Received: 25 May 2022 | Revised: 5 July 2022 | Accepted: 22 July 2022

DOI: 10.1049/ipr2.12602

IET Image Processing

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

A method of underwater bridge structure damage detection method based on a lightweight deep convolutional network

Xiaofei Li | Heming Sun | Taiyi Song | Tian Zhang | Qinghang Meng

College of Transportation Engineering, Dalian Maritime University, Dalian, China

Correspondence
Xiaofei Li, College of Transportation Engineering, Dalian Maritime University, Dalian 116026, China.
Email: lixiaofei@dlmu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 51608087

Abstract

The problem of the underwater structure disease of the bridge is increasingly obvious, which has seriously affected the safe operation of the bridge structure, so it is necessary to detect the underwater structure regularly. There are many kinds of bridge underwater structure diseases. This paper targets the bridge underwater structural crack diseases adopts multiple image recognition networks for verification, compares the advantages of different networks, and takes the YOLO-v4 network as the main body to build a lightweight convolutional neural network.Mobilenetv3 replaced CSPDarkent as the backbone feature extraction network, while the feature layer scale of Mobilenetv3 was modified, and the extracted preliminary feature layer was input into the enhanced feature extraction network for feature fusion. The PANet networks are replaced by the depthwise separable convolution. Using ablation experiments to compare the performance of four algorithm combinations in lightweight networks. At the same time, the disease identification accuracy of each network and the performance of the network are tested in various experimental environments, and the feasibility of the lightweight network is verified in the application of bridge underwater structure damage identification.

1 | INTRODUCTION

With the continuous enhancement of China's economic strength and the significant improvement of science and technology levels, water-related projects have developed rapidly. Bridge construction is developing rapidly in the direction of longer spans, deeper foundations, and higher bridge towers [1]. Among them, the construction of deep-water Bridges, such as the Hong Kong-Zhuhai-Macao Bridge [2], continues to develop, but the complexity of the underwater environment [3] makes the status of underwater structure detection of Bridges [4] more and more important. At present, there are few cases of bridge underwater structure detection, and the technical scheme preparation is insufficient. The conventional bridge underwater detection method is to visually detect underwater diseases by professional divers and record the disease size information, but this method is susceptible to the subjective influence of divers when measuring disease information. In the complicated underwater environment, divers have some hidden dangers [5]. Trying to use underwater robots instead of divers to identify and detect underwater bridge structural diseases. Therefore, it is necessary to build a lightweight identification network, and underwater robots can replace divers to identify and detect underwater structural diseases of Bridges. This paper compares the two commonly used detection methods. The advantages and disadvantages of the methods are listed in Table 1.

Underwater structure damage identification of bridges has always been a difficult point. Underwater environment complexity restricts the effectiveness of bridge structure damage identification, and an appropriate structure detection method is particularly important. The non-destructive testing (NDT) method plays an important role in bridge damage identification. Cerro and Ferrigno discuss the method and scope of the NDT [6].The most commonly used NDT methods include the following six: Ultrasonic pulse speed (UPV) method [7, 8], acoustic emission, ray inspection, infrared thermal imaging, sensor monitoring, and ground-detection radar (GPR) [9]. As an important

TABLE 1 Advantages and disadvantages of bridge detection methods

Detection method	Advantages	Disadvantages
Underwater visual inspection method	Carry out multiple inspections: foundation scour inspection, aquatic life inspection, riverbed measurement, bridge crack inspection	Subject to diver's subjective influence when measuring disease information.
Sonar technology detection	Large-scale detection, overall imaging, and obvious effect on large-scale diseases	It requires multiple people in the team to cooperate with the erection of the instrument, and the detection accuracy of minor bridge diseases is relatively low.
Underwater robot detection	Free and flexible, able to accurately identify and detect underwater diseases of bridges	It is greatly affected by the underwater environment, such as water turbidity, water flow speed

method of underwater non-destructive testing and identification, underwater detection and imaging technology has attracted more and more attention, and many countries are developing research on turbidity water detection and imaging technology. Bin et al. [10] used sonar technology to detect the underwater foundation damage of ancient Chinese stone arch Bridges, used a multi-beam echo sounder [11] to measure the terrain under the bridge, scanned the sediment under the bridge and submerged obstacles on both sides, and investigated the exposed state of wooden piles and the appearance of strip stone foundation. Combined with the results of multi-beam sounding, the erosion of the underwater foundation is analyzed. The underwater three-dimensional sonar imaging [12] is used to scan the underwater foundation and detect its state of the underwater foundation. Due to the high use conditions of instruments and equipment, the sonar technology and methods have certain limitations, the identification accuracy of bridge underwater structure cracks is low, and the application requirements are relatively high.

With the rapid development of computer technology, deep learning algorithms are constantly changing, making breakthroughs in image recognition, natural language processing, and data mining, and have gradually been applied to medical imaging [13], proteomics [14], physics [15] and other professional fields. The deep learning algorithm is also been gradually introduced into the professional field of bridge detection [16]. A deep learning algorithm introduces bridge disease detection to improve the efficiency of bridge disease detection to some extent. Li et al. [17] proposed an end-to-end SSENets model for accurately detecting bridge cracks, and applied the deep learning-based target detection method to bridge disease detection; Zhu et al. [18] proposed to use convolutional neural network for bridge disease detection. Gao [19] used Resnet architecture to identify concrete cracks on the bridge deck, with an accuracy of 93%. To obtain information on the length, width, and area of the disease, Ruan [20] first divided the disease image by semantic segmentation and then obtained the 2D information of the disease based on the image processing technology. The identification of crack width and length parameters on the dam surface was studied by Chen [21]. Ying [22] applied the UNet network with residual module to bridge crack detection, proposed a new method for bridge crack recognition, and proposed a measurement method of crack length and width in combination with the digital image. To sum up, many scholars have studied the application of deep learning in concrete disease detection [23], but there are few kinds of research on underwater structural damage. Meanwhile, the crack recognition algorithm of a convolutional neural network generally has problems such as complex network structure and too many training parameters.

Compared with previous generations of networks, Mosaic data enhancement, and Self-Adversarial Training(SAT) technology are added to the YOLO-v4 network [24]. Mosaic is a new data enhancement method, which blends four training images to enhance the robustness of the model. Cao et al. [25] applied the YOLO-v4 network to the real-time object detection of masks in the night environment, and achieved a relatively good detection effect, indicating that the YOLO-v4 network can be applied to the environment of weak underwater light. Hu et al. [26] modified the connection method of the feature pyramid network (FPN)+path aggregation network (PANET), and replaced the characteristic diagram of large-scale information in the original YOLO-V4 network with a finer-grained YOLO feature map. The modified network works well in identifying tiny objects in uneaten feed pellets underwater. Wang et al.[27] build a Triden-YOLO v4 efficient object detection network based on YOLO v4, which is designed for mobile devices with limited computing power.

Here, based on the YOLO-v4 neural network algorithm, Mobilenetv3 [28] is used to replace CSPDarkent as the backbone, and the ordinary convolution in the PANet network is replaced by 3 × 3 depthwise separable convolution, and the prior box is improved. Due to the limited computing capability of the underwater robot, it is possible to use the lightweight network to identify the damage to the bridge structure with a certain accuracy. The data set in this experiment is mainly derived from the materials provided by a bridge underwater inspection company and collected from the Internet. A total of 8780 pictures have been collected after processing. It can be divided into three types according to the different underwater fracture environments: clean water environment, muddy water environment, and deepwater environment, and verify the identification accuracy of the lightweight network under the condition of small cracks. At the same time, the performance of various networks under different image acquisition angles is verified.

2 | YOLO-V4 NETWORK IMPROVEMENT

2.1 | YOLO-v4 convolutional neural network

Currently, there are two types of commonly used target detection algorithms. One is the target detection networks based on regional recommendations, such as Mask R-CNN, and Faster R-CNN [29]; the other is the YOLO series based on the regression target detection network [30], which has higher detection accuracy and faster real-time detection speed compared with the first type of target detection algorithm.

The core idea of the YOLO series is to solve the target detection as a regression problem, and use an end-to-end network to input the target image into the model, which outputs the type of the target and marks the position of the object in the image [31]. YOLO-v4 is the fourth generation of the YOLO algorithm, which has a great improvement in detection accuracy and speed.

YOLO-v4 backbone feature extraction network is improved based on the YOLOv3 backbone(darknet-53) and proposes a CSPdarknet-53 feature extraction network. CSPnet divides darknet residual blocks into two parts, one of which continues to stack residual blocks as the backbone, and the other part is directly connected after simple processing. This improved method reduces the amount of network computation and avoids the problem of gradient disappearance

YOLO-v4 uses SPP and PANet structures as feature fusion networks, and the SPP structure maximizes the pooling of feature maps, converts them into feature maps of different scales, and then enters them into the PANet network to stitch together with the original feature map. This part is to upsample and downsample the three feature layers extracted through the backbone feature extraction network to obtain three optimized feature layers with more generalization.

The prediction network outputs three feature graphs, which are respectively used to detect the large target, medium target, and small target. Each point in the feature graph has three prediction boxes, and the offset, width, and height of the prediction box are set, as well as the type and position of the final output target.

2.2 | YOLO-v4 network improvement

Compared with other algorithms, the YOLO-v4 network has higher detection accuracy and speed, but the network model uses a large number of residual structures, which needs to calculate a large number of parameters in the process of image feature extraction, which needs to rely on powerful GPU computing resources. The research content of this paper is to use underwater robots to detect bridge underwater structural diseases. The underwater robot is shown in Figure 1. Due to the limited computing resources of underwater robots, it is difficult to embed the YOLO-v4 network. Therefore, it is of great significance to study a set of lightweight network models for underwater detection of Bridges.

2.2.1 | Depthwise separable convolution

The core idea of the Mobilenet network is to replace ordinary convolution with deep separable convolution, which greatly reduces the number of parameters and computation of the model and makes the model meet the real-time requirements in mobile devices. Depthwise separable convolution divides ordinary standard convolution into deep convolution and point-by-point convolution. Standard convolution uses a convolution kernel with the same number of input image channels to carry out convolution operations. Deep convolution uses different convolutions to check each channel of the input image for convolution operation. Point-by-point convolution combines the feature graph obtained by deep convolution with 1 × 1 convolution kernel to obtain a new feature graph, reducing the number of parameters and calculation of the model, and achieving the same effect as standard convolution.

