Article:

Liu, Z.-H., Chen, Q., Wei, H.-L. orcid.org/0000-0002-4704-7346 et al. (2 more authors) (2023) Channel-spatial attention convolutional neural networks trained with adaptive learning rates for surface damage detection of wind turbine blades. Measurement, 217. 113097. ISSN 0263-2241

https://doi.org/10.1016/j.measurement.2023.113097

Manuscript of the paper published in: Measurement 217, 113097, https://doi.org/10.1016/j.measurement.2023.113097

Channel-Spatial Attention Convolutional Neural Networks Trained with Adaptive Learning Rates for Surface Damage Detection of Wind Turbine Blades

***Abstract*—Damage detection of wind turbine blades can provide guidance for maintaining the wind turbine system and reduce maintenance costs. Although machine vision has made good progress in blade damage detection, the complex background information brings great challenges to blade damage detection. Existing methods treat all information or features of the image equally, which may result in insufficient attention to damage features. In this paper, a novel framework of channel-spatial attention convolutional neural networks, together with an adaptive learning rate scheme, is proposed for surface damage detection of wind turbine blades. It guides the attention of the feature extraction network to focus on the blade damage feature by embedding CBAM (Convolutional Block Attention Module) to enhance the blade damage features. To optimize the training process and make it reach saturation faster, a novel adaptive learning rate scheme is also proposed. The effectiveness of the proposed is verified on real wind turbine blade image database, containing three types of damage, manually collected from a commercial wind farm. Experimental results show that the proposed method can improve binary damage classification accuracy by 2.68% and multiple class damage classification accuracy by 5.36% in comparison with the compared state-of-the-art methods.**

***Index Terms*—Attention mechanism, damage detection, machine vision, adaptive learning rate, wind turbine blade**.

I. Introduction

Wind energy is one of the key strategies for addressing the global energy crisis because of its widespread distribution and lack of pollution [1]. The addition of 93.6 GW in 2021 brought the total installed wind power capacity to 838 GW, an increase of 12% year over year. [2]. The power generating capacity of wind turbines is strongly influenced by their wind turbine blades (WTB), which make up more than 20% of the equipment's overall cost. It is known that under harsh operating and environmental conditions, of wind turbines and the exposed blades, the blades are susceptible to wind erosion, acid rain corrosion, sun exposure and other problems. Statistics show that more than 40% of all wind turbine failures that result in power loss are due to blade failure. Accurate blade damage identification can help with realistic maintenance planning and save downtime [3]. Moreover, a damage that affects the blade operation is usually gradually intensified by small faults. This means that accurate detection of minor damage can prevent blade damage at low cost and extend the service life of the blade [4].

Currently, naked-eye detection by workers at high altitude is the most significant method of identifying blade damage in wind farms， mainly relying on their experience. It is undeniable that this has issues with high danger and poor efficiency. Other WTBs damage detection techniques primarily include ultrasound, vibration, acoustic emission (AE), thermal imaging, and machine vision;

they each have advantages and disadvantages in practical applications [5].

Rezamand *et al.* [6] analyzed WTB damage information in vibration signals to verify whether WTB was damaged. The proposed method is highly sensitive to the WTBs early failure, leading to improved wind energy generation efficiency and reduced maintenance costs. Joshuva *et al.* [7] fused the vibration signal with the histogram to realize the status monitoring of WTB. They analyzed the vibration signal with a histogram, and selected important features by using the J48 decision tree algorithm. Wang *et al.* [8] applied multi-channel convolutional neural networks to extract fault information from vibration signals for blade damage detection. The detection method based on the vibration signal has the advantages of simple implementation and non-destructive detection, but it has the disadvantages of being susceptible to environmental disturbance images and unable to detect slight surface damage. The application of AE in damage detection of WTBs is also a research trend [9]. Xu *et al.* [10] developed a waveform-based feature extraction strategy to acquire the features hidden in the original AE waveforms, so as to identify the blade damage mode. Tang *et al.* [11] classified the AE signals corresponding to different lesions through an unsupervised pattern recognition method. The AE-based method can confirm the location of the damage and is sensitive to the type of damage, but the detection cost is high and it is easily polluted by environmental noise. Ultrasound is a commonly used method in the detection of WTB damage. Oliveira *et al.* [12] utilized ultrasonic signal processed by PCA to judge the health of WTBs. Choung *et al.* [13] proposed a phased array platform for damage detection of WTBs. Ultrasound is excellent for internal damage detection in WTB, but its signal acquisition and processing are difficult. Thermography-based method is one of the most current methods to detect WTBs damage. Yang *et al.* [14] applied TWR principles to the detection and diagnosis of CFRP imaging. Galleguillos *et al.* [15] proposed a thermal imaging method for non-destructive detection of WTB composites. Thermography-based technology has different sensitivities to different materials and can detect fatigue damage, but there are few types of damage that can be detected.

The above methods are not well suitable for wind turbine blade damage detection in wind farms due to their drawbacks such as high cost, long downtime, and low detection accuracy. Usually further visual detection of blade damage details needs to be performed to determine a repair plan, so vision-based detection solutions have become the mainstream choice for intelligent blade damage detection. Peng *et al.* [16] used multi-directional Gabor transformation to address the issue of non-uniform illumination image. This work effectively preprocesses WTB images but does not optimize the detection model. Wang *et al.* [17] classified blade features for damage using trained classifiers and Haar-like features to describe blade damage areas. Guo *et al.* [18] used Haar-AdaBoost to screen out areas that may contain damage, and then classified damage through CNN. Since this method needs to manually design Haar features, the generalization ability of the model is relatively poor. With the development of computing performance, methods based on deep learning show better generalization. Reddy *et al.* [19] used deep convolutional neural networks (CNN) to classify WTB damage, and the experimental results showed excellent practicability. This method adjusts some of the hyperparameters of CNN instead of optimizing the network model structure; its classification accuracy is still below the desired goal. Foster *et al.* [20] used YOLOv5 to detect WTB damage by optimizing the WTB characteristics; but the overall performance of the method is still unsatisfactory. Zhang *et al.* [21] proposed a new Mask R-CNN based pipeline to detect damages, and designed new performance evaluation measures. This method can detect WTB damage relatively accurately, but its feature extraction network effectively fully extract the information contained in the image, thus its detection accuracy still needs to be improved.

These methods are efficient for blade damage detection, but their performance is not very prominent. The reason is that the background information of the real wind farm blade image is complex, which leads to the distraction of the "attention" in the feature network. Therefore, one way to improve the performance of damage detection is to introduce an attention mechanism to make the network focus on damage features [22]. The essence of attention can be regarded as weight, if a feature is attached great importance, it means that the weight is very high. Most recently attention mechanism has been applied to wind turbine blade damage detection. For example, Chen *et al.* [23] applied an attention module to WTB damage detection and demonstrated a great potential of attention mechanism for dealing with the WTB damage detection problem. Such a method has good performance in the detection of minor damage, but its model training efficiency still needs to be improved. The application of the attention mechanism in the WTB damage detection tasks is still in its infancy, and its potential needs to be further explored.

