Quest Journals
Journal of Software Engineering and Simulation
Volume 10 ~ Issue 3 (2024) pp: 33-39
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

Bridge crack detection based on Yolox-s

Ziyi Cheng[1]
1 School of Control and Computer Engineering, North China Electric Power University, Baoding, 071051, China
Corresponding Author: Weifeng Xu[1]

ABSTRACT: *Bridge crack detection is an important means to evaluate the safety and reliability of bridges. Insufficient detection will cause major safety hazards. Current research on crack detection algorithms is based on improving accuracy and does not consider factors such as detection speed and model parameters. This article combines the Yolox-s algorithm to propose a lightweight bridge crack detection algorithm. On this basis, an attention module is added to improve accuracy, and the loss function is replaced by Focal loss, which strengthens the learning of positive samples and improves the stability of the model. Experimental results show that compared with current mainstream bridge crack detection algorithms, this method achieves a balance between accuracy and speed.*

KEYWORDS: *Bridge crack detection, Object detection, Yolox, Deep Learning*

R*eceived 08 Mar., 2024;* ***R****evised 20 Mar., 2024;* ***A****ccepted 22 Mar., 2024* © *The author(s) 2024.*
P*ublished with open access at* ***www.questjournals.org***

I. INTRODUCTION

Bridges are an important part of modern transportation. As the scale of bridge construction continues to expand, regular safety inspections of bridges have become increasingly important. Inadequate inspections and assessments of bridges can create significant safety hazards. Among the bridge evaluation indicators, cracks are an important indicator for evaluating the safety and reliability of highway bridges. The occurrence of bridge cracks will lead to a reduction in the bridge's bearing capacity. Therefore, the detection of bridge cracks is very important for the safe maintenance of bridges. At present, the method of detecting bridge cracks mainly relies on manual inspection. Manual inspection mainly uses inspection equipment such as ladders or telescopes to assist in detecting cracks. When inspecting large bridges, the inspection personnel are mainly sent to the designated inspection location through bridge inspection vehicles. The bridge inspection vehicle can then safely and quickly detect bridge cracks by inspection personnel. However, it requires long-term road occupation during the inspection process, which can easily cause traffic paralysis. It is also very expensive, with high regular maintenance costs and high inspection costs. Therefore, a safer, efficient, flexible and low-cost bridge crack detection method has become an urgent need in the industry.

With the breakthrough success of the AlexNet[1] image classification algorithm, deep learning-based technology has gradually attracted attention from all walks of life. Today, many efficient and accurate convolutional neural network models have evolved from the original convolutional neural network. derived. These models are widely used in the fields of image classification and object detection. Among them, detection equipment deployed with target detection algorithms can perform detection tasks safely and efficiently. For example, after a drone is equipped with a detection algorithm, it can take advantage of its lightness and convenience to easily reach the location to be detected without affecting traffic, and use the camera and deployed detection algorithm to complete the detection task. Therefore, the introduction of target detection algorithm can become an important choice for bridge crack detection tasks.

Currently, object detection algorithms such as Faster-Rcnn[2], SSD[3], YoloV3[4], and YoloV4[5] series are used in bridge crack detection. Long[6] et al. apply fully convolutional neural networks to pixels. For level-level bridge crack detection, it is proposed to use deconvolution to improve the model, and use upsampling to compensate for the loss of detailed information during the convolution process. Kang[7] et al. proposed a crack detection method based on Faster R-CNN. This algorithm can realize automatic crack identification and positioning operations, and then extract the crack target from the crack location to complete the quantitative operation of on-site cracks to provide a basis. Zhao[8] et al. proposed a crack detection method based on

PSPNet, which uses pyramid pooling to obtain feature information at different scales to improve the accuracy of crack detection.

The above-mentioned research is basically conducted to improve the accuracy of model detection, without considering the factors of detection speed and model parameter quantity. However, actual bridge crack detection tasks generally require the target detection model to be deployed on edge devices with limited computing and small storage space. and for crack images, different cracks will have various forms of multi-scale distribution on the digital image, making it difficult for the preset anchor boxes obtained using the aggregation algorithm to effectively fit the data set. Therefore, an anchor-free detection algorithm is needed. Therefore, an anchor-free and lightweight target detection algorithm is needed.