2.2.2 | Inverted residual

In the traditional residual network structure, the input features are dimensionally reduced by 1 × 1 convolution, and the general dimension is reduced to 1/4. Then, the 3 × 3 standard convolution is used to conduct convolution operation on the dimensionally reduced features, and finally, the convolution is used to achieve dimensionally increased, that is, there are more channels at both ends and fewer channels in the middle. In this way, feature extraction by reducing the dimension reduces the amount of calculation and improves the calculation speed, but it cannot extract enough overall feature information in the low dimension. To extract enough feature information from the residual network, an inverted residual structure is proposed, which is improved as follows.

1. The 1 × 1 convolution dimension raising is used for the input feature Layer, and BM (Batch Normalization Layer) is used to normalize the obtained features, which is conducive to the accelerated convergence of the network and the improvement of the generalization performance of the network. ReLu6 with better performance is used as the activation function.
2. Use 3 × 3 depthwise separable convolution instead of 3 × 3 standard convolution. Here, depthwise separable convolution is used to reduce the amount of computation. Set the dimension-raising hyperparameter to 6, and change the number of channels in this layer to six times the original, so that enough feature information can be extracted in high dimensions.
3. Using 1 × 1 convolution to reduce the dimension, the nonlinear ReLu6 activation function will cause the loss of feature information in the low-dimensional space, and the linear activation function is used instead.

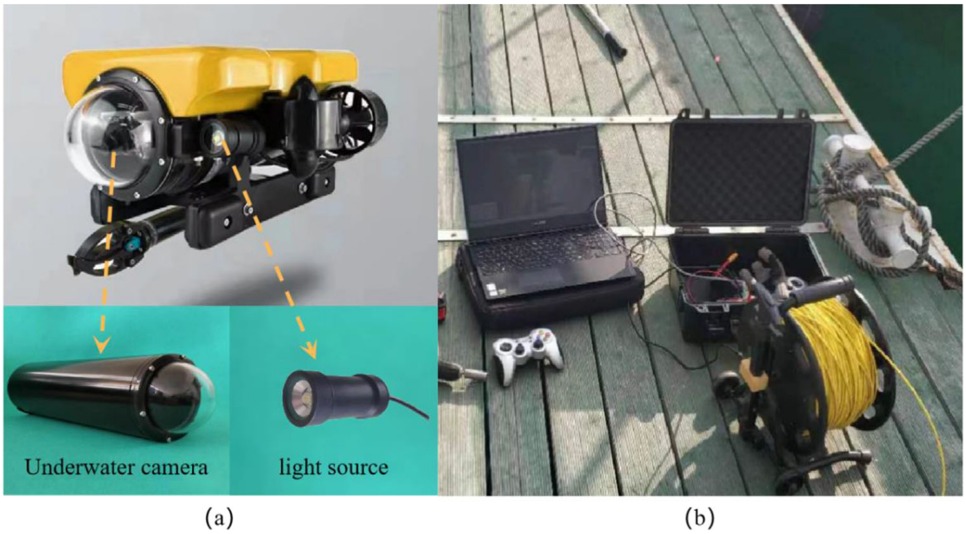


FIGURE 1 Data collection based on underwater robots: (a) Underwater robots. (b) Laptop computers used for monitoring

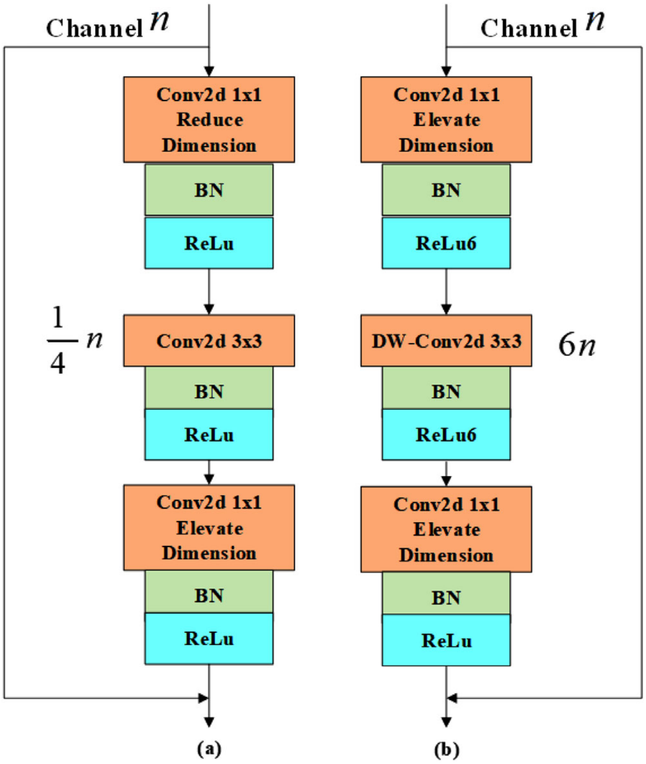

Channel n
Conv2d 1x1 Reduce Dimension
BN
ReLu
$\frac{1}{4}n$
Conv2d 3x3
BN
ReLu
Conv2d 1x1 Elevate Dimension
BN
ReLu
(a)
Channel n
Conv2d 1x1 Elevate Dimension
BN
ReLu6
DW-Conv2d 3x3
$6n$
BN
ReLu6
Conv2d 1x1 Elevate Dimension
BN
ReLu
(b)

FIGURE 2 Residual structure :(a) Ordinary residual structure. (b) Improved inverted residual structure

The residual structure and inverted residual structure are shown in Figure 2 below:

2.2.3 | Introduce the SE attention mechanism

The traditional network model considers that each channel of the feature layer has information of the same importance. Some channels contain less information, and each channel still needs to be calculated, which increases the calculation amount of the network. To make the performance of the network better, this improved network introduces a lightweight attention mechanism SE module, which is placed after the depthwise convolution in the inverted residual structure.

The SE structure diagram is shown in Figure 3. It mainly includes two parts: Squeeze and Excitation. The importance of each feature channel can be obtained through compression and excitation, and weight is given to each channel according to its importance. Make the network model pay attention to certain channels in a targeted manner, limit useless feature channels, and maximize network performance.

In Figure 2, the size of a feature layer after depthwise separable convolution is F × F × C. The traditional structure directly reduces its dimension and output. This improvement is based on the traditional network and introduces the attention mechanism through the following steps.

1. The first is the compression operation: a global average pooling is used to compress the feature channel F × F × C into 1 × 1 × C, and all the feature values on each channel are compressed into one value, which is calculated from all the feature values and has a global receptive field.
2. Next, perform the excitation operation: assign each feature channel a different weight value. The excitation part consists of two fully-connected layers. The first fully-connected layer has C × SERadio neurons, where SERadio represents the scaling parameter. The purpose of setting this parameter is to reduce the feature channel and thus reduce the amount of calculation. The input is compressed to get 1 × 1 × C features, the output is 1 × 1 × C × SERadio; the second fully connected layer has C neurons, the input is 1 × 1 × C × SERadio, and the output is 1 × 1 × C.
3. Finally, the scaling operation is carried out, and the weight value of each feature channel obtained in the previous step

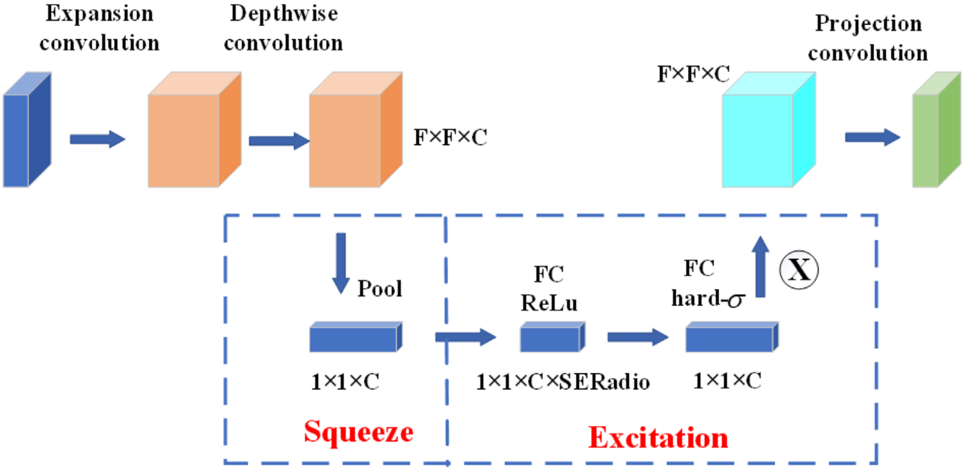


FIGURE 3 Squeeze-and-Excitation (SE) structure diagram

is multiplied by the corresponding feature channel of the original feature graph, and finally, the feature layer with the attention mechanism is obtained.

2.2.4 | Improvement of backbone feature extraction network

The CSPDarkent53 backbone feature extraction network has high detection accuracy, The number of parameters and the numerical calculations is large, to meet the low computing power of mobile devices. This paper uses the lightweight Mobilenetv3 instead of CSPDarkent53 as the backbone feature extraction network of YOLO-v4.

Since the input features of the subsequent network are fixed values, which do not match the size of the feature graph in the middle of the original Mobilenetv3 network, it cannot be directly used for replacement. Therefore, the Mobilenetv3 network needs to be improved. Layers 7, 13, and 17 in the Mobilenetv3 network were selected for extraction to replace the three effective feature layers in the original YOLO-v4 network. Modules after layer 17 were deleted. Point convolution operation was performed on the extracted three-layer modules to change the feature dimension to match the subsequent network.

2.2.5 | Improve the prior box

After PANet network feature fusion, the three output feature layers are 52 × 52, 26 × 26, and 13 × 13, which are used to detect small objects, medium objects, and large objects, respectively. The target detection system studied here is bridge underwater crack detection. Due to the complex underwater environment, long-distance shooting cannot display specific information due to the turbidity of the water body. It is too close to the underwater crack and cannot perceive the overall picture. Therefore, appropriate range detection is needed for underwater vehicles. Under the condition of a certain distance, the underwater cracks photographed have different shapes, but there is no significant difference in the proportion of the images. In addition, there is a single target type. With the original prediction method, different cracks may be divided into different layers for detection, and the detection scale may not match the perceived field of view. Most of the cracks need to be detected by the large output layer, which will cause the medium output layer and the small output layer to be unable to be trained, resulting in a decrease in the detection accuracy of the model.

