process large datasets, has been used in visual detection, which has a large computational complexity, but it improves the yolo algorithm's insufficiency in dealing with a large number of datasets through its efficient and excellent detection accuracy. Firstly, the swin-transformer [16], which represents a very large volume of target detection algorithm, obtained very high detection accuracy in coco datasets. Subsequently to improve the inference speed of visual transformer-based target detection algorithms, many lightweight detection networks based on multi-head attention with FFN feed-forward networks represented by efficientNet [17] were proposed. This greatly facilitates the development of visual transformer model-based deployment in edge devices. Baidu proposed an RTDETR [18] real-time detection model, which only encodes deep features, greatly reducing the computational complexity and volume of the model, which also makes the deployment of DETR detection models in edge devices a reality. Xinlin Liu took the lead in utilizing the RTDETR detection model to detect insulator defects [19], and obtained excellent detection speed and accuracy. Guemas et al. [20] used the RTDETR detection model to detect and classify Plasmodium falciparum with excellent results. Therefore, in the field of wind turbine blade defect detection, the types of blade defects will become more and more diverse in the future, which puts forward higher requirements on the number of datasets, and also puts forward higher requirements on the long-term memory of the algorithms, so the application of RTDETR in the detection of defects in wind turbine blades is also particularly important, and it will also become a future development trend.

1. This paper proposes a new DETR real-time detection model to solve the problems of difficult deployment, slow inference, and low detection accuracy of large transformer-based models.

2. In this paper, the FasterNet network is improved by fusing the heavy parameter operation with it, which substantially improves the model's ability for feature extraction.

3. In this paper, a lightweight feature selection and fusion attention module, SMLCA, is proposed to improve the edge feature extraction of wind turbine blade defects and enhance the completeness of large target detection.

2. Methods

RTDETR model, which encodes only the last layer of deep input features, thus greatly improving the processing speed of the algorithm and reducing the computational complexity. Shallow input features have rich target information but less feature extraction for background information. However, the last layer of deep features contains more complex and detailed semantic information, which significantly enhances the ability of the transformer to optimize background recognition and feature performance.

A new multi-head attention module, AIFI, is proposed to enhance the linkage of global feature information, which encodes features to generate tokens through a sliding window and integrates global feature relevance through softmax function the calculation process is shown in Eq. (1), where QKV are the model query, key values, and weights, respectively, and subsequently generates positional encoding to further store the feature information in the long term using feature mapping, thus realizing the long term memory of defective features. A new perceptual query has been proposed to optimize RTDETR for target feature extraction and detection. This approach introduces IoU (Intersection over Union) perceptual constraints, which help generate high classification scores for features with high IoU scores while assigning low classification scores to features with low IoU scores during training. As a result, this method enhances detection accuracy. The anchor frame threshold filtering and NMS post-processing are eliminated in the inference stage, improving the inference speed of RT-DETR [21].

$$Attention(Q, K, V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

Fig. 1 shows the improved RTDETR network structure, and a lighter backbone network is chosen for the network redundancy problem to reduce the computational requirements of RTDETR with edge detection devices and increase the inference speed of RTDETR. Meanwhile, we believe it is more effective to use Reparametric Convolution RepConv for feature extraction in the backbone network and introduce RepConv for feature extraction in FasterNet. In addition, to further improve the algorithm's detection performance, a new module is proposed for the final selection and fusion of features in the neck network. The final detection output of the target box is still realized by the combination of perceptual query and multi-head attention mechanism, which transforms the detection task into an unordered sequence output by matching Q, K and V in a process like the DINO detection header.

The AIFI Multihead Attention Module first processes the Head for deep features, which are subsequently processed by the Efficient Feature Selection and Fusion Module to be fed to the final Decoder Detection Head.

2.1. FRNet backbone

To further increase the inference speed of RTDETR, reduce the computational complexity of RTDETR, and improve the edge deployment capability of RTDETR, the lighter-weight Fasternet is adopted as the backbone network of RTDETR. The structure of Fasternet [22] is shown in Fig. 2. First, PConv and pointwise convolution form PWConv, then BN normalization and GLEU neuron activation are performed, which improves the generalization ability of the model in feature extraction and enhances the feature representation. In PConv, target and background features are extracted by slicing the number of channels and convolving some of them, which is insufficient for feature extraction using ordinary convolution. Therefore, using REPConv for branching feature extraction in PConv seems more efficient. PConv selectively extracts a subset of features while leaving the other features unprocessed. This method diminishes the computational demands on the neural network and reduces memory access during processing, lowering latency. In comparison, RepConv begins by utilizing a multi-branch convolutional layer to capture diverse features throughout training, with each branch focusing on learning particular characteristics. We combine the two so that the main branch network fully extracts features while reducing the number of parameters of the neural network, reducing the computational redundancy of the network, and increasing the inference speed of the network. A unique set of parameters is learned and continuously optimized during training to minimize the loss function, thus improving training results. Reparametrizing multiple branches as master branches during the inference process ensures the accuracy of inference, avoids the additional computational overhead of multiple branches, and effectively guarantees low latency.

