YOLO-SO: Small Object Detection Research Based on YOLOv5s for Crack Detection

Jiaqi Wu[1], Jingjing Zhou[2], Tian Zhang[2]*

[1] *School of Civil Engineering Architecture and Environment, Hubei University of Technology, China*
[2] *School of Artificial Intelligence, Hubei Business College, China*
jiaqi2025@163.com, zhoujingjing0514@163.com, ztian0209@163.com

Abstract

Addressing issues such as slow speed, low accuracy, and insufficient generalization performance of traditional small target detection algorithms, this paper proposes a novel real-time detection method for small targets named YOLO-SO based on the YOLOv5s deep learning object detection algorithm, and building cracks are taken as the research target. This method optimizes and improves the YOLOv5s object detection neural network. Firstly, CBM (Conv + BN + Mish) depth separable convolution modules are introduced into the backbone network layer, and lightweight CA (Coordinate Attention) is added to the output feature map of the backbone network to focus more on crack features, thus enhancing detection performance. Secondly, the dense connection concept is introduced, replacing the feature fusion network with the PADNet network to reuse feature information. Finally, the Complete Intersection over Union (CIoU) is introduced as the target localization loss function. Experimental evaluations are conducted on a crack dataset using Mosaic data augmentation, and comparisons are made with various existing object detection neural networks. The experimental results demonstrate that the improved model, compared to the original model, reduces parameter volume by 43.28%, reduces computational load by 47.47%, and improves detection accuracy by 2.18%, validating the superiority of the proposed algorithm in this paper.

Keywords: Small object detection, YOLOv5s, Crack detection, Deep learning, Coordinate Attention

1 Introduction

Since the 21st century, artificial intelligence and computer vision technologies have rapidly developed, becoming one of the fastest-growing technologies in the field of computer applications. Real-time object detection is a crucial research topic in computer vision and an indispensable component of computer vision systems. As one of the core tasks in computer vision, object detection aims to enable computers to automatically identify various objects in images or video frames and draw bounding boxes around them to indicate the position of each object.

YOLO is a deep learning-based object detection algorithm proposed by Joseph Redmon in 2016 [1]. Its main feature is the ability to accurately detect the position and category of objects using only one neural network. Over the years, the YOLO series of algorithms has become an essential part of the computer vision field [2-3], with advantages such as better detection performance and stability. The YOLOv5 algorithm was introduced in 2020, including several models such as YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. With its excellent detection performance, YOLOv5 has become one of the dominant algorithms in the field of object detection.

Compared to the original version of YOLO, YOLOv5 improves the detection accuracy of small objects while maintaining detection speed. However, it still cannot meet the demand for small object detection in multiple scenarios. Addressing this issue, optimizing and improving object detection algorithms is one of the significant challenges currently faced. Therefore, designing a model with strong scalability, fast detection speed, and excellent and stable detection performance is crucial for meeting the demand for small object detection in various scenarios.

The article selects building cracks as the research target and optimizes and improves the YOLOv5s deep learning object detection algorithm, which is currently mainstream and mature. A novel YOLO-SO (You Only Look Once - Small Object) network is proposed, and corresponding research on its detection issues is conducted.

The main contributions are as follows:

- Proposed a novel small-object real-time detection method named YOLO-SO, addressing the issues of insufficient detection accuracy and generalization in multi-scenario small-object detection.
- Adopted the CBM depth-wise separable convolution module in the backbone network instead of the traditional CBL (Conv + BN + ReLU) standard convolution module, effectively resolving the problem of network overfitting.
- Introduced the CA mechanism to allocate attention weights to the output feature maps of the backbone network, enhancing the network's focus on target features and further improving the recognition and detection performance of targets.
- Introduced the concept of dense connections and proposed a lightweight network called PADNet to replace the feature fusion network, allowing for

*Corresponding Author: Jingjing Zhou; E-mail: zhoujingjing0514@163.com
DOI: https://doi.org/10.70003/160792642025052603014

the reuse of feature information and increasing the accuracy of target detection across multiple scales and categories.

2 Related Work

2.1 Deep Learning Techniques

Zheng et al. [4] applied the Mask R-CNN deep learning model for building crack recognition, enhancing the recognition rate and proposing an improved algorithm for skeleton extraction based on the K3M algorithm, further improving the model's recognition accuracy. Xie et al. [5] used a fully convolutional network to extract crack boundary features, achieving accurate crack identification with good precision and robustness, as validated by comparison with manually measured results. Yang et al. [6] studied an intelligent crack detection algorithm based on computer vision, improving the detection effectiveness of fine cracks by enhancing the U-net network. Yu et al. [7] proposed a new three-branch spatio-temporal feature extraction network (TBSFENet), which can achieve the accurate classification and identification of targets.

