

Article

Surface Defect Detection for Automated Tape Laying and Winding Based on Improved YOLOv5

Liwei Wen ^{1,*}, Shihao Li ¹ and Jiajun Ren ²

¹ College of Material Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; lishihao@nuaa.edu.cn

² Haiying Aerospace Materials Research Institute (Suzhou) Co., Ltd., Suzhou 215100, China; renjiajun@nuaa.edu.cn

* Correspondence: wenliwei@nuaa.edu.cn

Abstract: To address the issues of low detection accuracy, slow detection speed, high missed detection rate, and high false detection rate in the detection of surface defects on pre-impregnated composite materials during the automated tape laying and winding process, an improved YOLOv5 (You Only Look Once version 5) algorithm model was proposed to achieve the high-precision, real-time detection of surface defects. By leveraging this improvement, the necessity for frequent manual interventions, inspection interventions, and subsequent rework during the automated lay-up process of composite materials can be significantly reduced. Firstly, to improve the detection accuracy, an attention mechanism called “CA (coordinate attention)” was introduced to enhance the feature extraction ability, and a Separate CA structure was used to improve the detection speed. Secondly, we used an improved loss function “SIoU (SCYLLA-Intersection over Union) loss” to replace the original “CIoU (Complete-Intersection over Union) loss”, which introduced an angle loss as a penalty term to consider the directional factor and improve the stability of the target box regression. Finally, Soft-SIoU-NMS was used to replace the original NMS (non-maximum suppression) of YOLOv5 to improve the detection of overlapping defects. The results showed that the improved model had a good detection performance for surface defects on pre-impregnated composite materials during the automated tape laying and winding process. The FPS (frames per second) increased from 66.7 to 72.1, and the mAP (mean average precision) of the test set increased from 92.6% to 97.2%. These improvements ensured that the detection accuracy, as measured by the mAP, surpassed 95%, while maintaining a detection speed of over 70 FPS, thereby meeting the requirements for real-time online detection.

Keywords: automated tape laying and winding; surface defect detection; YOLOv5



Citation: Wen, L.; Li, S.; Ren, J. Surface Defect Detection for Automated Tape Laying and Winding Based on Improved YOLOv5. *Materials* **2023**, *16*, 5291. <https://doi.org/10.3390/ma16155291>

Academic Editor: Enrique Casarejos

Received: 24 June 2023

Revised: 24 July 2023

Accepted: 24 July 2023

Published: 27 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As advanced composite materials are widely applied in fields such as aerospace, the automotive industry, and medicine, related automation technologies such as composite laying and winding are also constantly developing and maturing [1–3]. Due to the processes of automated tape laying and winding, such as tension control, laying trajectory, initial positioning, and the adhesive properties of the pre-impregnated material, various defects can occur during the automated tape laying and winding process. These defects can compromise the structural integrity of the components and impact their performance. Inclusion defects arise when the recovery device fails to separate the pre-impregnated tape and the backing film due to excessive viscosity. As a result, both the tape and film are laid on the mold together during the automated tape laying and winding process. Typically, clearance defects are accompanied by overlap defects, and two types of clearance/overlap defects can be observed in composite component manufacturing: inconsistent widths of the pre-impregnated tape during production and tape misalignment caused by issues with the laying trajectory. Fold defects can be attributed to three main factors. Firstly, folds may

originate from the pre-impregnated tape itself during its production. Secondly, failure of the bond strength between the pre-impregnated material and the mold can lead to folding. Lastly, variations in the curvature radius along the laying trajectory result in different curvatures on each side of the pre-impregnated tape. As the lay-up machine undergoes turning, tension on one side and compression on the other side contribute to the formation of folds through the combined effects of these forces.

Currently, most defect detection during automated laying and winding processes, as well as the final product inspection, still relies on manual inspection. The full automation of composite material manufacturing has not been achieved yet, and the bottleneck lies in the essential defect detection process. In order to meet the high-performance requirements of composite components, the machine must stop after each layer is laid down for visual inspection and confirmation by the inspection personnel. Moreover, frequent rework is required to meet the quality assurance and requirements. At the same time, manual inspection has many drawbacks, such as high labor costs, a slow inspection speed, a low efficiency, and a tendency to miss defects. Additionally, manual inspection is subject to physiological and psychological factors, and the results can be greatly influenced by subjective factors. Different operators may have different opinions and understandings of defects, resulting in inconsistent inspection results and product quality. For a large-scale composite component that requires hundreds of layers, inspection and rework have a huge impact on production speed.

In the field of automated surface defect detection for composite materials, some non-manual detection methods have been developed over time; however, they still have limitations. Initially, specific measurement devices are used to obtain feature change maps (with simplistic color information), which are then subjected to detection algorithms for further analysis. Christopher Sacco et al. [4] used a profilometer to measure the point cloud of the layer surface and transformed it into a grayscale image. They designed a machine learning algorithm based on fully convolutional neural networks to classify defects in the grayscale image. However, point cloud data are typically voluminous and require complex algorithms and high-performance computing devices, making real-time detection challenging. Carsten Schmidt et al. [5] developed an AFP (Automated Fiber Placement) defect detection system based on infrared thermography. A heating lamp was used to heat the surface of the pre-impregnated material in front of the pressure roller, while a thermal imager was placed behind the pressure roller to obtain a thermal image of the surface. A simple CNN (Convolutional Neural Network) network was then utilized to analyze the thermal image for defect classification. However, due to the non-uniform distribution of the heat, the temperature of the pre-impregnated material near the heat source was higher than that farther away, resulting in different temperature field variations. Moreover, the temperature data were highly influenced by temperature, resulting in a low detection accuracy. Sebastian Zambala et al. [6] utilized a laser triangulation sensor to scan the surface of the laminate to obtain a height feature image. When there were defects on the surface, the height feature image would differ, and then the U-Nets algorithm [7] was used to segment the height feature image to achieve defect detection. This system could only detect flat composite components, and could detect changes on the mold surface due to height variations.

In recent years, deep learning has developed rapidly, and its excellent performance in object detection has provided a new approach for research on the online detection of surface defects in automated tape laying and winding. Common object detection algorithms in deep learning include Faster R-CNN (Faster Region-based Convolutional Neural Network), YOLO, SSD (Single Shot MultiBox Detector), and Transformer, among others [8–10]. YOLO, proposed by Joseph Redmon in 2016, has attracted attention, and YOLOv3, proposed by him in 2018, is favored by the industry due to its excellent detection speed and accuracy. In April 2020, Alexey Bochkovskiy combined the recent outstanding ideas in the field of computer vision with YOLOv3 to form YOLOv4 [11,12], which had a higher speed and accuracy than the other outstanding detection algorithms at that time. In June 2020, Glenn

Jocher released YOLOv5, which further improved the detection speed and accuracy, and significantly improved the inference speed. However, compared to other industrial fields that have successfully applied deep learning [11,13], there are currently two reasons why deep learning has not been applied to detect the surface defects in composite materials during automated laying and winding. First, there are a lack of high-quality datasets for surface defects in composite materials during automated laying and winding. Second, a large number of small defects are generated during the automated laying and winding process, and the current object detection algorithms have a relatively low detection accuracy for small targets.

To address these issues, this paper proposes an improved object detection algorithm based on YOLOv5 to achieve the end-to-end detection of the defects generated during the automated laying and winding process of composite materials, directly processing images or videos of the laying and winding process. The main contributions of this paper are as follows:

- (1) The CA attention mechanism has been improved in the embedded Separate CA structure, which not only achieves long-range dependence in the spatial direction, but also enhances the positional information and improves the ability to extract features. Additionally, the use of interval embedding further enhances the detection speed.
- (2) A new SIOU_loss regression box loss function has been proposed to replace the original CIOU_loss loss function, introducing considerations for the matching direction and using the angle loss as a penalty term. This further accelerates the regression speed of the bounding box and improves the detection accuracy, especially for small objects.
- (3) Based on the proposed SIOU_loss regression box loss function in this paper and combined with the Soft-NMS regression box filtering method, a new non-maximum suppression method called Soft-SIOU-NMS has been proposed for the post-processing of the model. By using a more gentle pre-selection box removal method, redundant boxes are removed while retaining more effective boxes, which improves the detection accuracy for overlapping coverage defects.

2. Detection Principle of YOLOv5

The YOLO models (YOLOv1–YOLOv5) are one-stage detection algorithms that can achieve end-to-end detection. They treat object detection as a regression problem and do not use the sliding window method. Instead, the images are directly input for training, allowing for a clear distinction between the background and the objects being detected, greatly improving the detection speed. Among them, YOLOv5 is currently the fastest and most accurate model, with great application prospects. YOLOv5 is an improvement on YOLOv4, with optimization in four parts: input end, backbone network, neck layer, and output end, further improving the detection accuracy and speed [13,14].

The input stage refers to the component of the algorithm responsible for receiving images as inputs. In the case of YOLOv5, it accepts images as inputs to facilitate object detection. On the input side, YOLOv5 uses the same Mosaic data augmentation as YOLOv4 [15,16], but adds adaptive anchor box calculation and adaptive image scaling. YOLOv5 incorporates anchor box generation into the code so that it can calculate the most suitable aspect ratio anchor boxes for the current dataset. In terms of the images of surface defects in the composite material winding components, the size of these images cannot be guaranteed to be consistent during the shooting process, but it is necessary to ensure consistency in the inputs to the network training. The most common practice is to resize the images to a uniform size before training the network. Upon scaling, due to the varying aspect ratios, gray borders are employed for padding. However, conventional resizing with fixed input dimensions may result in excessive gray borders, leading to cluttered information and a compromised speed. Hence, an adaptive image scaling approach is adopted to adjust the size of these gray borders. The main steps involve calculating the scaling ratio as $r = \min(640/w, 640/h)$, resizing the dimensions accordingly based on this scaling ratio, and padding the other dimension to be divisible by the stride (e.g., 32). As

depicted in Figure 1, this operation significantly reduces the presence of gray borders. Simply scaling the image and forcefully padding it may lead to a loss of image information and an ineffective utilization of the receptive field information.