The learning rate is a crucial hyperparameter in deep learning networks, which directly affects the training result of the model [24]. Dong *et al.* [25] trained a deep learning network with a constant learning rate to predict electrical loads, but the training method used cannot be well adapted to the network training process. During the training process, a relatively large learning rate can usually increase the convergence speed in the early stage, but a smaller value is usually required when the model approaches the saturation point in the later stage. To solve this problem, Mvoulana *et al.* [26] trained a deep learning network using a decay learning rate scheme. This method can automatically change the value of the learning rate according to the training epoch, but it cannot match different learning rates to the steps in each epoch. To overcome this issue, an adaptive learning rate approach is designed, which can significantly improve the convergence speed and computational efficiency of the network.

In this paper, channel-spatial attention convolutional neural network for surface damage detection of WTBs using adaptive learning rate were proposed, which places more emphasis on damage features, and less emphasis on background and healthy region features. In this manner, the weight of damage features is increased to enhance the recognition of minor damage and eliminate the interference of inherent structures (drain holes, lightning arresters, etc.). Moreover, a novel adaptive learning rate design method was proposed to help the model reach saturation faster. The following is this paper's main contributions:

1) Design of new channel-spatial attention convolutional neural networks trained with adaptive learning rate for surface damage detection, which provide an end-to-end detection framework for blade damage.

2) Design of damage feature enhanced channel-spatial attention, which assigns higher weights to damage-sensitive channels and

Manuscript of the paper published in: Measurement 217, 113097, https://doi.org/10.1016/j.measurement.2023.113097

enhances the spatial information of damage features, and which enables the feature extraction networks to pay more attention to the damage features rather than the whole image.

3) A novel adaptive learning rate design method, which enables the model to reach the global optimum faster and more accurately.

The rest of this essay is structured as follows. In Section II, the proposed damage detection method including attention mechanism, feature extraction network and the adaptive learning rate is proposed. In Section III, a comprehensive case study of WTBs damage detection is carried out, and detailed analysis and discussions on the case results are provided. Finally, the conclusions are presented in Section IV.

II. THE PROPOSED CHANNEL-SPATIAL ATTENTION CONVOLUTIONAL NEURAL NETWORKS FOR DAMAGE DETECTION

The vision-based WTB damage detection framework mainly consists of three phases: feature extraction network, attention mechanism and adaptive learning rate. The damage detection framework of WTBs is depicted in Fig. 1. The feature extraction network comprises 5 blocks, where the channel attention module (CAM) and the spatial attention module (SAM) are embedded between blocks 1 and 2, 2 and 3, and so on. More details of the Block structure are provided in Section A below. The extracted and enhanced feature map is flattened into a vector of size (c*w*h, 1) through the "Flatten" operation first, and then mapped to a vector of the specified size by a linear function. Finally, the classification result is produced by the classifier. Note that the novel adaptive learning rate is applied to all parameter updates in the network training process.

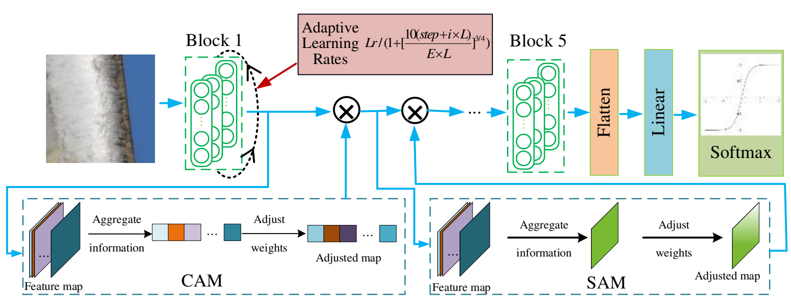


Fig. 1. WTBs damage detection flow chart.

A. Feature Extraction Network

The feature extraction network occupies an important position in the damage detection framework, because it is directly related to the acquisition of information contained in the image. At present, due to the complex design and insufficient generalization ability of traditional image feature extractors, the first choice for feature extraction has been replaced by deep learning-based feature extraction networks. The feature extraction network is composed of multiple CNN (Convolutional Neural Network) stacks, which has the characteristics of simple design and powerful functions. Among them, VGGNet [27] and ResNet (Residual Network) [28] have been confirmed by many researchers to be effective. In this paper, the backbone of VGGNet-16 is used for feature extraction. To fully demonstrate the role of adaptive learning rate and the role of the proposed attention module, this paper also uses Resnet as a supplementary feature extraction network.

VGGNet was proposed by researchers from Oxford University in 2014 and is a neural network model with an excellent structure [27]. In order to extract features, it combined several 3×3 convolution kernels instead of using large convolution kernels. The use of this design can both enhance the nonlinear mapping ability of the network and reduce the parameters of the network. At the same time, it also used multiple convolutional layers for stacking to enhance the feature expression ability of the network by increasing the network depth within a certain range. ResNet is also a network framework with an outstanding structure [28]. To address the issue of gradient explosion and gradient disappearance as the network deepens, it suggested a residual module.

According to the output feature map size, VGGNet is divided into 5 blocks, each block uses 2 or 3 convolutional layers to extract WTB features, and then uses a pooling layer to process the features. The structure of the feature extraction block is shown in Fig. 2. First, the convolution kernel is used to filter the feature map to extract the deep features, then batch normalization (BN) is used to process the features to speed up the convergence of the network. Finally, the features are mapped to the specified interval through the rectified linear unit (ReLU) activation function. The main parameters are set as follows: the stride is set to 1, and the size of the convolution kernel is set to 3, so that the convolution kernel can traverse all the positions of the feature map. In addition, stacking smaller convolution kernels instead using larger convolution kernels for each block not only achieves a larger receptive field, but also has fewer parameters. The pooling layer adopts the maximum pooling operation, and the pooling kernel is 2, so that the most prominent features of the feature map can be retained to compress the model.

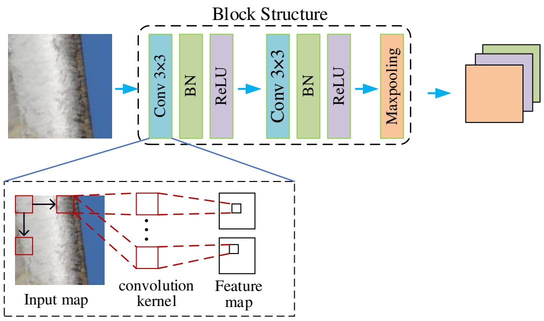


Fig. 2. The structure of the feature extraction block.