Huang et al. [8] proposed an accurate edge detection algorithm to identify potential cracks. Yu et al. [9] introduced the Phase Contour Enhanced Attention (PCEA) module to enhance the object edge information of the encoder and proposed the extended convolution pyramid (DCP) module, which has a significant ability to capture complex profiles and effectively overcome the problem of fine-grained resolution of targets. Hao et al. [10] proposed a hollow pyramid DenseNet sub-block crack detection and classification algorithm for preliminary positioning of crack regions, demonstrating good segmentation results for cracks of different widths and strong robustness. Zhang et al. [11] proposed a real-time and visible crack detection scheme based on the MRL method, transforming cracks into fluorescent information through computer vision and machine learning methods for crack classification and measurement, showing great potential in detecting cracks on large complex structural components.

2.2 YOLO Algorithm Improvements

Reference [12] proposed a YOLO algorithm incorporating the CBAM attention mechanism and leveraging the Random Paste Mosaic (RPM) small object data augmentation module introduced in Mosaic. This YOLO algorithm demonstrated significant advantages in detection speed. Yu et al. [13] addressed issues such as missed detections, false positives, and low accuracy in YOLO-v3 for object detection by introducing enhanced element layers fused with small-scale features and incorporating aspect ratios into the loss function for better adaptation to object detection tasks. Yan et al. [14] modified the residual units to dense connections and added channel attention mechanisms in YOLO's CSP module to improve the algorithm, mitigating gradient vanishing issues and reducing parameter count. Wang et al. [15] proposed a multi-target intelligent recognition method based on the YOLO deep learning network and an optimized transformer structure (YOLO-T). Liu et al. [16] proposed a Cross-Stage Partially Dense YOLO (CSPD-YOLO) model, based on YOLO-v3 and cross-stage partial networks, which is more suitable for insulator detection.

Hurtik et al. [17] proposed a deep learning model called poly-YOLO, addressing some shortcomings of YOLOv3. Ju et al. [18] addressed the low detection rate of small targets by proposing an improved YOLOv3 algorithm, which demonstrated significantly improved detection performance. Adou et al. [19] explored the model based on YOLOv3 for the detection of insulators and their defects, and the experimental results showed that the model was very efficient and could process 45 frames per second. Liu et al. [20] proposed an improved YOLO-micro insulator (MTI-YOLO) network, incorporating multi-scale feature detection heads, multi-scale feature fusion structures, and Spatial Pyramid Pooling (SPP) models; this approach achieves ideal results in insulator detection. Reference [21] enhanced the network's perception of small targets using a Flip-Mosaic algorithm, improving vehicle detection accuracy and reducing false positives with this enhanced YOLOv5 model.

3 Method

3.1 Overall Framework

The YOLO-SO proposed in this study adopts CSPDarknet as the backbone network. Firstly, the input images are processed by the backbone network to extract features, resulting in three feature maps of different sizes. Secondly, these feature maps are input into the attention mechanism module for attention weight allocation. Thirdly, the new feature maps are input into the PADNet (Pyramid Attention DenseNet) for multi-scale feature fusion. PADNet utilizes dense connections to reuse feature maps in the channel dimension, enhancing gradient flow and obtaining significantly strengthened feature maps. Finally, the enhanced feature maps are input into the YOLO-Head decoupled head to obtain the prediction results.

As shown in Figure 1, the YOLO-SO algorithm structure consists of five main components: the input layer, backbone layer, attention mechanism module, feature fusion layer, and output layer. The input layer preprocesses images using the Mean Shift (MS) filtering algorithm and the Contrast-Limited Adaptive Histogram Equalization (CLAHE) algorithm. The backbone layer comprises a series of modules, including CBM (Conv + BN + Mish), CSPNet, and SPPB modules. The attention mechanism module allocates attention weights to the output feature maps of the backbone network to enhance the network's focus on target features. The feature fusion layer optimizes the PANet network by introducing a lightweight DenseNet feature extraction network, resulting in a novel multi-scale feature fusion network (PADNet), which reduces network parameters while ensuring detection performance. The output layer adopts the YOLO-Head decoupled head form, separating classification and regression tasks to improve network convergence speed and further enhance detection performance.