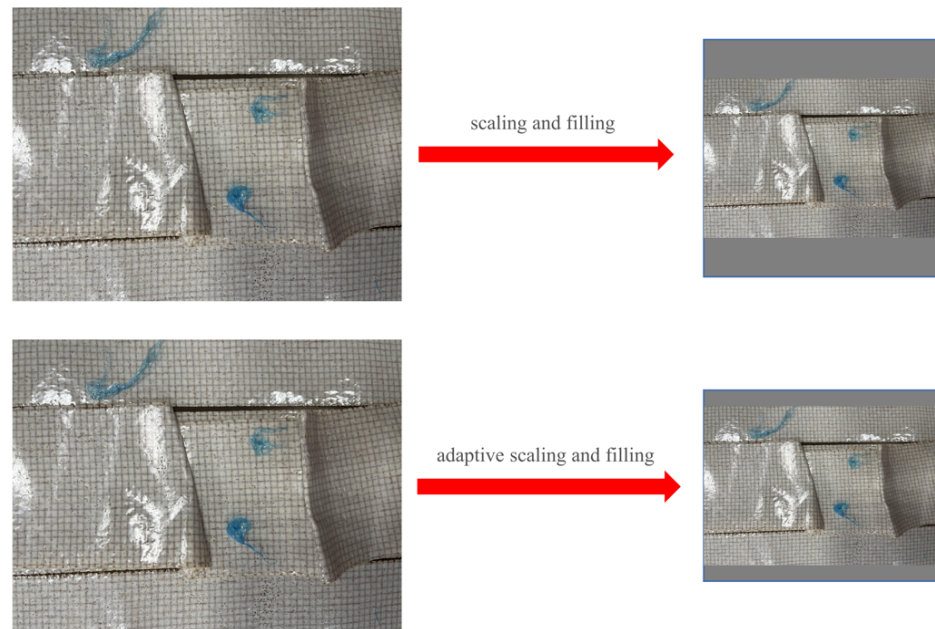


Figure 1. Adaptive scaling with padding.

The backbone network is a combination of convolutional and pooling layers designed to extract features from the input images. In YOLOv5, a backbone network is utilized to extract both low-level and high-level features from the input images, enabling a better understanding of the image content. In the backbone network, a convolutional layer with a 6×6 kernel is used instead of the Focus structure. On the one hand, this replaces the original three convolutional layers and improves the computational speed, while maintaining the same receptive field. On the other hand, replacing the Focus structure reduces the number of Concat operations and lowers the memory overhead.

The neck layer is the component located between the backbone network and the output layer. It plays a crucial role in connecting the backbone network and the output layer, while performing feature fusion and dimensionality reduction. In YOLOv5, the neck layer aids in transforming the features extracted from the backbone network into a format suitable for object detection. The neck also consists of the FPN (Feature Pyramid Network) + PAN (Path Aggregation Network) [17] structure, but with the addition of the CSP2 (Cross Stage Parital Network2) structure to enhance the network's feature fusion ability. In this section, the SPP (Spatial Pyramid Pooling) structure is replaced with the SPPF (Spatial Pyramid Pooling—Fast) structure, which involves passing the input through multiple 5×5 max pooling layers in series, as shown in Figure 2. Serially applying two 5×5 max pooling layers is equivalent to applying one 9×9 max pooling layer, and applying three in series is equivalent to one 13×13 max pooling layer. This approach not only achieves the same effect as the SPP, but also further improves the computational speed.

The output layer is the final layer of the network responsible for generating the results of the object detection. In YOLOv5, the output layer generates predicted bounding boxes and corresponding class confidences, which determine the presence of objects in the image and their respective locations. In the output layer of YOLOv5, the CIoU_Loss function is utilized as the bounding box loss function, while the NMS (non-maximum suppression) method is employed for handling overlapping objects.

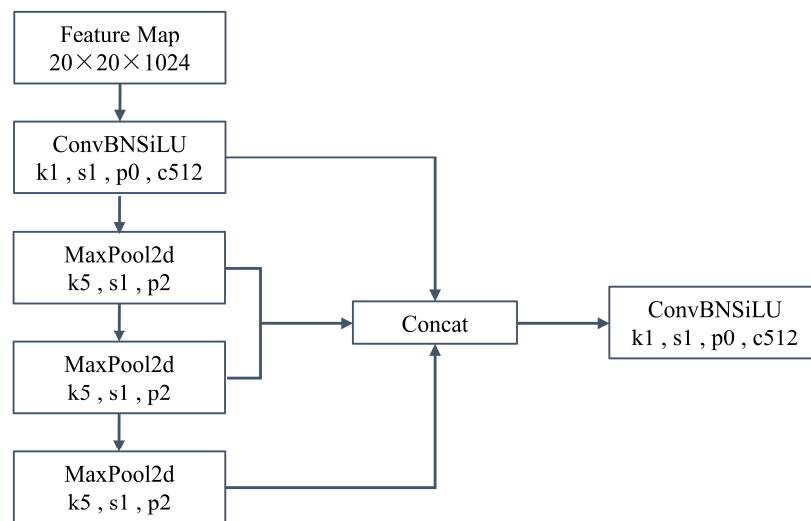


Figure 2. Sequential Passing of SPPF Structure Through Multiple 5×5 MaxPool Layers.

3. The Improvement of the Model Based on YOLOv5

3.1. Separate CA-YOLOv5 Network Model

The attention mechanism [18] is a biomimetic model that imitates the human ability to focus only on important parts. In this section, the attention mechanism is introduced to mimic the surface defects that occur during the composite material laying process observed by human eyes. This enables the detection network to quickly focus on the defect target and find the region of interest in the image, thereby improving the detection accuracy. Additionally, the color and shape features of the inclusion defects and gap defects in composite materials are more distinct from the background, and the introduction of the attention mechanism can more effectively extract these features.

As shown in Figure 3, the CA [19] attention mechanism not only achieves long-term dependency in the spatial direction, but also takes into account the position information, enhancing the expression of the feature position information. Compared to CBAM (Convolutional Block Attention Module) [20], it combines spatial attention [21] and Channel attention [22].

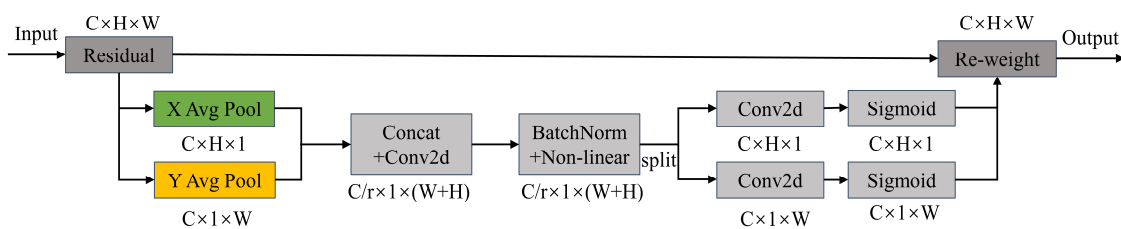


Figure 3. CA attention mechanism.

The CA module decomposes global pooling with the following formula:

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^w x_c(i, j) \tag{1}$$

When the height is h and the channel number is c , this can be represented by the following formula:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \tag{2}$$

When the width is w and the channel number is c , this can be represented by the following formula:

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (3)$$

After obtaining the feature maps separately on the width and height, they are concatenated together and then reduced in dimensionality through a 1×1 convolution. The resulting feature map is normalized and passed through the sigmoid activation function to obtain a feature map of $1 \times (W+H) \times (C/r)$.

$$f = \delta \left(F_1 \left(z^h, z^w \right) \right) \quad (4)$$

In the above equation, δ represents the non-linear activation function and $[*,*]$ denotes the spatial concatenation operation. After completing the aforementioned operations, the feature map undergoes a 1×1 convolution along both the height and width to obtain feature maps F_h and F_w , both with the same number of channels as the original. These two feature maps undergo a transformation using a 1×1 convolution, resulting in the attention weights g^h and g^w , defined as follows:

$$g^h = \sigma \left(F_h \left(f^h \right) \right) \quad (5)$$

$$g^w = \sigma \left(F_w \left(f^w \right) \right) \quad (6)$$

In the above equation, σ represents the sigmoid activation function. After obtaining the attention weights, a multiplication operation is performed on the feature map to obtain the weighting. The two feature maps with specific spatial information obtained through the above method complement each other. Ultimately, an attention-weighted feature map is obtained, defined as follows:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (7)$$

The CA module is a plug-and-play module that can be embedded behind any feature extraction module in the YOLOv5 structure. Commonly, the CA module is integrated into the backbone network, as shown in Figure 4.

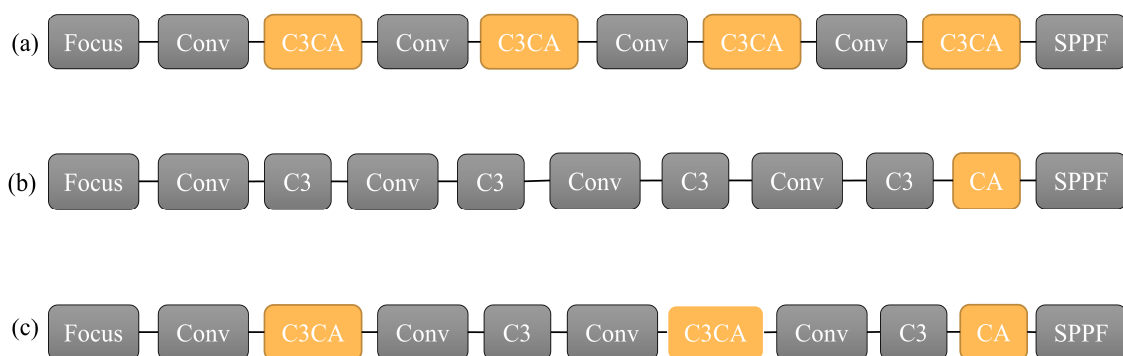


Figure 4. Comparison between common embedding methods of CA modules and the embedding method in this article. (a) C3CA structure (b) CA structure (c) Separate CA structure.