Yolox[9] is an improved version that combines the advantages of the Yolo[10] series of networks. It innovatively uses decoupling heads and anchor-free structures. The Yolox model follows the overall layout of the previous Yolo series, consisting of the backbone network, Neck layer and Yolox Head. The model structure of Yolox is shown in Figure 1. Its backbone network follows CspDarknet53 and uses the Silu activation function. The Silu activation function is an improved version of the Relu activation function. Compared with the Relu activation function, it has stronger nonlinear capabilities and inherits the advantage of the Relu function's fast convergence. . After passing the Resblock body module four times in the backbone network, three effective feature layers were obtained. In the last ResBlock body, an SPP (Spatial Pyramid Pooling) module was added to perform pooling operations on the image through different pooling kernels to extract more feature. After obtaining three effective feature layers, feature fusion is performed through the PANET structure of the neck, and finally classification and regression are performed through the Yolox Head to obtain the prediction results. Yolox is divided into nano, s, m, l, x and other versions according to the depth and width of different modules. Since Yolox-s can ensure a balance between accuracy and speed and its anchor-free strategy can be well utilized in bridge crack detection, this article selects Yolox-s as the baseline model.

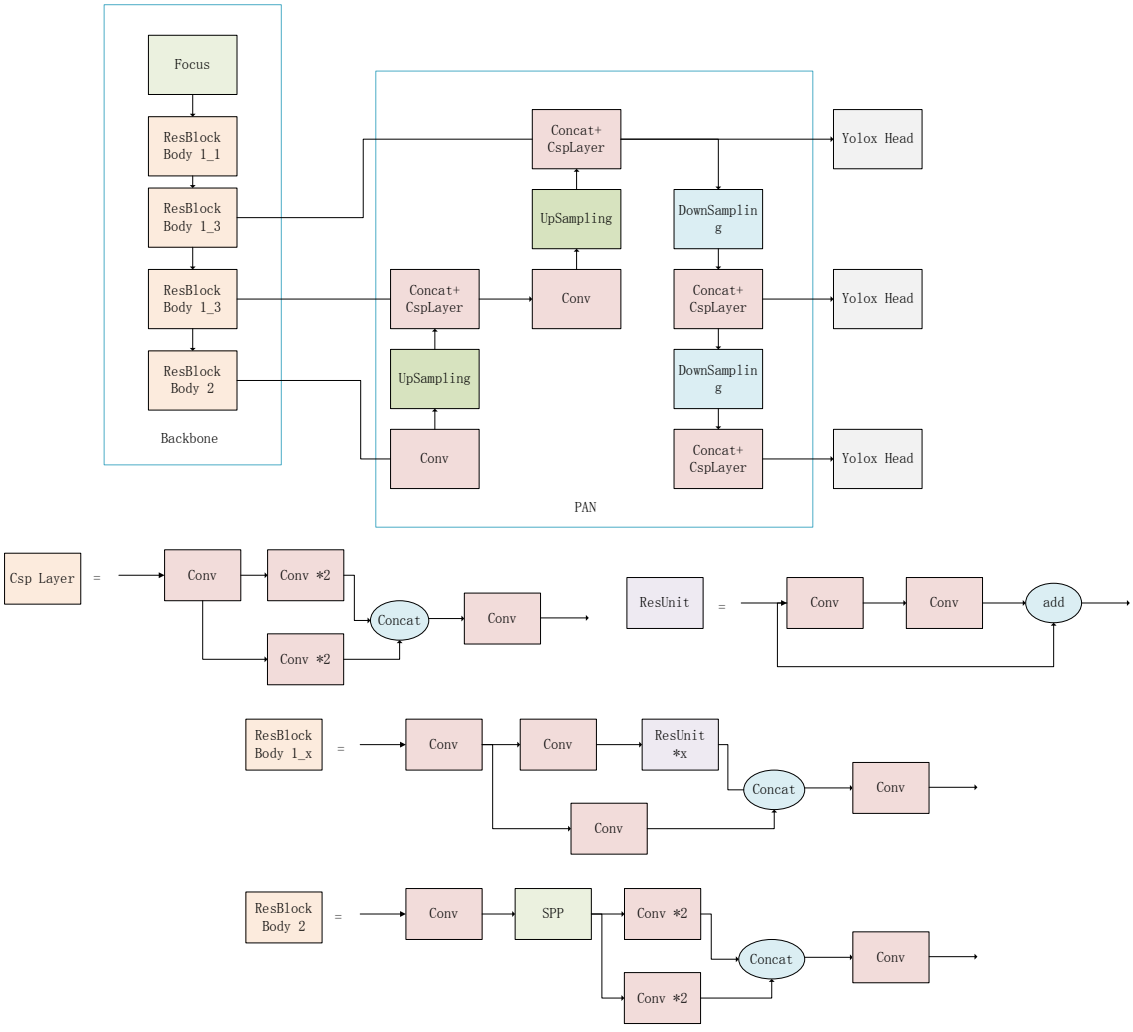

Figure1: Yolox model structure

II. IMPROVEMENTS TO YOLOX-S

2.1 ADD ATTENTION MODULE

Although Yolox-s can achieve a balance between speed and accuracy, there is a gap in accuracy compared to mainstream target detection algorithms. Therefore, this article adds an attention mechanism module to Yolox-s to improve the accuracy of the model in crack detection.

The attention mechanism module means that when the model processes input data, in order to achieve a specific task, it needs to pay attention to some important information from the input data and ignore some unimportant information. Therefore, the attention mechanism can adjust the model's attention to each input data by weighting different input data, thereby improving the performance of the model.

In order to ensure the balance between accuracy and speed, this article only adds the attention module in one place. First, decide where to add the attention mechanism module. Since the receptive fields on different feature maps are different, the receptive fields of shallow feature maps are smaller and suitable for small target detection. However, the semantic level of shallow feature maps is low and cannot be detected well. Small target, so this article adds the attention mechanism module to the input position of the shallow features in the feature fusion layer. The position of adding the attention module is shown in Figure 2. After the shallow features pass through the attention module, they are then combined with the deep semantic features. It can better improve the small target detection performance of the model.