2.2. Efficient feature extraction and fusion module

The features are stacked by convolution to stack the number of channels continuously. Thus, the feature map carries more and more feature information, so choosing how to utilize these maps efficiently is particularly important. To solve this problem, attention mechanisms have emerged to address the feature information clutter situation, ensuring that the neural network always focuses on the more critical information and ignores redundant information. Both traditional channel attention and spatial attention have their shortcomings. For example, the SE [23] attention mechanism compresses all channels into a single value, subse-

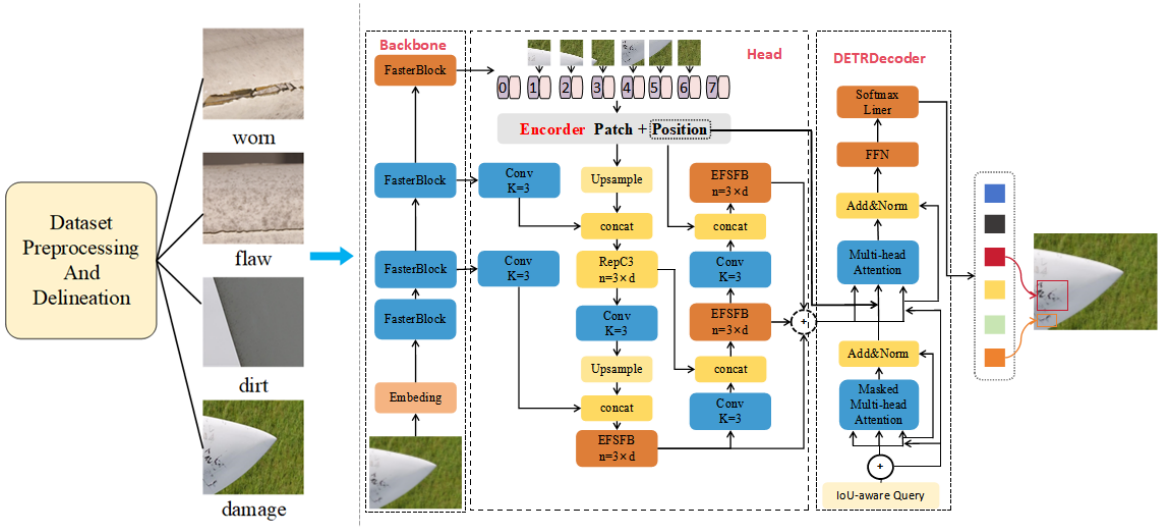

Fig. 1. FRE-DETR network structure. The backbone consists of a modified FasterBlock stack.

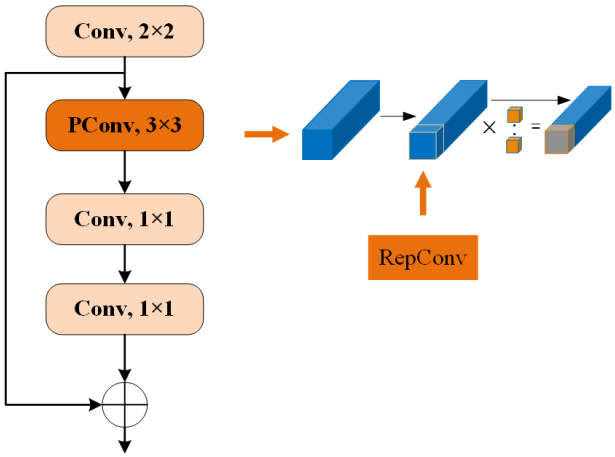

Fig. 2. Shows the composition of the FasterNet Block module on the left, while the structure of PConv and the computational flow are demonstrated on the right.

quently focusing only on information with higher weights, but this is undesirable in wind turbine blade defect detection, for which the edge information of blade defects is also very important in carrying out target detection. In contrast, for the extraction of edge information, the spatial attention mechanism seems more efficient to pay attention. Therefore, combining both characteristics for feature enhancement tends to provide superior performance. In MLCA [24], attention strengthens the channel feature information and introduces local spatial information to strengthen the edge feature extraction ability for defects. We introduce local feature extrusion learning in MLCA to enhance further the ability of the channel attention to focus on small targets in response to the characteristics of the category size span in the wind turbine blade defect data set.

The computational process of this channel attention is shown in Fig. 3(a). First, local spatial information is extracted through local pooling. One path sends this information to global pooling for one-dimensional feature extraction and local feature compression learning. In contrast, the other path reconstructs it to create ks*ks one-dimensional vectors for feature extraction from the local features. The information fusion was carried out after the inverse pooling operation of the two ways. The k-values in the feature extraction of all the one-dimensional convolutions are computed in the way proposed in ECA [25], the calculation process is shown in Eq. (2). The fused feature map is multiplied point by point with the original input feature map to get the final output feature map. The whole SMLCA calculation process is shown in Eqs. (2) and (3).

$$ks = \phi(C) = \left| \frac{\log_2^{(C)}}{r} + \frac{b}{r} \right|_{odd} \tag{2}$$

$$Fc = CA(F) \otimes F \tag{3}$$

$$\begin{aligned} CA(F) =& \sigma(LAvgPool(GAvgPool)) \\ &+ \sigma(LAvgPool(LAvgPOOL)) + S \end{aligned} \tag{4}$$

where C is the number of channels, k is the convolution kernel size, and b and r are hyperparameters. The value of k is k when k is odd and plus 1 when it is even. F is the input feature and Fc is the SMLCA output feature, where POOL is the pooling operation and S is the localized squeeze learning output.