Figure 4a shows the C3CA structure, which embeds the CA attention module into the C3 module and applies the attention mechanism to each generated feature map. Figure 4b shows the CA structure, which embeds the CA attention module after the last C3 module and applies the attention mechanism to the deeper feature maps. Figure 4c proposes a new embedded Separate CA structure, which is an interval insertion structure. Firstly, the image undergoes the attention mechanism during the initial feature extraction stage, and

the feature focus can enhance the positional information of the feature map. Then, the feature map with a small receptive field that contains positional information is subjected to another feature extraction, and after another round of the CA attention mechanism, a feature map with a larger receptive field and richer positional information is obtained. Lastly, a CA attention mechanism is added to the end to obtain a feature map with a larger receptive field and directional perception.

Compared to the CA structure, the Separate CA structure proposed in this paper adds the attention mechanism in the middle of the feature extraction process, greatly enhancing the feature extraction of the feature map, integrating positional information into the feature map, and obtaining more semantic information. Compared to the C3CA structure, the Separate CA structure reduces one attention module, not only reducing the complexity of the model, facilitating a speed improvement in the model, but also enriching the feature information by adding a separate CA module to the end of the feature map, which better improves the detection accuracy.

3.2. Regression Box Loss Function Improvement

In YOLOv5, the CIoU loss function [23] was originally used as the bounding box localization loss, which optimized the overlap area, distance between the center points, and aspect ratio between the predicted and ground truth boxes. The expression is as follows:

$$CIoU = IoU - \left(\frac{P^2(B,A)}{C^2} + \alpha v \right) \quad (8)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^A}{h^A} - \arctan \frac{w^B}{h^B} \right)^2 \quad (9)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (10)$$

When $w^B = kw^A$ and $h^B = kh^A$, and v takes a value of 0, the value of αv is 0, which renders the optimization based on the aspect ratio ineffective. This ultimately results in the predicted box being unable to fit the ground truth box. By taking the partial derivative of Formula (9), we obtain the following equation:

$$\frac{\partial v}{\partial w^B} = \frac{8}{\pi} \left(\arctan \frac{w^{7gt}}{h^{8gt}} - \arctan \frac{w^B}{h^B} \right) \times \frac{h^B}{(w^B)^2 + (h^B)^2} \quad (11)$$

$$\frac{\partial v}{\partial h^B} = -\frac{8}{\pi} \left(\arctan \frac{w^{10gt}}{h^{8gt}} - \arctan \frac{w^B}{h^B} \right) \times \frac{w^B}{(w^B)^2 + (h^B)^2} \quad (12)$$

As shown in the equation above, h^B and w^B cannot both decrease at the same time during the regression. For composite materials with laying and winding surface defects, the preset anchor box size often exceeds the size of the defect, leading to situations where two boxes may overlap during training. Additionally, except for CIoU (Complete-Intersection over Union), other loss functions such as GIoU (Generalized-Intersection over Union) [24] and DIoU (Distance-Intersection over Union) [25], etc., have not considered the directionality between the predicted and ground truth boxes. Specifically, when the center points of the predicted and ground truth boxes are close, these loss functions will all degrade to IoU (Intersection over Union) [26], resulting in a slow convergence. To address this issue, the SIoU (SCYLLA-Intersection over Union) loss function is introduced to improve the localization regression loss by incorporating the vector angle between the predicted and ground truth boxes, redefining the relevant loss functions, which mainly include four cost functions:

(1) Angle cost

As shown in Figure 5, the red rectangle A represents the ground truth box, the green rectangle B represents the predicted box, and the black dashed rectangle represents the

rectangle formed by the centers of the ground truth and predicted boxes. d is the distance between the two centers, c_w is the difference in width between the two centers, c_h is the difference in height between the two centers, α is the angle between d and c_w , and β is the angle between d and c_h .

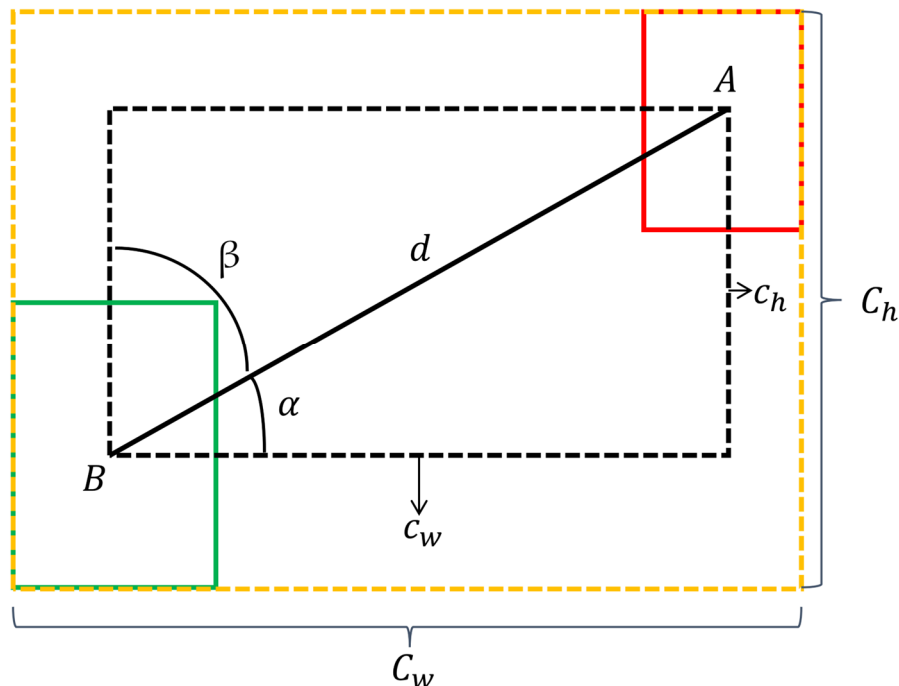


Figure 5. The Scheme for calculation of angle cost contribution into SIoU_loss.

When $\alpha \leq \pi/4$, the convergence process will first minimize α . When $\pi/4 < \alpha \leq \pi/2$, the convergence process will minimize β . To achieve this, the following calculation formula is introduced:

$$\Lambda = 1 - 2 \times \sin^2(\arcsin(\frac{c_h}{d}) - \frac{\pi}{4}) = \cos(2 \times (\arcsin(\frac{c_h}{d}) - \frac{\pi}{4})) \tag{13}$$

$$\frac{c_h}{d} = \sin(\alpha) \tag{14}$$

$$d = \sqrt{(A_x - B_x)^2 + (A_y - B_y)^2} \tag{15}$$

$$c_h = \max(A_x, B_x) - \min(A_y, B_y) \tag{16}$$

As shown in the above formula, (A_x, A_y) represents the center coordinate of the ground truth bounding box, while (B_x, B_y) represents the center coordinate of the predicted bounding box. When $\alpha = 0$ or $\alpha = \pi/2$, the predicted and ground truth bounding boxes are in a horizontal or vertical position, and the angle loss will be 0 in this case.

(2) Distance cost

Based on the angle loss defined above, the distance loss is redefined as follows:

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) = 2 - e^{-\gamma \rho_x} - e^{-\gamma \rho_y} \tag{17}$$

$$\rho_x = \left(\frac{A_x - B_x}{C_w} \right)^2 \tag{18}$$

$$\rho_y = \left(\frac{A_y - B_y}{C_h} \right)^2 \quad (19)$$

$$\gamma = 2 - \Lambda \quad (20)$$

As shown in Figure 5, the yellow dashed rectangle represents the minimum bounding box enclosing the ground truth and predicted boxes, where C_w and C_h , respectively, denote the width and height of the minimum bounding box. It can be observed from the formula that, as α approaches 0, the contribution of the distance loss decreases. When $\alpha = 0$ and $\gamma = 2$, it degenerates into the conventional distance loss. On the other hand, as α approaches $\pi/4$, the contribution of the distance loss increases.

(3) Shape cost

The definition of the shape loss is given by the following formula:

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^\theta = (1 - e^{-w_w})^\theta + (1 - e^{-w_h})^\theta \quad (21)$$

$$w_w = \frac{|B_w - A_w|}{\max(B_w, A_w)} \quad (22)$$

$$w_h = \frac{|B_h - A_h|}{\max(B_h, A_h)} \quad (23)$$

where (A_w, A_h) and (B_w, B_h) denote the width and height of the ground truth and predicted boxes, respectively. The importance of the θ value cannot be overstated, as it regulates the extent to which attention is paid to the shape loss. When $\theta = 1$, optimizing the shape loss would lead to the immediate optimization of the shape, thereby impeding its freedom of movement. To prevent an excessive focus on the shape loss and a decrease in the movement of the predicted boxes, a genetic algorithm is employed to calculate the value of θ for each dataset, which is determined to tend toward four. Consequently, the value of θ is restricted to the range [2,6].

(4) IoU cost

The IoU loss refers to the ratio of the intersection over union between the predicted bounding box and the ground truth bounding box.

In summary, the final definition of the SIoU loss function (L_{SIoU}) is given as follows:

$$L_{SIoU} = 1 - IoU + \frac{\Lambda + \Omega}{2} \quad (24)$$

By incorporating an angle penalty term into the SIoU loss function, the bounding box regression is further constrained, leading to an improved detection accuracy.