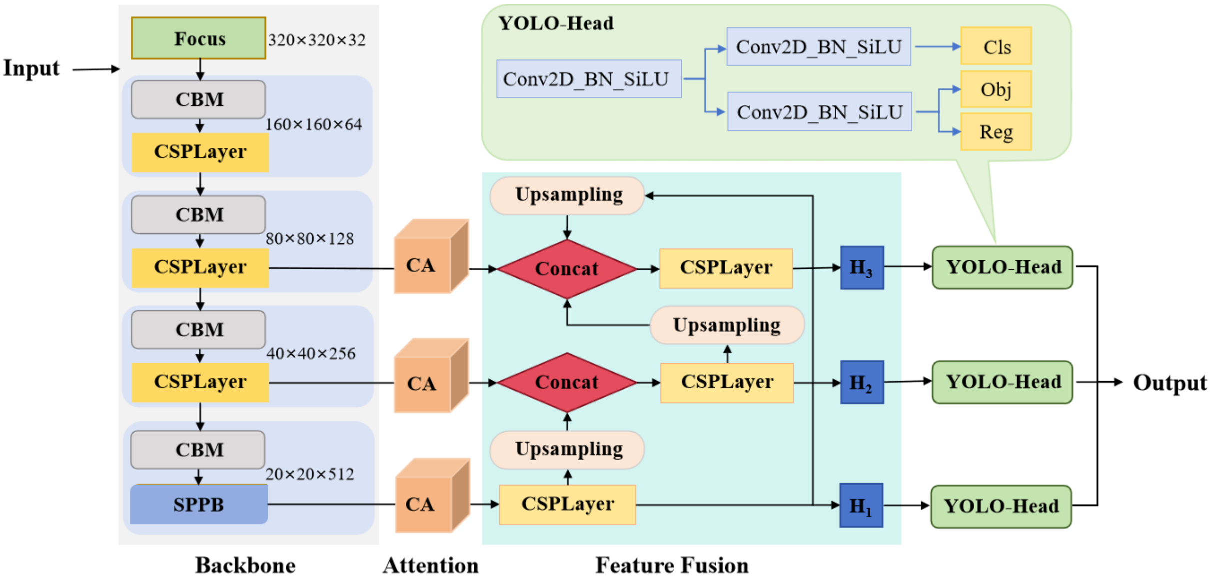

Figure 1. Overall structure diagram of the YOLO-SO algorithm

3.2 Input Layer

The input layer in this study employs Mean Shift filtering (MS) and Contrast-Limited Adaptive Histogram Equalization (CLAHE) algorithms for image preprocessing. These two methods effectively reduce background noise and enhance the information on crack edges. MS smoothens the image while preserving edge information, aiding in the removal of noise and unnecessary details from the image, and making cracks more clearly visible. Meanwhile, the CLAHE algorithm enhances image contrast, making crack edges clearer and improving the accuracy of crack detection. By combining these two preprocessing methods, image quality is effectively optimized, providing more reliable inputs for subsequent crack detection algorithms.

3.3 Backbone Layer

During the training process of deep learning network models, the feature information extracted is transmitted from shallow layers to deep layers, leading to an inevitable increase in computational complexity. Therefore, hardware devices must have high computational efficiency, particularly in network construction.

As shown in Figure 2, to optimize the network structure and improve the detection efficiency of the network model, this study adopts CSPDarknet as the backbone network. It replaces the traditional CBL (Conv + BN + ReLU) standard convolution modules with CBM (Conv + BN + Mish) depth separable convolution modules and replaces the original ReLU activation function with the Mish activation function, which has a more stable gradient flow during forward propagation, thus enhancing model stability.

As shown in Figure 1, the backbone layer consists of modules such as CBM (Conv + BN + Mish), CSPNet, and SPPB. These modules are used to perform operations such as convolution, batch normalization, and pooling. The Mish activation function is applied after batch normalization to reduce parameter volume and optimize computation. The Spatial Pyramid Pooling Block (SPPB) module performs multiple pooling operations on the feature maps and then concatenates the feature maps of multiple dimensions, enabling the network to extract higher-level semantic features.