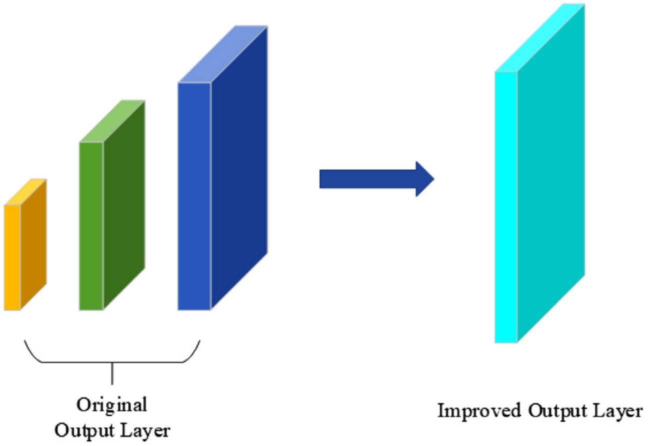


FIGURE 4 Output layer merge

To avoid the above problems, this paper improved on the original basis by fusing the three feature output layers into one output layer. The specific improvement method is shown in Figure 4 below. After the improvement, all cracks were detected by multi-feature fusion in the large output layer.

2.2.6 | Improve and strengthen the feature extraction network

PANet repeatedly samples the input effective feature layer to obtain more effective feature information. Since PANet adopts

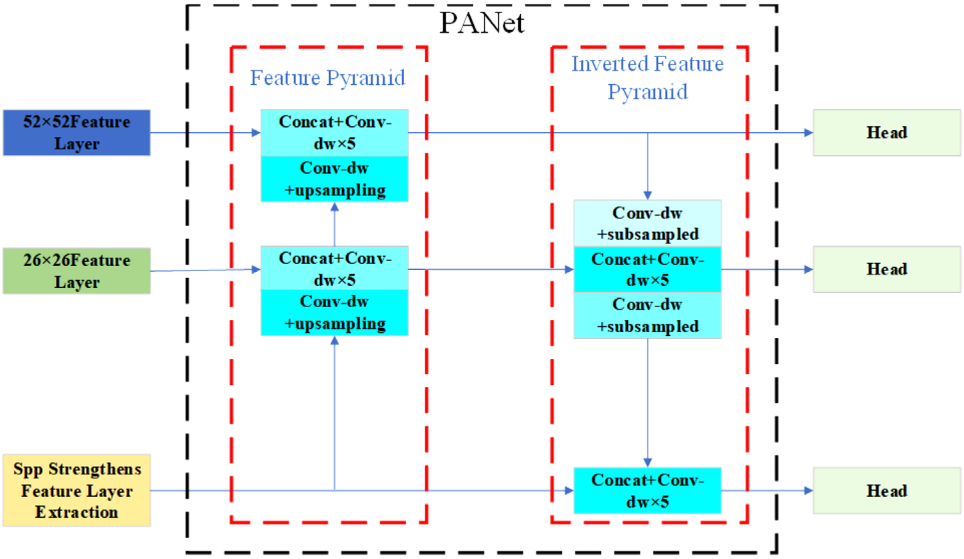


FIGURE 5 Optimizing the PANet structure

TABLE 2 Comparison of the network parameters

Network model	Number of parameters	Model size/MB
YOLO-v4	6,404,0001	244
Mobilenetv3-YOLO-v4	40,043,389	152
Lite-YOLO-v4	11,791,741	44.3
YOLO-v5l	47,056,765	179
YOLO-v5m	213,756,45	81.54
YOLO-v5s	7,276,605	27.76

a large number of 3 × 3 convolutions, the network computation is heavy. This paper uses depthwise separable convolution instead of general convolution to build a lightweight PANet network. The optimized structure is shown in Figure 5, where Conv-dw indicates that depthwise separable convolution is introduced.

When the improved PANet network splices and fuses the three input feature layers, it uses depthwise separable convolution to complete the upsampling and downsampling of the feature map. Finally, it inputs the fused feature information into the prediction network. The optimized PANet network significantly reduces the computation and improves the detection speed.

To verify whether the computational complexity of the improved network is reduced, the network parameters are compared with other networks in this paper. The specific situation is shown in Table 2. YOLO-v4 represents the original network; Mobilenetv3-YOLO-v4 replaces CSPdarkent53 with Mobilenetv3 as the backbone feature extraction network. Finally, Lite-YOLO-v4 is the network studied in this paper. Lite-YOLO-v4 represents an improved backbone feature extraction network and an enhanced feature extraction network for mobile devices.

Using Mobilenetv3 as the backbone feature extraction network can reduce the number of network model parameters to 62.5% of the original and the model size to 62.3%; On this basis, after PANet is optimized, the model parameters are reduced to 18.4% of the original, and the model size is the original YOLO-v4 of 18.2%. The number of parameters of the network here is 25.06% of YOLO-v5l, 55.16% of YOLO-v5m, and YOLO-v5s is 38.29% less than this network. It can be seen that the network studied in this paper can significantly reduce the amount of computation and meet the needs of devices with different configurations.

2.2.7 | Improved overall network model

In summary, based on YOLO-v4, this paper studies a lightweight convolutional neural network Lite-YOLO-v4, which can be used in mobile devices. The following specific improvements are made based on the original YOLO-v4:

1. Mobilenetv3 replaces CSPDarkent as the backbone feature extraction network and modifies the feature layer scale of Mobilenetv3 to connect it with the subsequent network. The extracted preliminary feature layer is input to the enhanced feature extraction network for feature fusion.
2. A large number of 3×3 ordinary convolutions are used in the PANet network. To reduce the amount of computation, this paper replaces ordinary convolutions with 3 × 3 depthwise separable convolutions.
3. The prior box is improved. The original three feature output layers are changed into one output layer.

The structure of the improved YOLO-v4 network model, namely the lightweight network model Lite-YOLO-V4, is shown in Figure 6 below:

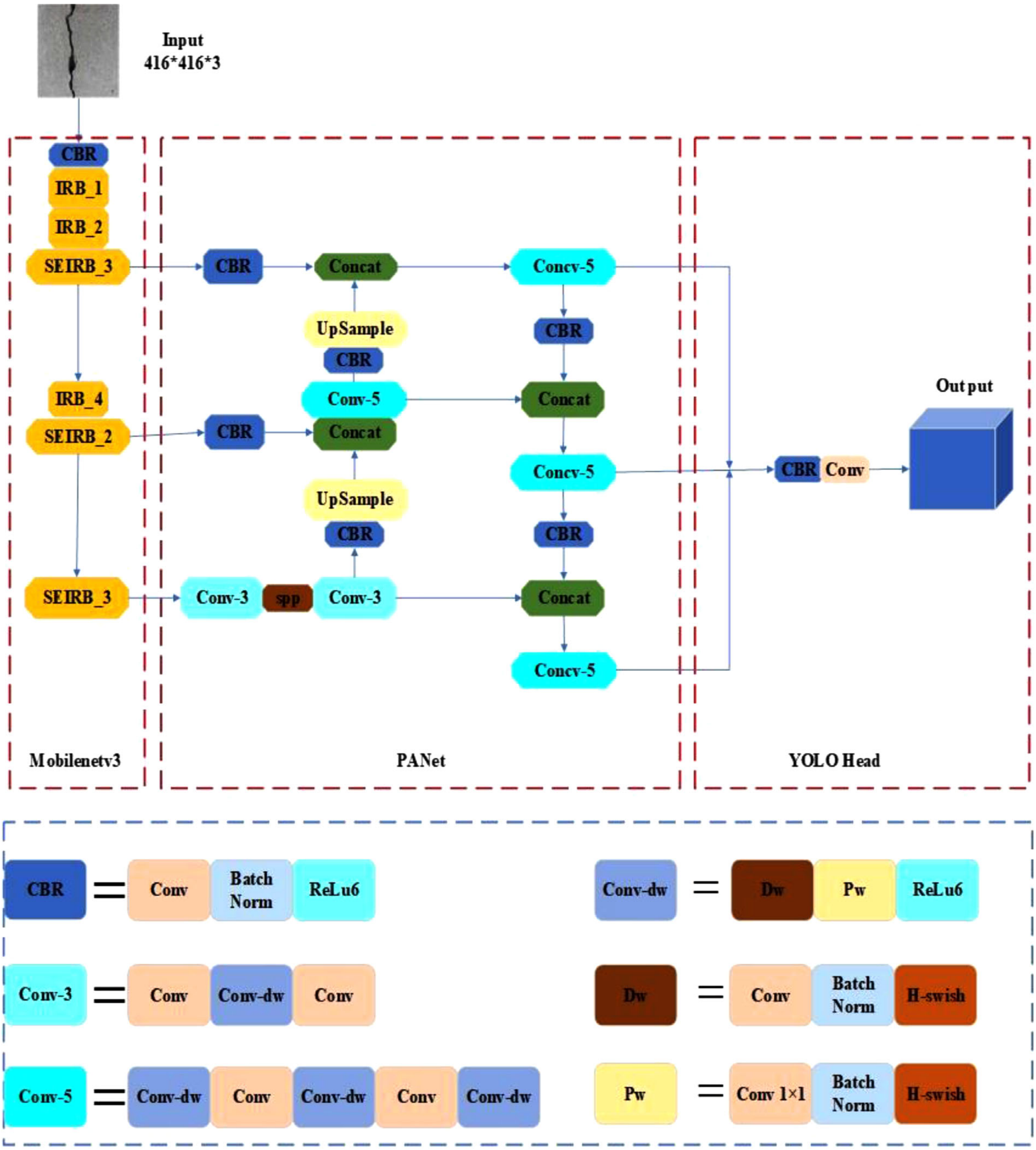


FIGURE 6 Structure of the Lite-YOLO-v4 crack detection algorithm

The IRB_X and SEIRB_X in above Figure 6 represent the inverted residual blocks and inverted residual blocks introducing Squeeze-and-Excitation Networks, respectively. The convolution step size is divided into two structures; When the convolution step size is 1, the structure of IRB_X and SEIRB_X is shown in Figure 7. The input information in the two structures will be added directly. When the convolution step size is 2, the structure of IRB_X and SEIRB_X is shown in Figure 8. The input information runs according to the structural process.

2.3 | Training techniques

2.3.1 | Mosaic

Mosaic data augmentation helps increase the richness of the samples in a similar way to CutMix data augmentation. The difference is that CutMix only selects two images for operation at a time. Mosaic selects four images at a time, and inputs the results after rotation and splicing them into the convolutional neural network for training.