3.3. Post-Processing Method Improvements

3.3.1. Post-Processing Method for YOLOv5

The bounding box regression filtering method used in YOLOv5 is NMS. After an image is input into the network, it is divided into $S \times S$ grids, and multiple bounding boxes with confidence levels are generated on each grid based on different object categories. The box with the highest score is selected from these boxes and the remaining boxes are discarded [27]. The specific process of this is as follows: (1) The regression boxes are divided according to the category and the background regression boxes are removed. (2) The regression boxes are sorted in descending order according to the classification confidence. (3) One type of target object is selected first. The regression box that ranks first in step 2 is removed from the input list and added to the output list. (4) The IoU value between all the regression boxes in the input list and the regression box with the highest confidence level is calculated. If the IoU value is greater than the threshold set, then the regression box is removed from the input list. (5) Steps 3–4 are repeated to distinguish

all the regression boxes of this type of object. (6) Steps 2–5 are repeated to identify all the object categories and finally obtain the output list.

The expression for selecting the regression box in NMS is as follows:

$$S_i = \begin{cases} S_i, IoU(A, B_i) < \varepsilon \\ 0, IoU(A, B_i) < \varepsilon \end{cases} \quad (25)$$

In Formula (25), A represents the regression box with the highest confidence, B represents the remaining regression boxes, and ε is the set threshold. As can be seen from the above filtering process, the method of discarding all the boxes with an IoU greater than the set threshold may lead to inaccurate predictions. Firstly, this will cause two different types of targets to be removed due to overlapping regression boxes; for example, when a fold defect and a foreign object defect overlap, the foreign object defect may be covered by the fold defect and then be removed. Secondly, this operation can also filter out one of two defects of the same type due to occlusion, leading to missed detection, which greatly affects the detection performance. At the same time, NMS increases the complexity of the network calculation, which requires higher hardware requirements.

3.3.2. Improving Non-Maximum Suppression Algorithm

Due to the suboptimal performance of NMS in regression box selection, this section proposes improvements based on two aspects.

The first improvement addresses the ‘all-or-nothing’ approach of NMS, which removes all the boxes with an IoU greater than the set threshold and is too aggressive. Therefore, the Soft-NMS approach is introduced, which uses a decay function $f(x)$ with values between 0 and 1. Soft-NMS takes the highest-confidence regression box A and the predicted box B_i 's IoU value as inputs to the decay function. The confidence score of B_i is then calculated by multiplying the decay function and the confidence score S_i , according to the following formula:

$$S_i = \begin{cases} S_i f(IoU(A, B_i)), IoU(A, B_i) \geq \varepsilon \\ S_i, IoU(A, B_i) < \varepsilon \end{cases} \quad (26)$$

The function $f(x)$ uses a Gaussian decay function that has a better decay effect, as shown in Formula (27) below:

$$f(IoU(A, B_i)) = e^{-\frac{IoU(A, B_i)^2}{\sigma}} \quad (27)$$

As described in the above formula, where σ is usually set to 0.5 and $f(IoU(A, B_i))$ is guaranteed to be continuously valued in the range from 0 to 1, the Gaussian penalty function ensures that a higher IoU value during the selection process results in a greater penalty and lower score. For boxes with an originally low confidence, their existing confidence may fall below the confidence threshold after being penalized. For boxes with a high initial confidence, their confidence remains high after the penalty and they can be retained. In this way, the removal of overlapping object detection boxes can be reduced.

To address the limitation of using the IoU as the evaluation metric, which only considers the overlapping region between the bounding boxes and may not fully account for positional factors, resulting in missed detections, we propose using the SIoU as a substitute for the IoU to further improve the detection accuracy by incorporating angle factors into the optimization. The equation for calculating the confidence score is as follows:

$$S_i = \begin{cases} S_i f(SIoU(A, B_i)), SIoU(A, B_i) \geq \varepsilon \\ S_i, SIoU(A, B_i) < \varepsilon \end{cases} \quad (28)$$

$$f(SIoU(A, B_i)) = e^{-\frac{SIoU(A, B_i)^2}{\sigma}} \quad (29)$$

4. Experimental Setup and Results Analysis

This section presents an exposition on the dataset, experimental framework, parameterization, and evaluation criteria adopted for conducting the experiments on the surface defect detection in composite materials. Subsequently, an analysis of the experimental results is discussed. To demonstrate and substantiate the effectiveness of the proposed method, a comparative study among five training models was conducted, with the models comprising the YOLOv5, YOLOv5 + SCA (Separate CA) with the integration of the CA attention mechanism, YOLOv5 + SIoU with an optimized regression box loss function, YOLOv5 + SSN (Soft-SIoU-NMS) with an improved post-processing approach, and a new model that consolidates all three improved methods proposed in this manuscript.

4.1. Dataset

Currently, there is no systematically compiled dataset of surface defects in composite material automated tape laying and winding. This dataset combines defects captured by industrial cameras and lenses during actual automated tape laying and winding processes, as well as defects obtained through artificial, simulated automated tape laying and winding, in order to obtain a larger and more comprehensive defect dataset. The pre-impregnated material used in this dataset is a new type of lightweight phenolic resin ablation material, produced by Hubei Sanjiang Aerospace Hongyang Electromechanical Co., Ltd. In this study, a total of 3200 images from the defect dataset were selected. The defect types were categorized as inclusions, clearances, and folds, as shown in Figure 6. The image annotation software LabelImg (<https://github.com/HumanSignal/labelImg>) was used to label each image in the dataset and generate XML files. These files contained information about the category, dimensions, and position of the target defects.

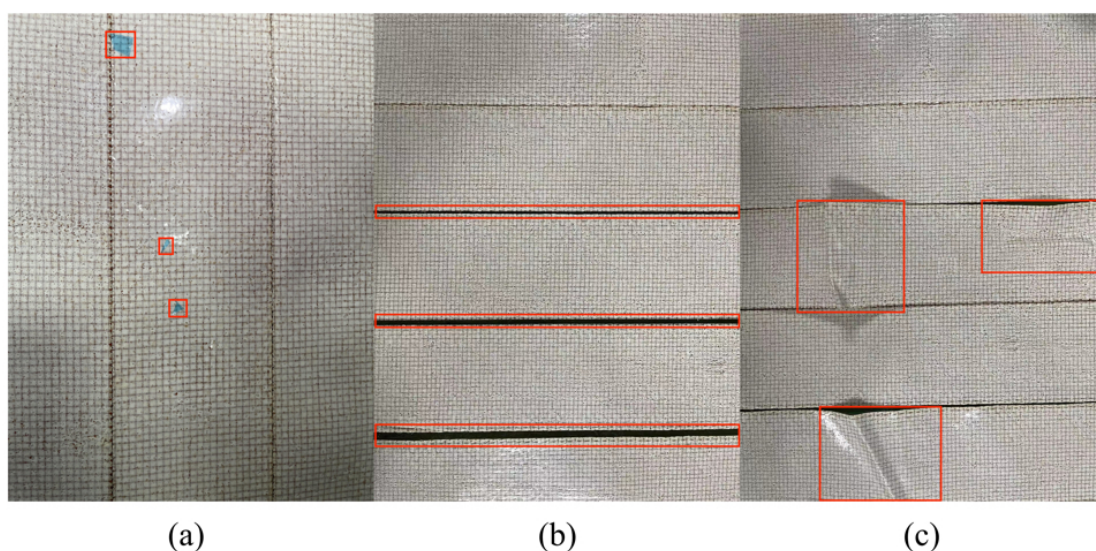


Figure 6. Partial Dataset (a) Inclusion (b) Clearance, and (c) Fold.

4.2. Experimental Environment and Parameter Configuration

The experimental environment and resource configurations are essential prerequisites for network model training. For this experiment, both the software and hardware configurations are shown in Table 1.

The experiment was conducted on a Windows 10 × 64 operating system, using Pycharm as the IDE, and the overall software platform consisted of Windows 10, Pytorch 1.10.1, Python 3.9, CUDA 10.2.89, cuDNN 8.3.3, and Pycharm 2020.1. The hardware configuration included an Intel(R) Core(TM) i7-10875H 2.30 GHz processor, 32 GB of memory, and an NVIDIA GeForce GTX 1650 Ti.

Table 1. Experimental Software and Hardware Configuration.

Name	Configuration/Version
Operating System	Windows 10 × 64
CPU	Intel(R) Core(TM) i7-10875H
GPU	NVIDIA GeForce GTX 1650 Ti
Memory	32 GB
Graphics memory	4 GB
IDE	Pycharm 2020.1
Deep Learning Framework	Pytorch 1.10.1
CUDA	CUDA 10.2.89
cuDnn	cuDNN 8.3.3
PythonVersion	Python 3.9

The defective image was input with a size of 640×640 (due to YOLOv5 using adaptive scaling and padding technology at the input end, the actual preprocessed size may not have been exactly 640×640 on both sides. Only one side was 640, and the padding size of the other side needed to be calculated). Considering the size of the dataset, the number of general iterations, and the training results of Section 3 for the model, the number of iterations in this section was set to 300 times. Taking into account the experimental software and hardware configuration, including the video memory and other reasons, the batch size was set to 3. At the same time, the initial learning rate and weight decay coefficient were set to 0.01 and 0.0005, respectively. The specific parameter settings are shown in Table 2.

Table 2. Model Training Parameter Settings.

Parameter	Setting
Initial Learning Rate	0.01
Epoch	300
Batch size	3
Momentum Size	0.937
Weight Decay Coefficient	0.0005
Input Image Size	640×640
Nc	3
Optimizer	SGD

4.3. Evaluation Indicators

In this experiment, the evaluation criteria included the recall, AP, FPS, mAP, and other indicators. Recall is used to describe the proportion of correctly detected defects to the total number of defects that should be correctly detected. mAP and AP are both indicators of the detection performance. mAP reflects the overall performance of the entire object detection network for all the categories in the dataset, while AP reflects the comprehensive performance of the individual category detection in the dataset. FPS represents the number of images that the detector can process per second.