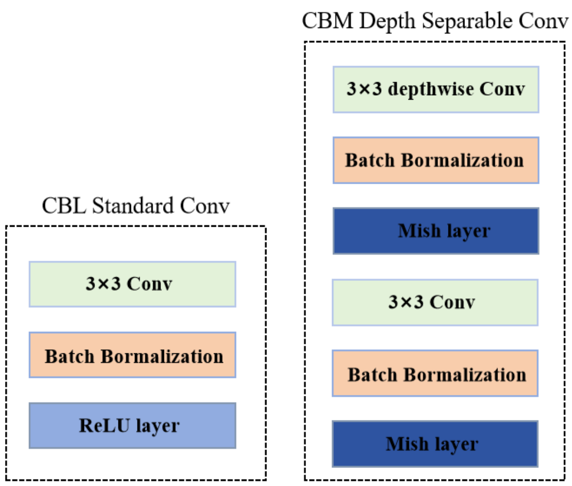

Figure 2. Depth-separable convolutional module and standard convolutional module

3.4 Attention Mechanism Module

To enhance the recognition and detection capabilities of YOLO-SO for cracks, this study incorporates a lightweight Coordinate Attention (CA) module into the output feature maps of the backbone network layer. The attention mechanism module allocates attention weights to the output feature maps of the backbone network, increasing the network's focus on crack features, and thereby improving the recognition and detection effectiveness of cracks.

Common channel attention modules in the visual domain include SE (squeeze-and-excitation) and CBAM (convolutional block attention module). SE calculates attention distributions by learning the relationships between different channels and improves features based on their importance, discarding features irrelevant for the current task. But, it ignores learning of positional information in feature maps. The CBAM attention mechanism consists of channel attention and spatial attention mechanisms. While focusing more on channel domain analysis based on traditional convolutional neural networks, CBAM introduces attention mechanisms in both channel and spatial domains. This design achieves a sequential attention structure from channel to spatial domains. However, the convolution in the CBAM attention mechanism can only extract local relationships, lacking the ability to extract long-distance relationships.

Comparatively, the CA mechanism combines channel attention and spatial attention, encoding channel relationships and long-term dependencies with precise positional information [22]. This comprehensive approach, considering both channel relationships and feature spatial positional information, addresses the shortcomings of SE and CBAM, contributing to improved accuracy in object detection. The specific operation of the CA attention mechanism includes two steps: Coordinate information embedding and Coordinate Attention generation. To ensure that the attention module captures distant spatial interactions with precise positional information, CA divides the global pooling operation into pooling operations in both width and height directions. The schematic diagram of the CA attention mechanism module is shown in Figure 3.

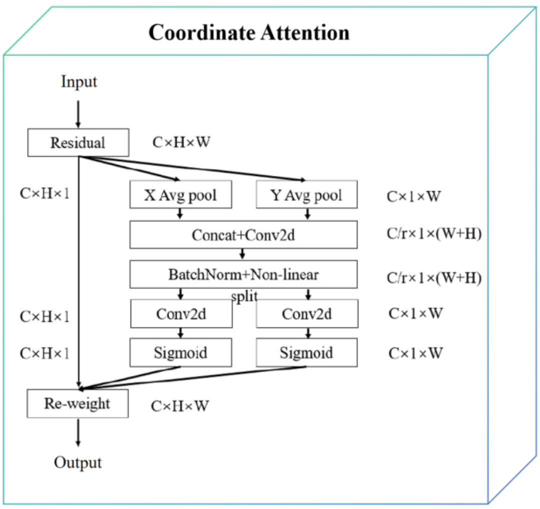

Figure 3. CA attention mechanism module

3.5 Feature Fusion Layer

To optimize the performance of object detection, this study introduces the DenseNet feature extraction network and optimizes the PANet network by combining its dense connection network concept. By concatenating the output of three different-sized feature layers, multi-scale detection of cracks is achieved, thus improving detection performance. Additionally, this study replaces the original Relu activation function in DenseNet with the Mish activation function, effectively addressing the issues of gradient vanishing and saturation in backpropagation, thereby further enhancing network performance.

As shown in Figure 4, based on the dense connection concept of the DenseNet network, a feature fusion network named PADNet is proposed. This network concatenates the feature output layers with each subsequent layer, forming a high-performance feature fusion network, thereby improving the recognition accuracy of the network model.