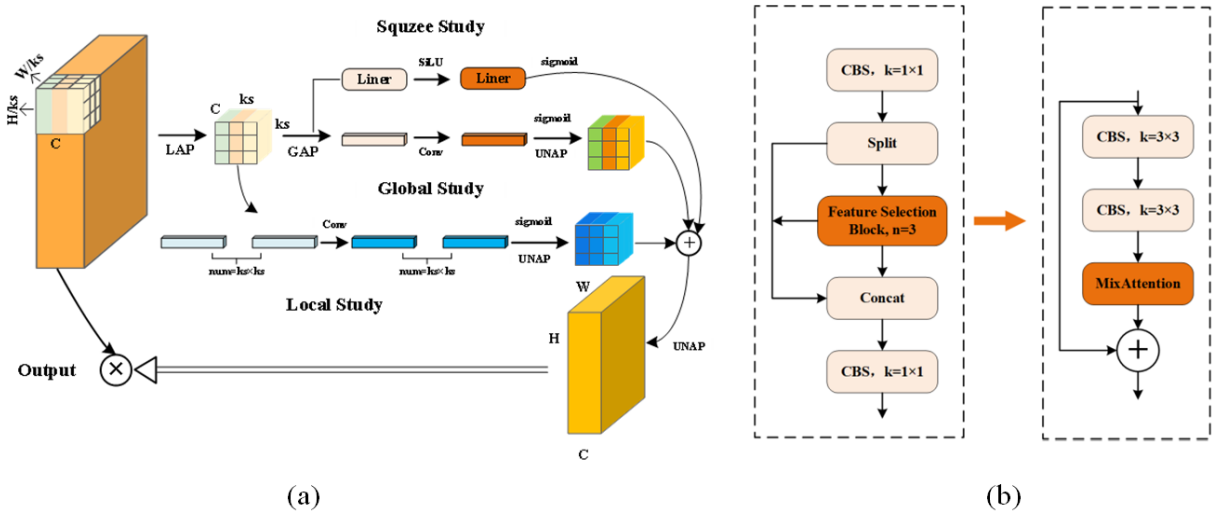

Fig. 3. Structure of the efficient feature extraction and fusion module and its internal structure. (a)SMLCA network structure. The features are learned in three ways to improve the ability to extract defective feature edges. (b)Structure of the efficient feature extraction and fusion module.

The structure of the efficient feature extraction and fusion module is shown in Fig. 3(b). We first use two 3×3 ordinary convolutions for deep feature extraction in the Feature selection Block and then introduce MLCA for feature selection to improve the attention to important features, reduce the influence of noise, and reduce the interference of other redundant information such as the background, and the residual connection is used at the end of the network to further improve the ability to utilize the features and to improve the expressive ability of the neural network [26, 27]. Subsequently, the three-layer Feature selection Block is connected in series to the Feature fusion Block to complete the network module, which is then used in the output fusion layer before the DETR decoder to enhance the decoder's attention to the important features of the blade defects as well as to enhance the network's ability to represent the features at a finer granularity level, Model representation capability.

3. Result discussions

3.1. Dataset

The dataset comes from a video of defective wind turbine blades taken by an unmanned aerial vehicle provided by Roboflow, for which we took screenshots and captured defective wind turbine blades in the blade chamber, some pre-processing operations [28–30] like image noise reduction are performed on the image, in addition to a dataset of defects in wind turbines presented by Foster et al. [31] and Shihavuddin et al. [32]. These two datasets have a total of 2900 images and the defect types include four categories: damage, dirt, troubles and defects. The ratio of training and validation sets is 4:1.

3.2. Training details

The experimental CPU for this paper's wind turbine blade defect dataset is a 16 vCPU Intel(R) Xeon(R) Platinum 8350C CPU @ 2.60GHz, GPU is an RTX 3090(24GB) * 1, PyTorch version is 1.11 .0 , Python version is 3.8 , and the operating system is Linux (ubuntu 20.04), Cuda version 11.3. the optimizer uses AdamW, batch is 20, epoch is 250, workers are 4 , initial learning rate is 0.0001 , momentum is 0.9 , and img size is 640×640. The optimizer uses a batch of 20 , an epoch of 250 , workers of 4 , an initial learning rate of 0.0001 , a momentum of 0.9 , and an image size of 640×640. The optimizer uses an RTX 3090 (24GB) * 1.

