Article | Received 26 December 2025; Revised 25 March 2026; Accepted 17 April 2026; Published 29 May 2026
https://doi.org/10.55092/ae20260001

# DWDM swin transformer-based multi-scale sampling aggregation network for defect image segmentation of composite materials

**Zhujun Wang and Yongjin Ling***

Automation of School, Shenyang aerospace of University, Shenyang, China

* Correspondence author; E-mail: lingyongjin@stu.sau.edu.cn.

## Highlights:

- Thanks to deep learning, the proposed solution has better neural learning methods and fewer errors. Compared to existing defect segmentation methods, this method has higher control conservatism.
- This article provides a new approach to solving the segmentation problem of composite material defects and improves segmentation accuracy.
- By utilizing the advantages of the activation function, the proposed scheme more clearly segments the defect edges, making it easier to solve the problem of unclear defect boundaries.

**Abstract:** Composite materials play a crucial role in aircraft manufacturing, as their performance and reliability directly affect the safety and efficiency of the aircraft. However, in practical applications, many composite material image segmentation tasks face problems such as highly similar shapes and colors between different types of materials, defects, and peripheral components, imposing significant limitations on the improvement of segmentation performance and thus restrict the development of composite material defect detection technology. Therefore, this paper proposes a multi-scale sampling aggregation network based on Dynamic Window Downsampling module (DWDM) Swin Transformer, which uses Swin Transformer as the basic architecture and introduces two innovative modules: DWDM downsampling module and cross entropy loop integral loss function. The DWDM downsampling module effectively reduces the dimensionality of the feature map by dynamically adjusting the window size, reducing the complexity of the model and the risk of overfitting. By reducing the spatial dimension of the feature map, downsampling significantly reduces the computational complexity of subsequent layers, thereby significantly improving the computational efficiency of the model. The cross entropy loop integral loss function is an innovation based on the traditional cross entropy loss function. By introducing the cyclic integration method, this loss function not only considers pixel level prediction accuracy, but also takes into account the cumulative effect along specific paths along the image boundary. This innovative loss function design makes the model more accurate in handling boundaries and complex shapes, effectively solving the problems of inaccurate edge segmentation and low segmentation accuracy in traditional methods. The experimental results indicate that the DWDM Swin Transformer network achieves advanced performance. Compared with U-Net, YOLOV8 and other

networks, the detection rate of this network has increased by 3.2%, and the missed detection rate and false detection rate have been reduced by 1% and 2%, respectively.

**Keywords:** compound material; image segmentation; swin transformer; multi-scale sampling; loop integral loss function

## 1. Introduction

Composite materials are becoming increasingly widely used in fields such as aerospace, automotive, and wind turbine blades due to their lightweight and high strength characteristics [1]. With the increase of applications, higher requirements have been put forward for quality control and defect detection of these materials. The advancement of Industry 4.0 and intelligent manufacturing has led to a growing demand for automated testing and quality control [2]. There are different types of composite materials, including glass fiber reinforced composites, carbon fiber reinforced composites [3], boron fiber reinforced composites, ceramic particle reinforced metal matrix composites, metal particle reinforced polymer matrix composites, *etc.*

Deep learning can currently automatically learn useful features from raw data without the need for manual features design. Hierarchical feature representation can be achieved through a multi-layered structure. Each layer of the network can learn different levels of features, from simple features at the bottom to complex features at the top. This hierarchical feature representation can better capture complex patterns in data. Deep learning can effectively process high-dimensional data through nonlinear transformations of multi-layer neural networks.

In the field of composite material defect image analysis, deep learning based image segmentation techniques are playing an irreplaceable and critical role [4–8]. Whether it is precise segmentation of images with different chemical material properties, detailed segmentation of image property differences caused by temperature changes, or direct focus on material defects in image segmentation [9], deep learning technology can efficiently and accurately identify and segment defect areas from composite material images [10–12] with its powerful capabilities. The efficiency and accuracy of this technology are attributed to the deep interdisciplinary integration of materials science, computer science, and data science. It is precisely this interdisciplinary collaboration that greatly promotes the widespread application and development of image segmentation technology in the field of composite material defect detection. The widespread use of composite materials in key fields such as aerospace, automotive manufacturing, and electronic equipment has further spurred a strong demand for defect detection technology in composite materials. Especially in safety critical applications such as aerospace structural components, there are extremely strict requirements for the accuracy and reliability of defect detection [13], which undoubtedly provides broad space and opportunities for the deep learning image segmentation technology to be deeply applied and continuously developed in this field. With the increasing attention to environmental impact and sustainability in technological development, the requirements for durability and reliability of materials in industrial manufacturing have also been raised. Especially in the aviation field, composite materials are widely used in key components of aircraft, such as wings, fuselage, *etc.* The structural integrity and performance stability of these components are crucial for flight safety. Therefore, even minor defects in composite materials can have a serious impact on the operation of the entire aircraft. In this case, achieving high-precision detection of defects in composite materials is particularly important.

It can not only ensure the safety of the aircraft, but also improve its reliability and durability. Improving detection rate and reducing missed and false detection rates have become particularly crucial. At present, ultrasonic testing [14], radiographic testing [15], and eddy current testing [16] are mostly used for defect inspection in composite materials. In addition, the computational efficiency of algorithms is also an important issue. In industrial production, rapid response is the key to improving production efficiency. However, some deep learning models, due to their complexity, have slow processing speeds in practical applications and cannot meet the needs of real-time detection. In addition, there are also issues with the interpretability of the algorithm. In defect detection, understanding the decision-making process of the model is crucial for improving user trust and making subsequent process improvements. However, deep learning models are often considered "black boxes" and their decision-making processes are difficult to explain. With the development of composite material technology, new types of defects and material properties continue to emerge, which requires algorithms to be continuously updated and adapted. However, updating and optimizing algorithms often require a significant amount of time and resources, which is a challenge for rapidly changing industrial environments. Therefore, how to design algorithms with high generalization ability, high efficiency, and good interpretability is an urgent problem to be solved in the field of composite material defect image processing. However, the above methods have significant limitations. The reinforcement of composite materials and different phases such as the matrix may have similar gray value ranges in the image, and the characteristics of defects may not be clear enough, making them susceptible to factors such as background noise and lighting conditions. Composite materials typically have complex multilayer and intersecting structures, which can result in multiple reflections and transmissions of detection signals, affecting their transmission characteristics and detection capabilities. For composite materials with increasingly complex microstructures, traditional detection models and algorithms relying on manual feature engineering or rule-based matching face inherent limitations in achieving accurate characterization. These conventional approaches often struggle to capture the non-linear, multi-scale correlations within complex microstructures, leading to compromised detection precision and robustness in practical applications. Against this backdrop, deep learning methods have emerged as a promising alternative, offering the capability to automatically learn hierarchical feature representations from raw microstructure data without explicit human intervention. In the inspection of aerospace composite wings, the false detection rate of ultrasonic image segmentation must be below 3% to meet installation requirements, whereas existing algorithms generally exceed 5%.