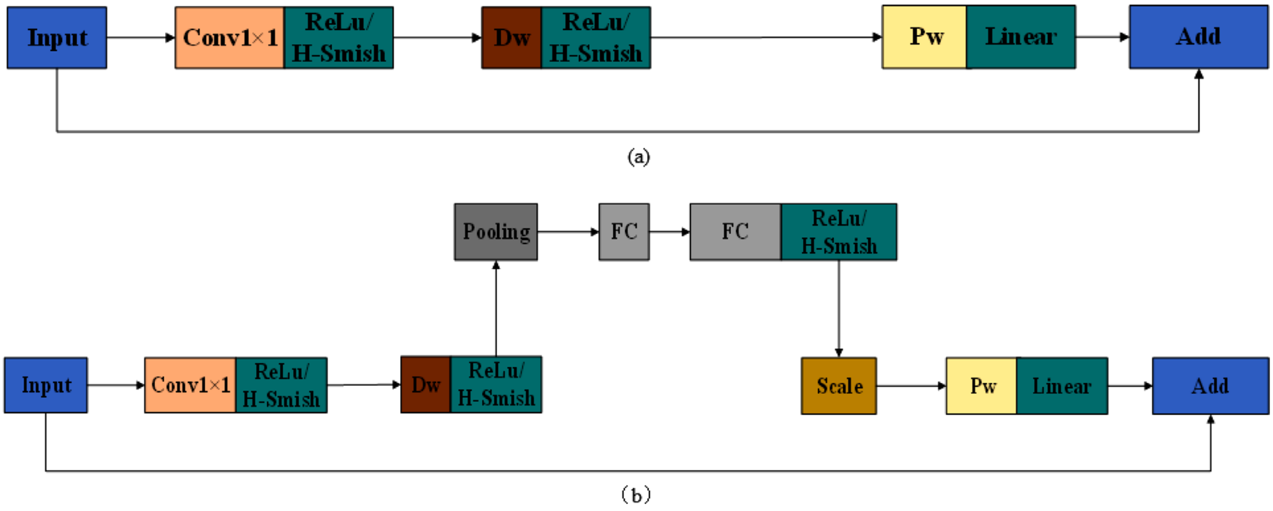


FIGURE 7 Inverse residual structure of stride = 1. (a) Ordinary inverse residual blocks (IRB). (b) SE inverted residual blocks (SEIRB)

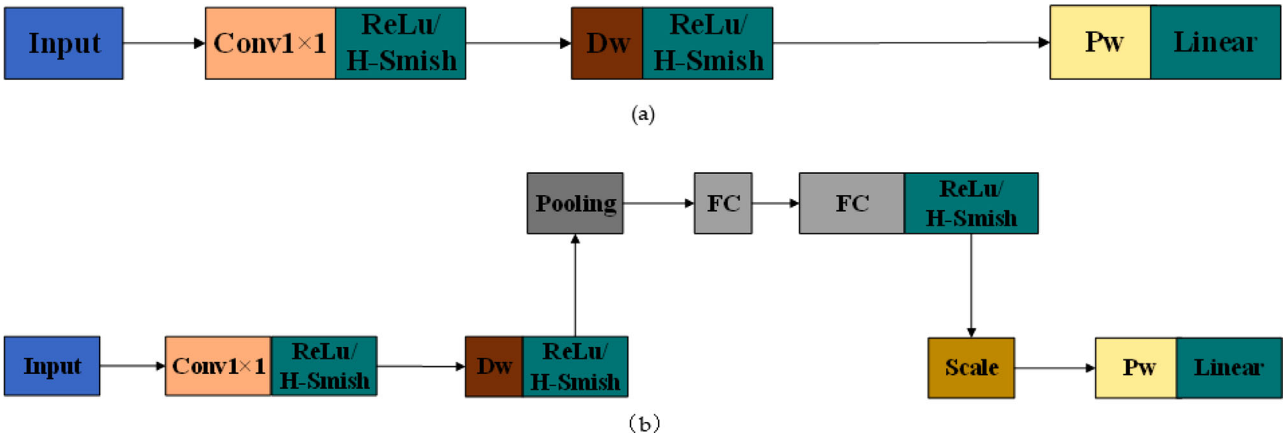


FIGURE 8 Inverse residual structure of stride = 2. (a) Ordinary inverse residual blocks (IRB). (b) SE inverted residual blocks (SEIRB)

2.3.2 | Cosine annealing learning rate

Deep learning uses a gradient descent algorithm to make the loss value of the model constantly close to the global minimum value and finally make the model converge. In model training, proper adjustment of the learning rate can accelerate model convergence and avoid oscillation near the minimum value. The original learning rate adjustment method easily falls into the local optimal solution. Hutter et al. [19] proposed a Stochastic Gradient Descent algorithm with Warm Restarts. There are many locally optimal solutions in the Gradient Descent process, and the algorithm periodically adjusts the learning rate. The local optimal solution can be jumped out. The optimal global solution can be approached by first slowing down, accelerating, then slowly decreasing, and then returning to the initial value when the decay is reduced to 0. This paper uses the cosine annealing attenuation algorithm provided by the TensorFlow framework, and the minimum learning rate is set as 10–5, changing once every ten epochs. The curve of the learning rate is shown in Figure 9 below:

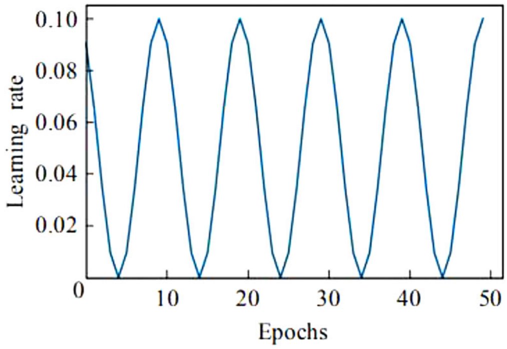


FIGURE 9 Learning rate change curve during training

2.3.3 | Transfer learning

Transfer learning applies the weight of a trained network to an untrained network, most of the data have a specific correlation, and the trained model has a strong generalization ability. Therefore, transfer learning can share the learned feature extraction

TABLE 3 Training parameter settings for the network model

Parameter name	Parameter size
Training	7024
Test	1756
Input	416 × 416
Epoch	100
Freeze-epoch	50
Freeze-batch size	4
Freeze-learning rate	1×10^{-3}
Unfreeze-epoch	50
Unfreeze-batch size	2
Unfreeze-learning rate	1×10^{-4}

ability with the untrained model by setting pre-training weights to speed up its training. For network models trained with small data sets, the feature extraction ability of pre-trained models trained on large data sets can be learned by using transfer learning to prevent over-fitting. To speed up the model training speed, the weight of Mobilenetv3 trained on the VOC data set is applied to the underwater crack detection of bridges in the method of transfer learning.

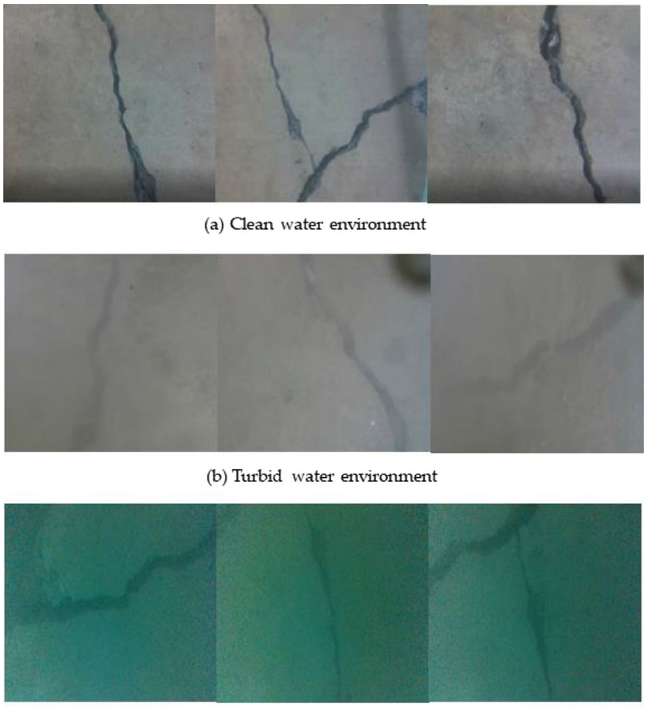
3 | MODEL TRAINING AND RESULT ANALYSIS

3.1 | Experimental environment

Bridge underwater crack detection model experiments using the TensorFlow framework to build the network. The processor running deep learning is Intel Xeon E3-1230, the NVIDIA Quadro K1200 model GPU with a video memory of 4GB, and using the deep learning platform TensorFlow-GPU = 1.13.2, Keras = 2.1.5.

The data set in this experiment mainly comes from materials provided by a bridge underwater inspection company and collected from the Internet, including Crack 500. After processing, a total of 8780 pictures were collected. The different underwater crack environments are mainly divided into the clean water environment, muddy water environment, and deepwater environment. Among them, the deepwater environment has an overall greenish image due to the absorption of light by water. The underwater crack pictures are marked by Labelimg software, and the prepared data set is divided into a training set and a test machine according to the ratio of 8:2. Figure 10 shows some photos of the three types of environments.

The network is trained for a total of 100 epochs. The backbone feature network is frozen in the first 50 epochs to speed up the training efficiency and is unfrozen in the latter 50 epochs for full network parameter training. The following Table 3 shows the specific parameters set during model training.



(a) Clean water environment

(b) Turbid water environment

(c) Deepwater environment

FIGURE 10 Partial underwater crack image. (a) Crack image in clean water environment. (b) Crack image in turbid water environment. (c) Crack image in deepwater environment

3.2 | Evaluation indicators

The commonly used evaluation indicators to evaluate the effect of a trained network model are Precision(P), Recall(R), Comprehensive index F1, Average precision (AP), Mean Average Precision (mAP), and Frames Per Second (FPS).

Since this network model is used to detect underwater cracks, the AP value is equal to the mAP value in this model.