3.3. Ablation experiments

The detection accuracy and model volume change performance of the FE-DETR proposed in this paper in the wind turbine blade defect dataset are shown in Table 1. Regarding detection accuracy, the detection accuracy of the improved FE-DETR model Precision is improved by 1.4%, and the average accuracy mAP0.5 is improved by 2%. Regarding model volume change, the number of parameters of the improved model is reduced to 63.15% of the original model. GFLOPS is also reduced to 66.84% of the original RT-DETR model, so the model's volume is greatly compressed. Higher detection accuracy is obtained with several parameters and computational complexity. The improved mixATTention method performs comprehensive feature ex-

traction for every section of the feature map. Incorporating partial spatial attention enhances the clarity of edge feature extraction. This improvement, in turn, boosts the detection of defects in wind turbine blades. The largest decrease in model size is achieved when only FasterNet is used as the backbone network, which also verifies that slicing the number of channels and performing feature extraction on only some channels can significantly reduce the computational and parameter counts of the model. However, decreasing the model's computational complexity will also impact the detection accuracy. As shown in Table 1, the model mAP50-95 decreased by 1.8%, so we optimized some of the convolutions in the PConv and used the REPConv for feature extraction. From the detection data, RepConv use multi-branch training, by widening the width of the network, to improve the model for the target learning ability, multiple convolutions at the same time on the target extraction effect is significantly better than the extraction effect of a Conv, and at the same time for the reasoning will be multiple convolutions merged into one which also ensures that the model inference speed.

Subsequently, targeted experiments on the effectiveness of the feature selection and fusion module were conducted to reflect the excellent performance of the improved model detection effect and comparative experiments were conducted using the previously proposed SE, ECA, and CBAM attention mechanisms. The results showed that the improved attention mechanism had the highest average accuracy, mAP50, and its parameters were lower than other attention mechanisms. It is proved that the improved attention mechanism greatly improves the average accuracy mAP50 of the model without increasing the number of high parameters. Compared with the original MLCA, we fuse the ideas of local squeeze learning and small scale fusion, so that the channels within the small scale are fully squeezed to learn a large amount of edge information of the image, which makes it easier for the model to learn the target information and reduces the interference of the background information.

The average accuracy mAP50 change of the model during training is shown in Fig. 4. Fig. 4(a) shows the comparison of model ablation experiments. After 200 rounds of training, the average accuracy mAP50 of FRE-DETR is consistently higher than that of the other models, which represents that FRE-DETR has the optimal detection accuracy. Fig. 4(b) shows the comparison test of the attention mechanism, and our attention mechanism with channel squeezing learning capability has the superior detection accuracy. All of the above are consistent with the final experimental results on the validation set. Fig. 4(c) and (d) shows the precision change curves of the model during the training process, which are not much different from each other except that a significant gap can be seen at the accuracy peak. Still, the model has a large Precision improvement compared to RTDETR18.

Table 3 shows the average accuracy inference results for each category. When using the Efficient Feature Selection and Fusion module, the detection model greatly improves the effectiveness of WORN and FLAW. The results show that the module learns the target's edge information through extrinsic learning and local space computation after extrinsic learning, and compresses the features in the small range channels so that each channel is enriched with feature information from all channels, while the small range channels contain a large amount of target edge information, which results in a substantial enhancement of the model's ability to focus on and process feature edge information. This enhancement improves the detection accuracy for larger targets and enhances the integrity of potentially defective targets.The detection results of target detection methods such as SSD and YOLO series are also shown in Table 3, where the computational volume of the model is measured using Gflops and the computational accuracy of the model is evaluated using the average precision mAP_50. It can be found that the proposed FRE-DETR exhibits excellent detection performance, and compared with the YOLO model with the same volume or large volume, our model is smaller in size and lighter in structure, and the model achieves the highest detection accuracy even with a small volume. Comparing YOLOv7 and YOLOv9, the detection accuracy is the same or even higher than that of YOLOv7 at 1/3 of the size of YOLOv7, which proves that our model is fully capable of detecting defects in wind turbine blades. We also conducted parameter optimization experiments on the local learning scale in the model attention module, and the experiments proved that the model detection effect is best when the local scale ks is 5.

3.4. FPS test experiment

To fully reflect the computational complexity of the model and evaluate the inference speed of the model, we tested the inference speed of the model at different step sizes, including 1, 4, 8, 16, and 32. thereby fully utilizing the graphics card's performance to obtain a more accurate inference speed. The inference speed of each improved model at various batch sizes is illustrated in Fig. 5. The original model demonstrates the highest frames per second (FPS) when the batch size is set to 1. However, when the batch size is 2 or greater, all the improved models exhibit an FPS that surpasses that of the RT-DETR18 model. Combined with the

Table 1. Results of model ablation experiments.

Methods	Parameters/M	GFLOPS	Precision	mAP50	mAP50-95
RT-DETR18	19876896	57.0	79.3	77.4	48.9
FasterNet-DETR	10811940	28.5	80.3	77.5	47.1
FR-DETR	10812770	28.5	80.8	78.3	48.8
FRE-DETR	12552464	38.1	82.9	79.4	48.9

Table 2. Attention comparison experiment.

