



## OPEN ACCESS

## EDITED BY

Yi Fang,  
New York University Abu Dhabi, United Arab  
Emirates

## REVIEWED BY

Kelvii Wei Guo,  
City University of Hong Kong, Hong  
Kong SAR, China  
Yushuai Yu,  
China University of Mining and Technology,  
China

## \*CORRESPONDENCE

Limei Zhang  
✉ [zhanglimei@btbu.edu.cn](mailto:zhanglimei@btbu.edu.cn)

RECEIVED 06 September 2025

REVISED 31 October 2025

ACCEPTED 16 December 2025

PUBLISHED 12 January 2026

## CITATION

Yin Z, Zhang L, Liu H, Du Q and Yu C (2026) EAC-YOLO: a surface damage identification method of lightweight membrane structure based on improved YOLO11.

*Front. Comput. Sci.* 7:1700167.

doi: 10.3389/fcomp.2025.1700167

## COPYRIGHT

© 2026 Yin, Zhang, Liu, Du and Yu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# EAC-YOLO: a surface damage identification method of lightweight membrane structure based on improved YOLO11

Zihang Yin, Limei Zhang\*, Huarong Liu, Qiuyue Du and Chongchong Yu

Department of Mechanical Engineering, School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing, China

Different surface damage can cause harm to membrane structures, and traditional manual inspection methods are inefficient and prone to missed detections and false alarms. At the same time, the current mainstream detection algorithms are highly complex, which is not conducive to deployment on resource-constrained devices. To achieve automatic identification of typical surface damage in membrane structures, we construct a dataset comprising five damage types based on common types of surface damage in membrane structures and propose a lightweight identification algorithm for membrane structure surface damage, specifically EAC-YOLO. Firstly, the SPPF module is reconstructed, and the ECA lightweight attention mechanism is introduced to enhance the model's ability to distinguish easily confused features. Secondly, ADOWN is introduced to replace the original down-sampling method, improving the retention ability of multi-scale damage features. Finally, the CGBlock and C3k2 modules are combined and reconstructed in the neck network to reduce the interference of damage background factors and capture more features of the damage and its surrounding environment. Experimental evaluation results on the established dataset show that the improved mAP<sub>50</sub> value reaches 87.5%, and the number of parameters, computational cost, and model size are reduced by approximately 28, 25, and 28%, respectively, compared with the original model, demonstrating the advantages of a small size and high accuracy.

## KEYWORDS

membrane structure, damage identification, YOLO11, target detection, deeplearning

## 1 Introduction

In recent years, membrane structures have gained widespread popularity due to their lightweight and aesthetically pleasing appearance.

However, under the appearance of its novel shape and unique appearance, it hides the inherent characteristics of high sensitivity to local damage. Small holes, stain corrosion and small folds on the surface of membrane materials may cause the failure of some or even the whole structure under the action of prestress, wind, rain and snow during use, which poses a threat to public safety that cannot be ignored. Affected by the long-term external complex environment, there are many types of damage on the surface of the membrane structure. It is very important to define and classify the types of damage in advance, collect many different damages on the surface of the membrane structure in use in an all-round way, and take timely repair and maintenance measures according to the types of damage.

Traditional damage detection methods rely on engineers' manual visual and empirical judgment. While the accuracy is poor, the data results are susceptible to engineers' subjective influence. It is difficult to provide an objective and quantifiable evaluation basis for low contrast or initial small damage (Tang et al., 2017; Psaromatis et al., 2020). At the same time, some large-scale membrane structure buildings with high height bring great difficulty to manual detection. This technical bottleneck reflects that the development of an efficient, accurate, and convenient damage identification system is no longer a simple technical pursuit, but an urgent need to ensure the long-term safety of such critical infrastructure.

With the development of computer and artificial intelligence technology, machine vision has begun to be applied in the field of surface damage detection. This method captures the surface image of the product using an appropriate light source and image sensor. It uses the corresponding image processing algorithm to extract the feature information of the image. Then, according to the feature information, the surface damage location, identification, grading, and other discrimination and statistics are carried out, including storage, query, and other operations. Compared to the traditional manual detection method, this method offers higher real-time performance and accuracy and exhibits specific working capabilities in harsh and extreme environments. At the same time, convolutional neural networks (CNNs) have received extensive attention in the field of industrial defect or damage detection due to their more powerful feature learning and feature expression capabilities. Surface defects or damage detection based on deep learning learns feature representation from a large amount of data by constructing a neural network model to realize automatic identification and classification of defects and damage. The rapid development of machine vision endows the machine with the ability of visual perception and realizes the efficient and stable acquisition of massive image information on the surface of the structure. The rise of deep learning means that on the basis of visual perception, it combines the ability to understand and judgment of the machine itself. The integration of the two essentially constructs a closed loop from perception to cognition. In the field of defect detection, it has been realized the transformation from artificial vision to machine perception, from subjective judgment to data analysis. It has laid a solid foundation for the rapid and accurate diagnosis of non-destructive, non-contact, and full-field, and has become an urgent evolutionary direction in the field of structural surface condition monitoring.

Some scholars combine the convolutional neural network method with machine vision to learn the application of surface defect detection in different fields. He et al. (2020) designed a steel plate defect detection system based on deep learning and achieved 74.8/82.3 mAP using 300 candidate sets on the baseline network ResNet34/50. The detection speed of 20 ft. per second on a single GPU achieves 92% of the above performance. Aiming at the problems of slow detection speed and low detection accuracy of traditional steel surface defect detection methods, Wang X. et al. (2022) proposed a steel surface defect detection algorithm based on improved YOLOv7 and obtained 80.2% mAP and 81.9% mAP on the GC10-DET dataset and NEU-DET dataset, respectively. Xue and Li (2018) proposed a fully convolutional network (FCN) classification model and implemented an automatic intelligent classification and detection method for tunnel lining defects. At a test time speed of 48 ms per image, the best set accuracy of the proposed model exceeds 95%. Aiming to address the

issues of low network accuracy, slow speed, and numerous model parameters in defect detection of printed circuit boards (PCBs), Tang et al. (2023) proposed an improved PCB surface detection algorithm based on YOLOv5, which achieved a 95.97% mAP at 92.5 FPS. Dong et al. (2020) proposed a pyramid feature fusion and global context attention network for pixel-level detection of surface defects. On the four datasets of NEU-Seg, DAGM 2007, MT defect, and Road defect, the average pixel accuracy reached 82.15, 74.78, 71.31, and 79.54%, respectively. Xu et al. (2023) proposed a YOLOv5-IMPROVEMENT model for intelligent defect recognition of weld X-ray films based on deep learning, achieving a precision and recall of 92.2 and 92.3%, respectively. To solve the problem of false positives and missed detection of casting defects in X-ray inspection. Lin et al. (2018) proposed a robust detection method based on a vision attention mechanism and feature map deep learning. The accuracy of defect detection is greater than 96%. Lv et al. (2020) proposed an end-to-end defect detection and classification network based on a Single-Shot Multibox Detector (SSMD) for metal surface defect detection. Luo et al. (2021) proposed a decoupled two-stage target detection framework based on CNN for non-significant defects on flexible printed circuit boards and the similarity between different defects, which achieves mAP of 94.15%. Liu et al. (2024) proposed a wind turbine damage detection algorithm based on a YOLOv8 design, which achieved 79.9% mAP.

However, there are still some issues when using the traditional target detection model to detect surface damage in membrane structures. For example:

- (1) The dataset of surface damage of membrane structures is scarce, and there are many types of surface damage and significant morphological differences. The contrast between the damage and the background is low, which can easily lead to missed detection.
- (2) The surface coverage area of the membrane structure is large. When using mobile devices such as drones to detect, the requirements for computer hardware and storage are high.

In the research related to handling complex background textures. Aiming at the complex background of tunnel surface and the poor performance of artificial feature extraction methods in detecting tunnel defects, Xu et al. (2021) proposed a tunnel defect detection method based on Mask R-CNN, based on path enhanced feature pyramid network (PAFPN) and edge detection branch. Based on the YOLOv4 model. Chen et al. (2022) embedded Gabor kernels into Faster R-CNN, a two-stage training method based on a genetic algorithm (GA) and backpropagation, which was designed to optimize the application of a CNN model in fabric defect detection. Li et al. (2023) proposed an automatic defect detection scheme for wire and arc additive manufacturing (WAAM) by combining a channel-level attention mechanism, multiple spatial pyramid pooling, and an exponential moving average, aiming to address the challenges of complex defect types and a noisy detection environment. Liu et al. (2023) proposed an efficient feature extraction (EFE) module, suggesting that the network achieves a satisfactory balance between performance and efficiency. Gao et al. (2022) proposed an improved Variant of the Swin Transformer, which enhances feature transmission between windows by designing a new window shift scheme, making the framework more capable of serving as the backbone for defect

detection. The scheme achieves 94.5% mAP at a rate of at least 42 frames per second. Zhao et al. (2021) proposed a multi-scale fusion training network using deformable convolutions to improve the detection of small and complex defects on steel surfaces, which increased mAP by 0.128% compared to the baseline. Tao et al. (2018) proposed a deep convolutional neural network cascade autoencoder (CASAE) for segmenting and positioning surface defects, addressing the problem of metal surface defect detection in complex industrial scenarios. Zhang et al. (2024) proposed an anchor-free network based on DsPAN for small target detection, given the characteristics of multi-scale, multi-type, and multi-small targets, and complex background interference on the surface defects of industrial products, which achieves mAP of 80.4, 95.8, and 76.3% on the three public datasets, respectively. To improve the recognition accuracy of defect targets that are difficult to distinguish under complex tunnel background and illumination conditions, Zhou et al. (2022) proposed a YOLOv4-ED model based on deep learning. The mAP, F1 score, model size, and FPS of the model reached 81.84, 81.99%, 49.3 MB, and 43.5 fps, respectively. To improve the detection performance of small targets. Zeng et al. (2022) proposed a lightweight enhanced multi-scale feature fusion method, ABFPN, for detecting small defects on the surface of PCBs, achieving a mAP<sub>50</sub> of 85.59%. Yuan et al. (2023) proposed an adaptive lightweight battery current collector defect detection model, DGNet, which achieved 91.8% mAP on the BCC surface defect dataset, the model size is 4.0 M, 3.7 GFLOPS. Li et al. (2022) proposed a lightweight salient object detection in optical remote sensing images (ORSI-SOD) solution called CorrNet, which reduced the number of parameters to 4.09 M, and solved the problem of high computational memory and high cost. Zhang et al. (2023) combined the inverse residual architecture with the coordinate attention (CA) mechanism and constructed the coordinate attention movement (CAM) backbone network for feature extraction, and designed an efficient, lightweight CNN model for surface defect detection of industrial products. Wang Y. et al. (2022) proposed a lightweight PCB defect detection model based on the improved YOLOX-MC-CA, which introduces CSPDarknet and coordinate attention (CA). Zhao et al. (2024) developed a lightweight defect detection model for turbine blades based on ShuffleNetv2 and coordinate attention (SN-CA-SSD) in the Internet of Things using single-shot multi-box detection, which achieves a balance between precision and efficiency. Shen et al. (2024) proposed a multi-scale interactive (MI) module and introduced it into a lightweight multi-scale interactive network (MINet) to perform real-time salient object detection on strip surface defects.

However, these methods are often targeted at solving the problem of processing a specific detection target, and there is still a lack of feature processing for surface damage in membrane structures. Based on several of the above methods, we utilize the representative YOLO model in the single-stage target detection algorithm to design our damage identification model. The YOLO model has more advantages in speed, accuracy, and ease of use, and has become one of the widely used target detection frameworks in industry and academia. By summarizing the morphological characteristics of the recognition target, the feature extraction module and the down-sampling operator in the original model are improved, aiming to enhance the model's learning ability in recognizing the surface damage characteristics of various complex membrane structures. At the same time, taking into account the actual application scenarios, the model structure is simplified while ensuring accuracy, making it more conducive to the deployment of mobile

terminals. In the future, the algorithm can be deployed on mobile devices such as drones to replace manual detection methods to perform surface damage detection and identification of membrane structures. Combined with computer vision, it can realize automatic identification, accurate positioning, and quantitative intelligent evaluation of membrane structure surface damage. We construct a dataset of membrane structure surface damage and propose a membrane structure surface damage identification algorithm based on EAC-YOLO. The main contributions of EAC-YOLO are included as follows.