3.3 | Ablation experiment

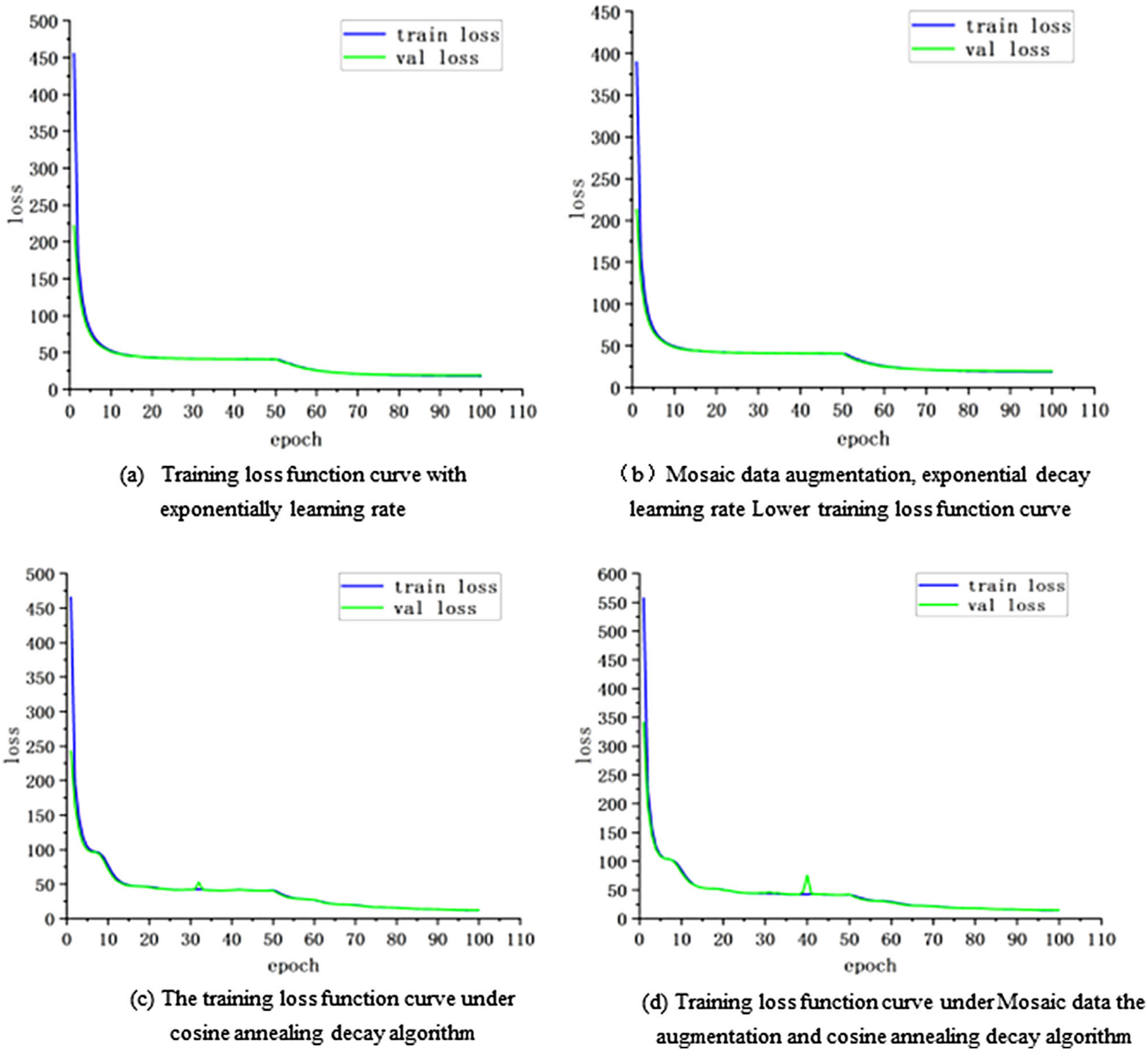
The network proposed a method to improve the performance in this paper, different combinations of training skills are used to train the network here, and the training results are compared and verified to find the training skills suitable for the network here.

The Lite-YOLO-v4 network proposed here is used as the experimental network to verify the performance improvement brought by different training techniques. The design of the ablation experiment is shown in Table 4 below:

1. The experimental network only uses an Exponential Decay Rate;
2. The experimental network uses Mosaic data enhancement and Exponential Decay Rate;
3. The experimental network only uses the cosine annealing decay;

TABLE 4 Ablation experiment design and evaluation index values

Serial number	Mosaicdata augmentation	Cosine annealing decay	Exponential decay	Recall	Precision	mAP
1	×	×	√	46.89%	83.85%	72.3%
2	√	×	√	38.21%	89.33%	71.6%
3	×	√	×	47.98%	93.97%	77.07%
4	√	√	×	39.23%	91.12%	73.05%



(a) Training loss function curve with exponentially learning rate

（b）Mosaic data augmentation, exponential decay learning rate Lower training loss function curve

(c) The training loss function curve under cosine annealing decay algorithm

(d) Training loss function curve under Mosaic data the augmentation and cosine annealing decay algorithm

FIGURE 11 Loss function curves under four experimental conditions

4. The experimental network uses Mosaic data augmentation and cosine annealing decay.

The four experimental conditions' loss of network training function curves is shown in Figure 11. Exponential decay of learning rate.

As can be seen in above Figure 11, the loss function of the training set and the loss function of the validation set converge in a consistent manner, indicating that the improved network model in this paper has good performance. Each

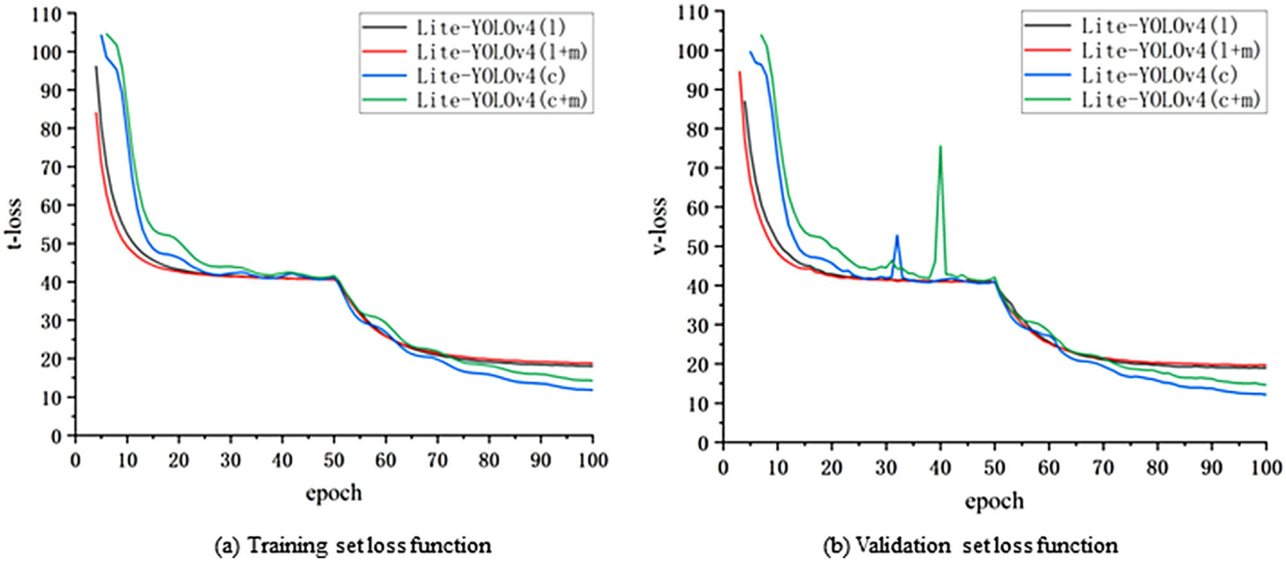


FIGURE 12 Comparison of loss function curves of detection models

network model can quickly converge in the first ten epochs and decrease slowly in 10–50 epochs, indicating that the network has been trained well at this time. At 50 epochs, the value of the loss function suddenly drops significantly, indicating that the network is thawed and the overall parameters are started to be trained. In the range of 70–100 epochs, the loss function curve does not decrease significantly, indicating that the network has converged at this time.

To further verify the impact of different training techniques on model training, this paper compares the training set loss functions and validation set loss functions of the four networks. Due to the large loss value of the first five epochs, the comparison effect of putting the four networks together is not apparent. The loss error of the first five values is large, which has little practical significance to the model. The loss values of the first five epochs had a large error. After removing the loss values of the first five epochs of the four networks, the comparison effect became more obvious and intuitive. Figure 12 shows the comparison of the network loss function curve of the training set and verification set in Figure 12. The numbers in the figure correspond to the above four cases respectively.

In the loss function descent curve, it can be found that at 50 epochs, the four network models can converge. However, the two networks using the exponential decay algorithm can converge faster and tend to converge at 20 epochs, (1), (2) The model loss function curve is relatively consistent; In the range of 50 to 100 epochs, it can be seen that the loss function curve of the cosine annealing decay algorithm is more volatile than the exponential decay algorithm, but it can reach a lower value. The loss function value is 2/3 of the exponential decay algorithm. It shows that using the cosine annealing decay algorithm can effectively jump out of the local optimal solution; observing the loss curves of (3) and (4), it can be seen that the loss function of the model without Mosaic data enhancement can reach a lower value, and the loss function value is about 4/5 of the Mosaic algorithm. It is inferred that this model uses Mosaic Data enhancement is not stable. To verify this conjecture, this paper evaluates the above four kinds of networks. The specific evaluation index values are shown in the following Table 3, where the serial numbers correspond to the above four kinds of networks.

Observe the above table and compare the first two networks. On the premise of using the exponential decay algorithm, the network enhanced by Mosaic data has a 0.7% lower mAP value than the former. Comparing the latter two networks, on the premise of using the cosine annealing decay algorithm, the mAP value of the network enhanced by Mosaic data is reduced by 2.02%. It can be seen that the stability of the Mosaic data enhancement algorithm is not good, and its performance in the lightweight network proposed in this paper is mediocre.

Through (1), (3) network comparison; (2), and (4) network comparison, it can be found that the mAP values of the network using the cosine annealing decay algorithm and the network using the exponential decay algorithm are increased by 2.77% and 1.45%, respectively. The results show that the cosine annealing algorithm achieves the optimal performance of the network by jumping out of the local optimal solution by jumping the learning rate.

The following compares the detection effects of cracks to verify the actual application of the above four networks. To make the comparison results sufficiently reliable, the following comparisons are made from different working conditions. Figure 13 shows the test results.