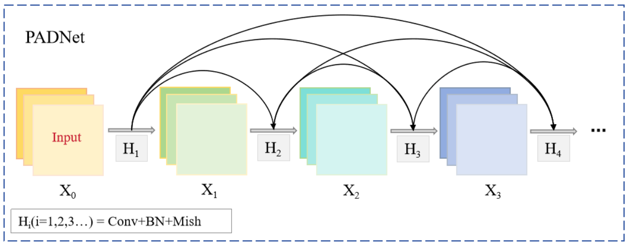

Figure 4. The feature fusion network of PADNet

As shown in Figure 1, the PADNet network first takes the feature layer X_0 with a size of 20×20 outputted by the SPPB module. Then, the X_0 feature layer is upsampled by a factor of 2 and concatenated with the feature layer with a size of 40×40 to form the concatenated 40×40 feature layer, denoted as the X_1 feature layer. Next, the X_1 feature layer and X_0 feature layer are upsampled by factors of 2 and 4, respectively, and concatenated with the feature layer with a size of 80×80 to obtain the X_2 feature layer. Due to the feature map stacking approach used in dense connections, which increases in the number of feature maps with the increase in the number of layers, H_i (i=1,2,3...), modules are added to each output layer to reduce the number of feature maps and accelerate network computation speed. Ultimately, the X_0 feature layer, X_1 feature layer, and X_2 feature layer are respectively outputted for detecting small, medium, and large targets. This study mainly enhances the feature information of deep feature layers and adopts the dense connection approach, ensuring that each feature fusion layer contains rich feature information. This design achieves feature reuse in the channel dimension, enhancing gradient flow and obtaining richer semantic features.

3.6 Output Layer

The decoupling head extracts the target position and category information separately, learning them through different network branches, and finally fuses them. This effectively reduces the number of parameters and computational complexity while enhancing the generalization ability and robustness of the model. YOLO-SO adopts the form of YOLO-Head decoupling head to separate the classification and regression tasks, which not only improves the convergence speed of the network but also enhances the detection performance of the network.

As shown in Figure 1, Reg, Obj, and Cls represent the prediction results. Among them, Reg is used to indicate the regression parameters for each feature point, while Obj and Cls are used to determine whether an object is present and the category of the object, respectively.

3.7 Loss Function

Bounding box regression is a critical step in object detection tasks. To address issues such as slow convergence speed and inaccurate regression associated with the Intersection over Union (IoU) loss function, YOLO-SO adopts the CIoU (Complete IoU) loss function. By incorporating the normalized distance between predicted boxes and target boxes and considering three geometric factors—overlapping area, center point distance, and aspect ratio of bounding box regression—this approach aims to improve the convergence speed and detection performance of the object detection neural network.

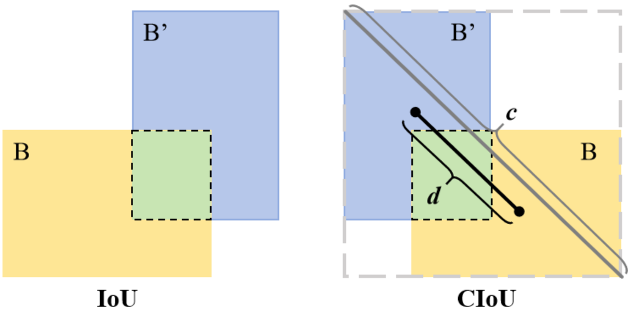

Figure 5. IoU and CIoU loss functions

The diagrams of IoU and CIoU loss functions are shown in Figure 5. In the diagrams, B represents the predicted box, B' represents the target box, c represents the diagonal length of the minimum enclosing box covering both boxes, and d represents the distance between the center points of the two boxes.

IoU and CIoU are calculated as follows:

$$IoU = \frac{|B \cap B'|}{|B \cup B'|}$$

$$CIoU = IoU - \left(\frac{d^2}{c^2} + av\right)$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w'}{h'} - \arctan\frac{w}{h}\right)^2$$

$$a = \frac{v}{(1 - IoU) + v^2}$$

In this context, IoU represents the ratio of the overlapping area between the predicted box and the target box to the total area. CIoU, building upon IoU, considers the distance between center points and aspect ratio to accelerate bounding box regression convergence speed and enhance network performance. Here, α serves as a weight parameter, and v is used to measure the consistency of the aspect ratio. w' and h' denote the width and height of the target box, while w and h represent the width and height of the predicted box, respectively.

The CIoU loss function can be expressed as:

$$L_{CIoU} = 1 - CIoU$$

4 Experiment

4.1 Dataset

The crack dataset used for network training consists of multiple sources, including publicly available datasets such as Concrete Crack Images for Classification, Crack500 for pavement cracks, and self-collected crack datasets using mobile devices.

The initial crack dataset established comprises a total of 2186 sample images sourced from various datasets. This includes 500 sample data from the Crack500 dataset, 1270 sample data selected from the Concrete Crack Images for Classification dataset, and 416 self-collected sample data. These sample images are categorized into different classes, including 885 images of transverse cracks, 873 images of longitudinal cracks, 291 images of map cracks, and 137 images of potholes. The distribution of each category of samples is shown in Table 1, while examples of images from different sample categories are shown in Figure 6.