In recent years, convolutional neural networks (CNNs) represented by ResNet [17] have exhibited exceptional feature extraction performance. On this basis, the Fully Convolutional Network (FCN) [18] innovatively substituted fully connected layers in conventional CNNs with convolutional operations, enabling the retention of image spatial information and adapting to semantic segmentation tasks. Drawing on the design concept of FCN, U-Net [19], a classic U-shaped network framework, was developed to address the segmentation challenges of defect images in composite materials. Equipped with an encoder–decoder architecture and skip connection mechanisms, this model effectively captures multi-scale contextual features and achieves promising results in various composite material segmentation scenarios. Consequently, it has gradually become the mainstream baseline for state-of-the-art segmentation algorithms in this field. Driven by the success of U-Net, a series of improved U-shaped variants have been continuously proposed and validated with favorable performance, including UNet++ [20], SegNet [21], DenseUNet [22], KiU Net [23], and CAB U-Net [24].

Building upon the remarkable achievements of emerging approaches like the Swin Transformer, this study introduces such advanced techniques into the domain of composite material image segmentation. Meanwhile, taking the Swin Transformer as the backbone, two dedicated modules are embedded into the U-shaped network architecture to address the remaining bottlenecks in composite defect segmentation. In practical composite defect datasets, target defects vary drastically in scale, and they also present distinct color distributions and contrast characteristics compared with surrounding background regions. Nevertheless, most existing algorithms suffer from excessive time complexity and fail to achieve precise segmentation for defects with large morphological and feature discrepancies. The pyramid structure [25] captures multi-scale contextual cues through dilated convolutions with varying rates or multi-resolution feature inputs. DeepLab [26] adopts dilated convolutions with diverse receptive fields to strengthen the model's capability in mining multi-scale semantic features. SSPP [27] leverages pooling operations with different spatial ranges to construct multi-resolution feature representations. Traditional detection methods struggle to effectively identify small local features and fail to adjust their receptive fields optimally for defects of varying sizes in composite materials, a challenge exacerbated by the ambiguous definition of defect scales. Defect scales in composites can be defined by absolute size, relative proportion to microstructural components, or functional impact on material performance. The inability of traditional methods to adapt to these diverse scale definitions across detection stages results in suboptimal feature extraction and compromised detection accuracy.

Inspired by the achievements of Transformer in image segmentation, numerous deep learning networks adopting Transformer as the backbone have emerged for computer vision (CV) tasks [28]. Visual Transformer (VIT) segments images into multiple small blocks, which are then input as sequences into Transformer models for processing, producing convincing results on datasets. Efficient Transformer [29] aims to improve the efficiency of Transformer by introducing conditional computing attention and dynamic pre threshold to reduce computational complexity. Although the Transformer based structures mentioned above have achieved excellent performance in certain image recognition scenarios, such models generally impose high demands on computational hardware. The core reason why efficient Transformers still face strict hardware requirements is due to the inherent characteristics of their underlying mechanisms, which require attention to the computational complexity of self attention mechanisms. The multi branch feature interaction process requires the simultaneous storage of a large number of intermediate feature maps, which places extremely high demands on GPU memory capacity. To address these issues, Swin Transformer proposes a shift window based self attention mechanism that improves the model's generalization ability. It also substantially lowers computational overhead by means of cross-window interactions, and delivers outstanding performance across a wide range of visual tasks, including image classification, object detection, and semantic segmentation. Due to the common problem of ignoring semantic boundaries and treating all spatial regions equally in the downsampling process of Transformer series networks, important details may be lost during the downsampling process, and important details on object boundaries may be lost during the downsampling process, which increases the complexity of the decoder in recovering accurate boundaries.

The difficulty in obtaining data in processing defect images of composite materials is also a challenge. In practical industrial applications, obtaining high-quality defect images often relies on expensive detection equipment and complex acquisition processes. Due to the diversity of composite materials and the complexity of defect types, the collected image data often suffers from significant

differences and imbalanced samples. These factors not only increase the difficulty of data preprocessing, but also pose higher requirements for the training and generalization ability of deep learning models. In addition, due to the fact that defects in composite materials often occur at the microscale, traditional imaging techniques may struggle to capture subtle defect features, limiting the significance of deep learning models in defect detection and classification. At the same time, privacy and security issues of data are also important factors to consider in the process of data acquisition, especially when it comes to sensitive industries and enterprise data. The composite material board is shown in Figure 1.

**Figure 1.** Composite material sheet.

Based on the above description, we name this method Swin Dynamic Window Downsampling Aggregation Network (DWDM Win Transformer). The proposed DWDM Win Transformer introduces two innovative points: defect segmentation in composite material images, which improves segmentation accuracy.

The basic contributions of the article are mainly summarized as the following two aspects:

We propose a DWDM module capable of fusing multi-scale semantic features to cascade multiscale contextual representations, which enables the network to adaptively focus on feature information at diverse scales. This design not only suppresses noise interference introduced by shallow features, but also substantially enhances the overall feature representation capacity. Furthermore, the presented DWDM module exhibits superior edge perception for defects of varying sizes, effectively alleviating the segmentation difficulties brought by drastic scale variations of defect targets.

We found that the traditional cross entropy loss function, due to its logarithmic operation characteristics, is difficult to fully abstract the overall features of defects when dealing with composite material defect problems. To solve this problem, we suggest an innovative loop integration algorithm. This algorithm first extracts the pixel points at the defect boundary, and then treats the predicted vector of each pixel as a vector field. Through loop integration operation, it not only considers the prediction accuracy at the pixel level, but also adds the cumulative effect along the image boundary or specific path. This method that comprehensively considers local and global information can more accurately reflect the overall characteristics of defects, thereby improving segmentation accuracy. By introducing DWDM module and cross entropy loop integral loss function, the network architecture proposed in this paper performs well in the task of composite material defect image segmentation, significantly improving segmentation performance and providing new ideas and methods for the development of this field.

The rest of this article is organized as Figure 2. The second section introduces the image segmentation algorithm proposed in this article. In the third section, segmentation experiments on composite material datasets were introduced. Finally, we draw a conclusion in the fourth section.