4.4. Experimental Results and Analysis

4.4.1. Training Results and Analysis

As shown in Figure 7, the top-down sequence represents the localization loss, object confidence loss, and classification loss. The experiment was conducted for 300 iterations, and after the training, the three types of losses gradually decreased, all reaching the minimum value at the last epoch.

Based on Figure 7, it can be observed that the model exhibited a faster fitting speed in the first 50 iterations and tended to stabilize around 300 iterations, indicating the convergence of the model. In terms of the localization loss, the order of the loss magnitude from smallest to largest was as follows: Improved YOLOv5 < YOLOv5 + SSN < YOLOv5 + SCA < YOLOv5 + SIOU < YOLOv5. The proposed model achieved the lowest loss, indicating the

smallest error between the predicted and annotated bounding boxes. YOLOv5 + SSN also exhibited a relatively low loss due to the better generalization regression method of SSN compared to the initial NMS. As for the object confidence loss, compared to the base model, YOLOv5, the proposed model achieved the lowest loss of 0.01364, indicating the smallest cross-entropy loss between the probability of the predicted boxes containing objects and the probability of the annotated boxes containing objects, which represented the highest probability of objects being contained within the bounding boxes. In terms of the category loss, the proposed model had the lowest loss, followed by YOLOv5 + SIoU. Similarly, the improved models had lower losses compared to YOLOv5, indicating a higher probability of containing target categories in the proposed model and improved model performance.

From Figure 8, it can be seen that the model had a faster fitting speed in the first 50 iterations, and tended to be stable when it reached 300 iterations. The model basically converged at this point. For the localization loss, the loss from small to large was: Improved YOLOv5 < YOLOv5 + SSN < YOLOv5 + SCA < YOLOv5 + SIoU < YOLOv5. Among them, the loss of this paper's model was the lowest, indicating that the error between the predicted box and the calibrated box was the smallest. The loss value of YOLOv5 + SSN was also low, because SSN had a more general regression method than the initial NMS. For the target confidence loss, the loss value of this paper's model reached the lowest, at 0.01364. The results of the various models compared with the basic model YOLOv5 showed that the improved network structure had a better accuracy than the original model. For the category loss, the loss value of this paper's model was the lowest, followed by YOLOv5 + SIoU. Similarly, the loss value of the improved model was lower than that of YOLOv5, indicating that the performance was improved.

4.4.2. Test Results and Analysis

To verify the detection effect of the model, five models were used to detect and recognize 120 composite material winding surface defect images in the test set. Figure 9 shows the AP curves of each category obtained after the detection. It can be observed that, compared to the other two defects, the detection accuracy of inclusions was higher, and the AP curve tilts to the right. Among the equivalent models, the detection of clearance defects was also higher than that of folded defects, because the color characteristics of the inclusion defects were more obvious and the number of inclusion defects in the dataset was relatively large. At the same time, the features of the clearance defects generated during the composite material winding process were obvious, usually spanning the beginning and end of an image.

Among them, the YOLOv5 + SSN model had a higher detection accuracy for clearance and folded defects than the original YOLOv5, indicating an improvement in the regression box selection method by Soft-SIoU-NMS for overlapping targets. The YOLOv5 + SIoU model improved the detection accuracy of defect targets by improving the loss function, and its detection AP for three types of defects exceeded that of the YOLOv5 model. However, there was a class imbalance in the detection accuracy, with a higher accuracy for inclusion defects and a lower accuracy for the other two defects, indicating that this improvement still needs to focus more on feature attention.

The YOLOv5 + SCA model had a high detection AP for all three types of defects, and the accuracy of the detection categories was relatively balanced. Its detection AP for inclusion defects was the highest, reaching 95.4%, and the difference in the detection AP between the other two defects was not significant, which were 94.7% and 94.6%, respectively. Introducing the attention mechanism enhanced the model's feature extraction and learning ability and improved its detection accuracy.

This paper's model is YOLOv5 + Separate-CA + SIoU + Soft-SIoU-NMS, which integrated the three improvements and improved the AP of all three types of defects, which were 97.0%, 98.3%, and 96.3%, respectively. Compared to YOLOv5, it increased the detection AP by 4.7%, 3.5%, and 5.7%, respectively, and the detection of each class of defect was more balanced.

Figure 10 shows the performance evaluation of the different models on the test set, including mAP, Recall, and FPS. Experiment 1 was the original YOLOv5 algorithm, with an mAP of 92.6%, a recall rate of 87.5%, and an FPS of 66.7, with an overall mediocre performance. Experiment 2 embedded the SCA module in the main network of YOLOv5. Since the module realizes long-range dependency in the spatial direction and enhances position information, it improved the feature extraction ability. Compared to the YOLOv5 model, its mAP value increased by 2.3% and its recall rate increased by 6.3%. Meanwhile, because the module was inserted using interval embedding, the detection speed also increased, with an FPS of 71.9. Experiment 3 modified the regression box loss function to SIoU. Although the network structure remained unchanged, the introduction of the angle loss improved the accuracy of the back-propagation. The mAP value increased by 1.1% compared to YOLOv5 to a final value of 93.7%, with the recall rate being increased to 92.7%. It had little effect on the FPS. Experiment 4 replaced the original NMS box selection method with Soft-SIoU-NMS. The mAP increased by 0.3 percentage points, but the recall rate increased significantly from 87.5% to 94.2%. However, due to computational constraints, the FPS decreased from 66.7 to 65.3, equivalent to detecting 1.4 fewer images per second. Experiment 5 was the improvement set in this paper, which integrated the improvements in experiments 2, 3, and 4. The improved network had an mAP of 97.2%, a recall rate of 96.1%, and an FPS of 72.1. This shows that the improved YOLOv5 algorithm in this paper not only effectively improved the detection accuracy, but also improved the detection speed. It was more accurate for identifying the surface defects of complex or small composite materials, and had the best overall performance.

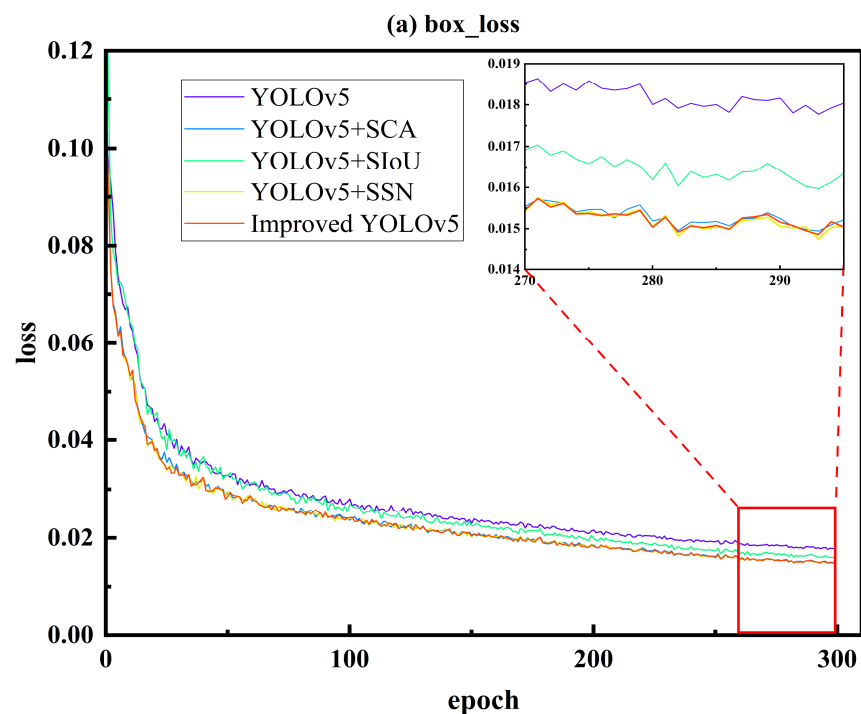


Figure 7. Cont.