Table 1. Crack data set sample distribution

Types	Quantity
Transverse cracks	885
Longitudinal cracks	873
Map cracks	291
Potholes	137

During the experiment, 80% of the entire dataset was used as the training set, while the remaining 20% was allocated for the validation set. Mosaic and MixUp data augmentation methods were employed to enrich the training dataset. The Mosaic data augmentation method involves randomly selecting four images, cropping them, and then merging them into a new image. On the other hand, the MixUp data augmentation method combines two randomly selected images to create a new image through blending. Both of these data augmentation methods serve to enrich the training dataset, enhance the generalization capability of the network model, and thereby strengthen the model's ability to detect targets in different scenarios and improve its robustness.

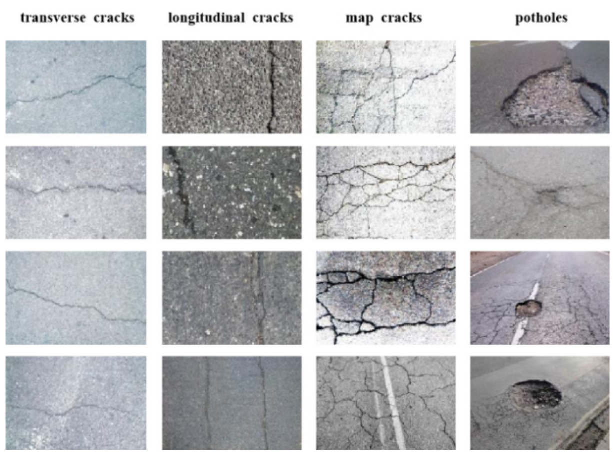

Figure 6. Example of crack sample image

4.2 Ablation Study

To verify the effectiveness of introducing the CA mechanism and the PADNet feature fusion network and their impact on model performance, sensitivity analysis was conducted. Table 2 presents the experimental results.

Table 2. Results of ablation experiments

PADNet	CA	Avg precision /%	Calculation /GB	Parameter /MB
–	–	87.3	15.8	13.4
√	–	84.5	8.4	7.3
–	√	88.6	16.3	13.9
√	√	89.2	8.7	7.6

(**Note:** "√" indicates add, "–" indicates no add)

From Table 2, it can be observed that with the introduction of the PADNet feature fusion network, compared to the original YOLOv5s, the model parameter count decreased by 45.52%, while the computational workload reduced by 46.84%. Although the reduction in model parameters increased efficiency and achieved lightweight characteristics, there was a slight decrease in detection accuracy.

To counteract the impact of the decrease in parameters and computational workload on accuracy, the efficient CA mechanism was introduced. CA effectively suppresses irrelevant feature information, allowing better focus on important features and further enhancing detection performance. With the introduction of the CA attention mechanism alone, the model's parameter count increased by 3.73%, the computational workload increased by 3.16%, and the average precision increased by 1.49%. Building upon PADNet, the introduction of the efficient CA attention mechanism led to a 4.11% increase in parameter count, a 3.57% increase in computational workload, and a 5.56% improvement in detection accuracy.

4.3 Quantitative Analysis

To verify the effectiveness and superiority of YOLO-SO, this study compares its performance with FasterR-CNN, SSD, YOLOv3, YOLOv4, and the benchmark network YOLOv5s. The above target detection neural networks are trained until the loss function converges and the network achieves its best performance. The network performance is tested based on the test set. The average Precision (mAP), frames per second (FPS), and model volume were compared to analyze the detection accuracy and speed of the algorithm. The results are shown in Table 3.

From the table, it can be observed that some algorithms such as YOLOv3, YOLOv4, and SSD have larger model sizes and slower detection speeds. Although FasterR-CNN achieves high detection accuracy, its large number of model parameters makes it unsuitable for deployment on mobile devices. On the other hand, the algorithm proposed in this paper maintains detection accuracy while reducing the model's parameter count and size. It meets the accuracy requirements of crack detection while achieving lightweight effects, reducing model complexity, and making it suitable for running on mobile devices.

Table 3. Performance index of different models

Model	Avg precision /%	Frame /FPS	Model size /MB
YOLOv3	83.4	54	123.4
YOLOv4	86.1	55	100.5
SSD	81.5	23	95
Faster R-CNN	95.3	9	48.5
YOLOv5s	87.3	75	53.6
YOLO-SO	89.5	73	30.9

4.4 Qualitative Analysis

On the test set, various pre-trained object detection neural networks were used to detect images, and some detection results are shown in Figure 7.