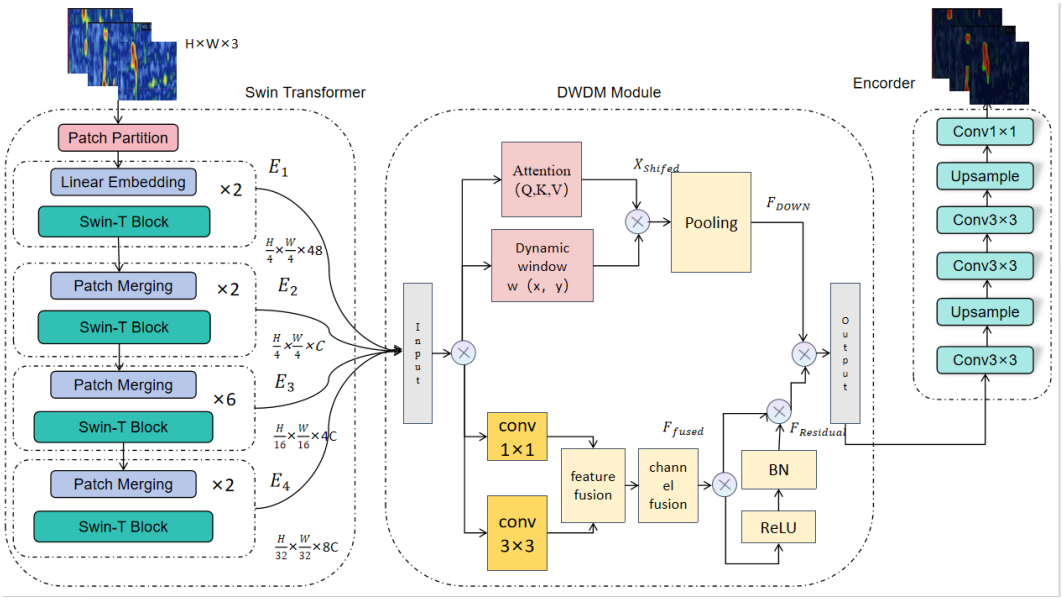


**Figure 2.** This work is built upon the newly developed DWDM Swin Transformer framework. Specifically, composite material images are input into a Swin Transformer-based encoder to extract multi-scale feature maps with varying dimensions. The dynamic window downsampling module is adopted to achieve effective feature aggregation, where the network's inherent attention mechanism is further leveraged to refine fused features. Afterwards, the extracted high-level features are progressively upsampled to restore the resolution of the original input image, ultimately yielding the final segmentation prediction.

## 2. Methods

### *2.1. Overview*

In this section, the proposed DWDM Swin Transformer, a decoder and encoder based on a U-shaped structure, is described in detail, as shown in Figure 2. Firstly, the Swin Transformer structure used in the model was introduced, followed by other important implementation details such as how the DWDM module effectively integrates the contextmulti-scale feature information and how the cross entropy loop integral loss function combines aggregation characteristics.

### *2.2. Swin Transformer encoder*

The Swin Transformer model [30] is a Transformer model proposed in the field of computer vision in 2021, which combines the idea of convolutional localization and constructs a network based on self attention mechanism. It can capture long-distance dependencies in images and has good response performance.

The Transformer structure first divides the input image into a series of non overlapping blocks, and then linearly flattens the pixels of these blocks into several one-dimensional vectors. Then, these vectors are processed through a multi head self attention (MSA) module to calculate the self attention

mechanism, and finally processed using Gaussian Error Linear Unit (GELU) nonlinear and Multi Layer Perceptron (MLP) layers, with each module preceded by a Layer Normalization (LN) layer processed through a residual connection.

Due to dividing the image into small blocks, Transformer needs to calculate the relationships between these blocks, which means that the computational cost continues to increase. Therefore, Swin Transformer adopts window based MSA (W-MSA) and shift window based MSA (SW-MSA), which change the conventional window partitioning method and use a more effective shift window partitioning strategy. It not only maintains effective computation of non overlapping windows, but also establishes connections between different windows. Two consecutive Swin Transformer blocks use W-MSA and SW-MSA modules respectively to establish connections between different windows. $\tilde{Z}^l$and $Z^l$represent the outputs of W-MSA and MLP in the l-th layer, which can be expressed as the following Equation:

$$\tilde{Z}^l = W - MSA\left(LN\left(Z^{l-1}\right)\right) + Z^{l-1} \quad (1)$$

$$Z^l = MLP(LN(\tilde{Z}^l)) + \tilde{Z}^l \quad (2)$$

By using a shift window partitioning strategy, the outputs of SW-MSA and MLP modules can be represented as:

$$\tilde{Z}^{l+1} = SW - MSA\left(LN\left(Z^l\right)\right) + Z^l \quad (3)$$

$$Z^l = MLP(LN(\tilde{Z}^{l+1})) + \tilde{Z}^{l+1} \quad (4)$$

In the article, considering complexity and time efficiency, we use Swin-T as the encoder for feature extraction. Divide the input image $X \in R^{H \times W \times 3}$ into $(H/s) \times (W/s)$non overlapping small blocks, where s represents the size of the small blocks, which are flattened and projected onto the c-dimension through a linear embedding layer. Next, these patch tokens are fed into Swin Transformer and undergo four stages in sequence to extract features. Each stage has several consecutive Swin Transformer blocks, including W-MSA and SW-MSA, and then uses patch merging modules to downsample and double the dimensionality of features. The outputs of each level are $(H/s) \times (W/s) \times C$, $(H/2s) \times (W/2s) \times 2C$, $(H/4s) \times (W/4s) \times 4C$, and $(H/8s) \times (W/8s) \times 8C$.

*2.3. DWDM module*

The diversity of composite materials and the wide range of defect types require algorithms to have extremely high generalization ability to adapt to different material and defect characteristics. Therefore, a Dynamic Window Downsampling Module is proposed, which combines the idea of shifting windows with the downsampling module, and combines the shifting window with the downsampling stage to make the sampling method flexible and achieve the goal of feature extraction. By reducing the spatial dimension of feature maps, downsampling [31] can significantly reduce the computational complexity and parameter count of subsequent layers, thereby improving the computational efficiency of the model. Reducing the dimensionality of feature maps can effectively reduce the number of parameters and computational complexity of the model, thereby reducing the risk of overfitting and improving the model's generalization ability to unseen data. As the feature map passes through the downsampling layer, the model is able to extract higher-level and more abstract features from the original input, which

is crucial for understanding complex patterns. In the Swin Transformer network architecture, the use of downsampling can achieve the syncretism of multi-dimension features, which helps to improve the performance of the model. The overall block diagram is shown in Figure 3.