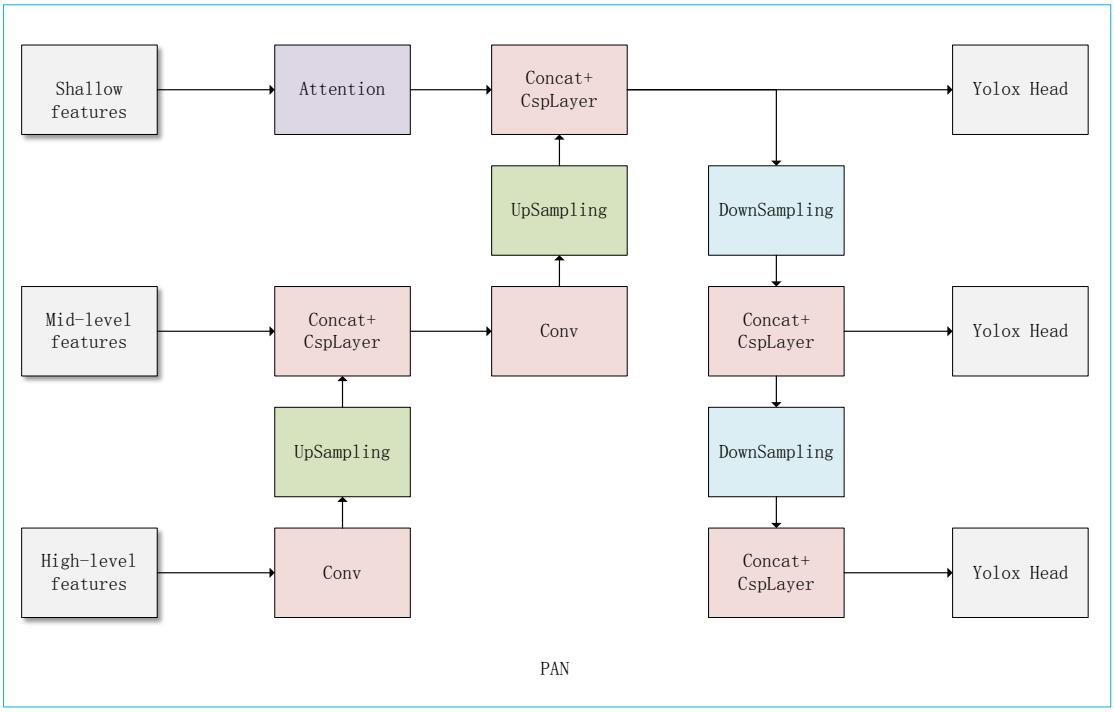

Figure2: Add attention module

This paper uses the embedded coordinate attention mechanism (Coordinate Attention, CA)[11] to assign different weights to the multi-scale channels of the feature fusion layer. It not only takes into account channel information, but also direction and position information, and is lightweight and flexible. Compared with the traditional method of randomly assigning weights, the weighted method of embedding coordinate attention can further increase the receptive field of the model, the attention of small targets, and the position sensitivity. As shown in Figure 3, the input feature map is segmented and compressed using the pooling kernels of (H, 1) and (1, W), and the input feature map is averagely pooled in the X direction and Y direction respectively, resulting in two The sizes are C×H×1 and C×1×W feature maps of independent direction perceptual attention respectively. The feature maps with direction information are then spliced through concat, and the process feature map f is generated through 1×1 convolution, normalization and nonlinear activation. Split f into two independent vectors fh and fw in the spatial dimension, and then adjust them to the same number of channels as the input feature map through 1×1 convolution, and then use the Sigmoid activation function to obtain two independent spatial directions. The attention weight is finally expanded and applied to the input features to obtain an output feature map that is sensitive to the target space dimension position information.