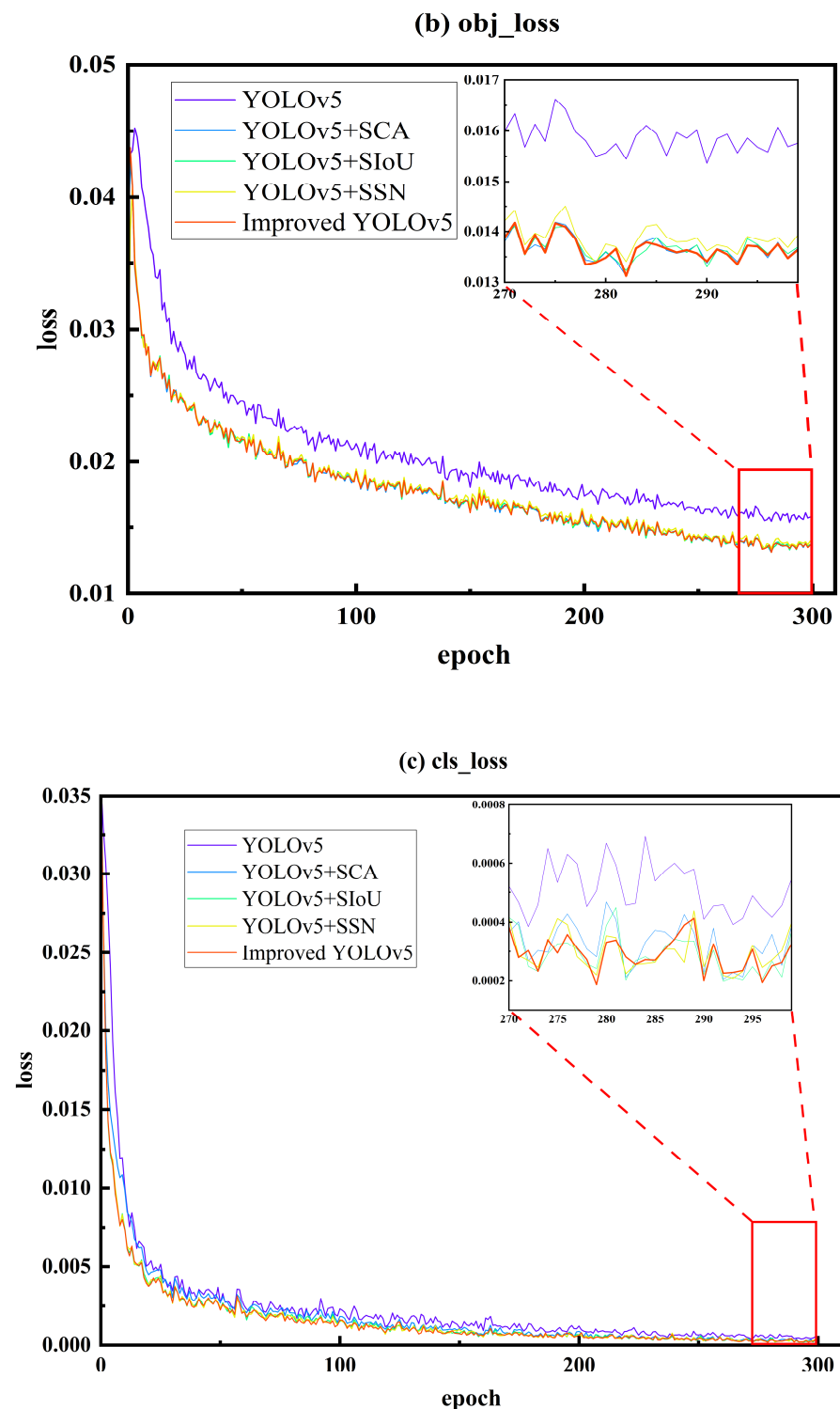


Figure 7. Training Loss Curve Graph ((a) box_loss, (b) obj_loss, and (c) cls_loss).

4.4.3. Application Results and Analysis

This detection model needed to be applied in the actual composite material laying and winding site. Therefore, the detection effect of the model was verified by detecting actual images and comparing YOLOv5 with the improved model in this paper. In the composite material winding site, the winding process was photographed using an industrial camera. The original YOLOv5 model and the model in this paper were used for video detection. Some detection images from the video were captured, and the recognition effect is shown

in Figure 11. Considering the possibility of low-confidence defects being filtered out, which would lead to incomplete and inaccurate comparisons, this article did not set a threshold for detection and comparison. In practical application, the determination of the threshold size needs to be initially attempted based on the results of the model testing for setting the confidence threshold, and then continuously fine-tuned during long-term actual detection processes, in order to achieve the best detection efficiency.

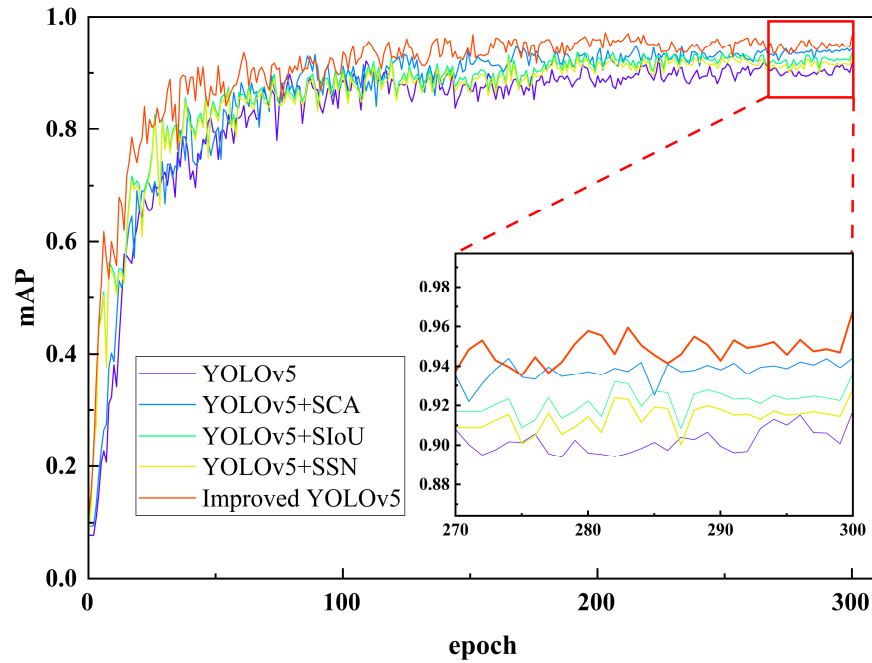


Figure 8. Comparison of Training mAP Graph.

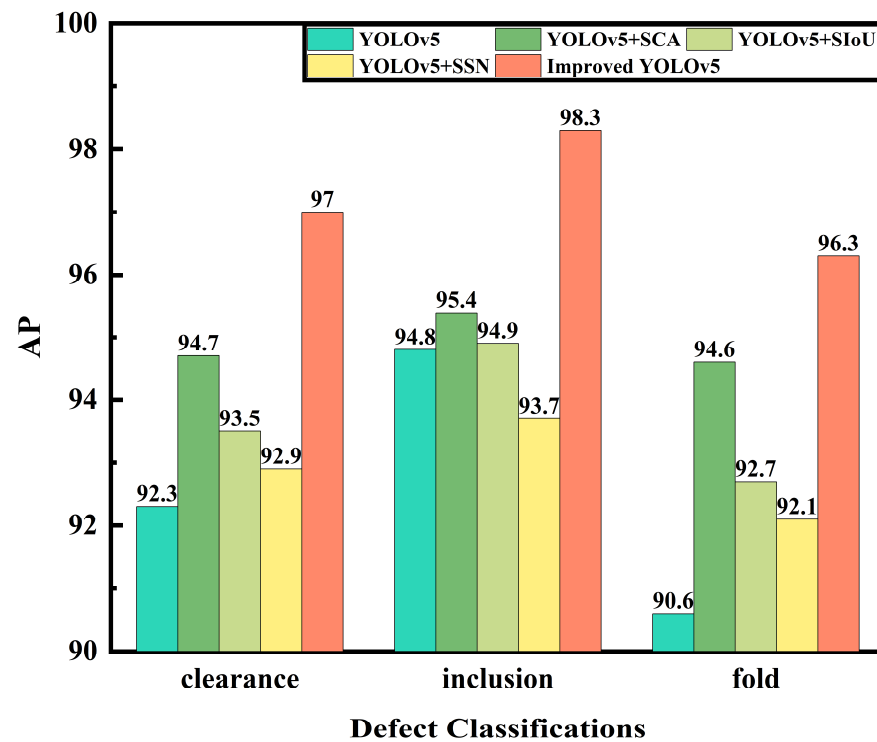


Figure 9. Comparison of Detection AP for Different Types of Surface Defects across the Given Models.

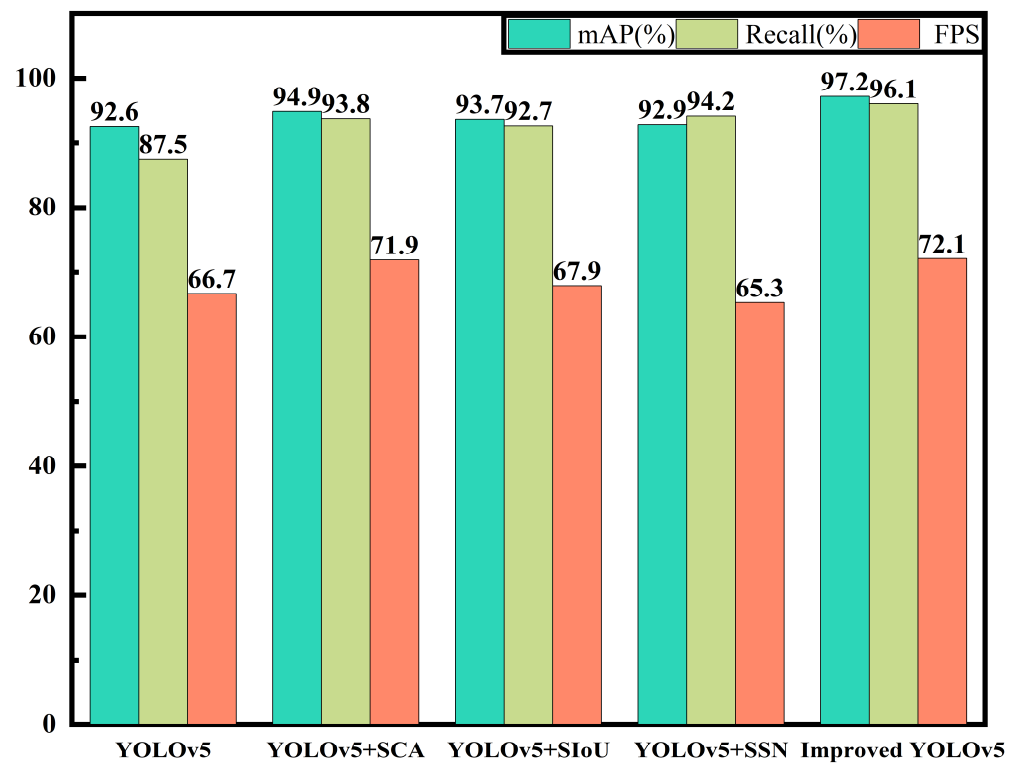


Figure 10. Performance Evaluation of Different Models on the Test Dataset.