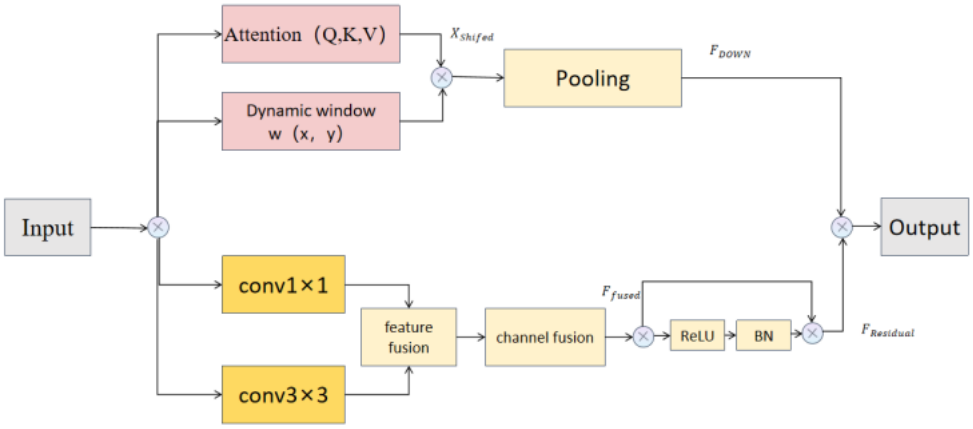


**Figure 3.** Structural diagram of dynamic window downsampling.

Firstly, consider the input feature map $X \in R^{B\times C\times H\times W}$, where B is the batch size, C is the number of channels, H and W are the height and width of the feature map, respectively. The dynamic window downsampling module divides the feature map into a series of overlapping or non overlapping windows, each with a size of M × M. For each window, a window function is defined to represent whether the position $\mathrm{w}(x,y)$ belongs to the current window. The window function can be a rectangular window, a circular window, or any other shape of window. For dynamic windows, the size and position of the window can be adaptively adjusted based on the content of the input feature map. Within each window, a self attention mechanism is applied to capture the dependency relationships within the window. The self attention mechanism can be expressed as:

$$Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

Among them, Q, K, and V are respectively the query, key, and value, and $d_k$ are scaling factors used to control the scale of attention scores. In consecutive layers, the position of the window will be shifted. The shift operation can be expressed as:

$$X_{shifted} = X + \Delta \tag{6}$$

Among them, Δ is the offset calculated based on the shift strategy. The shift operation helps the model capture cross window dependencies. Next, we will sort out the downsampling stage. Downsampling operations are usually implemented through convolution or pooling layers, which can be expressed as:

$$F_{down} = DOWNSAMPLEX_{shifted} \tag{7}$$

Among them, let $F$ be the original feature map, define the downsampled feature map as $F_{\text{down}}$, $DOWNSAMPLEX_{shifted}$ represents offset pooling downsampling operation.As mentioned earlier, our input feature map is $X \in R^{B\times C\times H\times W}$, where B is the batch size, C is the number of channels, H and W are the height and width of the feature map, respectively. The max pooling operation selects the maximum value in each local region as the $k\times k$ output. For each channel, max pooling can be expressed as:

$$Y(i,j)=\max_{m=0}^{k-1}\max_{n=0}^{k-1}X(i\cdot s+m,j\cdot s+n) \tag{8}$$

Among them, $Y(i,j)$ is the value of the output feature map at position (i, j), and $X(i\cdot s+m,j\cdot s+n)$ is the value of the input feature map at position.

The original feature map is processed using multi-scale feature fusion and residual operation. The feature map $H\times W$ is passed through two convolution kernels of different sizes 1 × 1 and 3 × 3, with outputs of $F_1$ , $F_2$, and channel numbers are $C_1$ , $C_3$ , respectively, to extract multi-scale features, as shown in Equations (9) and (10):

$$F_1=\mathrm{Conv}_{1\times1}(X) \tag{9}$$

$$F_2=\mathrm{Conv}_{3\times3}(X) \tag{10}$$

Concatenate the above two feature maps in the channel dimension, as shown in Equation (11):

$$F_{\mathrm{Concat}}=\mathrm{Concat}(F_1,F_3) \tag{11}$$

The size of the concatenated feature map $F_{concat}$ is $H\times W$ , and the number of channels is $C_1+C_3$ . To ensure consistency in the number of channels, a 1x1 convolution is used for channel fusion, as shown in Equation (12):

$$F_{\mathrm{fused}}=\mathrm{Conv}_{1\times1}(F_{\mathrm{concat}}) \tag{12}$$

The size of the output feature map $F_{\mathrm{fused}}$ is $H\times W$ , and the number of channels is $C$ . The residual part contains two convolutional layers, with batch normalization and ReLU activation function used in between. The convolutional layers used are shown in Equations (13) and (14):

$$F_{\mathrm{conv1}}=\mathrm{ReLU}\left(BN\left(\mathrm{Conv}_{3\times3}(F_{\mathrm{fused}})\right)\right) \tag{13}$$

$$F_{\mathrm{conv2}}=BN\left(\mathrm{Conv}_{3\times3}(F_{\mathrm{conv1}})\right) \tag{14}$$

After residual connection, the output is $F_{\mathrm{res}}$ obtained with the number of channels $C$ , as shown in Equation (15):

$$F_{\mathrm{res}}=F_{\mathrm{fused}}+F_{\mathrm{conv2}} \tag{15}$$

Merge the downsampled feature map with the original feature map to preserve multi-scale information, as shown in Equation (16):

$$F_{out}=F_{down}+F_{res} \tag{16}$$

Among them, $F_{res}$ represents the result of the original feature map after continuous residual connections.

The DWDM module can extract the features of defects in the dataset through the operation of shifting windows, and then obtain the defect features of the feature map.

### *2.4. Enconder*

The decoder integrates convolutional blocks, upsampling, and skip connections to gradually restore the resolution of the feature map, as shown in Figure 2. At each level of the decoder, the current intermediate feature map is first fused with the cascaded feature information output by the DWDM module at the same level; Subsequently, the output features of this level are generated through a 3 × 3 convolution

operation to optimize the quality of the upsampled results. In addition, in the final stage of the encoder path, the feature map processed in the previous stage will be upsampled first, and then the number of channels will be adjusted through 1 × 1 convolution to restore the feature map to the same size as the input image.

### *2.5. Loss function*

It is particularly important to design a targeted loss function when dealing with defect segmentation problems in composite functions. This loss function not only effectively addresses the issues of imbalanced datasets and unclear segmentation categories, but also drives the model to learn more generalized features, thereby reducing the risk of overfitting. Integrating domain knowledge or prior constraints into the loss function can better adapt the model to specific application scenarios. For data with complex structure or uneven distribution, a carefully designed loss function can help the model understand and process these data more accurately. In addition, some improved loss functions can reduce computational complexity and improve the efficiency of model training.

Due to the fact that the image of composite material defects is not regular but contains multiple different defect pixels, traditional loss functions may miss some key data when collecting information, resulting in biased network analysis results. Therefore, we propose an improved loss function called the cross entropy loop integral loss function. This loss function can capture defect information more comprehensively, improving the accuracy and reliability of segmentation.

The calculation of the cross entropy loop integral loss function is shown in Equation (17):

$$L_{total} = \lambda L_{CE} + (1-\lambda) L_{line-integral} \tag{17}$$

Among them, $\lambda$ is a weight parameter used to balance the contributions of two loss terms. As the problem to be solved varies in different situations, the value can be set according to the problem faced by oneself.

For classification tasks, the cross entropy loss function [32] is defined as:

$$L_{CE} = -\sum i y_i \log(p_i) \tag{18}$$

Among them, $y_i$ is the individual encoding of the real label, which $p_i$ is the probability predicted by the model. This Equation calculates the difference between the model prediction and the true label. Due to the convergence characteristics of logarithmic functions, they tend to monotonically increase in data volume. However, in cases where the data volume is too large, the convergence may be too gradual, resulting in unclear comparison between predicted and actual results.