- (1) The SPPF module in the backbone is improved, and SPPF-ECA is proposed to enhance the identification of small targets and confusing damage features.
- (2) ADown operator is introduced to optimize the original down-sampling method and reduce the number of parameters.
- (3) The CGBlock module in CGNet is introduced into the neck network to construct C3k2-CGBlock, which reduces the number of parameters and suppresses the influence of complex background factors on the membrane surface. It achieves a better feature expression ability for surface damage types of membrane structures and reduces the model's complexity while ensuring the accuracy requirements.

## 2 Related works

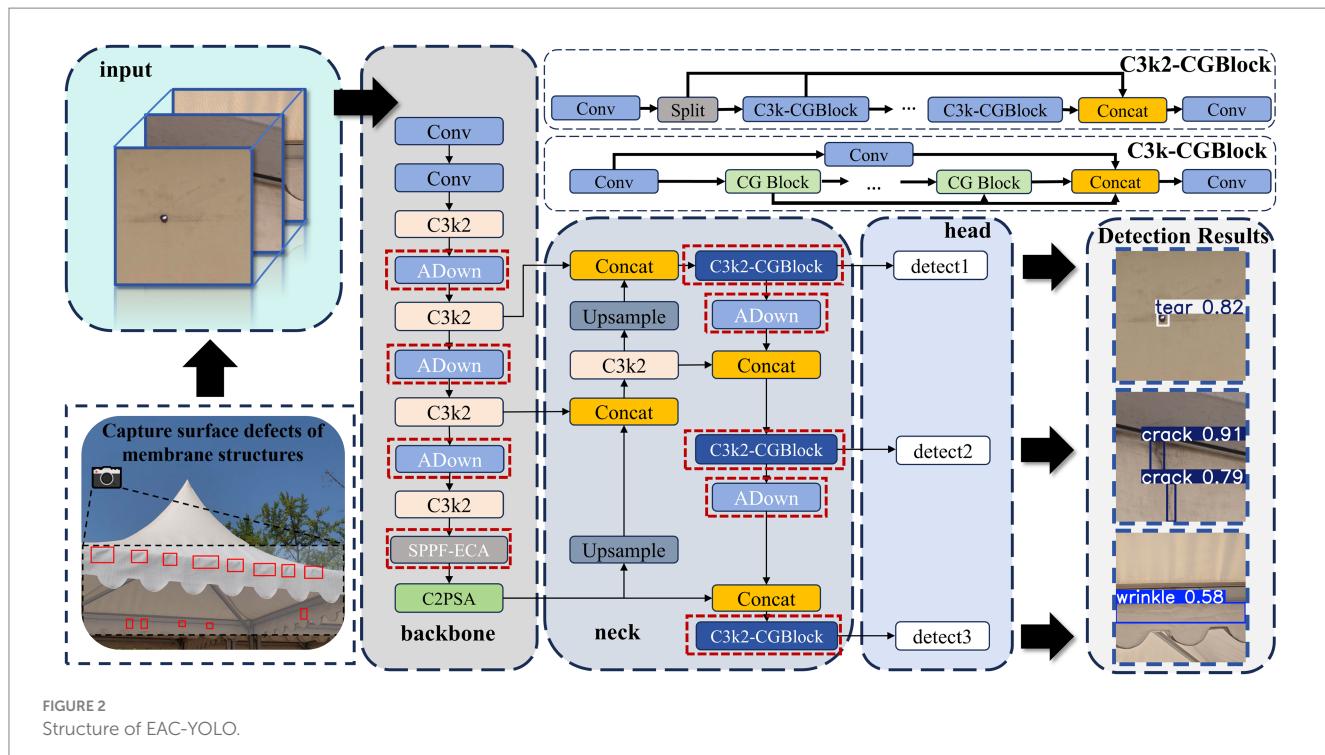
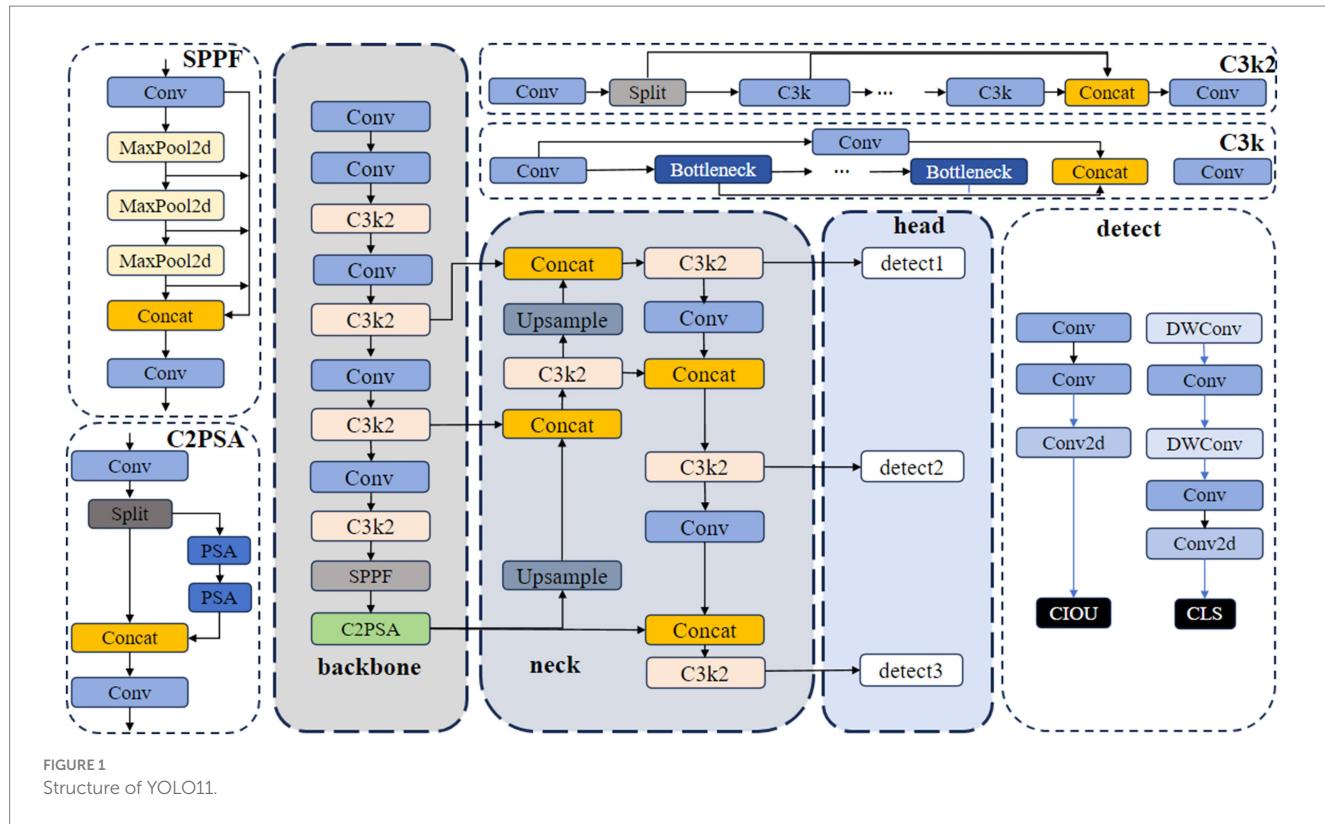
The YOLO (Redmon et al., 2016) series is a crucial algorithm in the field of target detection, known for its efficient Single-stage detection architecture. YOLO11 (Jocher and Qiu, 2024) is the latest generation of real-time target detection model, launched by the Ultralytics team in 2024, which has been comprehensively optimized in terms of accuracy, speed, and ease of use.

The structure of YOLO11 is shown in Figure 1. In the backbone and neck network of YOLO11, the latest C3k2 module is introduced to extract features. C3k2 is inherited from the C2f module in YOLOv8 (Jocher and Qiu, 2024). The choice of whether to use the C3k layer is added, which makes the model more flexible in scenes requiring more flexible feature extraction, such as those involving different receptive fields—advantages, while retaining the characteristics of C2f's ability to perform fast feature fusion. SPPF is a pyramid pooling module proposed in YOLOv5. It is mainly based on the idea of Spatial Pyramid Pooling (SPP) (He et al., 2015) and obtains multi-scale feature information through pooling operations at different scales. The C2PSA in its backbone network combines the C2f and Position-Sensitive Attention (PSA) attention mechanisms, dynamically adjusting the attention degree of various positions through these mechanisms, which improves the model's detection accuracy in complex scenes. The head part introduces a deep separable convolution to the CLS branch, reducing redundant calculations and improving efficiency.

In this paper, we use YOLO11 as the baseline model to improve the network structure, and a model suitable for surface damage identification of membrane structure is proposed.

## 3 Improved EAC-YOLO structure

The structure of EAC-YOLO is shown in Figure 2. The modules within the red dotted box in the figure represent the parts that have



been modified compared to the original model. The main improvements of EAC-YOLO include:

(1) SPPF reconstruction. We introduce the ECAttention mechanism in the Concat layer to enhance the model's discriminative ability,

allowing it to easily distinguish between features and extract small target features, such as small target tears and cracks.

(2) A Down operator, which improves the efficiency of down-sampling while enhancing the ability to retain multi-scale features and further reduces the number of model parameters.

(3) CGBlock is introduced into the neck network. This module is added to the C3k2 module and reconstructed to help the model capture the damage location and its surrounding environment characteristics, thereby reducing the impact of the complex background around the damage.

### 3.1 SPPF-ECA structure

Efficient Channel Attention (ECA) is a channel attention mechanism proposed by Wang et al. (2020) to solve the performance and complexity of the model. This module is improved based on Squeeze-and-Excitation Networks (SEAttention) (Hu et al., 2018), which avoids the dimensionality reduction operation of the fully connected layer in the channel attention module. Through the interaction of complete channels, the unique high-frequency features such as edges and textures of small damage are strengthened. ECA also employs a local cross-channel interaction strategy to develop an adaptive method for determining its kernel size  $k$ , thereby avoiding the need for manual tuning of  $k$  through cross-validation in fast 1D convolution operations. This approach enables efficient calculation of channel attention. The overall calculation process is as follows:

(1) The input feature map is transformed from  $(H, W, C)$  to  $(1, 1, C)$  by Global Average Pooling (GAP) aggregation, a convolution feature map without reducing the dimension. The output feature vector is:

$$Z_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j) \quad (1)$$

Where  $N$  represents the batchsize, while  $H$ ,  $W$ , and  $C$  denote the input height, width, and channel number, respectively.  $U_c$  is the input feature vector, and  $Z_c$  is the output feature vector.

(2) The kernel size  $k$  is determined by an adaptive method for the one-dimensional convolution operation. The size of the resulting feature map is still  $(1, 1, C)$ . The adaptive adjustment of the convolution kernel size satisfies:

$$k = \varphi(c) = \left\lceil \frac{\log_2(C) + b}{\gamma} \right\rceil_{odd} \quad (2)$$

Where  $k$  represents the size of the adaptive convolution kernel,  $\varphi$  represents the mapping function,  $\lceil t \rceil_{odd}$  represents the nearest odd value to  $t$ ,  $\gamma$  and  $b$  are hyperparameters which are used to map the adjustment function. We take their values as 2 and 1, respectively (Wang et al., 2020).

(3) The adaptive convolution kernel is applied to the one-dimensional convolution operation to obtain the adaptive weight of each channel in the feature map. A one-dimensional convolution operation with a convolution kernel size of  $k$  is performed. The convolution result is mapped to 0–1 by the sigmoid activation function of the output, and the weight of each channel is obtained as follows:

$$\omega = \sigma(C1D_k(y)) \quad (3)$$

Where  $y$  is the input feature map, and  $C1D_k$  is a one-dimensional convolution. The weighted feature map is obtained by multiplying the attention weight  $\omega$  by the original input feature map.

The entire calculation method significantly reduces the complexity of the model while maintaining performance, determines the coverage of local cross-channel interactions, and enhances the feature learning and representation capabilities of the neural network.