According to the detection results in Figure 13, it can be seen that the model based on the exponential decay algorithm is terrible at detecting in the muddy water environment and deepwater environment, as no cracks were detected; although the model using Mosaic data enhancement and exponential

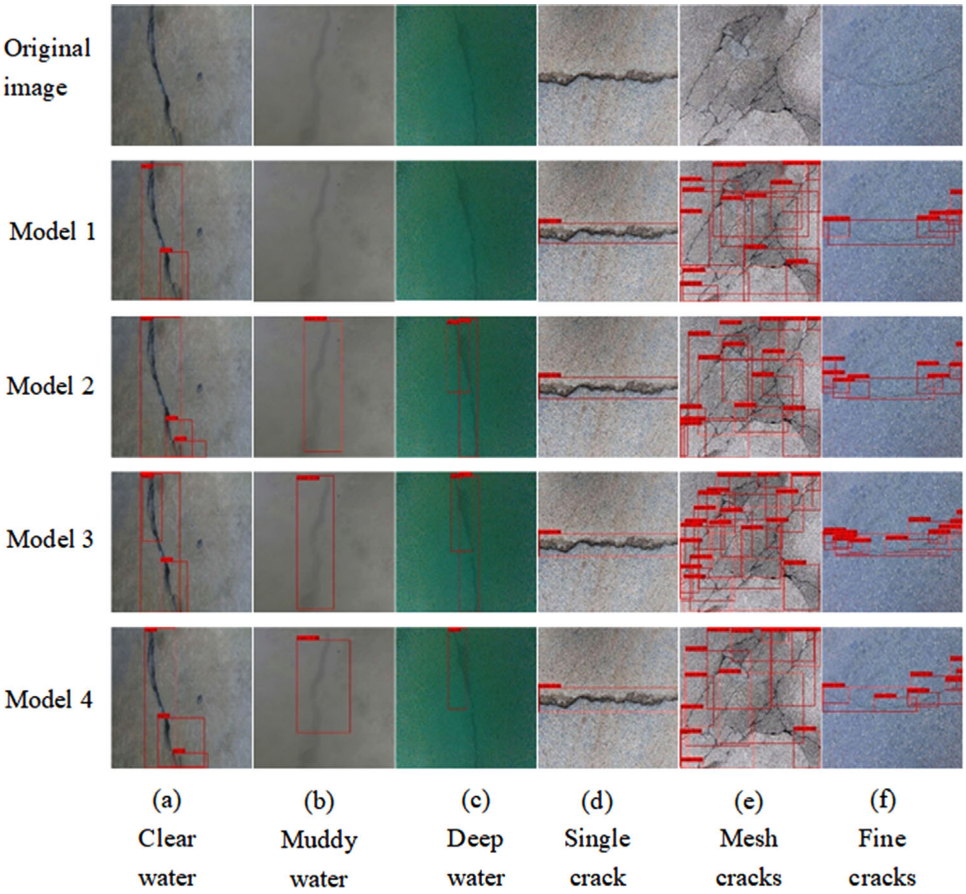


FIGURE 13 Crack detection results of four models under six working conditions

decay algorithm can identify underwater cracks under complex working conditions, the score is low. Using the learning annealing decay algorithm and Mosaic data enhancement to detect in harsh environments, there is a phenomenon of missed detection and incomplete crack detection; Instead of using the Mosaic algorithm, model 3, whose learning rate adopts the cosine annealing decay algorithm has the best detection performance. The detection results show that the crack is completely wrapped, and the detection score is the highest, which verifies the correctness of the inference according to Table 4.

3.4 | Experiment results and analysis

The experiment in the previous section proves that the model using the cosine annealing attenuation algorithm has better performance. In this section, to verify the improvement of detection accuracy and speed of the model studied here, the model is compared with CenterNet, YOLO-v4, YOLO-v4-tiny, Mobilenetv3-YOLO-v4, YOLO-v5l, YOLO-v5m, and YOLO-v5s algorithms. All models were trained with the same training parameters and data sets. The loss function curves of the eight models are shown in Figure 14. It can be seen that both the loss function of the training set and the loss function of the training set of all networks converge rapidly with 15 epochs. They indicate no problem with the data set made in this paper and the training parameters of the model set. The training effect of the eight models is good, without an over-fitting phenomenon, and the training results of each model are reliable, which can be used for comparative experiments.

To have a more precise evaluation of the training results of the above eight network models, indicators are evaluated for the above models in this paper. The specific evaluation index values of each network are shown in Table 5.

As shown in Table 5, when conducting underwater crack detection, the model size of the improved Lite-YOLO-v4 network is 20% of the original, and the average detection accuracy is 7.13% lower than that of the YOLO-v4 network. However, the detection speed is 178% higher than that of the YOLO-v4 network. At the same time, the training speed is greatly improved compared with the YOLO-v4 network. It can meet the requirements of real-time detection of mobile equipment with low computational power. Compared with the CenterNet network, the detection accuracy is 0.03% lower, but the model size is only 40% of CenterNet. The detection frame rate increases by 66.7%, and the training speed is also greatly improved. Compared with the YOLO-v4-tiny network, the model size and detection speed are not different, and the average detection accuracy is improved by 10.21%. Based

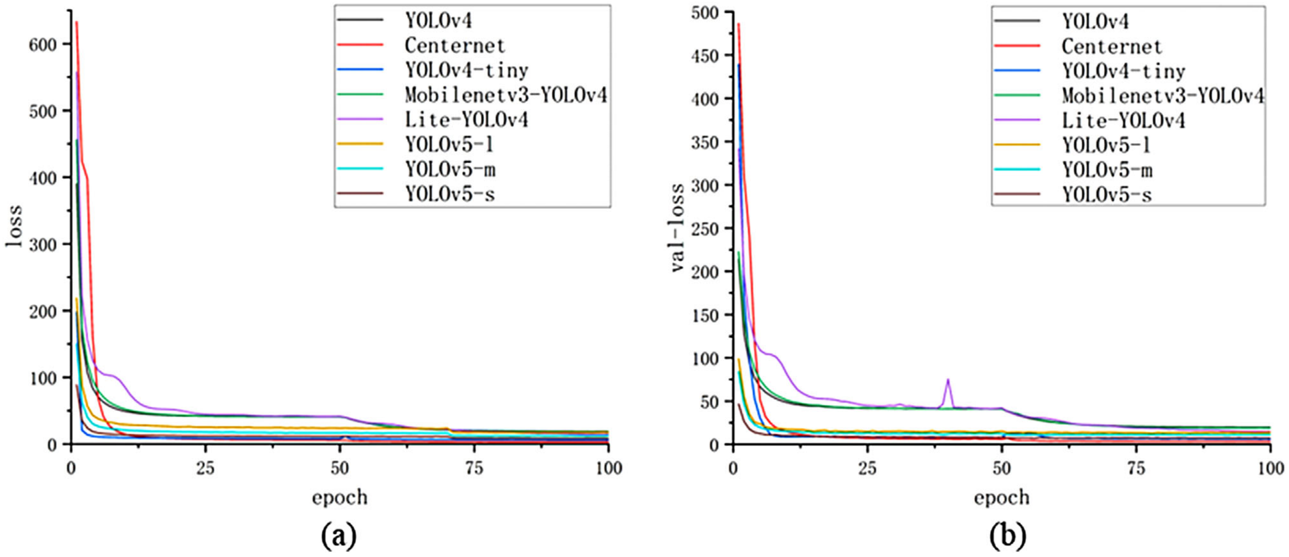


FIGURE 14 Model loss function plot. (a) Comparison curve of the loss function of the training set. (b) Validation set loss function contrast curves

TABLE 5 Values of the eight model indicators under the crack dataset

Model	Recall	Precision	mAP	FPS	Model Size/MB	Training time/epoch
YOLO-v4	53.61%	95.62%	84.2%	9	244	15 min
CenterNet	48.53%	94.31%	77.1%	15	125	8 min
YOLO-v4-tiny	40.03%	87.64%	66.86%	27	22.5	1.5 min
Mobilenetv3 -YOLO-v4	41.03%	91.2%	75.13%	14	152	9 min
Lite-YOLO-v4	47.98%	93.97%	77.07%	25	44.3	2 min
YOLO-V5l	53.69%	92.96%	82.49%	14	178	7 min
YOLO-V5m	52.25%	89.04%	80.41%	18	81.5	4 min
YOLO-V5s	47.96%	87.69%	75.87%	23	27.76	1.5 min

on Mobilenetv3-YOLO-v4, the PANet network structure and prior frame are improved here. Compared with the original network, the average detection accuracy is improved by 2.94%, but the detection accuracy is not significantly improved. The detection speed is increased by 78%, and the training time is 25% of the original, significantly improving the training speed. YOLO-v5 series network training duration is relatively short. Compared with YOLO-V5L and YOLO-v5m, the detection speed increased by 78.6% and 38.9%, respectively. Compared with YOLO-v5s, the accuracy of the network in this article is 6.28% higher. In comprehensive comparison, the network here ensures detection accuracy, improves detection speed, and reduces training time.

The above table proves that the network here has noticeable improvement compared with other networks, but there is no actual detection effect for comparison. To make the experiment more complete, the detection results of the eight models are compared from different detection environments, different detection angles, and different crack shapes.

3.4.1 | Model detection results under three detection environments

The underwater test environment is divided into the clean water environment, turbidity water environment, and deepwater environment, and the detection effect of the network model is compared under the three water environments. Figures 15, 16, and 17 show the detection effect of concrete cracks in clean water environment, turbidity water environment, and deepwater environment respectively.

In the clean water environment, cracks can be detected in all networks, but repeated detection exists in YOLO-v4-tiny and Mobilenetv3-YOLO-v4 networks. YOLO-v4-tiny and Centernet apparent false detection occur in the first group of images. The network of this article network, YOLO-v4, the YOLO-v5l, and the YOLO-v5m network can better identify cracks in a turbidity water environment. Centernet, YOLO-v4-tiny, Mobilenetv3-YOLO-v4, and YOLO-v5s have poor crack identification effects, which indicates that Lite-YOLO-v4 can be