Vector field prediction tasks [33] typically occur in fields such as physics, engineering, and computer vision, with the aim of predicting vector values at a given point in a given space. In computer vision, vector field prediction typically involves predicting the vector field in images and defects. Here, we apply it to predict the vector composed of composite material defects [34]. For vector field prediction tasks, the cyclic integral loss function is defined as:

$$L_{line-integral} = \oint_C F \cdot dr \tag{19}$$

*F* is usually a vector field predicted by a model, and in image segmentation, object detection, or other computer vision tasks, F can be a feature map or gradient map. In this case, we define F as the output layer from the model, which is the output of the last convolutional layer of the convolutional neural network:

$$F = CNN_{Output}(X) \tag{20}$$

Among them, X is the input image, $CNN_{Output}$ representing the output layer of the convolutional neural network.

*P* is the path of the loop integral [35], which can be a straight line, curve, or closed path. Define the path based on the specific geometric shape and structure of the problem. In image segmentation algorithm processing, the path can be the contour along the edge of the object. So we define it as the counterclockwise path of the defect, keeping the defect itself on the left side of the vector field:

$$P = \left\{ r(t) \middle| t \in (0,1) \right\} \tag{21}$$

Among them, r (t) is the parameterized representation of path P.

dr is an infinitesimal displacement vector on path P, used for calculating loop integrals. In numerical calculations, dr can be approximated by discretizing the path and calculating the difference between adjacent points. If path P can be described by a mathematical function, dr can be analytically calculated by taking the derivative of the path function:

$$dr = \frac{dr(t)}{dt} dt \tag{22}$$

So, the Equation for calculating curve integral is:

$$\oint_c F \bullet dr = \int_0^1 F\left(r(t)\right) \bullet \frac{dr(t)}{dt} dt \tag{23}$$

In numerical calculations, path C can be approximated by discretizing it into N points:

$$\oint_C F \bullet dr \approx \sum_{i=1}^{N} F\left(r(i)\right)\left(r_{i+1} - r_i\right) \tag{24}$$

Among them, are the discrete points on path C, and the curve integral is shown in Figure 4.

The cross entropy function provides an efficient parameter update approach for optimization algorithms due to its simplicity in gradient calculation. When dealing with probability values close to 0 or 1, the cross entropy function exhibits good stability, which helps prevent numerical instability during training. As a measure based on probability distribution differences, cross entropy loss not only provides intuitive probabilistic explanations, but also facilitates our understanding and interpretation of the model's behavior.

Given the advantages of the cross entropy function, we propose a novel loss function that combines cross entropy loss and loop integral loss. This combination aims to fully utilize the advantages of cross entropy function in probabilistic interpretation, while enhancing the model's ability to capture cumulative effects on image boundaries or specific paths through loop integral loss, thereby more effectively handling complex image segmentation tasks. Through this innovative loss function design, we expect the model to learn richer and more robust feature representations, thereby improving the accuracy and reliability of segmentation.

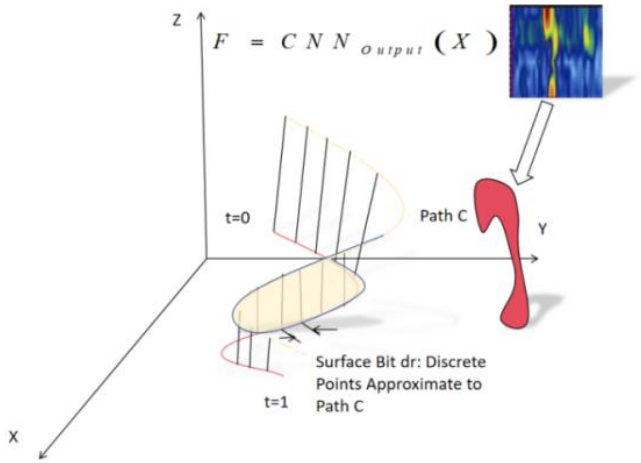


**Figure 4.** Schematic diagram of curve integral loss function.

## 3. Experiment

Training and validation were carried out on a composite material dataset, with mainstream evaluation metrics adopted to quantitatively assess the segmentation performance of the proposed method. Furthermore, comprehensive ablation experiments were implemented to verify the effectiveness of each component.

### *3.1. Dataset and implementation*

This dataset consists of 540 defect images of carbon fiber composite panels. 80% of them are used for training (a total of 432 images), 10% are used for validation (a total of 54 images), and 10% are used for testing (a total of 54 images). Data augmentation strategies including multi-dimensional random flipping and shifting are adopted to expand the training dataset and boost segmentation performance. The proposed network architecture is developed based on the PyTorch framework, and all experimental tests are completed on workstations equipped with A100 computing servers. The testing environment ensures room temperature to prevent the composite material board from being affected by external factors such as temperature and pressure, in order to improve image quality. The experimental environment is shown in Figure 5. Set the learning rate to 0.0001, optimizer Adam, and weight decay to 0.05.

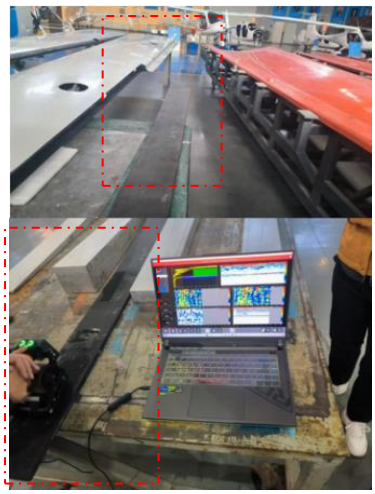

**Figure 5.** Experimental environment.

For the tested board, the defect location was marked by the production workers in order to make our detection form more accurate. An example of a composite material defect image is shown in Figure 6. The defect is pore defect.

**Figure 6.** Composite material defects.

In this inspection, the surface of the composite material board was placed on the workbench. By uniformly spraying coupling agent onto the surface of the composite material board, the detection instrument passed through the inspected area at a constant speed. The dataset consists of 540 carbon defect images to form a fiber composite board. Figure 7 shows the dataset, and the red box represents the defect.

In order to improve the robustness and generalization ability of the algorithm, various data augmentation techniques have been adopted, including random rotation, scaling, and flipping. These enhancement methods expand the data size to 540 pieces. The dataset is randomly divided into a training set of 432 pieces, a validation set of 54 pieces, and a test set of 54 pieces in a ratio of 8:1:1. Table 1 shows the instance distribution of defects in the dataset.