The structure of ECA network is shown in Figure 3, and the structure of SPPF-ECA is shown in Figure 4. Because the features of small targets, such as holes, wrinkles, and cracks on the membrane surface, are sparse, the dimensionality reduction operation of the conventional attention mechanism will further weaken their identifiability. ECA directly utilizes one-dimensional convolution to preserve the original information of all channels, thereby strengthening the unique edges, textures, and other high-frequency features of small damage through the interaction of complete channels. The original SPPF structure mainly includes three  $5 \times 5$  pooling layers and one Concat layer, and two Conv structures. The size of the Conv1 structure is  $(c1, c\_1, 1)$ , and the size of the Conv2 structure is  $(c\_4, c2, 1, 1)$ , where  $c1$  and  $c2$  are the number of input channels and output channels, respectively, and  $(c\_ = c1/2)$ . Therefore, to combine the advantages of SPPF and ECA, ECA is introduced after the Concat layer in SPPF, further improving the model's feature expression ability through efficient channel attention calculation and an effective receptive field. At the same time, the lightweight attention structure has an almost negligible additional calculation amount for introducing the SPPF structure, which has advantages in the scene of membrane structure surface damage detection that requires considering both accuracy and speed.

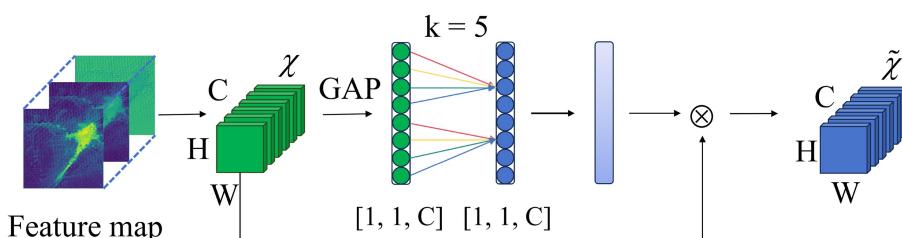


FIGURE 3  
Structure of ECA network.

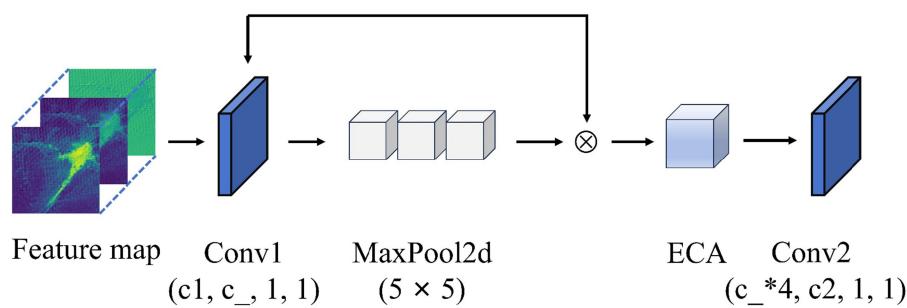


FIGURE 4  
Structure of SPPF-ECA.

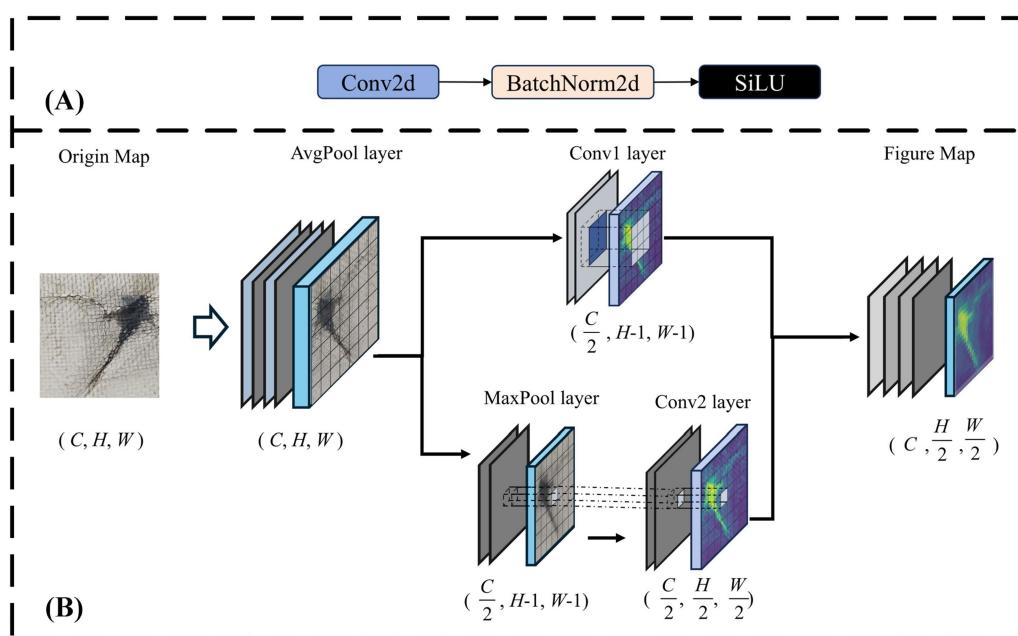


FIGURE 5  
Structure of down-sampling operators. (A) Conv; (B) ADown.

### 3.2 ADown operator

The down-sampling operator in YOLO is used to reduce the spatial size of the feature map, thereby reducing the computational requirements and the number of parameters. The down-sampling operator in the original model is a convolution with a step size of 2, including Conv2d, BatchNorm2d, and SiLU layers. The structure is relatively simple, but when dealing with small or complex texture damage in damage identification work, its single convolution kernel may not be able to capture multi-scale features, resulting in a loss of information. At the same time, its large parameter quantity will have a greater impact on the deployment of mobile terminals.

ADown (Wang et al., 2024) is a lightweight down-sampling operation that enhances the model's sensitivity to various damage features through multi-branch processing, combining maximum pooling and average pooling methods. Additionally, reducing the number of parameters may improve computational efficiency and

make it suitable for use in resource-constrained environments. The feature map first undergoes an average pooling operation to minimize noise interference and retain the overall texture features of the membrane surface; after that, the output feature map is evenly divided along the channel dimension. One part of the feature map, after equalization, captures the main features of the damage through convolution with a kernel size of 3. In contrast, the other part retains the input features after maximum pooling, reducing the amount of data. The result is merged with the feature map of the other part after a  $1 \times 1$  convolution fusion of feature information.

After the above improvements, the computational complexity of the model is significantly reduced, and the feature expression ability is enhanced. While maintaining down-sampling efficiency, the feature retention ability of multi-scale surface damage is preserved, making it suitable for membrane structure detection tasks that require high-precision positioning. The structure of Conv and ADown is shown in Figure 5.

### 3.3 C3k2-CGBlock structure

Context-Guided Network (CGNet) is a lightweight neural network designed for semantic segmentation, proposed by Wu et al. (2020). It understands the scene by simulating the context information of the human visual system and then performs effective feature extraction and fusion of context information. Context-Guided Block (CGBlock) is an important module in CGNet; the structure of CGBlock is shown in Figure 6. The input damage feature map of the membrane structure is compressed by the conv channel, and the dual-path feature extraction of  $f_{loc}$  and  $f_{sur}$  is performed.

The data is extracted in the  $f_{loc}$  through the standard  $3 \times 3$  convolutional layer to extract the local features in the feature map. At the same time, the size of the  $3 \times 3$  dilated convolutional receptive field is utilized in  $f_{sur}$ , thereby enhancing the capture of damage features and background factor information from the surrounding membrane surface, and compensating for the limitations of the Bottleneck model in deep networks. It may not effectively capture the shortcomings of continuous folds and cracks on the membrane surface, as well as the background interference caused by the warp and weft weaving of the fabric fibers within the membrane material. Then the data is spliced by the joint feature extractor  $f_{joi}$  through the Concat layer and subjected to batch normalization and PReLU operations. These features are used to capture the details of the input data and the information in the broader area. Finally, the information output by  $f_{joi}$  is captured by  $f_{glo}$  and utilized to enhance the global information of the entire input image, which suppresses the noise channel and enhances the key channel, retaining features sensitive to low-contrast damage and improving the detection ability of small wrinkles, slight smudges, and scratches. The context information is extracted by global average pooling and two fully connected layers and further processed. Finally, it is combined with the original information output by  $f_{joi}$  to enhance the learning of damage features by the joint feature learner.

## 4 Experiments

### 4.1 Dataset

Membrane structures are typically used as roofing materials for outdoor buildings. Due to long-term exposure to the natural environment, they will naturally withstand the influence of complex external environmental factors such as wind, rain, and snow, resulting in different types of surface damage. The damage comes in various types, sizes, and shapes. According to the inspection index in the technical specification.

Based on the membrane structure (CECS, 2015; Yu et al., 2023) and on-the-spot investigations of tensile membrane building structures in various locations, the five common damage types crack, wrinkle, scratch, tear, and smudge are selected as the research object, as shown in Table 1.

Since there is no publicly available dataset for identifying surface damage in membrane structures, 2,175 original damage samples were collected through various methods, including actual shooting and network downloads, according to the five types of damage listed in Table 1. The samples were removed from invalid picture samples, such as blurring and repetition, and a total of 2035 damaged picture samples were finally obtained. The pattern is shown in Figure 7.

Subsequently, the collected image samples are cropped, and the image is uniformly adjusted to  $1,080 \times 1,080$  pixels. To enrich the dataset, enhance its robustness, and improve the network model's generalization ability in detecting surface damage features, we simulate these features—the state of the membrane structure in more complex environments. The sample groups in the dataset have been changed. Luminance, random part clipping, random angle rotation, random pixel loss, random scale scaling, vertical mirror flipping, horizontal mirror flipping, and other data enhancements expand the dataset to 3,930 pictures. Different data augmentation methods are shown in Figure 8.

LabelImg software is used to perform image annotation and label generation on the collected images. The annotation interface is shown in Figure 9. The rectangular box is used to mark the position

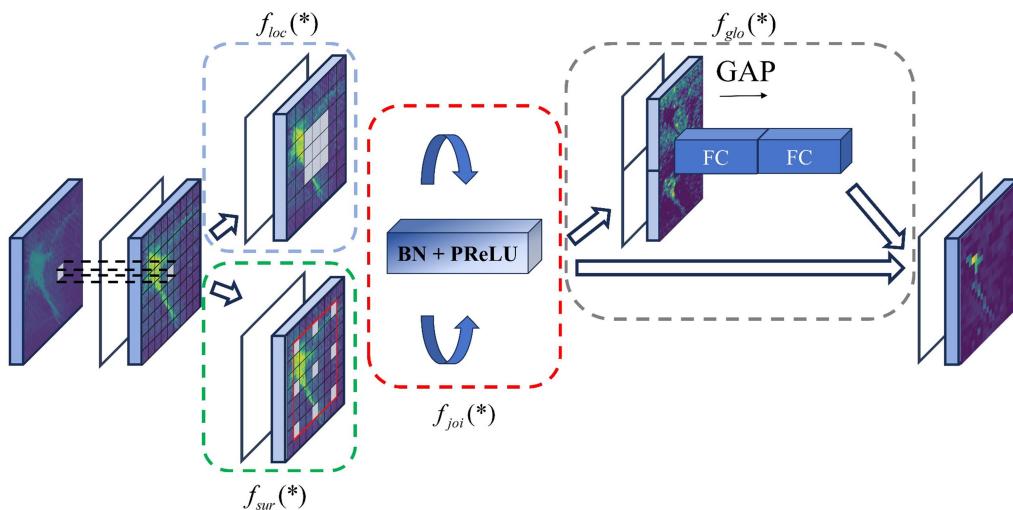


FIGURE 6  
Structure of CGBlock.

TABLE 1 Examples, characteristics, and causes of common damage types of membrane structures.

Types	Examples	Characteristics	Causes
Crack		Linear cracks inside the membrane material	Stress, Fatigue
Wrinkle		The irregular folding shape of the membrane material	Prestress loss, compression, or folding during transport or storage of membrane materials
Scratch		Lines or depressions	Surface impact treatment
Tear		Irregular-shaped holes, accompanied by internal fracture fiber exposure	Local fracture, separation, or internal crack propagation occurs under the action of external force
Smudge		Flakes and strips of dirt on the surface	Pollutant attachment

containing the target damage type. Some of the images contain multiple annotations of multiple surface damage types. The text file in YOLO format is used to form a sample set. After the annotation is completed, all images are divided into a training set, a verification set, and a test set. The specific numbers are 2,700 images for the training set, 615 images for the verification set, and 615 images for the test set.