Methods	Parameters/M	GFLOPS	Precision	mAP50	mAP50-95
SE	12552374	38.1	80.8	78.3	48.6
ECA	12533987	38.1	82.0	78.6	47.2
CBAM	12552824	38.1	81.1	78.7	50.1
MLCA	12552464	38.1	80.7	78.7	49.0
Mixatt	12552464	38.1	82.9	79.4	48.9

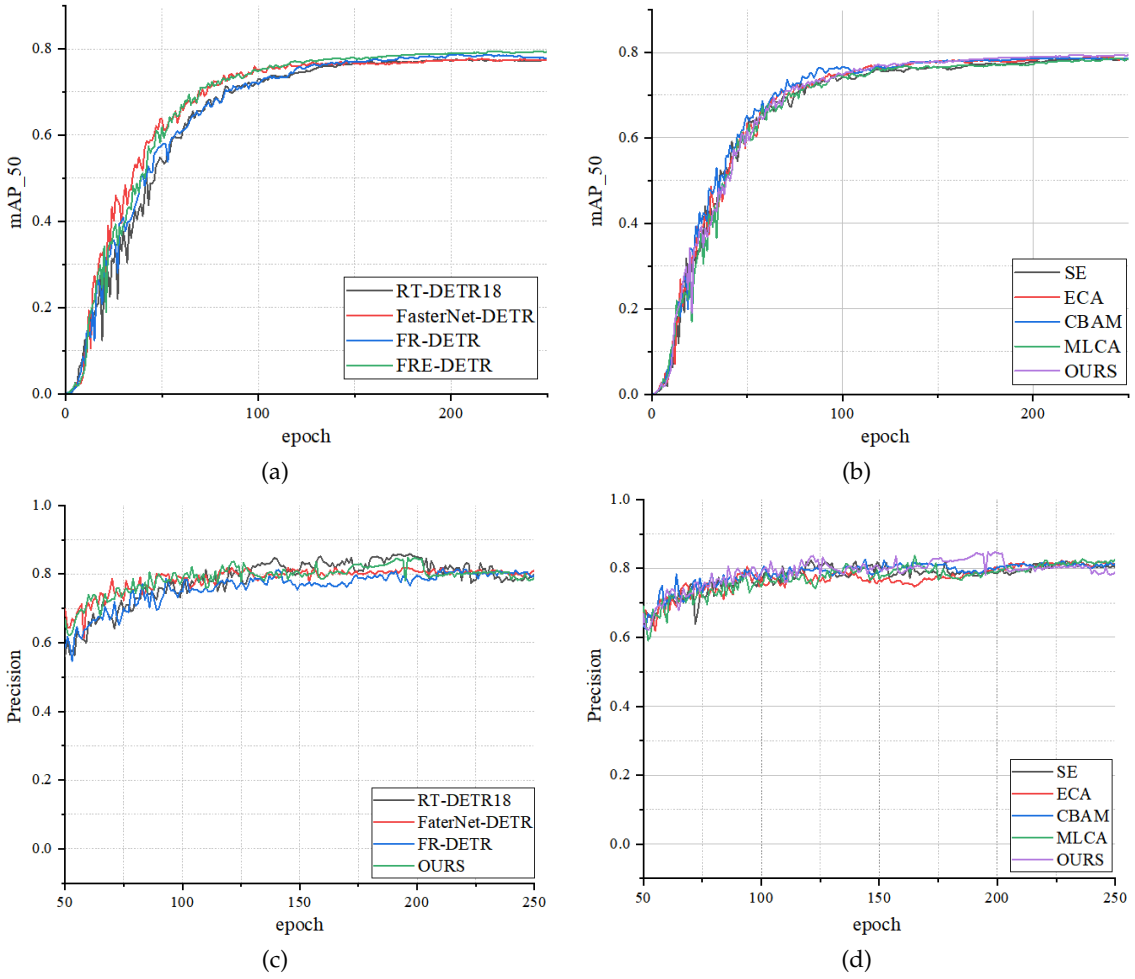

Fig. 4. Variation of model training process.

above reasoning accuracy experiments, it is proved that our detection models are already much higher than the original model in terms of detection accuracy and detection speed. Subsequently, inference speed tests were conducted for different convolutions. At batch size 1, the FPS is already higher than that of CBAM. With the increasing batch size, the improved attention mechanism has a similar inference speed to other classical attention mechanisms. Combined with the above inference accuracy test, the improved model has higher detection accuracy when the inference speed is

Table 3. Comparison of detection accuracy between categories.

Methods	#Params	mAP_50	damage	dirt	worn	flaw
Ks=5	38.1	79.4	73.3	86.2	80.9	77.2
Ks=7	38.1	79.1	72.8	86.7	79.4	77.5
Ks=9	38.2	78.7	72.4	85.9	79.6	76.9
Ks=13	38.2	79.0	73.1	86.0	79.8	77.1
YOLOv5	50.2	74.1	70.1	84.2	73.4	68.7
YOLOv7	102.6	79.2	72.8	86.8	80.4	76.8
YOLOv8	79.3	79.2	72.1	86.9	81.0	76.8
YOLOv9	103.2	79.6	73.5	86.2	80.5	77.6
YOLOv10	64.1	79.1	73.1	85.7	80.2	77.4
Faster RCNN	183.9	71.8	70.4	72.2	72.4	72.2
SSD	62.8	70.7	69.5	71.1	70.4	71.8
RT-DETR	59.4	77.4	71.9	88.6	76.4	72.6
FasterNet-DETR	32.8	77.5	72.1	85.9	79.5	72.5
FR-DETR	34.2	78.3	73.1	87.3	76.9	75.9
FRE-DETR	38.1	79.4	73.3	86.2	80.9	77.2

guaranteed to be the same. FRE-DETR improves FPS by 34.5 compared with RT-DETR at step size 16, and FasterNet-DETR improves by 82.5 compared with RT-DETR, which is a huge improvement and proves that the improved model has a faster inference speed compared with the original model.