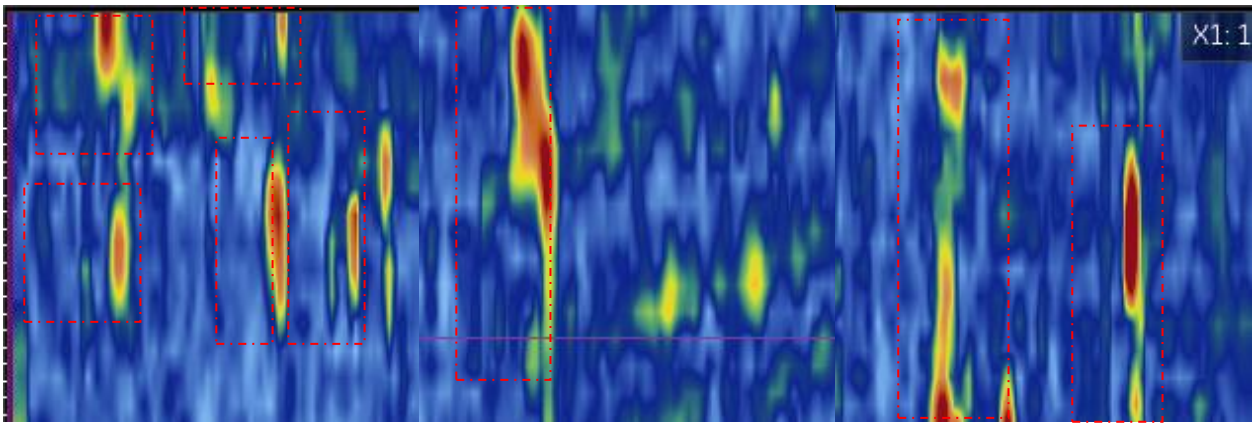


**Figure 7.** Dataset.

**Table 1.** Training set, validation set, and test set distribution of defect examples in composite materials.

| Dataset | Training set | Validation set | Test set | Total |
|---|---|---|---|---|
| Number | 432 | 54 | 54 | 540 |

We use commonly used evaluation metrics to quantify the segmentation performance of the proposed algorithm. Provide a composite material defect dataset with some indicators, including average Dice coefficient (mDice), mean intersection point (mIoU), false positive rate (FNR), and false positive rate (FPR). The detailed definition of indicators is shown in Equations (25)–(28):

$$mDICE = \frac{2 \times TP}{FN + FP + 2 \times TP} \tag{25}$$

$$mIoU = \frac{TP}{FP + FN + TP} \tag{26}$$

$$FNR = \frac{FN}{FN + TP} \tag{27}$$

$$FPR = \frac{FP}{FP + TN} \tag{28}$$

In Equations (25)–(28), TP (true positive) represents the number of defects correctly identified by the algorithm; TN (true negative) represents the number of correctly identified non defects; FP (false positive) and FN (false negative) refer to the number of non defects misclassified as defects and the number of defects missed, respectively.

### *3.2. Ablation study*

In this section, in order to verify the performance, the comparison between DWDM Swin and the scores of Transformer and source network Swin Transformer are used to evaluate the segmentation performance in ablation research. The specific performance is shown in Table 2.

We can see that the proposed model greatly improves mDice,and reduces the missed detection rate and false detection rate, meeting our requirements. Figure 8 shows the specific segmentation results, where the red part represents the segmented defects and the green part represents the noise in the image.

**Table 2.** Results of ablation experiment.

| **Method** | **mDICE** | **mIoU** | **mIoU** | **mIoU** |
|---|---|---|---|---|
| Swin-T | 0.899 | 0.869 | 0.062 | 0.051 |
| Swin-T+DWDM | 0.911 | 0.864 | 0.061 | 0.055 |
| Swin-T+loss | 0.920 | 0.868 | 0.061 | 0.042 |
| Ours | **0.922** | **0.872** | **0.051** | **0.031** |

The curve in Figure 9 clearly indicates that the improved algorithm outperforms other mainstream object detection algorithms in key metrics. The DWDM Swin Transformer not only quickly achieved high performance in the early stages of training, but also maintained stable and excellent performance throughout the entire training process, further proving that the DWDM Swin Transformer has significant performance improvement and faster convergence speed in composite material defect detection tasks. After adding the DWDM module, the missed detection rate decreases by 0.5%, which in industrial practice translates to 5 fewer missed delamination defects per 1,000 inspected products, thereby reducing service risk.

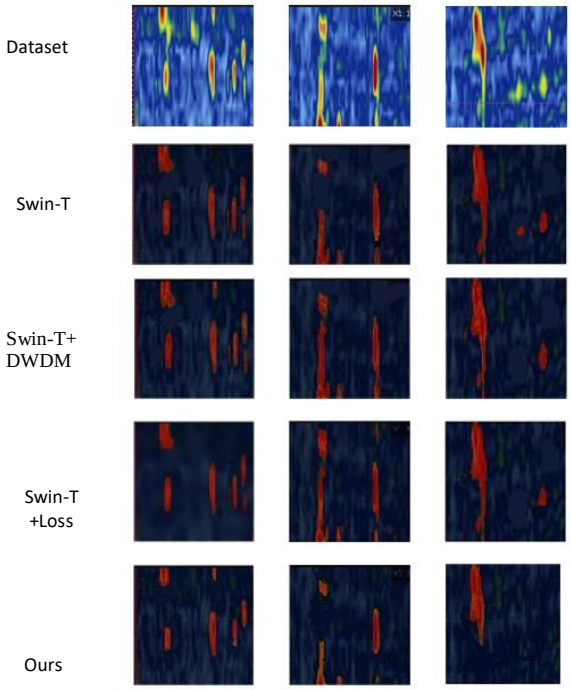


**Figure 8.** As shown in the figure, the original image is divided into Swin-T and improved DWDM Swin-T segmentation results.

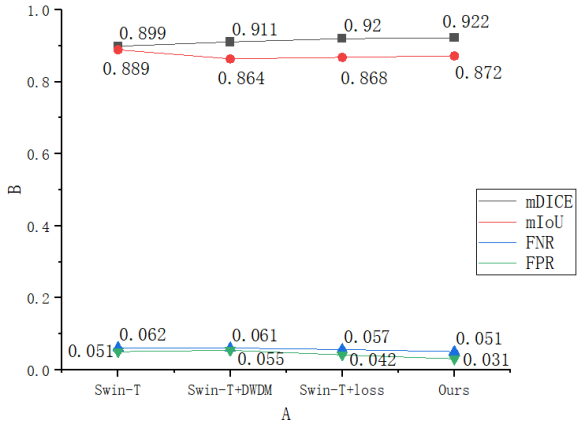


**Figure 9.** Curves of different algorithms.

*3.3. Contrast test*

In order to verify the superiority of the proposed algorithm performance, U-net, Segformer, and other methods were selected as comparison objects for comparative analysis. A detailed analysis was conducted on the performance of these algorithms in key performance indicators such as mDice, mIoU, FNR and FPR. The experimental results are shown in Figure 7 and Table 3, respectively.

To verify the performance indicators of different algorithms, networks such as U-net and Segformer were selected for validation, and the experimental results are shown in Figure 10. Among them, it was found that the performance of our network in mDICE and mIoU indicators was 92.2% and 87.2%, far superior to Transformer. Not only did it quickly achieve high performance in the early stages of training,

but it also maintained stable and excellent performance throughout the entire training process, further proving other networks.