## 4.2 Training details

According to the depth of the model and the amount of calculation, YOLO11 is divided into five different models: n, s, m, l, and x (Jocher and Qiu, 2024). In this experiment, YOLO11s is selected

as the baseline model. YOLO11s is a small-scale model in the YOLO11 series. It has a small model volume and calculation amount while ensuring a high detection accuracy. It is suitable for the real-time detection environment required for industrial membrane structure detection. According to the experimental content, ablation experiments, module comparison experiments, and detection model comparison experiments were carried out. The dataset used in the experiment is a self-built membrane structure surface damage (MSSD). To ensure the fairness of the comparison, all the comparison algorithms and ablation experiments in this experiment are completed under the same experimental conditions and training parameters.

In this experiment, the Windows 11 operating system is used. The GPU model used in the experiment is the NVIDIA GeForce

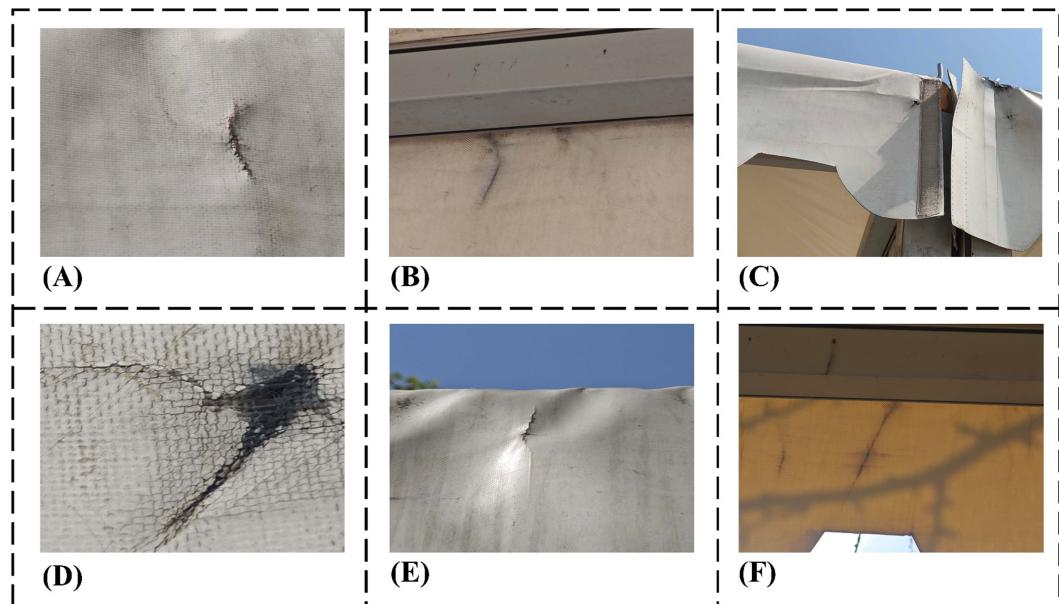


FIGURE 7

Sample acquisition image. (A) Front; (B) back; (C) distant view; (D) close view; (E) front lighting; (F) back lighting.

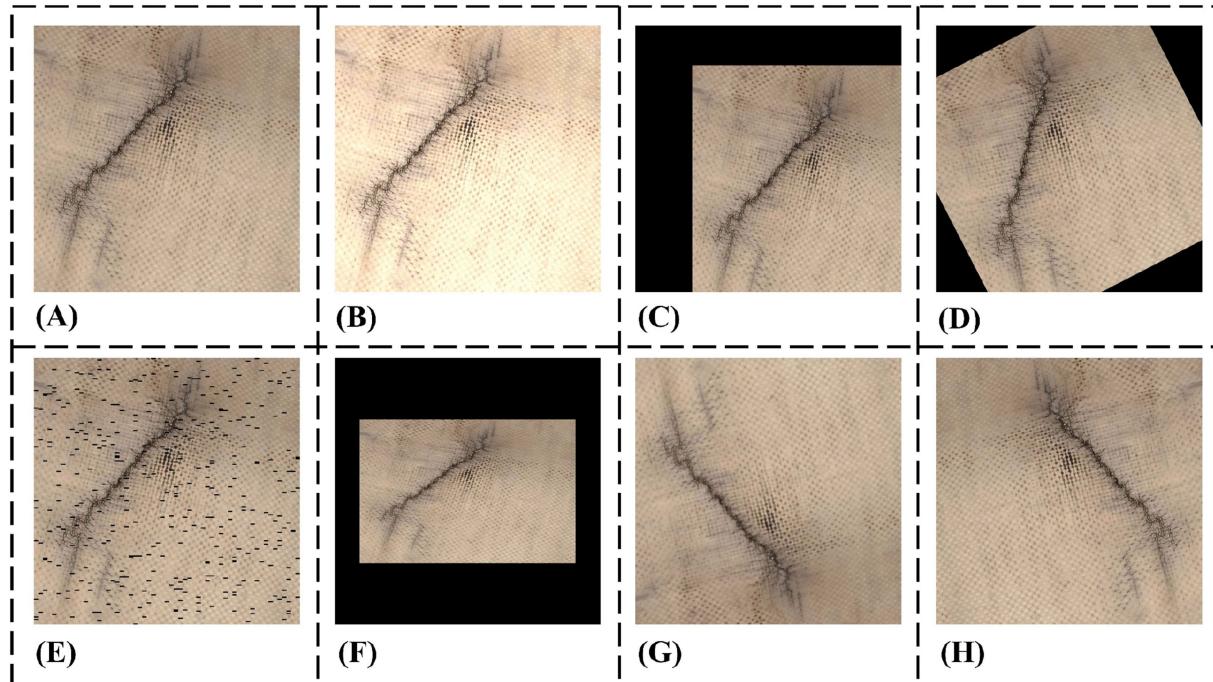


FIGURE 8

Different data augmentation methods: (A) original figure; (B) changing brightness; (C) random cropping; (D) rotating; (E) random pixel missing; (F) random scale; (G) vertical flip; (H) horizontal flip.

RTX 4070 Super, with 12 GB of memory. The CPU model is Intel(R) Core (TM) i9-14900K, 3.2 GHz; the Python version is 3.9.20. The deep learning framework and version used are PyTorch 2.0.0, and the CUDA version is 11.8. The experimental parameters are shown in Table 2.

### 4.3 Performance evaluation index

To objectively evaluate the experimental results of the improved model presented in this paper, precision (P), recall (R), model parameters, as well as Mean Average Precision (mAP) and

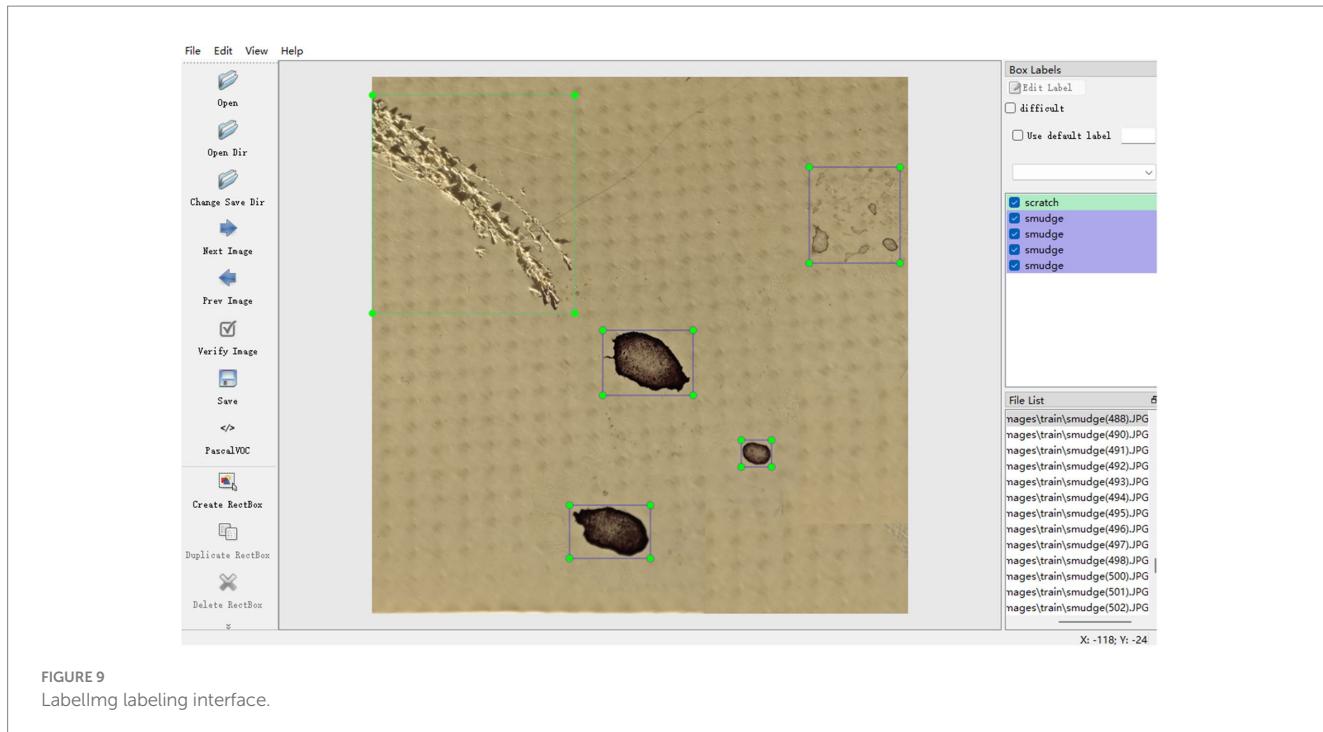


FIGURE 9  
Labelling labeling interface.

TABLE 2 Experimental related training parameters.

Parameter	Value
Optimizer	SGD
Epochs	250
Batch Size	16
Workers	8
Learning Rate	0.01
Images Size	640
Weight Decay	0.0005
Momentum	0.937

floating-point operations per second (FLOPS), are used as indicators to assess the model. As shown in Equation 4, the  $P$  refers to the proportion of correctly detected positive samples among all predicted positive samples, as shown in Equation 5, the  $R$  is the proportion of correctly detected positive samples among all actual positive samples., as shown in Equations 6, 7, mAP<sub>50</sub> refers to the mAP of all categories when the IoU threshold is 0.5 in the detection task, reflecting the model's generalization ability in multi-category scenarios. mAP<sub>50-95</sub> measures the average precision of the model across IoU thresholds from 0.5 to 0.95, and FLOPS represents the number of floating-point operations performed by the model, used to evaluate computing resource requirements and assess the complexity of the algorithm model. The calculation formula of some indexes is as follows:

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$AP = \int_0^1 P(R) dR \quad (6)$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i \quad (7)$$

In the formula,  $TP$  is the number of samples that the model correctly predicts the positive samples,  $FP$  is the number of negative samples misjudged as positive samples,  $FN$  is the number of positive samples misjudged as negative samples,  $C$  is the total number of categories in the detection task, and  $AP_i$  is the average precision of every category.

The loss function is used to measure the difference between the predicted value and the real value of the model in the target detection task, which directly affects the performance of the model. The loss evaluation indicators in YOLO include three types: Bounding Box Regression Loss (box\_loss), Objectness Loss (obj\_loss), and Classification Loss (cls\_loss). It can be seen from Figure 10 that after 250 rounds of training, the loss function curves of the two groups tend to converge. In the classification loss statistics of the validation set, the value of EAC-YOLO convergence is slightly lower than that of the YOLO11 model, which has a slight advantage.