The detection boxes of YOLO-SO are shown in blue, while the detection boxes of other networks are shown in red. From Figure 7, it can be observed that YOLO-SO detected all cracks in test images 1 to 4, and the detection boxes completely covered the crack areas. In contrast, the other 5 networks exhibited inaccurate detection results. In test images 2 and 4, YOLOv5s, YOLOv4, YOLOv3, and SSD all missed cracks, resulting in poor detection performance. In test images 1 and 3, except for FasterR-CNN, the detection boxes of the other networks did not completely cover the crack areas. Although FasterR-CNN detected all cracks, there were many overlapping detection boxes, leading to inaccurate results. Therefore, the comprehensive performance of YOLO-SO surpasses that of other classical object detection neural networks, demonstrating stronger detection capabilities for images of cracks.

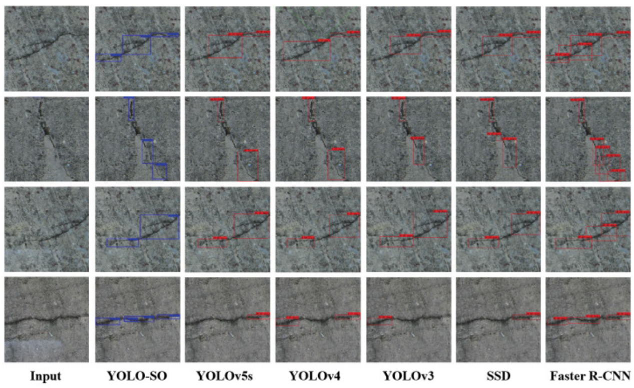

Figure 7. Detection results of different models (part)

5 Conclusion

Aiming at crack detection, this paper improves the YOLOv5s network and proposes a real-time crack detection method based on object detection, named YOLO-SO. And it solves the problem of insufficient accuracy and generalization of small target detection in multiple scenes. The results show that the proposed method has the ability for efficient, accurate, and real-time crack detection. Through comparative experiments with other object detection neural networks and multi-scene tests, the following conclusions are obtained:

1) Compared with other target detection neural networks, YOLO-SO reduces the number of parameters and size of the model while maintaining the detection accuracy, which meets the accuracy requirements of crack detection. The mAP of YOLO-SO on the test set is 89.5%, and the model size is 30.09MB, which can be deployed on the mobile terminal for real-time crack detection.

2) Based on the YOLOv5s network, the PADNet network is introduced, the CA attention mechanism is integrated, and the localization loss function is modified to CIoU, which can effectively improve network performance and enhance the detection ability of the network. Compared with YOLOv5s network, YOLO-SO reduces the parameter volume by 43.28%, reduces the computational load by 47.47%, and improves the detection accuracy by 2.18%.

3) For crack images in different scenes, YOLO-SO can adapt to different lighting environments and has strong robustness. The detection results of high-resolution crack images also show that the network has strong detection performance.

References

[1] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788.

[2] P. Singh, R. Krishnamurthi, AgriGuard: IoT-Powered Real-Time Object Detection and Alert System for Intelligent Surveillance, *International Journal of Performability Engineering*, Vol. 20, No. 4. pp. 232-241, April, 2024.

[3] A. Kumar, G. S. Lehal, Layout Detection of Punjabi Newspapers using the YOLOv8 Model, *International Journal of Performability Engineering*, Vol. 20, No. 3, pp. 186-193, March, 2024.

[4] Z.-N. Zheng, *Research on building crack detection and measurement methods based on deep learning and image processing technology*, MA. Sc. Thesis, Nanjing University of Science and Technology, Nanjing, China, 2021.

[5] Z. Xie, *Recognition of Concrete Surface Cracks Based on Fully Convolutional Network*, MA. Eng. Thesis, Northwest A&F University, Xianyang, China, 2022.

[6] Q.-Y. Yang, *Research on concrete crack detection based on U-Net variant network*, MA. Sc. Thesis, Changzhou University, Changzhou, China, 2022.

[7] F. Yu, C.-Y. Yu, Z.-Y. Tian, X.-X. Liu, J.-C. Cao, L. Liu, C.-H. Du, M.-H. Jiang, Intelligent Wearable System With Motion and Emotion Recognition Based On Digital Twin Technology, *IEEE Internet of Things Journal*, Vol. 11, No. 15, pp. 26314-26328, August, 2024.