From Figure 11, it can be seen that the left side A1~F1 shows the detection results using YOLOv5 before improvement, and A2~F2 shows the detection results using the improved model proposed in this paper. The detection capability of the improved YOLOv5 algorithm was significantly enhanced. Specifically, in A1 and A2 of Figure 11, the image brightness was normal, and both algorithms detected all defects. However, the improved YOLOv5 had a much higher confidence in detecting targets than the original YOLOv5. In the B images, which were captured when the light source was blocked during the winding process, the overall image was darker, but both detection models were not affected by the brightness and accurately detected all the defects. Similarly, the improved YOLOv5 had a greater confidence. In C1 and C2, a large blue inclusion defect partly covered the clearance defect. The original YOLOv5 only detected the inclusion defect, but not the clearance defect. The improved YOLOv5 not only detected all the defects, but also had a higher confidence. In the image D, there were many defects, and the folded defect and inclusion defect overlapped. The original YOLOv5 had instances of false detection and missing detection. It detected the clearance defect on the left side of the image as an inclusion defect and missed the clearance defect on the right side. In contrast, the improved YOLOv5 algorithm effectively detected all the defects. Comparing E1 and E2 shows that the improved YOLOv5 algorithm not only had a higher confidence in detecting targets, but also detected small defects that were undetected in E1. Moreover, for dense inclusion defects, the original YOLOv5 detected both defects as one, while the proposed algorithm distinguished them, resulting in more precise detection results. In F1 and F2, both detection algorithms detected all defects, and the confidence in detecting the folded defect was the same. However, by examining the figure, it can be seen that the bounding box of the improved YOLOv5 was more accurate.

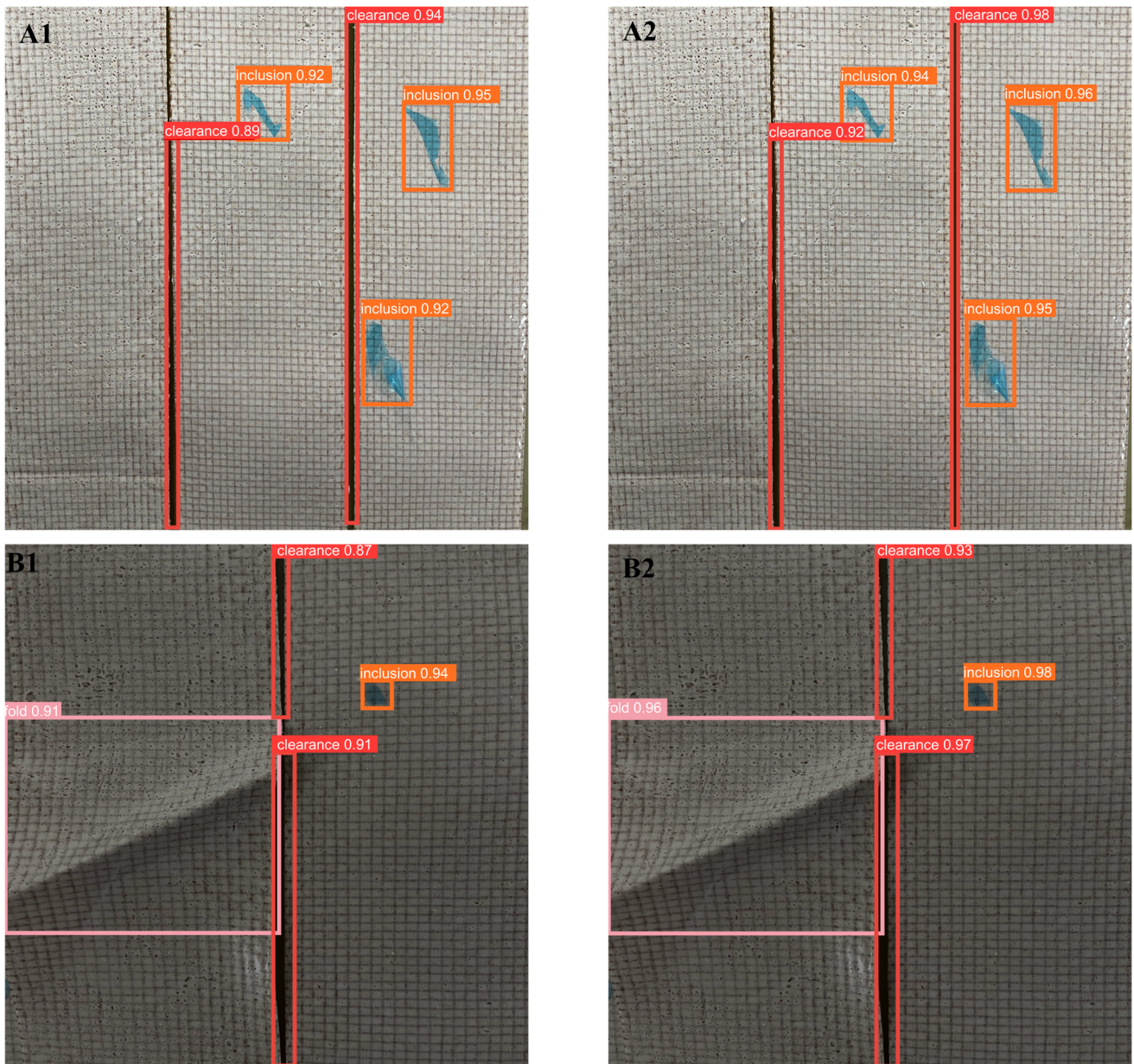


Figure 11. Cont.

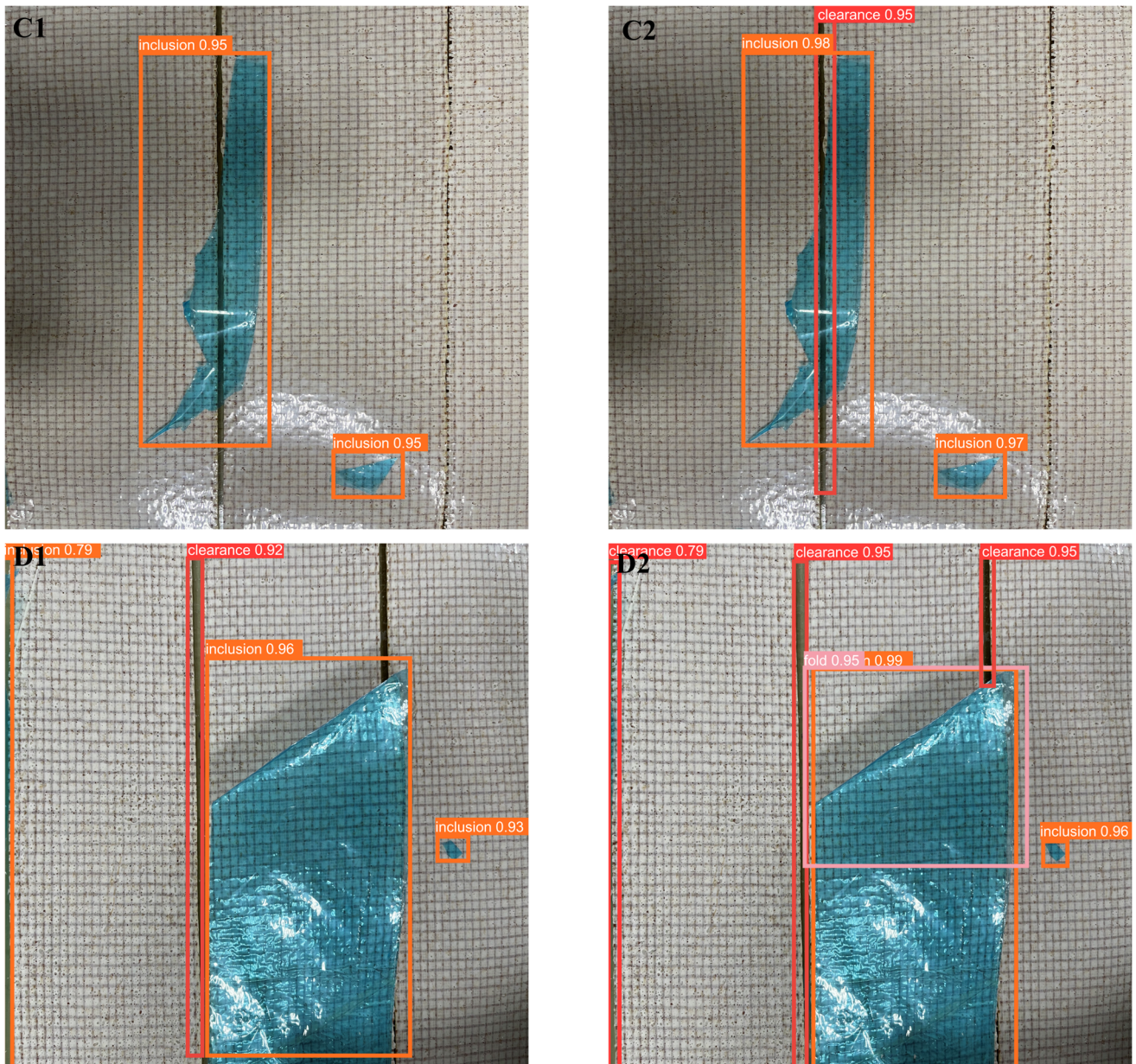


Figure 11. Cont.