Each convolution kernel can be regarded as a filter, and different convolution kernels will produce different responses to the input feature. The response of each convolution kernel to the input feature is a 2D feature map, and the responses of multiple convolution kernels are aggregated into a 3D feature map. The number of convolution kernels is the third dimension of the feature map: channels. In this way, each channel represents the response of a convolution kernel, and the sensitivity of different convolution kernels to different features creates differences between channels and therefore distinguishes different channels. The different positions of each channel also indicate that different spatial positions of the feature map contain different feature information. This is the theoretical basis for the design of CAM and SAM.

To detect the damage features in these two dimensions, we sequentially insert a CAM and a SAM after each block. The parameters of the two modules are updated by the adaptive learning rate to complete the automatic search and weighting of the damage features by the attention module, so that the feature extraction network pays more attention to the damage features. The specific implementation steps of these two modules will be described section II-B.

B. The Proposed Channel-Spatial Attention Modules

The attention mechanism can be seen as a technique for dynamically adjusting the weight of input image features in decision-making, that is, to reduce the importance of unimportant information and concentrate on important information [29]. In the human visual system, humans do not remember and process the entire scene, but process parts of the scene or thing of interest. Similarly, when applying attention technique in the image field, the weights of some features of interest are adjusted, so as to intensify these features. At present, the attention mechanisms can be roughly divided into the following four groups: channel attention [30-31], spatial attention [32-33], the combination of channel and spatial attention [34], and self-attention [35-36].

Channel attention and spatial attention intensify the target features from the channel and space dimensions, respectively, but they both have a main drawback: the feature information contained in the space or channel will be lost. Different from other attention mechanisms, self-attention directly operates on each patch of the input image instead of using convolution to extract features and derive the relationship between patches. Self-attention usually requires the use of very large datasets and huge calculations to achieve good results, so it is rarely used at present.

The attention mechanism combining channel and space takes advantage of the two and overcomes the shortcomings of both; it becomes the main method to enhance the target features. Convolutional Block Attention Module (CBAM) [34], proposed by Woo *et al.*, is an excellent representative of channel and spatial attention. The attention method employed in this paper, which is developed based on CBAM, consists of two independent modules: CAM and SAM. These two modules process the original feature map in turn, so that the damage features are enhanced at both the channel and space levels. The combination of CAM and SAM is shown in Fig.1.

As described in section II-A, each convolution kernel can be regarded as a filter, and each channel is generated after the feature map is filtered by a convolution kernel. Different convolution kernels have different sensitivities to the damage features in the feature map, and the convolution kernel with a high sensitivity can generate a 2D feature map containing a large number of damage

Manuscript of the paper published in: Measurement 217, 113097, https://doi.org/10.1016/j.measurement.2023.113097

features, so that the damage features are transferred to the next-level feature map. Therefore, we need to find channels that are sensitive to damage signatures and enhance them. This is the role of channel attention in the framework of this paper.

The implementation process of channel attention includes three steps. The first step is to collect the information of each channel in the input feature, the second step is to establish the relationship between channels, and to obtain the weight of each channel, and the third step is to weight the original feature map. The implementation can be expressed as (1):

$$\begin{aligned}\hat{M} &= M \times F(A(M)) \\ &= M \times F(C) \\ &= M \times W_C\end{aligned} \tag{1}$$

where $M \in {}^{C\times H\times W}$ denotes a feature map, $A(\bullet)$ means to aggregate channel information, $F(\bullet)$ represents the perceptual function of the channel relationship, $C \in {}^{C\times 1\times 1}$ represents aggregated channel information, $W_C \in {}^{C\times 1\times 1}$ represents the weight of the channel.

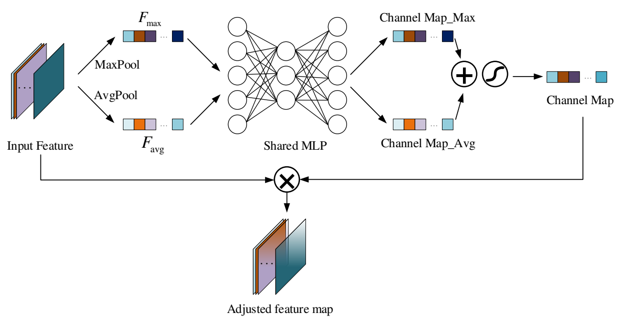


Fig. 3. Detailed structure of channel attention.

The workflow of CAM is shown in **Fig. 3**, we adopt global max pooling and average pooling to aggregate the feature information of each channel in parallel, this can greatly reduce the network computational complexity including parameter estimation. CAM generates two sequences $F_{\max}$ and F_{avg} , which stand for max-pooled features and average-pooled features, respectively. The use of maximum pooling is to retain the most significant features in the original feature map, and the use of average pooling is to collect overall features; the two combined features can enhance one another. Then, both sequences are passed to a shared network to create an elementary channel attention map $M_c \in {}^{C\times 1\times 1}$. Multi-layer perceptron (MLP) makes up the shared network. It is set to ${}^{C/r\times 1\times 1}$ for the hidden activation size, where r is the reduction ratio, which is set to 16 in this paper to reduce parameter overhead. $F_{\max}$ and F_{avg} are respectively input into MLP to generate Channel Map_max and Channel Map_avg, then they are fused into a vector by adding corresponding elements, and finally mapped to [-1,1] through the sigmoid function. In this way, the weight vector of the channel is obtained. In a nutshell, the channel attention is calculated as follows:

Manuscript of the paper published in: Measurement 217, 113097, https://doi.org/10.1016/j.measurement.2023.113097

$$\begin{aligned}\mathrm{M}_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{avg})) + W_1(W_0(F_{\max})))\end{aligned} \tag{2}$$

where σ represent the sigmoid function, $W_0 \in {}^{C/r\times C}$, $W_1 \in {}^{C\times C/r}$, and F represent the feature map.

As mentioned in Section II-A, the feature processed by the convolutional layer contains not only the information extracted by the channel, but also a large amount of spatial information in the 2D feature map generated by each channel. Therefore, to fully search the damage feature in the feature map and enhance it, we also need to design spatial attention. Similar to channel attention, the implementation procedure of spatial attention is as follwos. The first step is to collect the channel information of each position on the feature map, the second step is to establish the spatial relationship of each position and find out the spatial information of the target features, and the third step is to weight the 2D feature map of each channel of the original feature map. The procedure can be expressed as:

$$\begin{aligned}\hat{M} &= M \times F(A(M)) \\ &= M \times F(S) \\ &= M \times W_S\end{aligned} \tag{3}$$

Where $W_S \in {}^{1\times H\times W}$ represents the weight of the spatial, other parameters are consistent with (1).