#### 4.4 SPPF contrast experiment

To verify the superiority of the reconstructed SPPF-ECA, Convolutional Block Attention Module (CBAM) (Woo et al., 2018), Efficient Multi-Scale Attention (EMA) (Ouyang et al., 2023), Coordinate Attention (CA) (Hou et al., 2021), Large Separable Kernel Attention (LSKA) (Lau et al., 2024), and SPPF fusion are compared with SimSPPF (Li et al., 2023), BasicRFB (Liu and Huang, 2018), FocalModulation (Yang et al., 2022), and SPPF in the baseline model

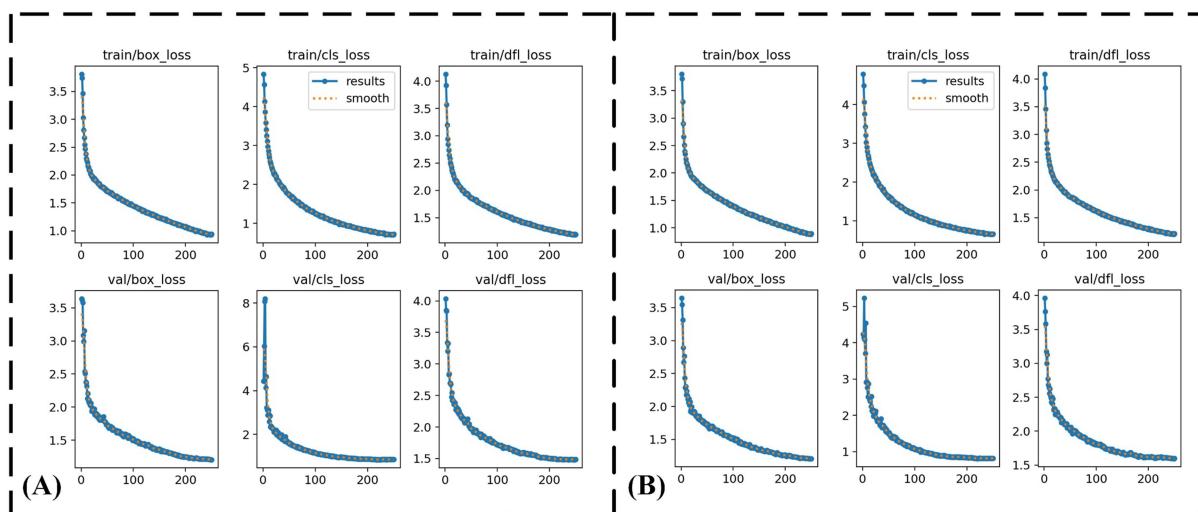


FIGURE 10  
Comparison of loss curves for training of the two models. (A) YOLO11 (B) EAC-YOLO.

TABLE 3 Adding experiments of different pyramid pooling modules.

Model	P/%	R/%	mAP <sub>50</sub> /%	FLOPS/G	Params/M
SPPF	88.2	81.5	86.5	21.3	9.41
SimSPPF	89.1	81.0	86.6	<b>21.2</b>	<b>9.42</b>
FocalModulation	87.0	75.1	82.5	21.6	9.83
BasicRFB	88.9	80.3	85.9	21.8	10.08
SPPF-CBAM	87.5	77.3	83.8	22.2	10.46
SPPF-EMA	90.1	80.8	86.1	21.6	9.43
SPPF-CA	89.8	81.8	86.9	21.4	9.52
SPPF-LSKA	88.5	<b>82.2</b>	86.5	22.1	10.35
SPPF-ECA	<b>90.8</b>	81.1	<b>87.2</b>	21.3	9.41

in the same way. A comparative experiment was conducted using the MSSD dataset. The evaluation indices were selected to compare the precision, recall, mAP<sub>50</sub>, parameter quantity, and FLOPS. The experimental results are shown in Table 3.

From the analysis results of Table 3, it can be seen that:

- (1) When the attention mechanism layer is added at the same position in the SPPF module of the baseline model, although the R of the SPPF-ECA model is slightly reduced, FLOPS and parameters remain unchanged. The P and mAP<sub>50</sub> reached 90.8 and 87.2%, representing improvements of 2.6 and 0.7% over the baseline values of 88.2 and 86.5%, respectively. FLOPS and parameters remain unchanged.
- (2) The SPPF-LSKA model has the highest R, which is 0.7% higher than the model, but the improvement effect of the P and the mAP<sub>50</sub> is not apparent enough, and the number of parameters and FLOPS increase by 0.94 M and 0.9G, respectively. Excessive parameters will increase the computational burden of the model.
- (3) The parameter quantity and FLOPS of the model SimSPPF are the lowest, but the optimization effect on the R and the mAP<sub>50</sub> is not ideal.

The indicators of the experimental results show the effectiveness of SPPF-ECA deployed in the model. To further verify the effect of each module in the practical application scenario of membrane structure surface damage detection, Grad-Cam (Selvaraju et al., 2020) was used to generate heat maps of the above modules for identifying small target damage features, and each improved module was visually evaluated. In the heat map, the darker the color of the red area, the greater the calculation weight of the part, and the higher the confidence of the test results. The results are shown in Figure 11.

It can be seen from Figure 11 that most of the improved modules have a small red area corresponding to the damage of the original image in the detection heatmap, indicating that the attention to the damaged parts is low. The yellow and green color distribution of the damage background is generally higher, and the suppression effect of background interference factors is poor.

There are some shortcomings in the application of small target damage detection, such as small holes and scenes with complex background interference. As shown in Figure 10J, the red area calculated by SPPF-ECA has the highest coincidence degree with the corresponding damage area in the original image. The color distribution of the damaged background is lower, indicating that the

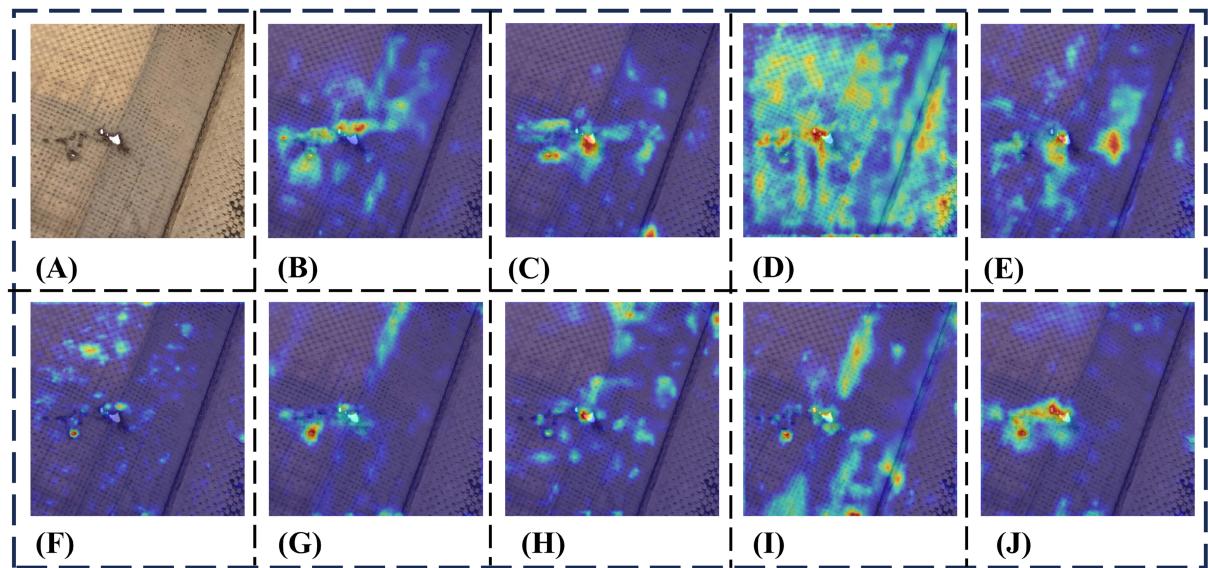


FIGURE 11

The heatmap drawing results of each model detection. (A) Original figure; (B) SPPF; (C) SimSPPF; (D) FocalModulation; (E) BasicRFB; (F) SPPF-CBAM; (G) SPPF-EMA; (H) SPPF-CA; (I) SPPF-LSKA; (J) SPPF-ECA.

TABLE 4 Ablation experimental results.

SPPF-ECA	ADown	C3k2-CGBlock	P/%	R/%	mAP <sub>50</sub> %	mAP <sub>50-95</sub> %	FLOPS/G	Params/M	Size/MB
-	-	-	88.2	81.5	86.5	58.0	21.3	9.41	19.2
✓	-	-	<b>90.8</b>	81.1	87.2	57.3	21.3	9.41	19.2
-	✓	-	90.1	80.5	86.5	58.4	17.2	7.50	15.4
-	-	✓	90.1	80.6	86.0	57.6	20.0	8.68	19.3
-	✓	✓	80.2	81.6	87.0	57.9	15.9	6.76	13.9
✓	-	✓	89.7	81.3	86.7	57.3	20.0	8.68	17.7
✓	✓	-	90.2	81.9	87.2	<b>59.1</b>	17.2	7.50	7.50
✓	✓	✓	89.4	<b>83.1</b>	<b>87.5</b>	58.4	<b>15.9</b>	<b>6.76</b>	<b>13.9</b>

module can focus more accurately on the damaged area and minimize background interference factors to the greatest extent.

#### 4.5 Ablation experiment

To evaluate the influence of each algorithm module on the surface damage identification and detection model of fabric membrane structure, this experiment carried out multiple sets of ablation experiments through the gradual introduction of SPPF-ECA, ADown, and CGBlock modules, and different combinations under the same configuration environment, and the performance evaluation indexes in the results were counted. It can be seen from the data in Table 4.

- (1) When the SPPF-ECA module is introduced, P and mAP<sub>50</sub> reach 90.8 and 87.2%, respectively, representing absolute improvements of 2.6 and 0.7% over the baseline model. At the same time, it almost does not bring about an increase in the

number of parameters and calculations. This is because the strategy of replacing the fully connected layer with a convolution in the ECA introduces fewer parameters while completing the channel attention calculation.

- (2) When the CGBlock and ADown are introduced separately, the number of parameters is reduced by 0.73 M and 1.91 M, and the P of both models reached 90.1%. At the same time, the mAP<sub>50-95</sub> is increased by 0.4% when the ADown module is introduced, with a slight decrease in R, demonstrating the effectiveness of introducing a single module.
- (3) When the SPPF-ECA and ADown modules are introduced at the same time, it can be seen that P of the model(90.2%) is increased by 2% by the baseline model(88.2%), the R and mAP<sub>50</sub> reached 81.9 and 87.2%, respectively, representing absolute improvements of 0.4 and 0.7% over the baseline model, which makes up for the deficiency of the R when SPPF-ECA is introduced alone. At the same time, the

mAP50-95 index increased by 1.1%, and the parameter quantity decreases by 1.91 M, which demonstrates the effectiveness of the combination of the two. The combination of the two combines the efficient channel attention mechanism with the high down-sampling efficiency application. The ability to retain the characteristics of minor surface damage on the fabric membrane surface is significantly improved.

(4) After introducing C3k2-CGBlock based on the two modules, although P is 0.8% lower than that of the first two modules, it is still 1.2% higher than the baseline model. At the same time, R and mAP50 reached 83.1 and 87.5%, respectively, representing absolute improvements of 1.6 and 1.0% over the baseline model. The best performance was observed across all groups of experiments, with mAP50-95 increasing by 0.4%. The ADown-CGBlock model and the ECA-ADown-CGBlock model had the lowest number of parameters in all groups. Compared with the former, the introduction of the number of individual parameters brought by ECA is almost negligible; the FLOPS, parameters, and model volume reached 15.9G, 6.76 M, and 13.9 MB, which decreased by 25, 28 and 28% compared with the original model, which met the requirements of a lightweight model.

To visually show the actual effect of the improved algorithm, this experiment compares the EAC-YOLO model with the YOLO11 model to visualize the selected multiple images. The results are shown in [Figure 12](#).

As can be seen from [Figure 12](#):

- (1) In the identification of crack damage, YOLO11 is missing in the minor target ridge damage, and a small number of transverse cracks in the dataset of MSSD are mistakenly detected as scratch damage. The EAC-YOLO model performs well, which shows the ability of the added ECA and the ADown down-sampling operator to extract the characteristics of minor target damage. At the same time, ECA optimizes the distinction between confusing features in the case of less distribution of transverse crack damage.
- (2) In the identification of wrinkle damage, the YOLO11 model has missed detection when identifying wrinkle damage with low background contrast, and there is a case of repeated identification of the same feature. At the same time, EAC-YOLO demonstrates a strong ability in low-contrast damage detection. The positioning of damage features is also more accurate, further demonstrating the effectiveness of the CGBlock module addition.
- (3) In the identification of scratches and tears, YOLO11 misdetects the background as a damage feature when there are complex background factors around the damage, while in EAC-YOLO, CGBlock plays a role in combining local and global features, and there are a few cases of background factors. In addition, the tear damage has the characteristics of multiple forms, and the randomness of its shape brings some difficulties to the feature extraction work. Especially in the state of image enhancement simulating a complex environment, its feature information is further affected, which leads to the situation that

the baseline model is easy to causing missed detection. However, we added the ECA to the baseline model to improve the feature expression ability of the model, focusing on the deep feature information of the tear part and reducing the missed detection.