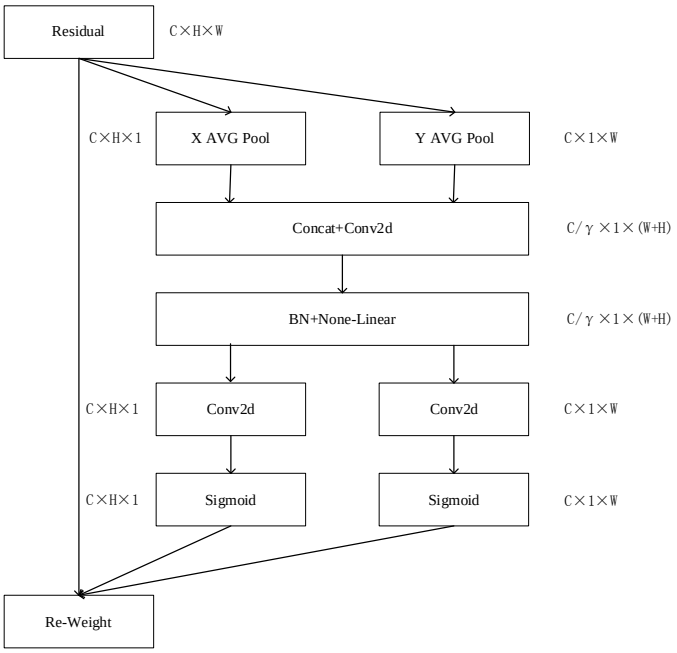

Figure3: Coordinate Attention

2.2 LOSS FUNCTION

Since the cracks to be detected are similar to the bridge background, when the cracks are small, it is difficult to distinguish between positive crack samples and negative bridge background samples. During the model training process, there are a large number of background negative samples, resulting in the model not being able to learn enough crack positive samples. The number of negative samples is too large, occupying most of the loss, so that the optimization direction of the model is not what we want, resulting in unstable model performance. In order to solve this problem, this article introduces the Focal loss function to replace the cross-entropy loss. Focal loss is built on the basis of the cross-entropy loss function. The focal loss function is defined as follows:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(\mathrm{p_t}) \ \# \rightarrow (1)$$

$$p_t = \begin{cases} p(x) & if\ y=1 \\ 1-p(x) & otherwise \end{cases} \# \rightarrow (2)$$

Pt reflects the difficulty of sample classification (the Pt that is difficult to classify is small), so (1-Pt)γ is used to attenuate the original cross entropy loss. γ is a hyperparameter used to control loss attenuation. αt is used to control the weight of positive and negative samples. The best effect is when γ=2 and αt=0.75.