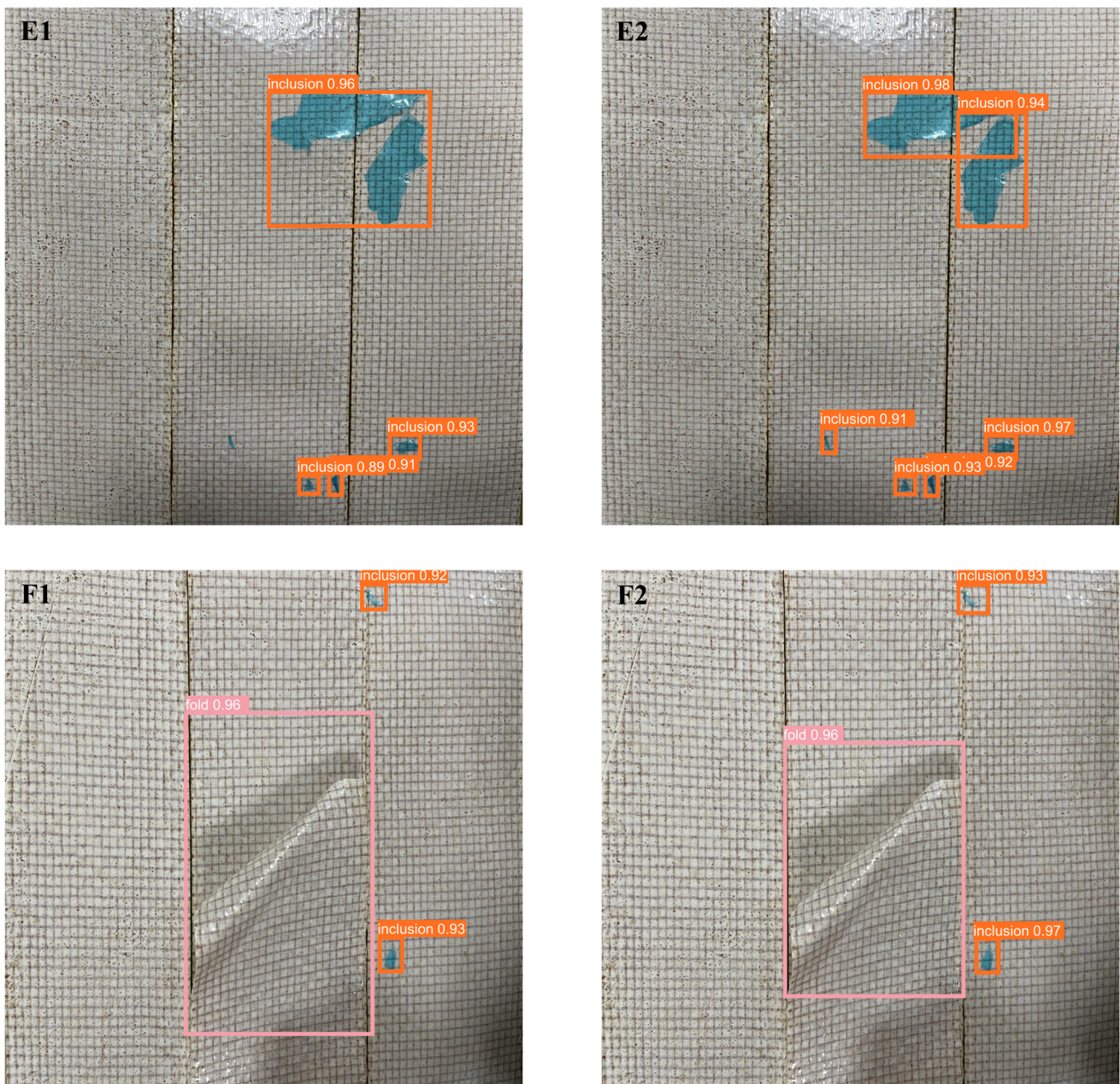


Figure 11. Detection performance of YOLOv5 on video streams before and after improvement (YOLOv5: (A1–F1), Improved YOLOv5: (A2–F2)).

5. Conclusions

An improved defect detection algorithm for the automated laying and winding process of composite materials was proposed in this paper, which combined machine vision and deep learning. The improved YOLOv5 algorithm, which integrated the Separate CA + SIoU + Soft-SIoU-NMS module, achieved accuracy improvements of 4.7%, 3.5%, and 5.7%, respectively, for the three types of defects: clearances, inclusions, and folds. The mAP was also improved by 4.6%, and the FPS was increased by 5.4% compared to the baseline algorithm. These improvements demonstrate the positive impact of the improved module on the model, and a further accuracy improvement was achieved under the condition of satisfying the detection speed. The improved YOLOv5 model had a greater accuracy and met the requirements of detection speed in industrial scenes, making it suitable for the defect detection of composite material automated laying and winding processes. However, the improved algorithm only worked for detecting clearances, inclusions, and folds, and

could not detect defects such as resin-rich regions. The next focus of the research is the detection of such defects.

Author Contributions: Conceptualization, L.W.; Methodology, L.W. and S.L.; Software, L.W. and S.L.; Validation, S.L.; Formal analysis, L.W.; Investigation, J.R.; Resources, J.R.; Data curation, J.R.; Writing—original draft, L.W.; Writing—review & editing, L.W. and S.L.; Visualization, S.L.; Supervision, S.L. and J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. He, K.; Nie, H.; Yan, C. The Intelligent Composite Panels Manufacturing Technology Based on Tape-laying Automatic System. *Procedia CIRP* **2016**, *56*, 610–613. [[CrossRef](#)]
2. Aabid, A.; Baig, M.; Hrairi, M. Advanced Composite Materials for Structural Maintenance, Repair, and Control. *Materials* **2023**, *16*, 743. [[CrossRef](#)] [[PubMed](#)]
3. Al-Furjan, M.; Shan, L.; Shen, X.; Zarei, M.; Hajmohammad, M.; Kolahchi, R. A review on fabrication techniques and tensile properties of glass, carbon, and Kevlar fiber reinforced polymer composites. *J. Mater. Res. Technol.* **2022**, *19*, 2930–2959. [[CrossRef](#)]
4. Sacco, C.; Radwan, A.B.; Beatty, T.; Harik, R. Machine Learning Based AFP Inspection: A Tool for Characterization and Integration. *SAMPE J.* **2020**, *56*, 34–41.
5. Schmidt, C.; Hocke, T.; Denkena, B. Deep learning-based classification of production defects in automated-fiber-placement processes. *Prod. Eng.* **2019**, *13*, 501–509. [[CrossRef](#)]
6. Zambal, S.; Heindl, C.; Eitzinger, C.; Scharinger, J. End-to-End Defect Detection in Automated Fiber Placement Based on Artificially Generated Data. In Proceedings of the Fourteenth International Conference on Quality Control by Artificial Vision, Mulhouse, France, 15–17 July 2019; p. 68.
7. Hsu, K.; Yuh, D.-Y.; Lin, S.-C.; Lyu, P.-S.; Pan, G.-X.; Zhuang, Y.-C.; Chang, C.-C.; Peng, H.-H.; Lee, T.-Y.; Juan, C.-H.; et al. Improving performance of deep learning models using 3.5D U-Net via majority voting for tooth segmentation on cone beam computed tomography. *Sci. Rep.* **2022**, *12*, 19809. [[CrossRef](#)] [[PubMed](#)]
8. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single Shot Multibox Detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14. Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
9. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *39*, 1137–1149. [[CrossRef](#)]
10. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
11. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.109342020.
12. Redmon, J.; Farhadi, A. Yolov3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767 2018.
13. Wang, H.; Yang, X.; Zhou, B.; Shi, Z.; Zhan, D.; Huang, R.; Lin, J.; Wu, Z.; Long, D. Strip Surface Defect Detection Algorithm Based on YOLOv5. *Materials* **2023**, *16*, 2811. [[CrossRef](#)] [[PubMed](#)]
14. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo Algorithm Developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [[CrossRef](#)]
15. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. Mixup: Beyond Empirical Risk Minimization. *arXiv* **2017**, arXiv:1710.094122017.
16. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization Strategy to Train Strong Classifiers with Localizable Features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6023–6032.
17. Jin, R.; Xu, Y.; Xue, W.; Li, B.; Yang, Y.; Chen, W. An Improved Mobilenetv3-Yolov5 Infrared Target Detection Algorithm Based on Attention Distillation. In *Proceedings of the Advanced Hybrid Information Processing, Pt I*; Liu, S., Ma, X., Eds.; European Alliance Innovation: Gent, Belgium, 2022; Volume 416, pp. 266–279. [[CrossRef](#)]
18. Wang, Y.; Du, H.; Zhang, X.; Xu, Y. Application of Visual Attention Network in Workpiece Surface Defect Detection. *J. Comput. Aided Des. Comput. Graph.* **2019**, *31*, 1528–1534.
19. Xuan, W.; Jian-She, G.; Bo-Jie, H.; Zong-Shan, W.; Hong-Wei, D.; Jie, W. A Lightweight Modified YOLOX Network Using Coordinate Attention Mechanism for PCB Surface Defect Detection. *IEEE Sens. J.* **2022**, *22*, 20910–20920. [[CrossRef](#)]
20. Li, Z.; Li, B.; Ni, H.; Ren, F.; Lv, S.; Kang, X. An Effective Surface Defect Classification Method Based on RepVGG with CBAM Attention Mechanism (RepVGG-CBAM) for Aluminum Profiles. *Metals* **2022**, *12*, 1809. [[CrossRef](#)]

21. Lin, D.; Cheng, Y.; Li, Y.; Prasad, S.; Guo, A. MLSA-UNet: End-to-End Multi-Level Spatial Attention Guided UNet for Industrial Defect Segmentation. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 441–445. [[CrossRef](#)]
22. Zhu, Y.; Ding, R.; Huang, W.; Wei, P.; Yang, G.; Wang, Y. HMFCA-Net: Hierarchical multi-frequency based Channel attention net for mobile phone surface defect detection. *Pattern Recognit. Lett.* **2022**, *153*, 118–125. [[CrossRef](#)]
23. Yuan, Y.; Tan, X. Defect Detection of Refrigerator Metal Surface in Complex Environment. *J. Comput. Appl.* **2021**, *41*, 270–274.
24. Yang, Y.; Liao, Y.; Cheng, L.; Zhang, K.; Wang, H.; Chen, S. Remote Sensing Image Aircraft Target Detection Based on GIoU-YOLO V3. In Proceedings of the 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 9–11 April 2021; pp. 474–478.
25. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *arXiv* **2019**, arXiv:1911.08287. [[CrossRef](#)]
26. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. UnitBox: An Advanced Object Detection Network. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 24–26 October 2016; pp. 516–520.
27. Xu, Y.; Jiang, M.; Li, Y.; Wu, Y.; Lu, G. Fruit Target Detection Based on Improved YOLO and NMS. *J. Electron. Meas. Instrum.* **2022**, *36*, 114–123.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.