(4) In the identification of smudge features, EAC-YOLO shows that it has certain advantages in the learning of different scale feature forms of the same category. In the face of the characteristics of strong randomness of state distribution and different morphological scales of smudges, ADown has stronger multi-scale feature learning ability than Conv.

## 4.6 Comparative experiments of YOLO series algorithms

To evaluate the effectiveness of the EAC-YOLO algorithm more objectively, this paper conducts performance comparison experiments with other YOLO algorithms, including YOLOv3-tiny, YOLOv5s, YOLOv6s, YOLOv8s, YOLOv9s, YOLOv10s, YOLO11s, and YOLO12s. The experimental results are shown in [Table 5](#).

From [Table 5](#), we can see that among all the model experimental comparisons, the P of EAC-YOLO is the highest, reaching 0.894, and R, mAP<sub>50</sub>, and mAP<sub>50-95</sub> are the best, reaching 83.1, 87.5, and 58.4%, respectively. In the comparison model, the YOLO11 model performs best in terms of R, mAP<sub>50</sub>, and mAP<sub>50-95</sub>, achieving 81.5, 86.5, and 58%, respectively. Meanwhile, the EAC-YOLO model increases by 1.6, 1.0, and 0.4% on this basis. And its computational complexity, measured in FLOPS, is 15.9G, second only to the YOLOv3-tiny model. In terms of the number of parameters and the model volume, EAC-YOLO has the lowest number of parameters, with both indicators approximately 28% lower than those of the YOLO11 baseline model. The smaller volume model achieves higher detection accuracy, making it more suitable for mobile terminal deployment and meeting the need for model lightweight improvement.

To further verify the superiority of the EAC-YOLO model, [Figure 13](#) illustrates the overall trend of precision, recall, loss, and mAP<sub>50</sub> as the epoch increases during the training process of each model. From the diagram, it can be seen that in the Loss change curve, the EAC-YOLO and YOLOv8 curves are closest to the lower end, indicating that the loss situation is the most ideal; in terms of precision, recall, and mAP<sub>50</sub>, EAC-YOLO is at the top compared with other relative positions, and the curve is in a steady upward trend, indicating that the model performs best under the relevant indicators and is applied to the surface damage identification of membrane structures. To comprehensively compare more indicators, such as the number of parameters and the computational complexity of the relevant models, comprehensive performance index radar charts are drawn for different models. As shown in [Figure 14](#), the figure is integrated with precision, recall, mAP<sub>50</sub>, mAP<sub>50-95</sub>, parameters, FLOPS, model size, and other relevant metrics. It can be seen more intuitively from the figure that the regional graph of the EAC-YOLO model is fuller, indicating that the number of parameters, the computational load, and the model size are smaller. At the same time, its precision, recall, and mAP<sub>50</sub> are higher, resulting in better overall performance.

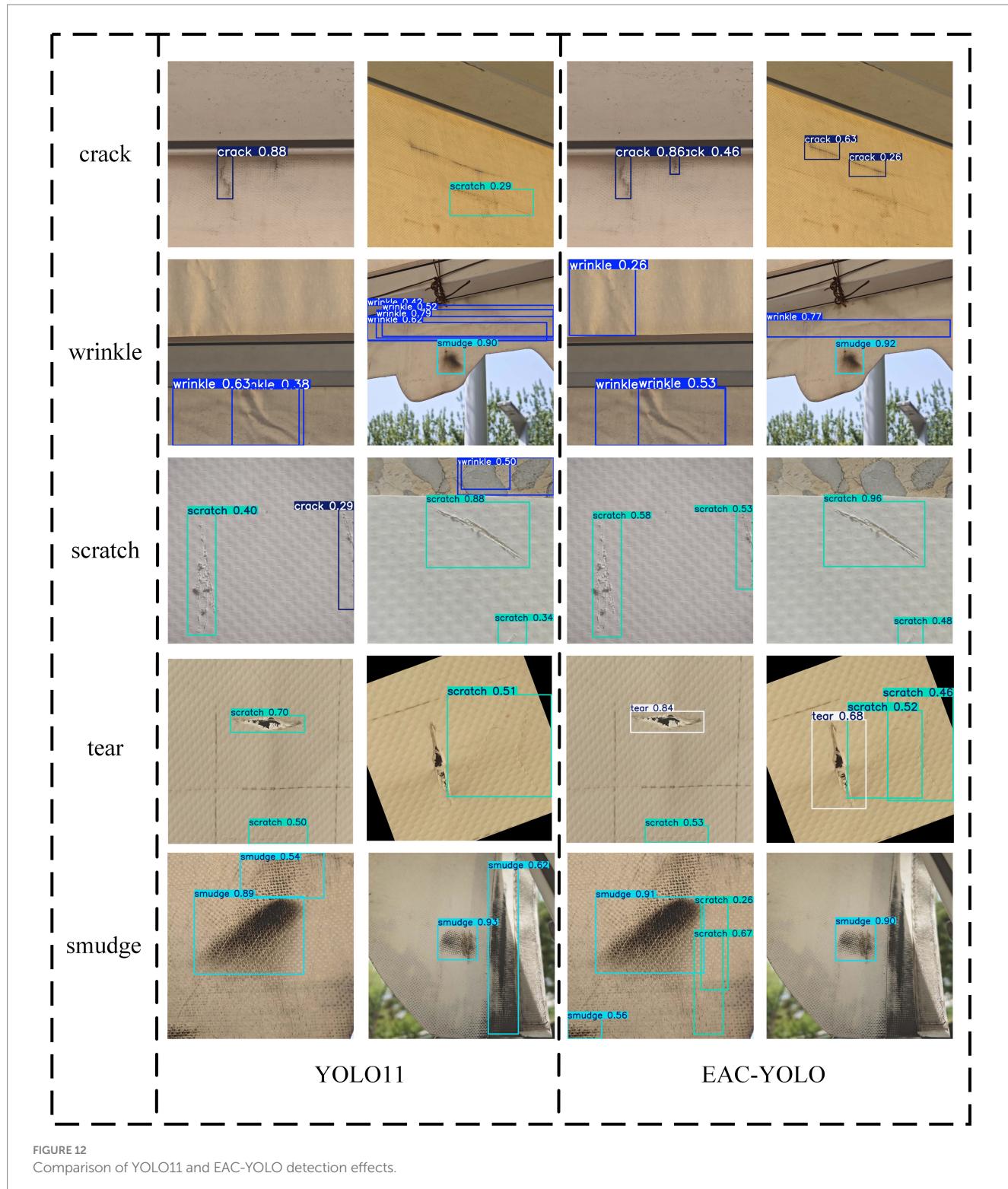


FIGURE 12  
Comparison of YOLO11 and EAC-YOLO detection effects.

To analyze the visual results of the EAC-YOLO model and the comparison model in the surface damage identification of the membrane structure, we used the files trained by each model to evaluate the effect of the test set. From the recognition results, we selected several groups of representative results, as shown in Figure 15. The area marked with a cross in the figure indicates that

there is a false detection in the area, and the area in the dotted box indicates that there is a missed detection. We can see from Figure 15:

- (1) In the results of group (a), most of the models have missed detection of cracks, and the (E) model and (H) model

TABLE 5 Experimental results of the YOLO series on MSSD.

Model	P/%	R/%	mAP <sub>50</sub> /%	mAP <sub>50-95</sub> /%	FLOPs/G	Params/M	Size/MB
YOLOv3-tiny	83.9	76.2	81.0	47.5	14.3	9.52	18.2
YOLOv5s	88.5	79.3	85.7	55.1	18.7	7.82	16.0
YOLOv6s	84.8	77.9	83.5	52.5	42.8	16.00	32.2
YOLOv8s	88.8	81.0	85.6	57.5	28.4	11.12	21.5
YOLOv9s	87.9	75.8	84.0	54.2	30.4	7.83	16.1
YOLOv10s	87.6	79.0	83.9	55.5	24.5	8.03	16.6
YOLO11s	88.2	81.5	86.5	58.0	21.3	9.41	19.2
YOLO12s	88.2	79.1	85.3	55.2	21.2	9.23	19.0
EAC-YOLO	<b>89.4</b>	<b>83.1</b>	<b>87.5</b>	<b>58.4</b>	15.9	<b>6.76</b>	<b>13.9</b>

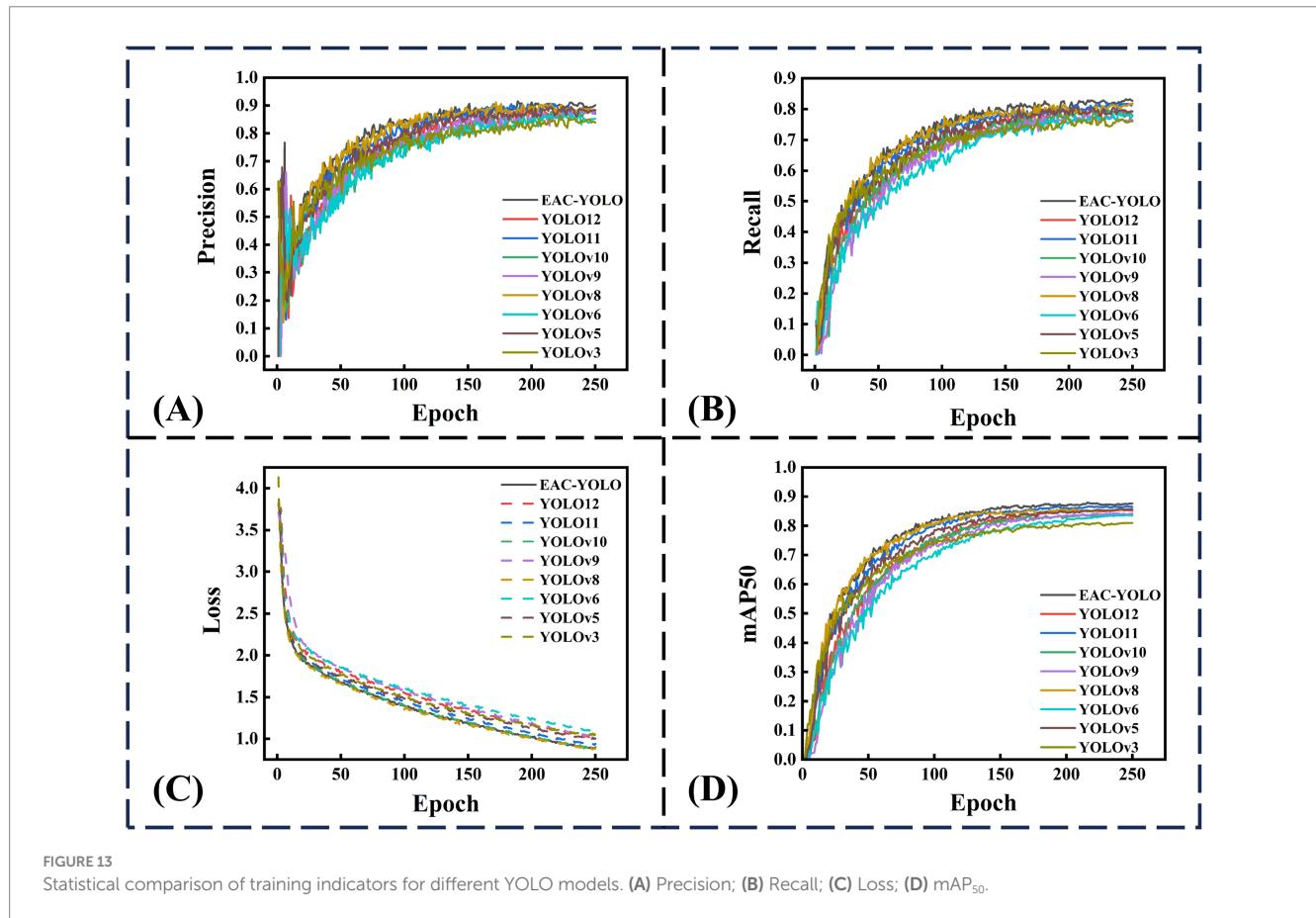


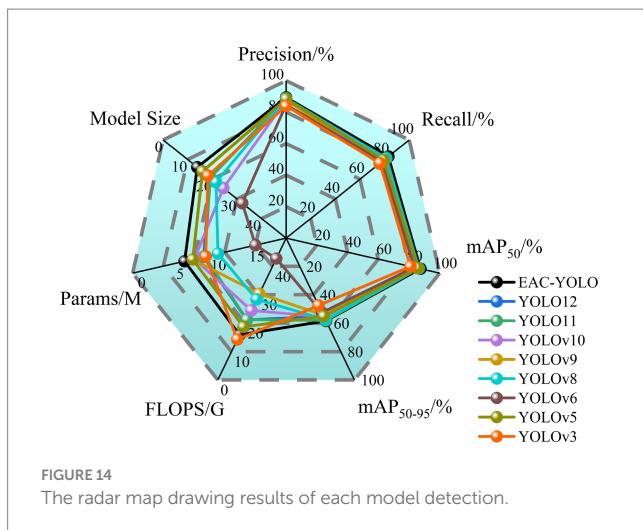
FIGURE 13

Statistical comparison of training indicators for different YOLO models. (A) Precision; (B) Recall; (C) Loss; (D) mAP<sub>50</sub>.

mistakenly damage the environmental objects other than the membrane structure. EAC-YOLO has a more correct identification of crack damage. This is because the crack itself has the characteristics of high shape randomness, multi-scale, small target features, etc., and ordinary feature extraction methods may be difficult to solve these problems in a targeted manner. The pooling method of ADOWN introduced in this paper improves the sensitivity to multi-scale and multi-morphological crack features in the down-sampling calculation process of the model. The problem of missed detection caused by different scales and variable

shapes of partial damage on the surface of the film is solved. At the same time, the model combined with the attention mechanism reduces the damage weight in the picture, reduces the interference effect, and improves the recognition ability of small target features.