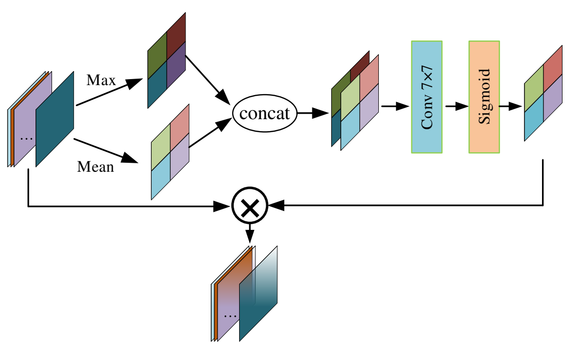


Fig. 4.Detailed structure of spatial attention.

In spatial attention module, we first use the mean and maximum values for the feature maps to aggregate the channel information at the same location, so that we can obtain two 2D features. We concatenate the two 2D feature maps in the channel dimension, and use a convolutional layer to reduce the dimension of the concatenated feature map so that its channel is 1. Then we send the dimensionality-reduced feature map to a convolutional layer to further extract the location information of the damage feature, and weight the 2D feature map of each channel of the original feature map. In a nutshell, the spatial attention is calculated as follows:

$$\begin{aligned}\mathrm{M}_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{avg})) + W_1(W_0(F_{\max})))\end{aligned} \tag{4}$$

where $W_0 \in {}^{C/r\times C}$, $W_1 \in {}^{C\times C/r}$, σ represent the sigmoid function, and F represent the feature map

The sequential application of CAM and SAM constitutes the attention mechanism proposed in this paper. By using the adaptive learning rate to update the attention module parameters, the channel and spatial attention modules that can automatically search for blade damage features are trained. The damage feature is enhanced from both channel and space dimensions, so that the damage feature can obtain higher scores in subsequent classifiers.

C. Training Process Based on Adaptive Learning Rate

In deep learning methods, an essential hyperparameter in model training is learning rate, which controls the learning procedure towards obtaining the best or optimal model parameters [37]. During the network training, the update rule of the network weight is $\theta=\theta-\alpha\frac{\partial}{\partial\theta}J(\theta)$,where α is the learning rate. It's critical to properly adjust the learning rate because if it is too large, the network may fail to converge, or even diverge from the optimal error point, but if it is too small, the network may converge very slowly or even be trapped in a local extreme point. A general rule is as follows: in the early stage of network training, the learning rate can be set a relatively large value, so that the network can converge quickly; in later stages, the learning rate should be relatively smaller to avoid missing the optimum of the network model parameters. In this context, the learning rate can be assumed to be a monotonically decreasing function as the training progresses. The training of the network can be decomposed into multiple epochs, and each epoch can be divided into multiple steps. Therefore, the values of epoch and step can be combined to form the independent variable of the learning rate function. It needs to take into account that the learning rate cannot be extremely close to zero, even at the end of training.

Commented [HW1]: Do not use '0'. Instead, use zero.

In order for the damage detection network to approach its optimal model parameters quickly and improve the classification accuracy of blade damage as much as possible, this paper proposes the following adaptive learning rate update scheme:

$$lr(i\times L+step)=Lr/(1+[\frac{10(step+i\times L)}{E\times L}]^{3/4}) \tag{5}$$

where E , L , Lr are all constants, representing the training epoch, the batches contained in each epoch and the preset maximum learning rate. $step$ and i represent the current epoch and batch values, respectively.

If we replace $i\times L+step$ with $x\in(0,E\times L)$, then we can simplify (5) to:

$$lr(x)=LR/(1+[\frac{10\cdot x}{E\times L}]^{3/4}) \tag{6}$$

The derivative of (6) is:

$$lr'(x)=-0.75LR\times(\frac{10}{E\times L})^{0.75}/x^{0.25}[1+(\frac{10\cdot x}{E\times L})^{0.75}]^{2} \tag{7}$$

By analyzing (7), we can know that the proposed adaptive learning rate has the following property: it gradually decreases as the training progresses, but is still non-zero at the final stage, which means that the network has always learning ability.

III. CASE STUDIES AND DISCUSSION

This study uses a dataset of real images, collected from a Chinese wind farm, to demonstrate and test the performance of the proposed method from several different aspects including: i) how the feature extraction network gives the damage features more consideration; ii) how the damage detection accuracy can be improved with the new method.

A. Dataset and Implementation Details

Given the fact that images of wind turbine blade damage are difficult to collect, the real dataset used in this study is not large; it contains only 395 images. We label these images according to the professional advice of the engineers, so the model can learn the expert knowledge in those images and give more accurate detection suggestions. For example, the model marks these images containing slight damage (that do not deteriorate) are considered normal images, thereby reducing maintenance costs for wind farms. To make good use of the images, morphological transformation and image adjustment were performed on the existing datasets. The process includes: reversing, color adjustment, rotation, contrast adjustment, brightness adjustment and sharpness adjustment.

Specifically, the purpose of flipping and rotating the images is to mimic the acquisition of damage image from various angles. Images under various lighting conditions can be simulated by adjusting brightness and chromaticity. The foreground information can be made more prominent and damage features in the image can be highlighted by adjusting the contrast and sharpness. These operations increase the number of images and complement the dataset. It is worth mentioning that the resolutions of these original images are not consistent, therefore, for convenience of further processing, those WTB images are all resized and unified to 224×224 pixels.

(a) Non-damage (b) Crack (c) Sand Hole (d) Skin Shedding

Fig. 5.Samples of the WTB image.

According to WTB's maintenance regulations, the images of this dataset are divided into the following: sand hole images, crack images, skin shedding images and non-damage images. Fig. 5 depicts some examples in the WTBs dataset. Table I lists the precise number of categories and how the training set and test set are split up.

Manuscript of the paper published in: Measurement 217, 113097, https://doi.org/10.1016/j.measurement.2023.113097

TABLE I

THE DISTRIBUTION OF BLADE IMAGE DATA

	Non-damage	Crack	Sand Hole	Skin Shedding	Sum
Train	2009	250	550	580	3389
Test	200	80	130	150	560

Experiments were conducted on a workstation installed with Inter i5-10600KF CPU, NVIDIA GeForce RTX2060 8G GPU, 16G Memory, Pytorch 1.8.1, and Cuda 10.2.89. For a fair comparison, all models are trained with consistent parameter settings and without using any pre-trained weights and. The maximum number of epochs is 50, the initial learning rate is 0.01, the batch size is 16.

B. Adaptive Learning Rate Evaluation

This section presents the actual performance of the proposed adaptive learning rate in this case. To measure the performance of the proposed learning rate, this paper uses the proposed learning rate and constant learning rate when training Resnet50 and VGG16, respectively. The constant learning rate is set to 0.01. The complete training process is shown in Fig. 6.