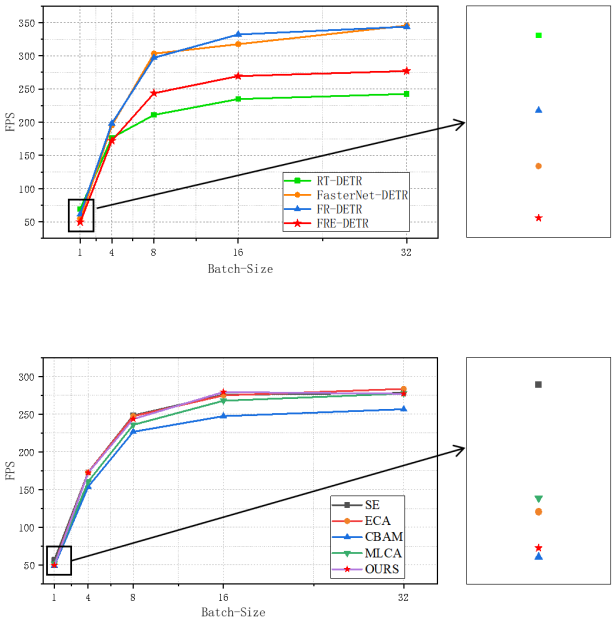

Fig. 5. Comparison of model inference speed.

3.5. Model comparison tests and detection experiments

For the efficiency and effectiveness of the proposed attention module for target edge reinforcement, we visualize the heat map of the output of the attention layer and compare it with the model without using this attention module. In Fig. 6, it can be found that the model after using the attention module becomes more focused on the defects, and it is easier to perceive the deep features, especially for the edge delineation of the features becomes clearer. It improves the model's ability to suppress the noise of the target with linear representation, which in turn improves the model's ability to express the tracking of the target [33, 34].

When the model weakens the NMS post-processing, it is very likely to cause the overlapping of the detection frame, which not only makes the detection effect messy but also easily causes the detection frame to be unable to fully align with the defects, which can be seen from Fig. 7, where the overlapping of the detection frame of the RT-DETR model is very serious, and the normal detection effect cannot be obtained. The FRE-DETR Due to the existence of the computational process of local pooling, the computational ability can be enhanced by pooling different regions, which in turn makes it easier to highlight the region where the target feature exists, so that the target region receives enhanced attention and other background regions receive less attention, which in turn makes the model pay more attention to the region with strong attention and reduces the generation of redundant detection frames when generating detection frames. detection model significantly reduces overlapping detection frames, enhancing accuracy. FRE-DETR demonstrates superior performance in detecting small targets, as illustrated in Fig. 7. The comparison highlights that the detection accuracy for small targets is significantly improved, effectively addressing the issue of target leakage, and enhancing the model's overall detection capabilities. Both models have a better detection effect for detecting a single target, and there is no leakage or misdetection phenomenon. Without post-processing operation, overlapping the prediction frames when performing dense detection is still easy, which is unavoidable, as seen in Fig. 7.

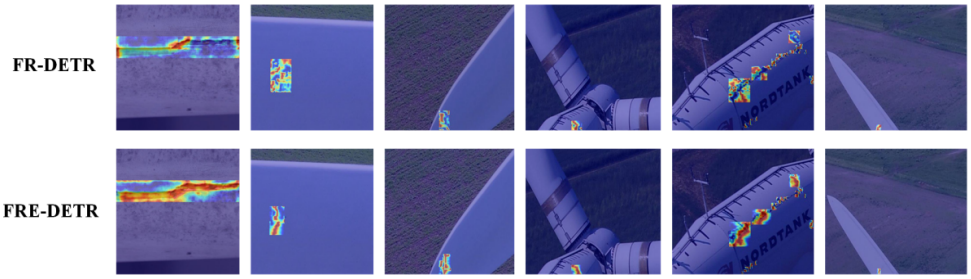

Fig. 6. Model Heat Map Visualization.

Fig. 7. Model detection experiment.

Therefore, further study is needed for the subsequent more FREE detection.