- (2) In the results of group (b), the tearing feature target in the image is larger, but the damage influences the background texture interference. In this environment, reducing the surface texture interference of the membrane material can bring higher confidence for recognition. Among the results of all groups, the (I) group has the highest confidence,



which is partly due to our introduction of the CGBlock structure, which simplifies the C3k2 structure while dynamically learning channel weights through group convolution and global average pooling and lightweight fully connected layers, suppresses irrelevant channels, and enhances key feature channels, reducing the negative background effects.

(3) In the results of group (c), we can see that, in addition to the (E) model and (I) model, most of the models have missed the identification of such damage as folds, and the identification of unidirectional folds is mainly due to the following points. (1) This type of damage has a high shape randomness, unidirectional folds, radial folds, and irregular accumulation folds that vary greatly in their manifestations, and the overall recognition confidence of the damage is not high; (2) the contrast between the damage and the material background is low, which is easily to cause missed detection. In this regard, we introduce the dilated convolution in the CGBlock structure to expand the receptive field, and extract the defect features and the background factors of the surrounding membrane surface, thus alleviating the low contrast problem caused by the multi-form wrinkles and poor lighting conditions.

## 5 Conclusion

To achieve the automatic identification of surface damage on membrane structures and address the complexity of current mainstream detection algorithms, which are unsuitable for deployment on resource-constrained devices, we construct a membrane structure surface damage dataset named MSSD and propose a lightweight EAC-YOLO algorithm. Firstly, the ECA mechanism is introduced into the SPPF module and restructured, with almost no change in parameters; the mAP<sub>50</sub> and precision increased to 90.8 and 87.2%, from 88.2 and 86.5%. Through the heatmap test results, it is found that the introduction of this module has a good ability to extract features of small targets, such

as holes, and reduces the interference of complex background textures. Secondly, to enhance the model's multi-scale feature learning ability and further reduce the number of parameters, the ADown down-sampling operator is introduced. Compared to previous work, mAP<sub>50-95</sub> was increased from 57.3 to 59.1%, the number of parameters was reduced by 1.91 M, and FLOPS were decreased by 4.1G, thereby optimizing the model's ability to extract features of multi-scale tears, smudges, etc. Finally, the CGBlock is introduced, and the C3k2 module is restructured. The number of parameters and FLOPS are reduced to 6.76 M and 15.9G, respectively, suppressing the texture background noise channels and enhancing the key information channels. Overall, the precision, recall, and mAP<sub>50</sub> of the EAC-YOLO model reach 89.4, 83.1, and 87.5%, respectively. The number of parameters, computational cost, and model size are reduced by 28, 25, and 28%, respectively. Compared with the YOLO series algorithms, the comprehensive performance of EAC-YOLO is the most outstanding, optimizing the feature extraction ability for multi-type, multi-scale, and low-contrast membrane structure surface damage.

In this paper, a new method for damage identification of membrane structures is proposed, which solves some problems in the process of identification. However, there are still many problems to be solved, such as:

- (1) A certain number of damaged samples of membrane structure are collected. However, due to the limited number of membrane structures in reality and the high maintenance frequency of some membrane structures, the surface damage of many structures is less, and the number of samples collected is limited. At the same time, the surface damage types of the membrane structure require further investigation and enrichment.
- (2) We have made some progress in the automatic identification of membrane structure surface damage. According to the current test results, among all the recognition types, common tears, cracks, and smudges can be accurately identified. This shows that the relevant modules we introduce can learn small target features, multi-morphological and multi-scale features, and low-contrast features. However, for the recognition of scratches and wrinkles, although there is a certain room for improvement, since these two types of damage are often dense when they occur, and the uncertainty of direction and shape is high. The probability of overlap is also high, which presents significant challenges to the labeling work, and the recognition accuracy needs to be improved. The recognition accuracy of these two types of damage needs to be further investigated. For the improvement of model detection accuracy, there are mainly the following aspects:
  - a Add more damage samples to enhance the generalization of the model.
  - b Select a more complex and advanced neural network training model. These models may have a more powerful learning ability but may bring more computation and burden to model training.

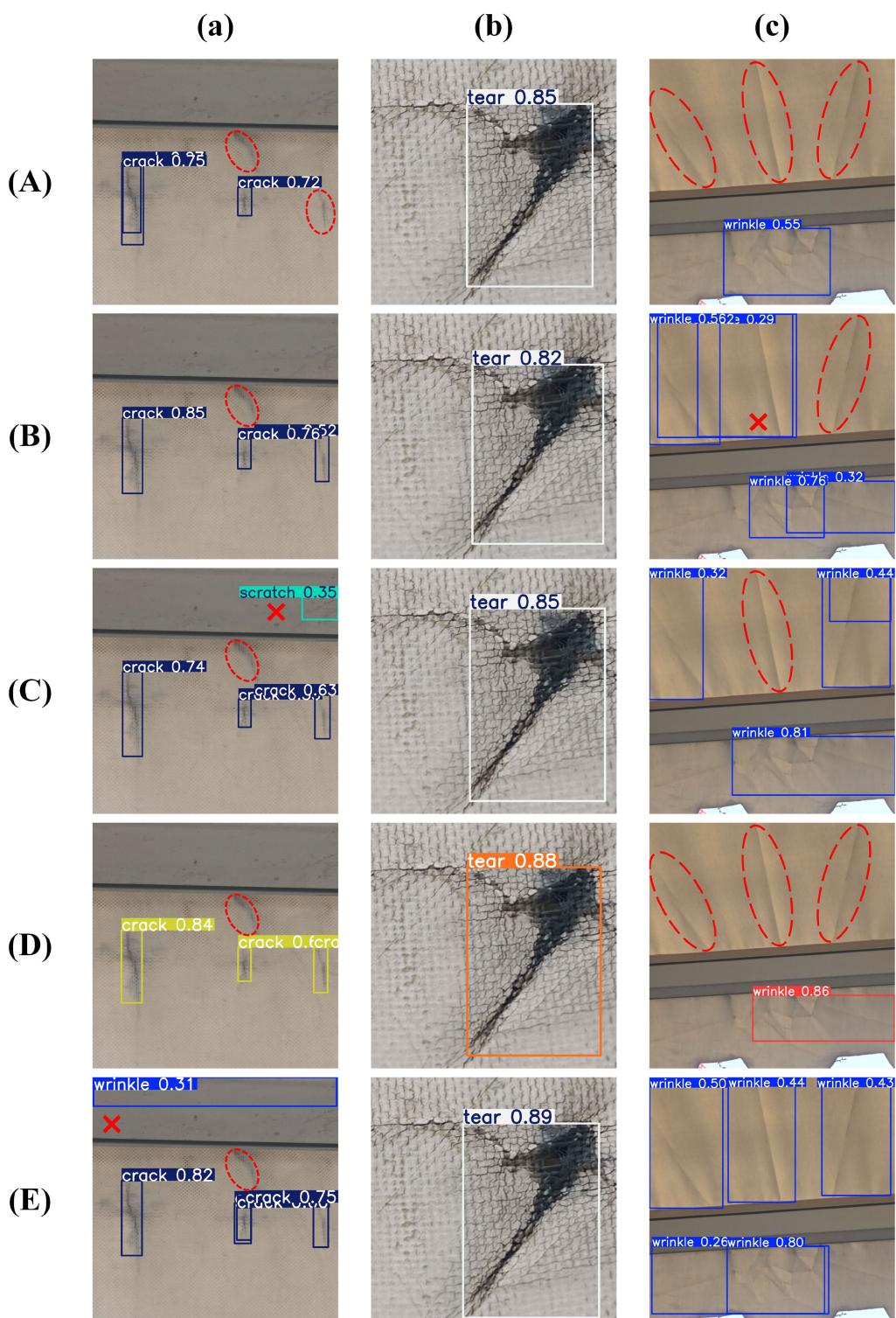


FIGURE 15 (Continued)

c In terms of the lightweight design of the model, the model we proposed reduces the calculation amount, parameter quantity, and model size by about one-third of the volume, respectively. However, from the perspective of the model as a whole, the number of model parameters of YOLO11s

reaches nearly 10 million, and the complexity is not low, which means that there is still a lot of room for improvement in lightweight processing, especially the head part, which is expected to propose better improvement strategies.

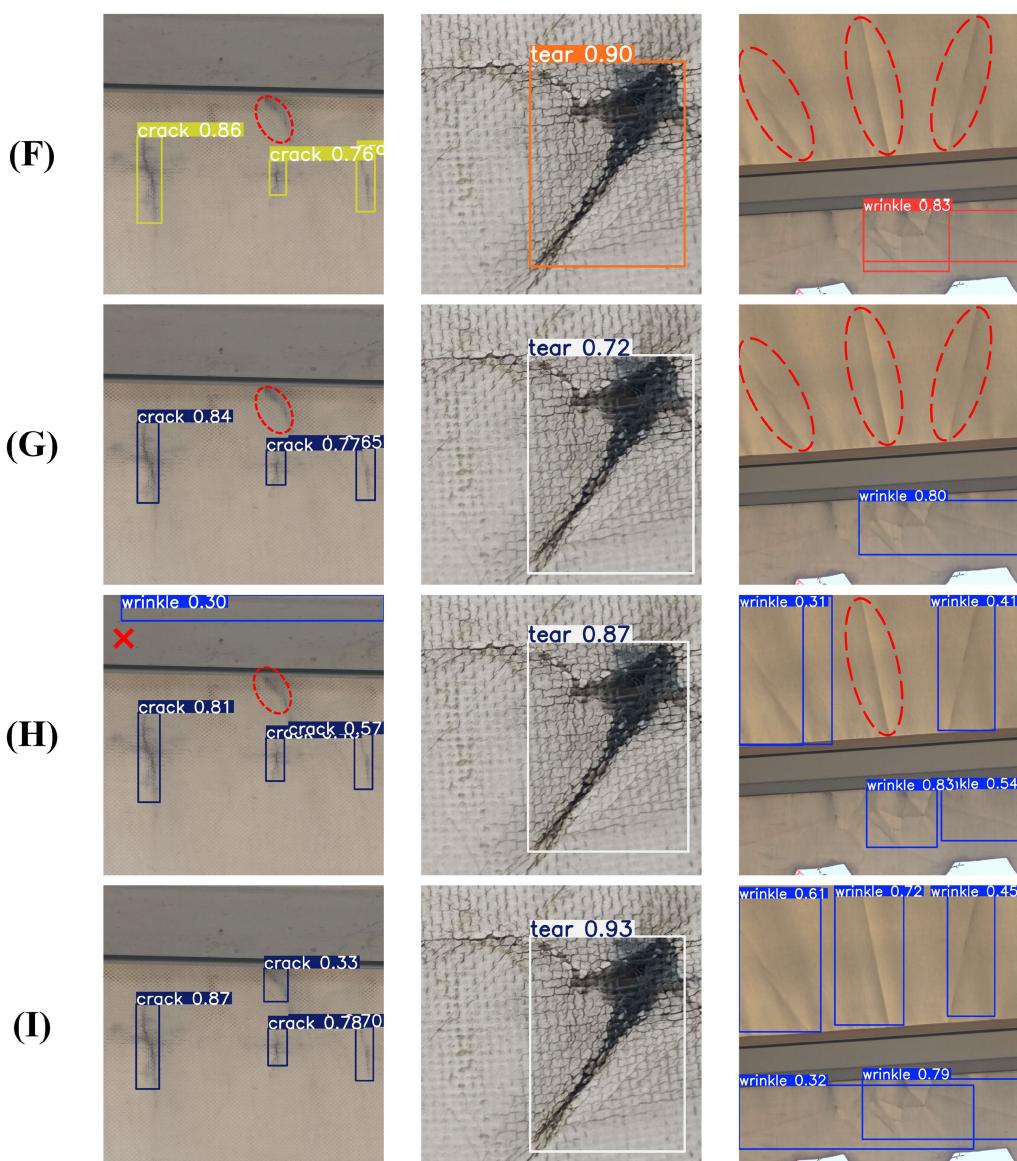


FIGURE 15

Visualization results of damage identification in comparative models. (A) YOLOv3-tiny; (B) YOLOv5s; (C) YOLOv6s; (D) YOLOv8s; (E) YOLOv9s; (F) YOLOv10s; (G) YOLO11s; (H) YOLO12s; (I) EAC-YOLO.