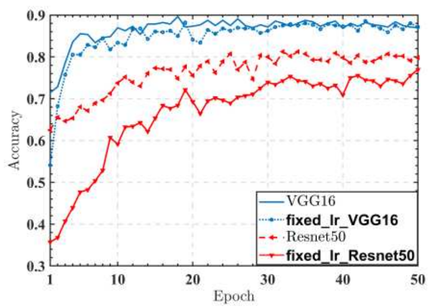


Fig. 6.Training process with changing learning rate.

Experimental result shows that the VGG16 and Resnet50 networks, with the proposed adaptive learning rate, can reach the peak faster and improve the classification accuracy. Because the adaptive learning rate performs well in the experimental data presented in this paper, the adaptive learning rate design method is used uniformly in all subsequent experiments.

C. Performance Evaluation

This section presents the experimental results produced by the proposed method, together with a comparison with results given by three state-of-the-art attention methods, namely, SE block [30] based on channel attention, Coordinate block (CA) [33] based on spatial attention, VIT [36] based on self-attention and SVM. The inputs of SVM are WTB images. At the same time, in order to verify the generalization ability of the proposed method in different models, these attention modules will be applied in both VGG16 and Resnet50. In doing so, the experiments were divided into two groups, each contains a baseline network and three networks with

Commented [HW2]: Not to start a new sentence with "And" embedded attention modules. Since VIT directly processes the input image without using CNN, it serves as a separate control group, as does SVM. In order to reduce the influence of other factors, all the networks were trained in the same environment using the same training strategy, and did not use the pre-trained weights (except VIT).

To perform a comprehensive analysis on WTB damage detection, this study considers two evaluation criteria: a) whether the WTB is damaged or not; b) what kind of damage is present on the WTB.

1) Criterion a) involves three indicators: classification accuracy (ACC), false positive rate (FPR), and true positive rate (TPR), which are defined as follows:

$$\mathrm{ACC} = \frac{TN + TP}{TN + TP + FN + FP} \tag{8}$$

$$TPR = \frac{TP}{FN + TP} \tag{9}$$

$$FPR = \frac{FP}{TN + FP} \tag{10}$$

where images with damage are considered as positive samples (P), and images without damage are considered as negative samples (N). TP (True Positive) and TN (False Negative) denote the number of correctly classified images with and without damage, respectively; FP (False Positive) is the number of images without damage but misclassified as images with damage, and FN (False Negative) is the number of images with damage but misclassified as images without damage.

Note that in the damage detection of WTB, the cost of misclassifying damage as no-damage are much higher than the cost of misclassifying no-damage as damage. Because the former can be passed through artificially, the latter may cause huge losses if the blade is not repaired in time. Therefore, the importance of TPR is greater than that of FPR.

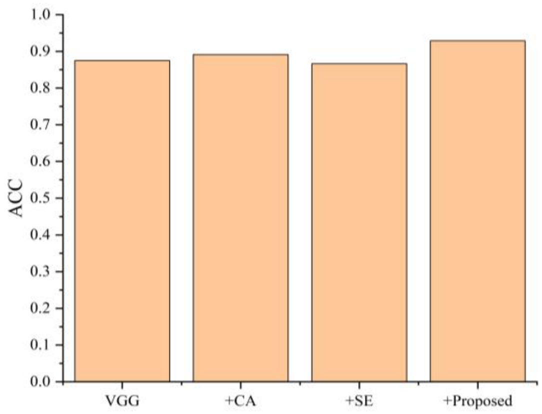


Fig. 7. WTB damages binary classification accuracy

The damage detection accuracy of Criterion a) is shown in Fig. 7. It can be seen that the damage detection performance of the proposed method is the best among the four methods. Table II presents more detailed information of the experimental results. The "Time" in the table is the time taken to test 560 images. As can be seen from Table II, the proposed method achieves the best scores in terms of ACC, TPR and FPR, which are the most important metrics for the case. Compared with VGG16 (without attention mechanism), this proposed method improves ACC by 2.68%, TPR by 3.17%, and improves the performance compared with other advanced attention mechanisms. But on Resnet50, we can note a strange phenomenon: introducing some attention mechanisms may have negative effect on network model optimization; we will do further research on this finding the future. In addition, the time cost of the proposed method is generally a little longer than the other methods, but it has less impact in real scenarios. In the process of blade damage detection, the speed of image acquisition is generally less than 30fps, and the judgment speed of the model is greater than the acquisition speed. In general, the attention mechanism can enhance the feature extraction ability of the network, and our proposed method has certain advantages compared with the compared existing methods. It was noted that in this experiment, using VGG structure as the feature extraction basic network can always produce a better result than Resnet.

TABLE II
THE TEST RESULT OF CLASSIFYING DAMAGE AND NON-DAMAGE

Network	TP	TN	FN	FP	ACC (%)	TPR (%)	FPR (%)	Time (s)
VGG	329	200	31	0	94.46	91.39	100	10.32
+CA	334	199	26	1	95.18	92.78	99.5	11.18
+SE	340	200	20	0	96.43	92.78	100	10.46
+Proposed	**344**	**200**	**16**	**0**	**97.14**	**94.56**	**100**	**11.38**
ResNet	326	200	34	0	93.92	90.56	100	9.92
+CA	316	199	44	1	91.96	87.78	99.5	10.86
+SE	309	200	51	0	90.89	85.83	100	10.46
+Proposed	**332**	**200**	**28**	**0**	**95**	**92.22**	**100**	11.01
VIT	329	196	31	4	93.75	91.39	98	17.26
SVM	335	160	25	40	88.39	93.06	80	268.57

In general, in the task of judging whether the blades are damaged or not, the embedded attention mechanism can effectively intensify the feature extraction ability of the network and the proposed method outperforms the three compared methods.

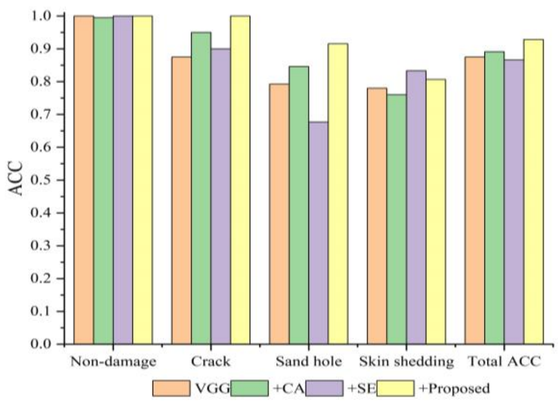


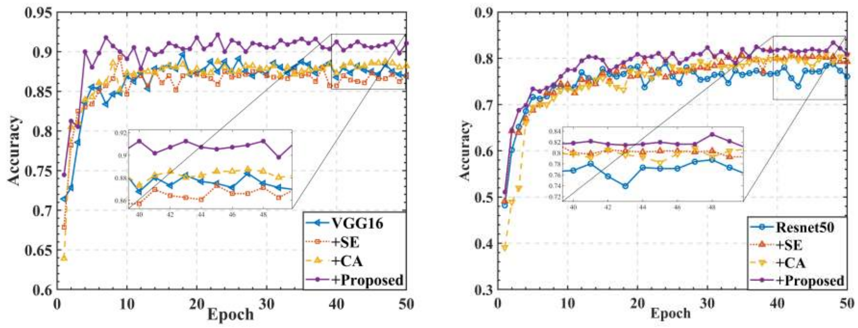
Fig. 8.WTB multiple damage detection accuracy

2) For criterion b), the WTB dataset are divided into follow: sand hole, skin shedding, crack, and non-damage. This experiment trains the model to classify what kind of damage in the blades. The classification accuracy of each category and total accuracy are used to measure the damage detection performance.