4. Conclusions

This experiment proposes a new FRE-DETR detection model for wind turbine blade defects by improving the hybrid DETR, which has faster detection speed, higher detection accuracy and lower model volume. FasterNet is chosen as the backbone network for defective feature extraction based on the lightweight feature extraction module, and the reparameterization module for feature extraction is introduced to address the problem that partial convolution is not effective enough for defective extraction, which greatly ensures the improvement of feature extraction efficiency without affecting the inference speed. Finally, a novel feature selection and fusion module is proposed by enhancing the detail features inter-channel features to solve the problem of cumbersome feature maps and the inability to effectively utilize more effective features, as well as the problem of fuzzy boundaries that are difficult to identify the defective boundaries of wind turbine blades, to further improve the recognition accuracy of FRE-DETR on the defects of wind turbine blades. From the final experiments, we can conclude that FRE-DETR outperforms RT-DETR in terms of inference speed at step sizes larger than 2, and the FPS outperforms RTDETR by close to 200 frames at a step size of 32, the average detection accuracy mAP_50 improves by 2%, and the computational complexity of the model is only 64.1%. Our model significantly outperforms RT-DETR in terms of detection accuracy and speed, and

it achieves the highest detection accuracy and the fastest detection speed, as well as exhibits greater robustness and environmental adaptability in the same volume. In the future, we aim to enhance the algorithm's ability to adapt to various environments by improving the dataset. This will involve increasing both the quantity and quality of data, specifically focusing on wind turbine blade defects in different scenarios. The deployability, inference accuracy, and speed of the transformer-based visual transformer inspection model will be further enhanced to ensure its effective application in wind turbine blade defect detection.

Acknowledgements

This work was supported by the China Electric Huachuang Wind Turbine Blade Anti-Icing Warning System and Custom Anti-Icing Coating Service Project, under project number 1-C-F-01-K03-10-2024-024.

References

[1] F. Jaramillo, J. M. Gutiérrez, M. Orchard, M. Guarini, and R. Astroza, (2022) *"A Bayesian approach for fatigue damage diagnosis and prognosis of wind turbine blades"* **Mechanical Systems and Signal Processing *174***: 109067. DOI: 10.1016/j.ymssp.2022.109067.

[2] K. Kong, K. Dyer, C. Payne, I. Hamerton, and P. M. Weaver, (2023) *"Progress and trends in damage detection methods, maintenance, and data-driven monitoring of wind turbine blades–A review"* **Renewable Energy Focus *44***: 390–412. DOI: 10.1016/j.ref.2022.08.005.

[3] Y. Wang, T. Peng, W. Wang, and M. Luo, (2023) *"High-efficient view planning for surface inspection based on parallel deep reinforcement learning"* **Advanced Engineering Informatics *55***: 101849. DOI: 10.1016/j.aei.2022.101849.

[4] L. Wang and Z. Zhang, (2017) *"Automatic detection of wind turbine blade surface cracks based on UAV-taken images"* **IEEE Transactions on Industrial Electronics *64***(9): 7293–7303. DOI: 10.1109/tie.2017.2682037.

[5] J. Zhu, C. Wen, and J. Liu, (2022) *"Defect identification of wind turbine blade based on multi-feature fusion residual network and transfer learning"* **Energy Science & Engineering *10***(1): 219–229. DOI: 10.1002/ese3.1024.

[6] J. Zhang, G. Cosma, and J. Watkins, (2021) *"Image enhanced mask R-CNN: A deep learning pipeline with new evaluation measures for wind turbine blade defect detection and classification"* **Journal of Imaging 7**(3): 46. DOI: 10.3390/jimaging7030046.

[7] Y.-h. Liu, Y.-q. Zheng, Z.-f. Shao, T. Wei, T.-c. Cui, and R. Xu, (2024) *"Defect detection of the surface of wind turbine blades combining attention mechanism"* **Advanced Engineering Informatics *59***: 102292. DOI: 10.1016/j.aei.2023.102292.

[8] X. Ran, S. Zhang, H. Wang, and Z. Zhang, (2022) *"An improved algorithm for wind turbine blade defect detection"* **IEEE Access *10***: 122171–122181. DOI: 10.1002/adts.202100631.

[9] L. Liu, P. Li, D. Wang, and S. Zhu, (2024) *"A wind turbine damage detection algorithm designed based on YOLOv8"* **Applied Soft Computing *154***: 111364. DOI: 10.1016/j.asoc.2024.111364.

[10] H. Yu, J. Wang, Y. Han, B. Fan, and C. Zhang, (2024) *"Research on an intelligent identification method for wind turbine blade damage based on CBAM-BiFPN-YOLOV8"* **Processes *12***(1): 205. DOI: 10.3390/pr12010205.

[11] Y. Yao, G. Wang, and J. Fan, (2023) *"WT-YOLOX: An efficient detection algorithm for wind turbine blade damage based on YOLOX"* **Energies *16***(9): 3776. DOI: 10.3390/en16093776.

[12] Y. Liu, Z. Wang, X. Wu, F. Fang, and A. S. Saqlain, (2022) *"Cloud-edge-end cooperative detection of wind turbine blade surface damage based on lightweight deep learning network"* **IEEE internet computing 27**(1): 43–51. DOI: 10.1109/mic.2022.3175935.

[13] Y. Zhang, Y. Yang, J. Sun, R. Ji, P. Zhang, and H. Shan, (2023) *"Surface defect detection of wind turbine based on lightweight YOLOv5s model"* **Measurement *220***: 113222. DOI: 10.1016/j.measurement.2023.113222.

[14] C. Zhang, T. Yang, and J. Yang, (2022) *"Image recognition of wind turbine blade defects using attention-based MobileNetv1-YOLOv4 and transfer learning"* **Sensors 22**(16): 6009. DOI: 10.3390/s22166009.