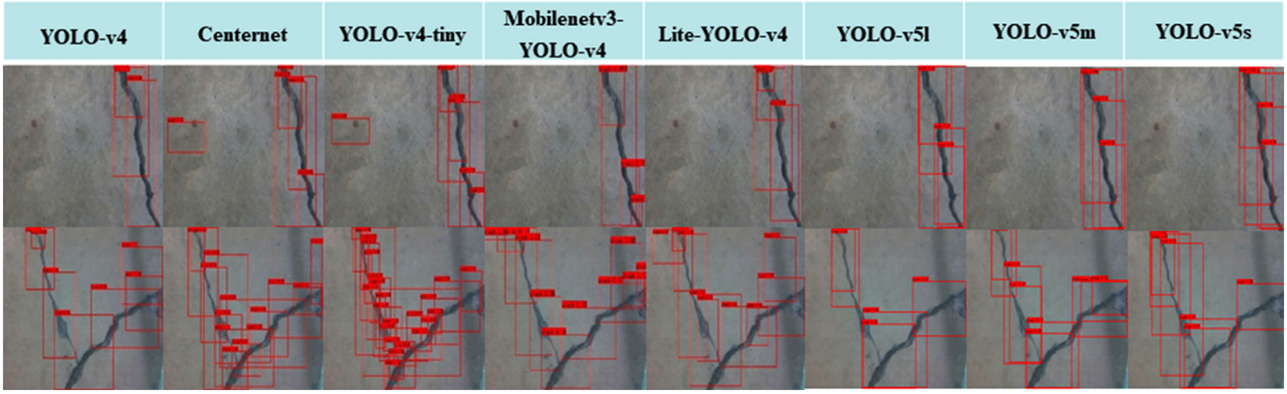


FIGURE 15 Comparison of model detection effects in clean water environment

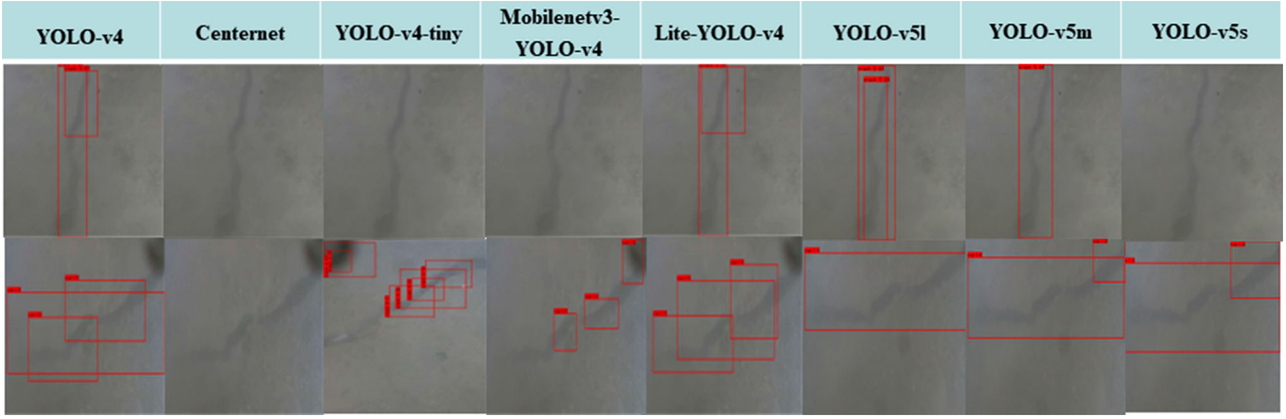


FIGURE 16 Comparison of model detection effects in turbidity water environment

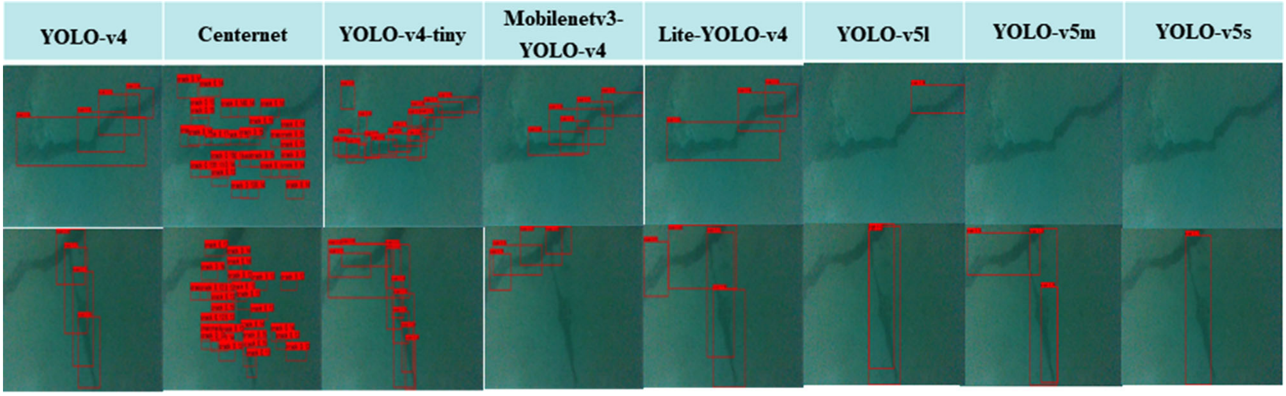


FIGURE 17 Comparison of model detection effects in deepwater environment

used for crack detection in a turbidity water environment. In the harsh green water environment, the crack detection effect of the Centernet network is not ideal, the prediction box is chaotic and cracks cannot be detected. Although Mobilenetv3-YOLO-v4 can identify the existence of cracks, the prediction frame is not completely wrapped, and there is a phenomenon of missed detection.YOLO-v4-tiny network can recognize cracks, but there are many predicted boxes that can cover information about cracks. And YOLO-v5 series detection effect is not ideal. In this paper, lite-YOLO-v4 and YOLO-v4 networks can identify fractures well, indicating that this network can be detected in harsh conditions and poor clarity in a deepwater environment.

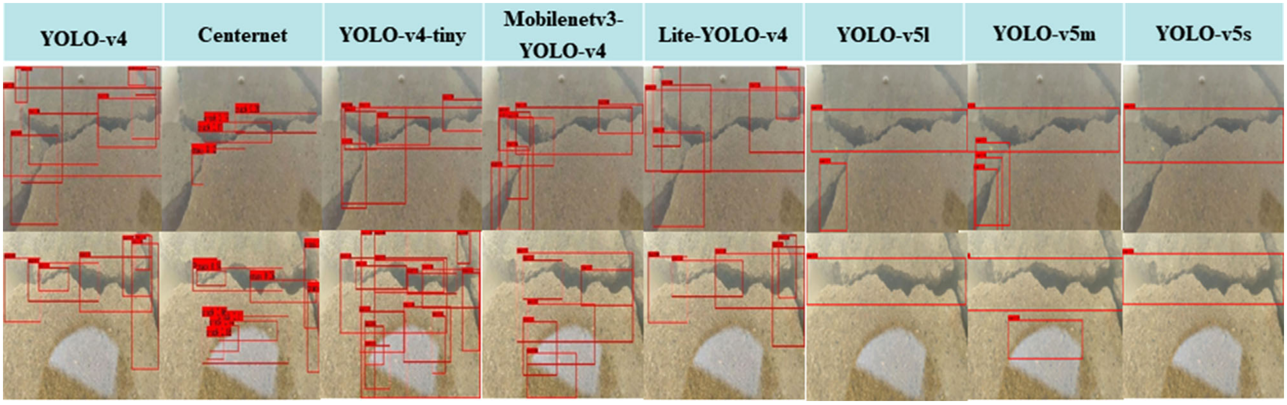


FIGURE 18 Comparison of detection effects from 30° detection angle

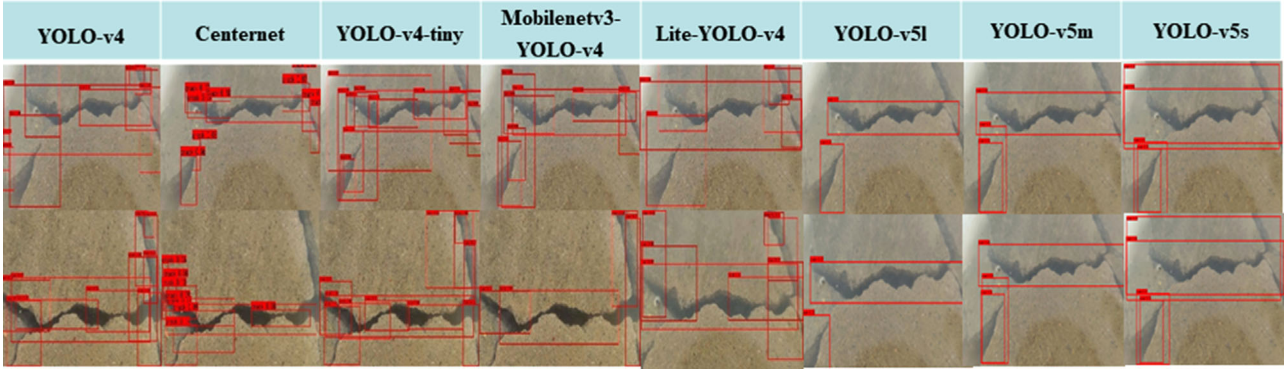


FIGURE 19 Comparison of detection effects from 60° detection angle

3.4.2 | Test results of the two test angles

In the inspection process of an underwater robot, there will be cases where the detection angle is not aligned with the underwater crack disease of the bridge., so it is proposed to verify the inspection effect under 30° and 60° inspection angles. The detection results are shown in Figures 18 and 19.

In 30° angle detection, although other networks can identify cracks, many prediction boxes are generated, while the prediction box of this network is less and more accurate in identifying cracks. In the 60° angle detection, although other networks can entirely cracks, the prediction is incomplete and there are omissions. This network can detect cracks more accurately.

3.4.3 | Tiny crack detection

In the above experiments, cracks are detected from different underwater environments and different angles to compare the detection effects. The characteristics of cracks are relatively prominent. The underwater structural diseases of bridges generally start from tiny cracks. The experiment in this section is to prove the detection effect of this network on tiny cracks. The detection results are shown in Figure 20.

The network model can detect the tiny cracks that are hard to see by the human eye in the underwater environment. The first four models can detect tiny cracks, but some tiny cracks cannot be detected in YOLO-v4-tiny, YOLO-v5 series, and Mobilenetv3-YOLO-v4 networks because of the missing phenomenon. The range of YOLO-v4 and CenterNet network detection boxes is too extensive, and the network here can accurately identify the location of cracks.

By comparing the crack detection experiments under different working conditions, it can be proved that this network can improve the network's overall performance compared with other networks while improving the detection speed and ensuring accurate crack detection under complex working conditions. This network can be applied to the actual detection of bridge underwater cracks.

4 | CONCLUSIONS

This paper proposes a lightweight neural network based on YOLO-v4, which removes the classification layer and output layer from Mobilenetv3 and replaces the CSPDarkent53 network structure as the backbone feature extraction network of YOLO-v4. In the residual network, a lightweight attention mechanism is introduced. A large number of ordinary

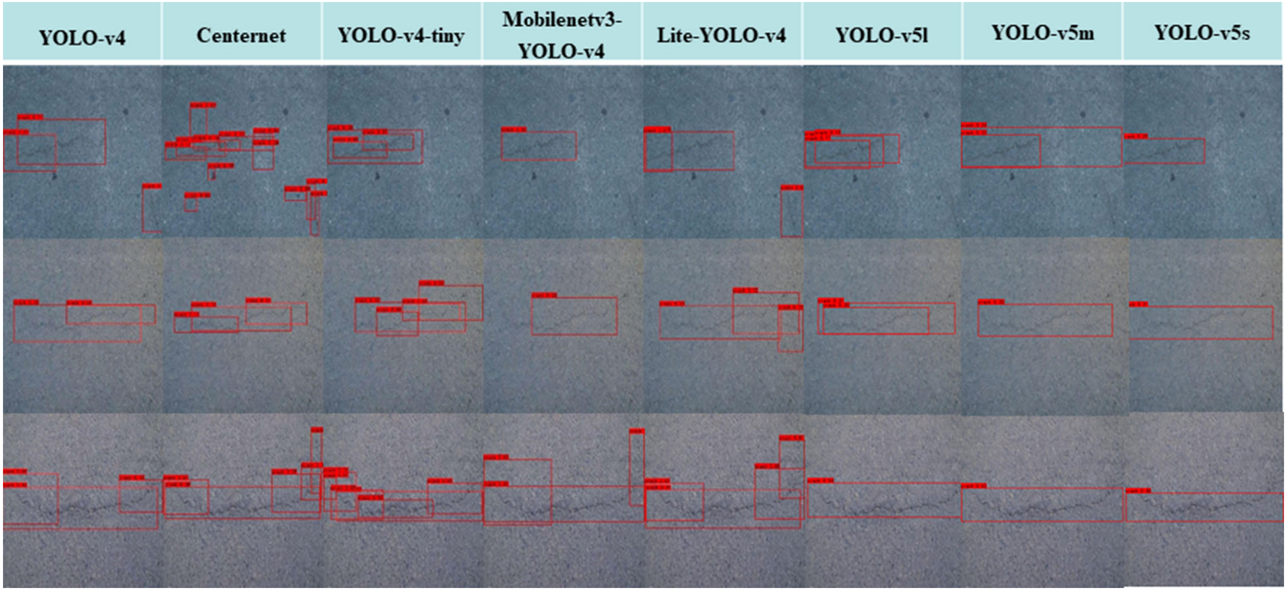


FIGURE 20 Comparison of detection effects of various models for tiny cracks

convolutions in the PANet structure are replaced with depthwise separable convolutions, and multi-feature fusion is performed on the prior box. The improved network model's parameters and calculations are significantly reduced. The model size is only 1/5 of the original, which improves the detection efficiency while ensuring accuracy.