To sum up, the improved Yolox-s structure is shown in Figure 4：

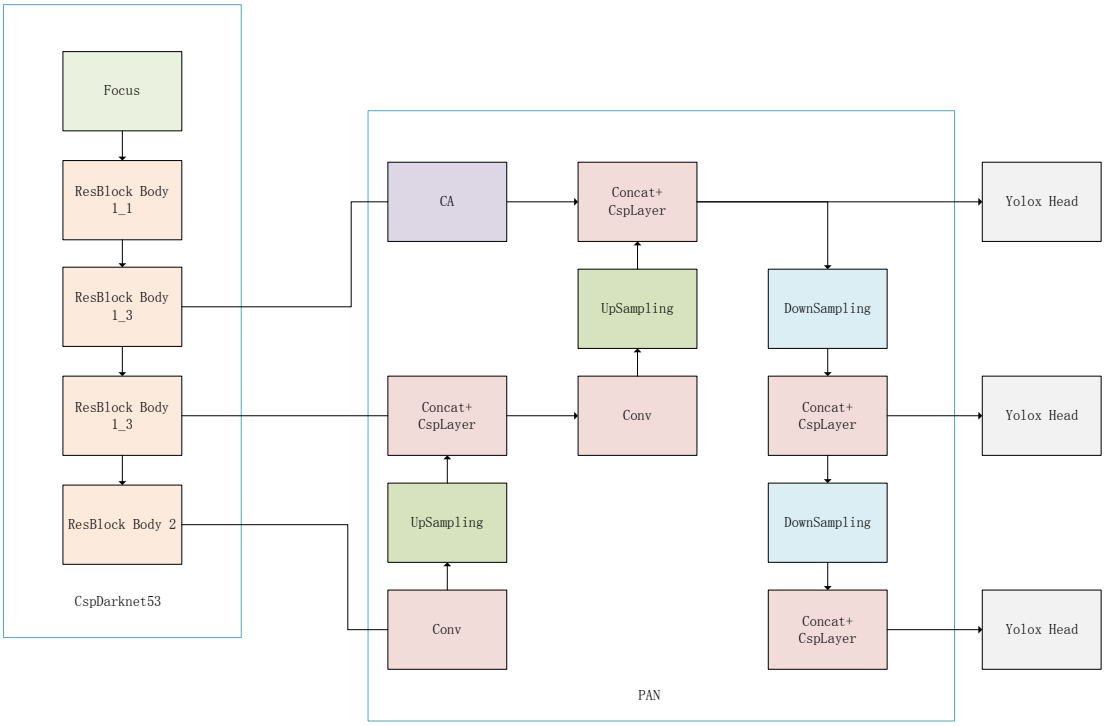

Figure4: Improved Yolox-s

III. EXPERIMENT

3.1 DATASET CONSTRUCTION

After determining the algorithm structure, a suitable data set needs to be constructed as a training data set for the model. Because concrete materials are easy to obtain, have good fire resistance, and are not easily weathered, they are often used as materials for bridge construction. Therefore, this experiment selected multiple concrete bridge crack images from the bridge crack open source data set, and performed rotation, translation, cropping and other operations on the acquired images to obtain more than 4,000 data sets containing various concrete bridge cracks. After obtaining the data set, use the Labelimg image annotation tool to annotate the data set. The annotation type is Crack, and divide the training set, verification set and test set according to the ratio of 8:1:1.

3.2 Evaluation index

In order to measure the model effect of target detection, detection accuracy and detection rate are introduced as model evaluation indicators. Detection precision (AP) is the area of the curve enclosed by precision and recall, where precision and recall are defined as follows：

$$Precison = \frac{TP}{TP + FP} \ \# \rightarrow (3)$$

$$Recall = \frac{TP}{TP + FN} \ \# \rightarrow (4)$$

In the formula, TP is the number of correctly predicted positive samples; FP is the number of incorrectly predicted positive samples; FN is the number of incorrectly predicted negative samples.

In terms of detection rate, in order to evaluate whether the model can meet the needs of real-time target detection, the number of images processed per second (FPS) is introduced as an evaluation index to evaluate the speed of the target detection model. The larger the value, the faster the target detection model.