In the future, the proposed method is expected to be deployed on high-definition camera groups or inspection drones surrounding large membrane structures, such as stadiums, which will automatically collect membrane surface image data regularly along preset routes. These data are transmitted in real time over the network to the cloud-based diagnostic center, where the algorithm is deployed. Efficiently complete a comprehensive physical examination of the entire building in a short period of time and automatically generate detailed reports including the type, location, size, and severity of damage. It lays a solid foundation for non-destructive, non-contact, full-field rapid and accurate diagnosis, and has certain significance for the maintenance of membrane structure.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

ZY: Conceptualization, Validation, Data curation, Writing – review & editing, Methodology, Resources, Investigation, Writing – original draft, Software, Formal analysis, Visualization. LZ: Writing – review & editing, Funding acquisition, Project

administration, Validation, Supervision. HL: Data curation, Writing – review & editing. QD: Writing – review & editing, Supervision, Validation. CY: Writing – review & editing, Supervision.

## Funding

The author(s) declared that financial support was received for this work and/or its publication. This work is supported by the National Natural Science Foundation of China under grant No. 52078005.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

CECS (2015). CECS 158–2015, Technical specification for membrane structures. Beijing: China Planning Press, 38–39.

Chen, M., Yu, L., Zhi, C., Sun, R., Zhu, S., Gao, Z., et al. (2022). Improved faster R-CNN for fabric defect detection based on Gabor filter with genetic algorithm optimization. *Comput. Ind.* 134:103551. doi: 10.1016/j.compind.2021.103551

Dong, H., Song, K., He, Y., Xu, J., Yan, Y., and Meng, Q. (2020). PGA-net: pyramid feature fusion and global context attention network for automated surface defect detection. *IEEE Trans. Ind. Inform.* 16, 7448–7458. doi: 10.1109/TII.2019.2958826

Gao, L., Zhang, J., Yang, C., and Zhou, Y. (2022). Cas-VSwin transformer: a variant swin transformer for surface-defect detection. *Comput. Ind.* 140:103689. doi: 10.1016/j.compind.2022.103689

He, Y., Song, K., Meng, Q., and Yan, Y. (2020). An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. *IEEE Trans. Instrum. Meas.* 69, 1493–1504. doi: 10.1109/TIM.2019.2915404

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1904–1916. doi: 10.1109/TPAMI.2015.2389824

Hou, Q., Zhou, D., and Feng, J. (2021). “Coordinate attention for efficient mobile network design,” in *IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 13708–13717.

Hu, J., Shen, L., and Sun, G. (2018). “Squeeze-and-excitation networks.” in *IEEE conference on Computer Vision and Pattern Recognition (CVPR) (IEEE, Salt Lake)*. pp. 7132–7141.

Jocher, G., and Qiu, J. (2024). Ultralytics YOLO11 (version 11.0.0) [computer software]. Available online at: <https://github.com/ultralytics/ultralytics>

Lau, K. W., Po, L.-M., and Rehman, Y. A. U. (2024). Large separable kernel attention: rethinking the large kernel attention design in cnn. *Expert Syst. Appl.* 236:121352. doi: 10.1016/j.eswa.2023.121352

Li, G., Liu, Z., Bai, Z., Lin, W., and Ling, H. (2022). Lightweight salient object detection in optical remote sensing images via feature correlation. *IEEE Trans. Geosci. Remote Sens.* 60, 1–12. doi: 10.1109/TGRS.2022.3145483

Li, W., Zhang, H., Wang, G., Xiong, G., Zhao, M., Li, G., et al. (2023). Deep learning based online metallic surface defect detection method for wire and arc additive manufacturing. *Robot. Comput.-Integr. Manuf.* 80:102470. doi: 10.1016/j.rcim.2022.102470

Lin, J., Yao, Y., Ma, L., and Wang, Y. (2018). Detection of a casting defect tracked by deep convolution neural network. *Int. J. Adv. Manuf. Technol.* 97, 573–581. doi: 10.1007/s00170-018-1894-0

Liu, M., Chen, Y., Xie, J., He, L., and Zhang, Y. (2023). LF-YOLO: a lighter and faster YOLO for weld defect detection of X-ray image. *IEEE Sensors J.* 23, 7430–7439. doi: 10.1109/JSEN.2023.3247006

Liu, S., and Huang, D. (2018). “Receptive field block net for accurate and fast object detection.” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 385–400.

Liu, L., Li, P., Wang, D., and Zhu, S. (2024). A wind turbine damage detection algorithm designed based on YOLOv8. *Appl. Soft Comput.* 154:111364. doi: 10.1016/j.asoc.2024.111364

Luo, J., Yang, Z., Li, S., and Wu, Y. (2021). FPCB surface defect detection: a decoupled two-stage object detection framework. *IEEE Trans. Instrum. Meas.* 70, 1–11. doi: 10.1109/TIM.2021.3092510

Lv, X., Duan, F., Jiang, J., Fu, X., and Gan, L. (2020). Deep metallic surface defect detection: the new benchmark and detection network. *Sensors* 20:1562. doi: 10.3390/s20061562

Ouyang, D., He, S., Zhan, J., Guo, H., Huang, Z., and Luo, M. (2023). “Efficient multi-scale attention module with cross-spatial learning.” in *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.1–5.

Psarommissis, F., Prouvost, S., May, G., and Kiritsis, D. (2020). Product quality improvement policies in industry 4.0: characteristics, enabling factors, barriers, and evolution toward zero defect manufacturing. *Front. Comp. Sci.* 2, 1–15. doi: 10.3389/fcomp.2020.00026

Redmon, J., Divvala, S. K., Girshick, R. B., and Farhadi, A. (2016). “You only look once: unified, real-time object detection.” in *IEEE conference. Computer Vision and Pattern Recognition*. 2015, pp. 779–788.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-cam: visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* 128, 336–359. doi: 10.1007/s11263-019-01228-7

Shen, K., Zhou, X., and Liu, Z. (2024). MINet: multiscale interactive network for real-time salient object detection of strip steel surface defects. *IEEE Trans. Ind. Inform.* 20, 7842–7852. doi: 10.1109/TII.2024.3366221

Tang, B., Kong, J., and Wu, S. (2017). Review of surface defect detection based on machine vision. *J. Image Graph.* 22, 1640–1663.

Tang, J., Liu, S., Zhao, D., Tang, L., Zou, W., and Zheng, B. (2023). PCB-YOLO: an improved detection algorithm of PCB surface defects based on YOLOv5. *Sustainability* 15:5963. doi: 10.3390/su15075963

Tao, X., Zhang, D., Ma, W., Liu, X., and Xu, D. (2018). Automatic metallic surface defect detection and recognition with convolutional neural networks. *Appl. Sci.* 8:1575. doi: 10.3390/app8091575

Wang, X., Gao, J., Hou, B., Wang, Z., Ding, H., and Wang, J. (2022). A lightweight modified YOLOX network using coordinate attention mechanism for PCB surface defect detection. *IEEE Sensors J.* 22, 20910–20920. doi: 10.1109/JSEN.2022.3208580

Wang, Y., Wang, H., and Xin, Z. (2022). Efficient detection model of steel strip surface defects based on YOLO-V7. *IEEE Access* 10, 133936–133944. doi: 10.1109/ACCESS.2022.3230894

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). “ECA-net: efficient channel attention for deep convolutional neural networks.” in *2020 IEEE conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 11534–11542.

Wang, C.Y., Yeh, I.H., Liao, and Hym (2024). “Yolov9: learning what you want to learn using programmable gradient information.” in *European Conference on Computer Vision (ECCV)*. pp. 1–21.

## Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Woo, S., Park, J., Lee, J-Y., and Kweon, IS. (2018). “CBAM: convolutional block attention module.” in *Proceedings of the European Conference on Computer Vision (ECCV)*. 11211, 3–19.

Wu, T., Tang, S., Zhang, R., Cao, J., and Zhang, Y. (2020). CGNet: a light-weight context guided network for semantic segmentation. *IEEE Trans. Image Process.* 30, 1169–1179. doi: 10.1109/TIP.2020.3042065

Xu, L., Dong, S., Wei, H., Ren, Q., Huang, J., and Liu, J. (2023). Defect signal intelligent recognition of weld radiographs based on YOLO V5-IMPROVEMENT. *J. Manuf. Process.* 99, 373–381. doi: 10.1016/j.jmapro.2023.05.058

Xu, Y., Li, D., Xie, Q., Wu, Q., and Wang, J. (2021). Automatic defect detection and segmentation of tunnel surface using modified mask R-CNN. *Measurement* 178:109316. doi: 10.1016/j.measurement.2021.109316

Xue, Y., and Li, Y. (2018). A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects. *Comput. Aided Civ. Infrastruct. Eng.* 33, 638–654. doi: 10.1111/mice.12367

Yang, J., Li, C., Dai, X., and Gao, J. (2022). “Focal modulation networks.” Conference. Neural Information Processing Systems (NeurIPS), 35, 4203–4217.

Yu, Q., Zhang, Y., Xu, J., and Zhao, Y. (2023). Two-stage intelligent detection system of pixel-level classification and damage type recognition for membrane materials. *Mater. Lett.* 347:134645. doi: 10.1016/j.matlet.2023.134645

Yuan, L., Chen, Y., Tang, H., Gao, R., and Wu, W. (2023). DGNet: an adaptive lightweight defect detection model for new energy vehicle battery current collector. *IEEE Sensors J.* 23, 29815–29830. doi: 10.1109/JSEN.2023.3324441

Zeng, N., Wu, P., Wang, Z., Li, H., Liu, W., and Liu, X. (2022). A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection. *IEEE Trans. Instrum. Meas.* 71, 1–14. doi: 10.1109/TIM.2022.3153997

Zhang, D., Hao, X., Wang, D., Qin, C., Zhao, B., Liang, L., et al. (2023). An efficient lightweight convolutional neural network for industrial surface defect detection. *Artif. Intell. Rev.* 56, 10651–10677. doi: 10.1007/s10462-023-10438-y

Zhang, Y., Zhang, H., Huang, Q., Han, Y., and Zhao, M. (2024). DsP-YOLO: an anchor-free network with DsPAN for small object detection of multiscale defects. *Expert Syst. Appl.* 241:122669. doi: 10.1016/j.eswa.2023.122669

Zhao, W., Chen, F., Huang, H., Li, D., and Cheng, W. (2021). A new steel defect detection algorithm based on deep learning. *Comp. Intel. Neurosci.* 202:5592878-13. doi: 10.1155/2021/5592878

Zhao, H., Gao, Y., and Deng, W. (2024). Defect detection using shuffle net-CA-SSD lightweight network for turbine blades. *IEEE Internet Things J.* 11, 32804–32812. doi: 10.1109/JIOT.2024.3409823

Zhou, Z., Zhang, J., and Gong, C. (2022). Automatic detection method of tunnel lining multi-defects via an enhanced you only look once network. *Comput. Aided Civ. Inf. Eng.* 37, 762–780. doi: 10.1111/mice.12836