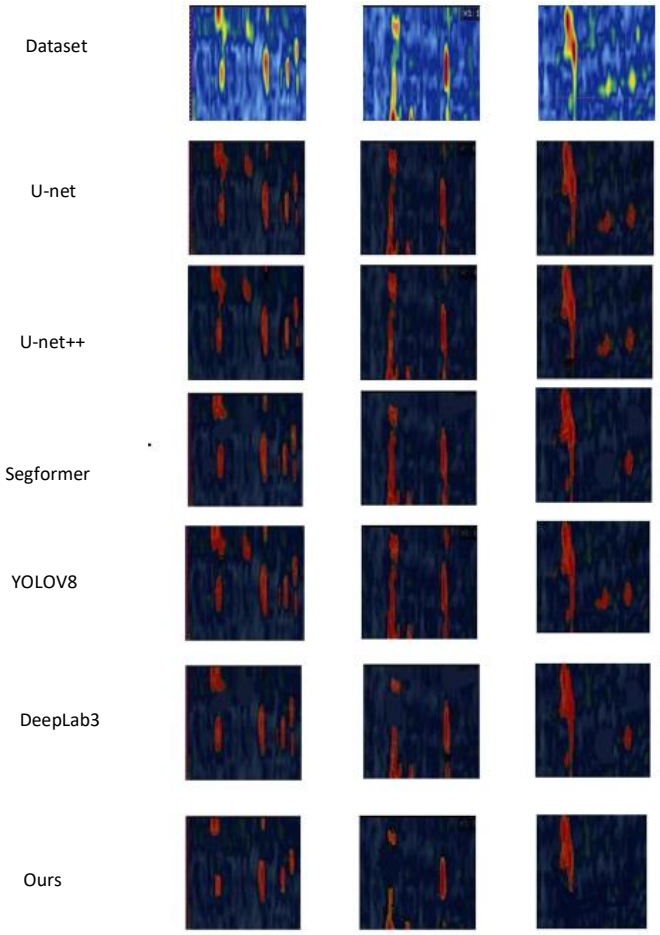


**Figure 10.** Comparative experimental results.

The curve in Figure 11 clearly indicates that the improved algorithm outperforms other mainstream object detection algorithms in key metrics. The DWDM Swin Transformer not only quickly achieved high performance in the early stages of training, but also maintained stable and excellent performance throughout the entire training process, further proving that the DWDM Swin Transformer has significant performance improvement and faster convergence speed in composite material defect detection tasks.

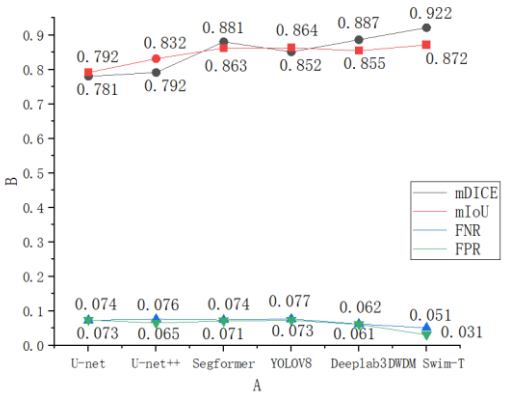


**Figure 11.** Comparative experimental results.

**Table 3.** The test results of different networks, bold indicates the best results.

| Method | mDice | mIoU | FNR | FPR |
|---|---|---|---|---|
| U-net [36] | 0.781 | 0.792 | 0.073 | 0.074 |
| U-net++ [37] | 0.792 | 0.832 | 0.076 | 0.065 |
| Segformer [38] | 0.881 | 0.863 | 0.074 | 0.071 |
| YOLOV8 [39] | 0.852 | 0.864 | 0.077 | 0.073 |
| Deeplab3 [40] | 0.887 | 0.855 | 0.062 | 0.061 |
| DWDM Swim-T | **0.922** | **0.872** | **0.051** | **0.031** |

## 4. Conclusion

This study proposes a novel multi-scale downsampling network termed DWDM Swin Transformer, which is constructed on the basis of the Swin Transformer and applied to defect segmentation tasks for composite material images. The core design of the DWDM Swin Transformer is to fully capture global contextual information and fine edge features of composite material defects through the Swin Transformer encoder and the embedded DWDM module, while eliminating invalid noise interference and redundant shallow feature information. The embedded module can effectively mitigate the problem of fuzzy adhesion between defects and background edges, which otherwise leads to poor segmentation accuracy. However, this method still requires a large number of manually labeled composite material defect samples for supervised training, and insufficient labeled data may cause potential overfitting problems. In future research, we will further strengthen the semantic characterization ability of the model to improve overall segmentation accuracy and reduce both missed detection and false detection rates. Meanwhile, considering the limitations of manual labeling, we will explore unsupervised segmentation strategies based on the existing DWDM Swin Transformer architecture to meet the high-precision application requirements of industrial composite material defect inspection.

## Data availability statement

No supplementary or additional data were generated in this study.

## Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used DeepSeek only for Page1–3 language polishing and editing. The authors take full responsibility for the content of the manuscript.

## Acknowledgments

This work was supported by Research on Rapid Prototyping Technology for Aerial Vehicles (Grant No. Zong 20240224), and Research and Development of Intelligent Landing Decision Auxiliary System for New Energy Manned Aircraft (Grant No. 2023JH26/10100001).

## Authors' contribution

Conceptualization, Z.W. and Y.L.; methodology, Z.W. and Y.L.; software, Z.W. and Y.L.; validation, Z.W. and Y.L.; formal analysis, Y.L.; investigation, Y.L.; resources, Y.L.; data curation, Y.L.; writing—original draft preparation, Z.W. and Y.L.; writing—review and editing, Z.W. and Y.L.;

visualization, Z.W. and Y.L.; supervision, Z.W. and Y.L.; project administration, Z.W. and Y.L. All authors have read and agreed to the published version of the manuscript.

## Conflicts of interest

The authors declare no conflicts of interest.

## References


[1] Balkan D. Delamination prediction in layered composites using optimized ANN algorithms: a comparative analysis. *Symmetry* 2025, 17:91.

[2] Herrera-Vidal G, Coronado-Hernández JR, Maheut J. Complexity management challenes in the industry 4.0 era: a systematic review in production systems. *Results Eng.* 2025, 26:105329.

[3] Shen L, Li X, Zhang Y, Yan W, Dai Y, *et al.* Research progress on SiCf/SiC composites and their cladding components. *AI Mater*. 2025, 1(1):0001.

[4] Endalew AM, Woo K, Cairns DS. Effect of drilling-induced defects on progressive damage of open-hole composite laminates under compression. *Adv. Compos. Mater.* 2022, 31(4):399–427.

[5] Zhang S, Cui X, Ma G, Tian R. Symmetric boundary-enhanced U-Net with mamba architecture for glomerular segmentation in renal pathological images. *Symmetry* 2025, 17:1506.