3.3 Experimental results

This experiment is programmed in python, and the environment used is ubuntu20.04, Pytorch version 1.11.0, CUDA version 11.6, and the GPU used is Nvidia 2080Ti.

The improved method in this article was trained on the Yolox-s crack data set for 300 rounds. The experimental results are shown in Table 1：

Table 1: Ablation experiments

Algorithm	AP/%	FPS/f/s
Yolox-s	90.85	66.7
+CA	92.27	65.1
+Focal Loss	91.22	66.7
Ours	92.83	65.1

As can be seen from the table above, the accuracy of Yolox-s with the added CA module is improved by 1.42 compared to the original method. The replacement loss function is Focal Loss, which also has a slight improvement in accuracy compared with the original method. Combined with the above-mentioned improved method in this article, the accuracy It has improved by 1.98 compared to the original, which proves that the improvement is effective.

In order to further verify the effectiveness of the improved method in this paper, the method in this paper was compared with commonly used mainstream target detection algorithms on the bridge crack data set. The experimental results are shown in Table 2：

Table 2: Comparative experiment

Algorithm	AP/%	FPS/f/s	Parameters /MB
Faster RCNN	87.52	8.7	523.7
SSD	84.15	24.1	100.2
YoloV3	82.65	28.2	236.6
YoloV4	92.71	22.3	244.7
Yolox-s	90.85	66.7	34.21
Ours	92.83	65.1	34.23

In the above table, the backbone used by Faster RCNN is Vgg16. Since Faster RCNN is a two-stage model, during the detection process, the target candidate area must be extracted first, and then classified and detected, so the detection speed is slow. The SSD and Yolo series algorithms are both one-stage target detection models, which omit the step of extracting candidate areas. Classification and regression are completed at the same time, so the detection speed is faster than the two-stage model. Among Faster-RCNN, SSD, YoloV3, YoloV4 and Yolox-s, the target detection model with the highest accuracy is YoloV4, but the model parameters are large and the speed is slow. The method in this article is more accurate than YoloV4 and much faster, achieving a balance between accuracy and speed. The detection results of this method on some data sets are shown in Figure 5.

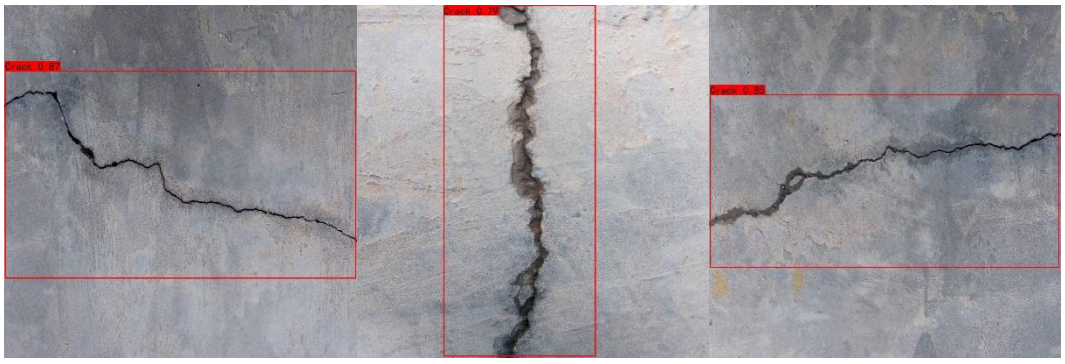

Figure5: Detection results on part of the data set

IV. CONCLUSION

This paper proposes a bridge detection algorithm based on Yolox-s. It adds a Coordinate Attention module to the feature fusion layer to improve detection accuracy. It replaces the loss function with Focal loss to improve the performance stability of the model. Experiments show that compared with the original algorithm and the current mainstream algorithm, the algorithm in this paper has better accuracy and achieves a balance between accuracy and speed.

REFERENCES

[1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[2] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[3] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14,

[4] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[5] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[6] Long J , Shelhamer E , Darrell T . Fully convolutional networks for semantic segmenta tion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.

[7] Kang D, Benipal S S, Gopal D L, et al. Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning[J]. Automation in Construction, 2020, 118: 103291.

[8] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

[9] Ge Z, Liu S, Wang F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[10] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[11] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.