[15] W. Chen, G. Zhu, M. Mao, X. Xi, W. Xiong, L. Liu, S. Wang, and Y. Chen. "Defect detection method of wind turbine blades based on improved YOLOv4". In: *2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD)*. IEEE. 2022, 1–4. DOI: 10.1109/ICSMD57530.2022.10058380.

[16] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. "Swin transformer: Hierarchical vision transformer using shifted windows". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, 10012–10022. DOI: 10.48550/arXiv.2103.14030.

[17] X. Liu, H. Peng, N. Zheng, Y. Yang, H. Hu, and Y. Yuan. "Efficientvit: Memory efficient vision transformer with cascaded group attention". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023, 14420–14430. DOI: 10.48550/arXiv.2305.07027.

[18] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen. "Detrs beat yolos on real-time object detection". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2024, 16965–16974. DOI: 10.48550/arXiv.2304.08069.

[19] X. Liu, Z. Rao, Y. Zhang, and Y. Zheng. "UAVs Images Based Real-Time Insulator Defect Detection with Transformer Deep Learning". In: *2023 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE. 2023, 1–6. DOI: 10.1109/ROBIO58561.2023.10354816.

[20] E. Guemas, B. Routier, T. Ghelfenstein-Ferreira, C. Cordier, S. Hartuis, B. Marion, S. Bertout, E. Varlet-Marie, D. Costa, and G. Pasquier, (2024) *"Automatic patient-level recognition of four Plasmodium species on thin blood smear by a real-time detection transformer (RT-DETR) object detection algorithm: a proof-of-concept and evaluation"* **Microbiology Spectrum *12***(2): e01440–23. DOI: 10.1128/spectrum.01440-23.

[21] H. Ouyang, (2023) *"Deyov3: Detr with yolo for real-time object detection"* **arXiv preprint arXiv:2309.11851**: DOI: 10.48550/arXiv.2309.11851.

[22] J. Chen, S.-h. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, and S.-H. G. Chan. "Run, don't walk: chasing higher FLOPS for faster neural networks". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023, 12021–12031.

[23] J. Hu, L. Shen, and G. Sun. "Squeeze-and-excitation networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, 7132–7141. DOI: 10.1109/tpami.2019.2913372.

[24] D. Wan, R. Lu, S. Shen, T. Xu, X. Lang, and Z. Ren, (2023) *"Mixed local channel attention for object detection"* **Engineering Applications of Artificial Intelligence *123***: 106442. DOI: 10.1016/j.engappai.2023.106442.

[25] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu. "ECA-Net: Efficient channel attention for deep convolutional neural networks". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, 11534–11542.

[26] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, (2023) *"UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios"* **Sensors *23***(16): 7190. DOI: 10.3390/s23167190.

[27] X. Jiang, X. Zhuang, J. Chen, J. Zhang, and Y. Zhang, (2024) *"YOLOv8-MU: an improved YOLOv8 underwater detector based on a large kernel block and a multi-branch reparameterization module"* **Sensors *24***(9): 2905. DOI: 10.3390/s24092905.

[28] A. K. Aggarwal, (2023) *"A review on genomics data analysis using machine learning"* **Wseas Trans. Biol. Biomed *20***: 119–131. DOI: 10.37394/23208.2023.20.12.

[29] A. K. Aggarwal, (2014) *"Rehabilitation of the blind using audio to visual conversion tool"* **Journal of Biomedical Engineering and Medical Imaging *1***(4): 24–31.

[30] S. J. Mambou, P. Maresova, O. Krejcar, A. Selamat, and K. Kuca, (2018) *"Breast cancer detection using infrared thermal imaging and a deep learning model"* **Sensors *18***(9): 2799. DOI: 10.3390/s18092799.

[31] A. Foster, O. Best, M. Gianni, A. Khan, K. Collins, and S. Sharma. "Drone footage wind turbine surface damage detection". In: *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE. 2022, 1–5. DOI: 10.1109/IVMSP54334.2022.9816220.

[32] A. Shihavuddin, X. Chen, V. Fedorov, A. Nymark Christensen, N. Andre Brogaard Riis, K. Branner, A. Bjorholm Dahl, and R. Reinhold Paulsen, (2019) *"Wind turbine surface damage detection by deep learning aided drone inspection analysis"* **Energies *12***(4): 676. DOI: 10.3390/en12040676.

[33] X. Wang, J. Xia, J. H. Park, X. Xie, and G. Chen, (2022) *"Event-triggered adaptive tracking with guaranteed transient performance for switched nonlinear systems under asynchronous switching"* **IEEE Transactions on Cybernetics *54***(1): 496–505. DOI: 10.1109/tcyb.2022.3223983.

[34] X. Wang, J. Xia, J. H. Park, X. Xie, and G. Chen, (2022) *"Intelligent control of performance constrained switched nonlinear systems with random noises and its application: An event-driven approach"* **IEEE Transactions on Circuits and Systems I: Regular Papers *69***(9): 3736–3747. DOI: 10.1109/tcsi.2022.3175748.