The experimental control settings are basically the same as the experimental settings in criterion a). The accuracy of multi-damage classification is shown in Figure 8. It can be seen from the figure that the proposed method has the best performance for both single type damage detection and multi-type damage detection. More experimental results are shown in Table III.

TABLE III
THE TEST RESULT OF BLADE DAMAGE CLASSIFICATION

	non-damage			**Crack**			**Sand hole**			**Skin shedding**				Time
	True	**False**	**ACC**	**True**	**False**	**ACC**	**True**	**False**	**ACC**	**True**	**False**	**ACC**	**Total ACC**	(s)
VGG	200	0	100%	70	10	87.5%	103	27	79.23%	117	33	78%	87.5%	10.4351
+CA	199	1	99.5%	76	4	95%	110	20	84.61%	114	36	76%	89.11%	11.3267
+SE	200	0	100%	72	8	90%	88	42	67.69%	125	25	83.33%	86.61%	10.5747
+Proposed	200	0	**100%**	80	0	**100%**	119	11	**91.54%**	121	29	80.66%	**92.86%**	**11.5052**
ResNet	200	0	100%	61	19	76.25%	82	48	63.08%	98	52	65.33%	78.75%	10.0241
+CA	199	1	99.5%	67	13	83.75%	92	38	70.77%	100	50	66.67%	81.78%	10.9606
+SE	200	0	100%	66	14	82.5%	90	40	69.23%	101	49	67.33%	81.60%	10.5857
+Proposed	200	0	100%	78	2	97.5%	77	53	59.23%	105	45	70%	**82.14%**	11.0733
VIT	196	4	98%	62	18	77.5%	95	35	73.08%	89	61	59.33%	78.93%	17.3276
SVM	148	52	74%	72	8	90%	66	64	50.77%	102	48	68%	69.28%	360.5759



(a) (b)

Fig. 9. Accuracy changes during network training. (a) Represents the training process of the attention module embedding in VGG16. (b) represents the training process of embedding Resnet50.

As can be seen from Table III, this proposed method also performs well, but also had the problem of taking longer in the multi-damage classification of WTB. Its total accuracy and per-class accuracy are higher than other attentions. Specifically, its overall accuracy improves by 5.36% compared to VGG16 (without the embedded attention module). Similar to the results measure by criterion a), the VGG16-based method is always better than the Resnet50-based method under the same conditions. Fig. 9 shows the performance of the proposed and comparative methods on VGG16 and Resnet50. As can be seen from **Fig. 9**, the proposed method outperforms other methods in the whole training process, and the same is true for embedding in Resnet.

D. Model Analysis

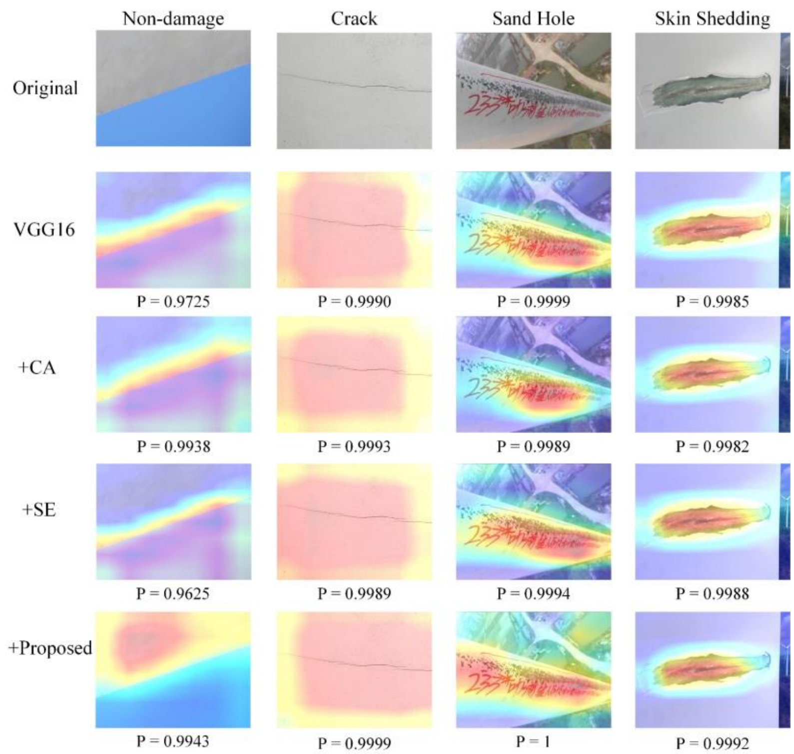


Fig. 10. Visualization with Grad-CAM.

To explore and understand how the proposed attention method works for WTB damage classification, this paper adopts Gradient-weighted Class Activation Mapping (Grad-CAM) [38] as a network visualization tool here. Grad-CAM is a recently proposed visualization tool, which uses the gradient in the convolutional layer to represent the weight of the feature map in the network. In this way, the distribution of the network's "attention" to the input image can be visualized. Fig. 10 shows the visualization of the last convolutional layer of Grad-CAM in the network, which uses color depth to indicate the degree of attention of the network.

In Fig. 10, we can see that when detecting images of healthy WTBs, the proposed model puts "attention" on the WTBs area instead of the background part, expanding the weight difference between the foreground and background. Other models do not produce a clear distinction between WTBs and background parts. When detecting WTBs images with cracks, due to the clearer features, all methods have better performance, and the confidences are all over 0.998. When detecting images of WTBs with sand hole, the proposed model has the most complete coverage of sand hole, and other methods do not pay attention to the sparse sand hole at the edge of the image. When detecting images of WTBs with skin shedding, it can be seen that other networks slightly distract from part of the "attention" on the right wind turbine of the image, except the proposed method.