[6] Fan D, Ji G, Zhou T, Chen G, Fu H, *et al.* PraNet: parallel reverse attention network for polyp segmentation. In *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Lima, Peru, October 4–8, 2020, pp. 263–273.

[7] Wei J, Hu Y, Zhang R, Li Z, Zhou SK, *et al.* Shallow attention network for polyp segmentation. In *Proceedings of the Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Strasbourg, France, September 27–October 1, 2021, pp. 699–708.

[8] Wu H, Pan J, Li Z, Wen Z, Qin J. Automated skin lesion segmentation via an adaptive dual attention module. *IEEE Trans. Med. Imag.* 2021, 40(1):357–370.

[9] Sadhin A, Hashim S, Rayhan R. Deep learning for road defect detection from aerial imagery. *Proc. Comput. Sci*. 2023, 1:0002.

[10] Wang X, Liu M, Wang Y, Fan J, Meijering E. A 3D tubular flux model for centerline extraction in neuron, volumetric images. *IEEE Trans. Med. Imag.* 2022, 41(5):1069–1079.

[11] Chen Z, Liu B. A high-accuracy PCB defect detection algorithm based on improved YOLOv12. *Symmetry* 2025, 17:978.

[12] Ma J, Tang L, Fan F, Huang J, Mei X, *et al.* SwinFusion: cross-domain long-range learning for general image fusion via swin transformer. *IEEE/CAA J.* 2022, 9(7):1200–1217.

[13] Sheela KS, Justus V, Asaad RR, Kumar RL. Enhancing liver tumor segmentation with UNet-ResNet: leveraging ResNet's power. *Technol. Health Care* 2025, 33(1):1–15.

[14] Feng J, Yan T, Hou Z. Experimental investigation of rock damage induced by ultrasonic high-frequency vibration under high confining pressure. *Symmetry* 2025, 17:372.

[15] Sun S, Lv L, Li C, Pan J, Huang H, *et al.* Defects identification of wind turbine blades based on radiographic testing. *Nondestr. Test. Eval.* 2025, 41(4):1942–1967.

[16] J. Cao, N. Li, P. Jiang, *et al.* Research on eddy current testing technology for lead seal crack defects of high voltage cable. In *Proceedings of the 2019 IEEE 3rd International Conference on Circuits, Systems and Devices (ICCSD)*, Chengdu, China, 2019, pp. 171–176.

[17] Yu Y, Yu J, Chen Z, Wu J, Yan Y. A universal routing algorithm based on intuitionistic fuzzy multi-attribute decision-making in opportunistic social networks. *Symmetry* 2021, 13:664.

[18] Lin S, Ali N, Zain A, Amin M, Alwee R. A review of human pose estimation algorithm based on convolutional neural network. *Proc. Comput. Sci*. 2023, 1(2):0047.

[19] Zhou G, Zhang Z, Yin W, Chen H, Wang L, *et al.* Surface defect detection of CFRP materials based on infrared thermography and Attention U-Net algorithm. *Nondestr. Test. Eval.* 2023, 39(2):238–257.

[20] Wang M, Zhou F, Wang H, He X, Hou Z. High precision reconstruction of seismic data based on improved U-Net++. *Coalfield Geol. Explor.* 2025, 53(9):15.

[21] Sahin N, Alpaslan N, Hanbay D. Robust optimization of SegNet hyperparameters for skin lesion segmentation. *Multimedia Tools Appl.* 2022, 81(25):36031–36051.

[22] Zhou Y, Chang H, Lu X, Lu X. DenseUNet: improved image classification method using standard convolution and dense transposed convolution. *Knowledge-Based Syst.* 2022, 254:109658.

[23] Valanarasu JMJ, Sindagi VA, Hacihaliloglu I, Patel VM. KiU-Net: overcomplete convolutional architectures for biomedical image and volumetric segmentation. *IEEE Trans. Med. Imaging* 2022, 41(4):965–976.

[24] Ding X, Peng Y, Shen C, Zeng T. CAB U-Net: an end-to-end category attention boosting algorithm for segmentation. *Comput. Med. Imaging Graphics* 2020, 84:101764.

[25] Liu Z, Jing D, Ji C. ProCo-NET: progressive strip convolution and frequency-optimized framework for scale-gradient-aware semantic segmentation in off-road scenes. *Symmetry* 2025, 17:1428.

[26] Huang ZH,Pan YF,Huang, WR, *et al.* Predicting microvascular invasion and early recurrence in hepatocellular carcinoma using deeplab v3+segmentation of multiregional mr habitat images. *Academic Radiology*, 2025, 32(6):3342–3357.

[27] Yang X. A three frequency independent reconfigurable notch filter based on SSPP. *Wireless Commun. Technol*. 2025, 34(1):49–53.

[28] Bai Y, Zhang Y. A technical study on improving the reasoning accuracy of deep learning networks in the CV field. *Commun. Manage. Technol.* 2024, 1:34–37.

[29] Guan K, Zhou Y, Wang L, Chang B, Du D. A self-configuring transformer segmentation method for welding radiographic defect detection in steel pipes. *Nondestr. Test. Eval.* 2026, 41(4):1968–1989.

[30] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows. In *Proceedings of the International Conference on Computer Vision*, Montreal, Canada, October 10–17, 2021, pp. 9992–10002.

[31] Liu M, Chen M, Wu M, Ye Y. GAE-YOLO: global perception enhanced target detection method for external hidden danger of transmission lines. *J. Instrum.* 2025, 46(2):267–278.

[32] Gao X, Wu H, Gao H, Qi Y. Fault diagnosis of chiller based on consensus loss generative adversarial network. *J. Instrum.* 2025, 46(1):285–297.

[33] Huang P. The reducibility of vector fields on infinite dimensional torus. *Chin. Sci. Math.* 2025, 55(4):803–812.

[34] Zhang X. Research on detection of surface defects in industrial products based on support vector data description. Master's Thesis, Central University of Finance and Economics, 2024.

[35] Gong Y, Zhao P, Du L. Stability analysis of grid type inverters considering control time delay. *Foreign Electro. Meas. Technol.* 2024, 43(10):179–189.

[36] Ronneberger O, Fischer P, Thomas Brox. U-Net: convolutional networks for biomedical image segmentation. In *proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, October 5–9, 2015, pp. 234–241.

[37] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N. Unet++: a nested U-Net architecture for medical image segmentation. In *Proceedings of Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Granada, Spain, September 20, 2018, pp. 3–11.

[38] Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, *et al.* SegFormer: simple and efficient design for semantic segmentation with transformers. *arXiv* 2021, arXiv:2105.15203.

[39] Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. *arXiv* 2017, arXiv:1706.05587.

[40] Sun H, Tan CH, Pang S, Wang H, Huang B. RA-YOLOv8: an improved YOLOv8 seal text detection method. *Electronics* 2024, 13(15):3001.