[8] N. Huang, G. Cao, Improved Algorithm for Edge Detection of Building Structures and Cracks Based on IFC, *2023 International Conference on Data Science and Network Security (ICDSNS)*, Tiptur, India, 2023, pp. 1-6.

[9] F. Yu, Y. Zhang, H.-Y. Li, C.-H. Du, L. Liu, M.-H. Jiang, Phase Contour Enhancement Network for Clothing Parsing, *IEEE Transactions on Consumer Electronics*, Vol. 70, No. 1, pp. 2784-2793, February, 2024.

[10] M. Hao, *Research on Concrete Crack Detection and Evaluation Based on Deep Learning*, MA. Sc. Thesis, Xi'an University of Architecture and Technology, Xi'an, China, 2021.

[11] L. Zhang, Z.-C. Wang, L. Wang, Z. Zhang, X. Chen, L. Meng, Machine learning-based real-time visible fatigue crack growth detection, *Digital Communications and Networks*, Vol. 7, No. 4, pp. 551-558, November, 2021.

[12] J.-Y. Liu, X.-F. Zhu, X.-Y. Zhou, S.-H. Qian, J.-H. Yu, Defect Detection for Metal Base of TO-Can Packaged Laser Diode Based on Improved YOLO Algorithm, *Electronics*, Vol. 11, No. 10, Article No. 1561, May, 2022.

[13] H.-Y. Yu, Y. Li, D.-X. Zhang, An Improved YOLO v3 Small-Scale Ship Target Detection Algorithm, *2021 6th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, Kunming, China, 2021, pp. 560-563.

[14] F.-X. Yan, Y.-X. Xu, Improved Target Detection Algorithm Based on YOLO, *2021 4th International Conference on Robotics, Control and Automation Engineering (RCAE)*, Wuhan, China, 2021, pp. 21-25.

[15] M.-X. Wang, B.-L. Yang, X. Wang, C. Yang, J. Xu, B.-Z. Mu, K. Xiong, Y.-Y. Li, YOLO-T: Multitarget Intelligent Recognition Method for X-ray Images Based on the YOLO and Transformer Models, *Applied Sciences*, Vol. 12, No. 22, Article No. 11848, November, 2022.

[16] C.-Y. Liu, Y.-Q. Wu, J.-J. Liu, Z. Sun, H.-J. Xu, Insulator Faults Detection in Aerial Images from High-Voltage Transmission Lines Based on Deep Learning Model, *Applied Sciences*, Vol. 11, No. 10, Article No. 4647, May, 2021.

[17] P. Hurtik, V. Molek, J. Hula, M. Vajgl, P. Vlasanek, T. Nejezchleba, Poly-YOLO: higher speed, more precise detection and instance segmentation for YOLOv3, *Neural Computing and Applications*, Vol. 34, No. 10, pp. 8275-8290, May, 2022.

[18] M.-R. Ju, H.-B. Luo, Z.-B. Wang, M. He, Z. Chang, B. Hui, Improved YOLO V3 Algorithm and Its Application in Small Target Detection, *Acta Optica Sinica*, Vol. 39, No. 7, Article No. 0715004, July, 2019.

[19] M. W. Adou, H.-R. Xu, G.-H. Chen, Insulator Faults Detection Based on Deep Learning, *2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID*), Xiamen, China, 2019, pp. 173-177.

[20] C.-Y. Liu, Y.-Q. Wu, J.-J. Liu, J.-M. Han, MTI-YOLO: A Light-Weight and Real-Time Deep Neural Network for Insulator Detection in Complex Aerial Images, *Energies*, Vol. 14, No. 5, Article No. 1426, March, 2021.

[21] T.-H. Wu, T.-W. Wang, Y.-Q. Liu, Real-Time Vehicle and Distance Detection Based on Improved YOLO v5 Network, *2021 3rd World Symposium on Artificial Intelligence (WSAI)*, Guangzhou, China, 2021, pp. 24-28.

[22] S.-K. Ma, H.-T. Jiang, L. Chang, C. Zheng, Lane line detection based on attention mechanism and feature aggregation, *Microelectronics and Computer*, Vol. 39, No. 12, pp. 40-46, December, 2022.

Biographies

Jiaqi Wu was born in Hubei, China. His major interests are in the areas of computational intelligence, chaos encryption, and information security.

Jingjing Zhou works at the School of Artificial Intelligence, Hubei Business College. Her research interests include cloud security, chaos encryption, and information security.

Tian Zhang works at the School of Artificial Intelligence, Hubei Business College. His major interests are in the areas of security, chaos encryption, and information security.