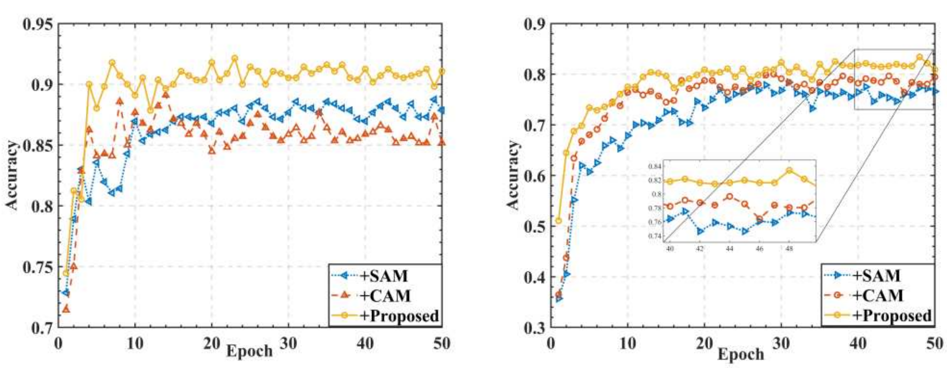
By visualizing the network, it is clear that the proposed approach outperforms the other two attention techniques in terms of enhancing the network's capacity to detect WTB damage features.

TABLE IV
THE TEST RESULT OF BLADE DAMAGE CLASSIFICATION

	non-damage			**Crack**			**Sand hole**			**Skin shedding**			
	True	**False**	**ACC**	**True**	**False**	**ACC**	**True**	**False**	**ACC**	**True**	**False**	**ACC**	**Total ACC**
VGG	200	0	100%	70	10	87.5%	103	27	79.23%	117	33	78%	87.5%
+Proposed	**200**	**0**	**100%**	**80**	**0**	**100%**	**119**	**11**	**91.54%**	**121**	**29**	**80.66%**	**92.86%**
+No-CAM	200	0	100%	72	8	90%	95	35	73.08%	114	36	76%	85.89%
+No-SAM	200	0	100%	78	2	97.5%	106	24	81.54%	117	33	78%	89.46%
ResNet	200	0	100%	61	19	76.25%	82	48	63.08%	98	52	65.33%	78.75%
+Proposed	**200**	**0**	**100%**	**78**	**2**	**97.5%**	**77**	**53**	**59.23%**	**105**	**45**	**70%**	**82.14%**
+No-CAM	199	1	99.5%	64	16	80%	75	55	57.69%	99	51	66%	78.03%
+No-SAM	200	0	100%	70	10	87.5%	76	54	58.46%	106	44	70.67%	80.71%

We remove the channel attention module (No-CAM) and the spatial attention module (No-SAM), respectively, in order to further investigate the respective roles of the components in the proposed method. We then repeat the experiments from Section III-B. The experimental results and training process are shown in Table IV-V and Fig.11, respectively.

As can be seen from Table IV-V, the lack of spatial attention module or channel attention module leads to performance deterioration for damage detection. In other words, the proposed method, by combining the spatial and channel attentions which complement each other, can enhance the network ability to capture damage features in multiple dimensions. The training process in Fig.11 also supports this conclusion.



(a) (b)

Fig. 11. An illustration of the training process of a single attention module of the proposed method. (a) Modules embedded in VGG16, (b) Modules embedded in Resnet50

TABLE V
THE TEST RESULT OF CLASSIFYING DAMAGE OR NON-DAMAGE

Network	TP	TN	FP	FN	ACC (%)	TPR (%)	FPR (%)
VGG	329	200	31	0	94.46	91.39	100
+Proposed	**344**	**200**	**16**	**0**	**97.14**	**95.56**	**100**
+No-SAM	337	200	23	0	95.89	93.61	100
+No-CAM	339	200	21	0	96.25	94.17	100
ResNet	326	200	34	0	93.92	90.56	100
+Proposed	**332**	**200**	**28**	**0**	**95**	**92.22**	**100**
+No-SAM	324	200	36	0	93.57	90	100
+No-CAM	316	200	44	0	92.14	87.78	100

IV. CONCLUSION

In this paper, a channel-spatial attention convolutional neural network framework, together with an adaptive learning rate scheme, has been proposed for surface damage detection of wind turbine blades. It decomposes the feature extraction network of VGG16 into 5 blocks and rescales the associated feature maps. In the recalibration stage, it combines channel and spatial attention modules to capture the feature information of damage and enhance the damage features, so that the damage area has a higher weight in the feature map. Moreover, we proposed a novel adaptive learning rate scheme, which can make the training process of feature extraction networks and attention modules reach saturation faster.

The experimental results for the real blade images acquired in the wind farm clearly show the effectiveness and the improvement in performance of the proposed damage detection method in comparison with other compared methods. For the damage binary classification task, the proposed method improves the accuracy rate by 2.68% compared with the existing methods. For the challenging multi-damage classification tasks, the proposed method improves the accuracy by 5.36% compared with the compared methods.

The proposed method can significantly improve both the efficiency and accuracy for wind turbine blade damage detection. In future work, more types of damage will be considered and the detection of damage location will also be included in the experimental tasks.

REFERENCES

[1] W. Yin, S. Feng and Y. Hou, "Stochastic Wind Farm Expansion Planning with Decision-Dependent Uncertainty Under Spatial Smoothing Effect," *IEEE Trans. Power Systems*, 2022, doi: 10.1109/TPWRS.2022.3184705.

[2] Global wind report 2022: The Data: 2021 was the wind industry's second-best year, Global wind energy council [online]. https://gwec.net/global-wind-report-2022/

[3] Y. Du *et al*. "Damage detection techniques for wind turbine blades: A review." *Mechanical Systems and Signal Processing*, vol.141, 2020, pp. 0888-3270, doi: 10.1016/j.ymssp.2019.106445.

[4] Wenyun Wang, Jingyun Yang, Juchuan Dai, Anhua Chen, "EEMD-based videogrammetry and vibration analysis method for rotating wind power blades," *Measurement*, vol. 207, Art no. 112423, 2023, doi: 10.1016/j.measurement.2022.112423.

[5] L. Wang and Z. Zhang, "Automatic Detection of Wind Turbine Blade Surface Cracks Based on UAV-Taken Images," *IEEE Trans. Industrial Electronics*, vol. 64, no. 9, pp. 7293-7303, Sept. 2017, doi: 10.1109/TIE.2017.2682037.

[6] M. Rezamand, *et al*. "A New Hybrid Fault Detection Method for Wind Turbine Blades Using Recursive PCA and Wavelet-Based PDF," *IEEE Sensors Journal*, vol. 20, no. 4, pp. 2023-2033, 15 Feb.15, 2020, doi: 10.1109/JSEN.2019.2948997.

[7] A. Joshuva, V. Sugumaran, "A lazy learning approach for condition monitoring of wind turbine blade using vibration signals and histogram features," *Measurement*, vol.152 pp. 107295, 2020, doi: 10.1016/j.measurement.2019.107295.