To verify that different combinations of training techniques can improve the network's performance here, an ablation experiment is designed, and different combinations of Mosaic, cosine annealing algorithm, and exponential decay are used to train the network. The results show that the network using the cosine annealing algorithm has the best performance.

This paper builds several commonly used target detection algorithms CenterNet, YOLO-v4, YOLO-v4-tiny, Mobilenetv3-YOLO-v4, and YOLO-v5 series on the device. The detection results of the above network and the network here are compared from with other networks, this network improves detection speed and improves network performance. It can also ensure accurate detection of cracks under complex working conditions and deploy it in embedded equipment to detect bridge underwater cracks.

CONFLICT OF INTEREST

All authors disclosed no relevant relationships.

DATA AVAILABILITY STATEMENT

Data subject to third party restrictions: The data that support the findings of this study are available from third party. Restrictions apply to the availability of these data, which were used under license for this study.

ORCID

Heming Sun https://orcid.org/0000-0003-1508-5767

REFERENCES

1. Huang, W., Pei, M.S., Liu, X.D., Wei, Y.: Design and construction of super-long span bridges in China: Review and future perspectives. Front. Struct. Civil Eng. 14(4), 803-838 (2019).https://doi.org/10.1007/s11709-020-0644-1
2. Yan, Y., Mao, X.Q., Wang, X., Yu, X.F., Fang, L.: Design and implementation of a structural health monitoring system for a large sea-crossing project with bridges and tunnel. Shock Vibr. 2019, 2832089 (2019). https://doi.org/10.1155/2019/2832089.
3. Zong, X.H., Chen, Z.H., Wang, D.D.: Local-CycleGAN: a general end-to-end network for visual enhancement in complex deep-water environment. Appl. Intell. 51, 1947–1958 (2021). https://doi.org/10.1007/s10489-020-01931-w
4. Jung, J.Y., YOON, H.J., Cho, H.W.: Research of remote inspection method for river bridge using sonar and visual system. J. Korea Acad. Ind. Cooperation Soc. 18(5), 330–335 (2017). https://doi.org/10.5762/KAIS.2017.18.5.330
5. Li, P.F., Ji, T.Y., Tang, Z.X., Xu, L., Hu, W.G.: Analysis and prospect of underwater structure detection in water-related engineering. Chin. Water Transp. 17, 301–302 (2017).
6. Cerro, G., Ferrigno, L., Laracca, M., Milano, F., Carbone, P., Comuniello, A., De Angelis, A., Moschitta, A.: An accurate localization system for nondestructive testing based on magnetic measurements in quasi-planar domain. Measurement 139, 467–474 (2019). https://doi.org/10.1016/j.measurement.2019.03.022
7. Bogas, J.A., Gomes, M.G., Gomes, A.: Compressive strength evaluation of structural lightweight concrete by non-destructive ultrasonic pulse velocity method. Ultrasonics 53, 962–972 (2013). https://doi.org/10.1016/j.ultras.2012.12.012.
8. ASTM C597-09: Standard test method for pulse velocity through concrete, American Standards for Testing Materials 04(02), 1–4 (2009). https://doi.org/10.1520/C0597-09
9. Lim, M.K., Cao, H.G.: Combining multiple NDT methods to improve testing effectiveness, Constr. Build. Mater. 38, 1310–1315 (2013).https://doi.org/10.1016/j.conbuildmat.2011.01.011.
10. Chen, B., Yang, Y., Zhou, J., Zhuang, Y.Z., McFarland, M.: Damage detection of underwater foundation of a Chinese ancient stone arch bridge via sonar-based techniques. Measurement 169, 108283 (2021). https://doi.org/10.1016/j.measurement.2020.108283.

11. Jung, J.W., Cho, K.H., Hong, S.S.: Applicability evaluation of multi beam echo sounder for inland water. J. Korean Soc. Surv. Geod. Photogramm. Cartogr. 36(6), 629–639(2018). https://doi.org/10.7848/ksgpc.2018.36.6.629.
12. Kong, W.Z., Yu, J.S., Cheng, Y., Cong, W.H., Xue, H.H.: Automatic detection technology of sonar image target based on the three-dimensional imaging. J. Sens. 2017, 8231314 (2017) https://doi.org/10.1155/2017/8231314
13. Lakhani, P., Gray, D.L., Pett, C.R., Nagy, P., Shih, G.: Hello world deep learning in medical imaging. J. Digital Imaging 31, 283–289 (2018). https://doi.org/10.1007/s10278-018-0079-6.
14. Zhao, X.Y., Qin, W.J., Qian, X.H.: Application of deep learning methods in biological mass spectrometry and proteomics. Adv. Biochem. Biophys. 45, 1214–1223 (2018).
15. Chen, J.Z., Yang, C.W., Ren, J.: Machine learning based on wave and diffusion physical systems. Acta Phys. Sinica 70(14), 144204 (2021). https://doi.org/10.7498/aps.70.20210879
16. Munawar, H.S., Ullah, F., Shahzad, D., Heravi, A., Qayyum, S., Akram, J.: Civil infrastructure damage and corrosion detection: An application of machine learning. Buildings 12, 156 (2022). https://doi.org/10.3390/buildings12020156.
17. Li, H.T., Xu, H.Y., Tian, X.D., Wang, Y., Cai, H.Y., Cui, K.R., Chen, X.D.: Bridge crack detection based on SSENets. Appl. Sci. 10, 4230 (2020). https://doi.org/10.3390/app10124230.
18. Zhu, J.S., Zhang, C., Qi, H.D., Lu, Z.Y.: Vision-based defects detection for bridges using transfer learning and convolutional neural networks. Struct. Infrastruct. Eng. 16, 1037–1049 (2020). https://doi.org/10.1080/15732479.2019.1680709
19. Gao, Q.F., Wang, Y., Liu, C.G., Guo, B.Q., Liu, Y.: Concrete bridge crack identification and location technology based on convolutional neural network algorithm. Highway 65, 268–274 (2020)
20. Ruan, X.L., Wang, B., Wu, J.F., Zhao, X.G., Chen, Y.: Deep learning based reinforced concrete bridge lost block exposed rib disease identification. World Bridge 48, 88–92 (2020)
21. Chen, B., Zhang, H., Wang, S., Wang, H.R., Liu, Z.W., Li, Y.L., Xie, H.: Research on crack detection method of dam surface based on full convolutional neural network. J. Hydroelectr. Power 39, 52–60 (2020)
22. Ying, J.J., Xia, F., Lu, G.Q., Wang, Y.: Research on bridge crack identification method based on deep learning. Water Resour. Plann. Des. 01, 75–80 (2021)
23. Zhang, Z.G., Zhang, Z.D., Li, J.N., Wang, H.Y., Li, Y.B., Li, D.H.: Detection of potato in complex environment using improved YoloV4 model. Trans. Chin. Soc. Agric. Eng. 37, 170–178 (2021)
24. Huang, C.P., Zhai, K.K., Xie, X., Tan, J.J.: Deep residual network training for reinforced concrete defects intelligent classifier. Eur. J. Environ. Civil Eng. (2021) https://doi.org/10.1080/19648189.2021.2003250.
25. Cao, Z.H., Shao, M.F., Xu, L., Mu, S.M., Qu, H.C.: MaskHunter: real-time object detection of face masks during the COVID-19 pandemic. IET Image Process. 14(16), 4359–4367 (2020). https://doi.org/10.1049/iet-ipr.2020.1119.
26. Hu, X.L., Liu, Y., Zhao, Z.X., Liu, J.T., Yang, X.T., Sun, C.H., Chen, S.H., Li, B., Zhou, C.: Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network. Comput. Electron. Agric. 185, 106135 (2021). https://doi.org/10.1016/j.compag.2021.106135.
27. Wang, G.B., Ding, H.W., Li, B., Nie, R.C., Zhao, Y.F.: Trident-YOLO: Improving the precision and speed of mobile device object detection. IET Image Process. 16(1), 145–157 (2021). https://doi.org/10.1049/ipr2.12340.
28. Chang, J., Guan, S.Q., Shi, H.Y., Hu, L.P., Ni, Y.Q.: Defect classification based on improved generative adversarial networks and MobileNetV3. Laser Optoelectron. Prog. 58(4), 1–6 (2021) https://doi.org/10.3788/LOP202158.0410016
29. Zhaoguo Zhang, Zhendong Zhang,Jianian LI, Haiyi Wang,Yanbin LI, Donghao LI.: Detection of potato in complex environment using improved YoloV4 model. Trans. Chin. Soci. Agric. Eng. 37(22), 170–178 (2021)
30. Sun, Y.P., Zhong, P.S., Liu, M., Cao, A.X., Li, L.: Defect Detection of stamping parts based on YOLOv4 algorithm. Forging Stamping Technol. 47, 222–228 (2022)
31. Shi, D.M., Tang, H.Y.: A new multiface target detection algorithm for students in class based on bayesian optimized YOLOv3 model. J. Electr. Comput. Eng. 2022, 4260543 (2022). https://doi.org/10.1155/2022/4260543.

How to cite this article: Li, X., Sun, H., Song, T., Zhang, T., Meng, Q.: A method of underwater bridge structure damage detection method based on a lightweight deep convolutional network. IET Image Process. 16, 3893–3909 (2022). https://doi.org/10.1049/ipr2.12602