[8] Wang, Meng-Hui, et al. "Fault detection of wind turbine blades using multi-channel CNN." Sustainability 14.3 (2022): 1781.

[9] Z. Liu, X. Wang and L. Zhang, "Fault Diagnosis of Industrial Wind Turbine Blade Bearing Using Acoustic Emission Analysis," *IEEE Trans. Instrumentation and Measurement* t, vol. 69, no. 9, pp. 6630-6639, Sept. 2020, doi: 10.1109/TIM.2020.2969062.

[10] D. Xu, P.F. Liu, Z.P. Chen, "Damage mode identification and singular signal detection of composite wind turbine blade using acoustic emission." *Composite Structures*, vol.255, 2021, pp 0263-8223, doi: 10.1016/j.compstruct.2020.112954.

[11] J. Tang, S. Soua, C. Mares, T.H. Gan, "A pattern recognition approach to acoustic emission data originating from fatigue of wind turbine blades, " *Sensors* 17 (11) (2017) 2507

[12] Oliveira, Moisés A., et al. "Ultrasound-based identification of damage in wind turbine blades using novelty detection." *Ultrasonics* 108 (2020): 106166.

[13] J. Choung, S. Lim, S. H. Lim, S. C. Chi and M. H. Nam, "Automatic Discontinuity Classification of Wind-turbine Blades Using A-scan-based Convolutional Neural Network," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 1, pp. 210-218, January 2021, doi: 10.35833/MPCE.2018.000672.

[14] R. Yang, Y. He, A. Mandelis, N. Wang, X. Wu and S. Huang, "Induction Infrared Thermography and Thermal-Wave-Radar Analysis for Imaging Inspection and Diagnosis of Blade Composites," *IEEE Trans. Industrial Informatics*, vol. 14, no. 12, pp. 5637-5647, Dec. 2018, doi: 10.1109/TII.2018.2834462.

[15] C. Galleguillos, A. Zorrilla, A. Jimenez, L. Diaz, Á. Montiano, M. Barroso, A. Viguria, F. Lasagni, "Thermographic non-destructive inspection of wind turbine blades using unmanned aerial systems, " *Plast. Rubber Compos*. 44 (3) (2015) 98–103.

[16] Y. Peng, W. Wang, Z. Tang, G. Cao, "Non-uniform illumination image enhancement for surface damage detection of wind turbine blades." *Mechanical Systems and Signal Processing* 170 (2022): 108797.

[17] L. Wang and Z. Zhang, "Automatic Detection of Wind Turbine Blade Surface Cracks Based on UAV-Taken Images," *IEEE Trans Industrial Electronics*, vol. 64, no. 9, pp. 7293-7303, Sept. 2017, doi: 10.1109/TIE.2017.2682037.

[18] Jihong Guo, Chao Liu,Jinfeng Cao, Dongxiang Jiang. "Damage identification of wind turbine blades with deep convolutional neural networks." *Renewable Energy* 174 (2021): 122-133.

[19] Abhishek Reddy, V. Indragandhi, Logesh Ravi, V. Subramaniyaswamy. "Detection of Cracks and damage in wind turbine blades using artificial intelligence-based image analytics." *Measurement* 147 (2019): 106823.

[20] A. Foster, O. Best, M. Gianni, A. Khan, K. Collins and S. Sharma, "Drone Footage Wind Turbine Surface Damage Detection," 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Nafplio, Greece, 2022, pp. 1-5, doi: 10.1109/IVMSP54334.2022.9816220.

[21] Zhang J, Cosma G, Watkins J. Image Enhanced Mask R-CNN: A Deep Learning Pipeline with New Evaluation Measures for Wind Turbine Blade Defect Detection and Classification. Journal of Imaging. 2021; 7(3):46. https://doi.org/10.3390/jimaging7030046

[22] E. Su, S. Cai, L. Xie, H. Li and T. Schultz, "STAnet: A Spatiotemporal Attention Network for Decoding Auditory Spatial Attention From EEG," *IEEE Trans. Biomedical Engineering*, vol. 69, no. 7, pp. 2233-2242, July 2022, doi: 10.1109/TBME.2022.3140246.

[23] Q. Chen, Z. -H. Liu and M. -Y. Lv, "Attention Mechanism-based CNN for Surface Damage Detection of Wind Turbine Blades," 2022 International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM), Xiamen, China, 2022, pp. 313-319, doi: 10.1109/MLCCIM55934.2022.00061.

[24] H. Iiduka, "Appropriate Learning Rates of Adaptive Learning Rate Optimization Algorithms for Training Deep Neural Networks," *IEEE Trans. Cybernetics*, doi: 10.1109/TCYB.2021.3107415.

[25] Yunxuan Dong, Xuejiao Ma, Tonglin Fu, Electrical load forecasting: A deep learning approach based on K-nearest neighbors, Applied Soft Computing, Vol. 99,(2021),106900,ISSN 1568-4946,doi:10.1016/j.asoc.2020.106900.

[26] A. Mvoulana, R. Kachouri, M. Akil, "Fine-tuning Convolutional Neural Networks: a comprehensive guide and benchmark analysis for Glaucoma Screening," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 6120-6127, doi: 10.1109/ICPR48806.2021.9412199.

[27] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[28] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. CVPR*, LV, USA 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[29] Meng-Hao Guo, *et al.* "Attention mechanisms in computer vision: A survey." *Computational Visual Media* (2022): 1-38.

[30] J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks," *Proc. CVPR*, 2018, pp. 7132-7141, doi: 10.1109/CVPR.2018.00745.

[31] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," *Proc. CVPR*, 2020, pp. 11531-11539, doi: 10.1109/CVPR42600.2020.01155.

[32] X. Zhu, D. Cheng, Z. Zhang, S. Lin and J. Dai, "An Empirical Study of Spatial Attention Mechanisms in Deep Networks," *Proc.* ICCV, 2019, pp. 6687-6696, doi: 10.1109/ICCV.2019.00679.

[33] Q. Hou, D. Zhou and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," *Proc. CVPR*, 2021, pp. 13708-13717, doi: 10.1109/CVPR46437.2021.01350.

[34] S. Woo, J. Park, J. Lee, and I. S. Kweon, “CBAM: convolutional block attention module,” in *Proc.* ECCV, Sept. 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1

[35] X. Wang, R. Girshick, A. Gupta and K. He, "Non-local Neural Networks," in *Proc. CVPR*, 2018, pp. 7794-7803, doi: 10.1109/CVPR.2018.00813.

[36] A.Dosovitskiy., *et al.* "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[37] Xueheng Hu, Shuhuan Wen, H.K. Lam, Dynamic random distribution learning rate for neural networks training, Applied Soft Computing, Vol.124, (2022) 109058, doi: 10.1016/j.asoc.2022.109058.

[38] R. R. Selvaraju et al, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proc.* ICCV, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74