International Journal of Science and Engineering Applications
Volume 14-Issue 11, 77 – 88, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1411.1014

Intelligent Detection Technology for the Quality of Old Residential Buildings

Liu Fu
School of Electronic
Information and
Electrical Engineering
Yangtze University

Abstract: This study focuses on addressing the technical bottlenecks of traditional manual inspection methods for old residential building defects, such as low efficiency, strong subjectivity, and high missed detection rates. Leveraging the advancement of artificial intelligence and computer vision technologies, we propose an intelligent detection approach based on the improved YOLOv11-seg model to identify common building surface defects (cracks, spalling, and algae).First, an independent dataset was constructed, containing 11,702 color images (640×640 pixels) with detailed manual annotations, which were randomly divided into training, validation, and test sets at appropriate ratios to ensure the reliability of model training and evaluation. Second, targeting the challenges of multi-category, multi-scale defects and complex backgrounds in old buildings, two optimization modules—DCA (Deformable Convolution and Channel Attention fusion structure) and DS_C3K2 (Backbone optimization structure integrating Dynamic Convolution and SE Attention)—were introduced into the original YOLOv11-seg architecture. These modules enhance the model's ability to perceive spatial features and express channel selectivity, thereby improving the accuracy of defect contour extraction and reducing missed and false detections. Experimental results show that the improved YOLOv11-seg model outperforms YOLOv5-seg and YOLOv8-seg in key metrics: in bounding box detection, its mAP50 reaches 0.845 and precision is 0.872; in mask segmentation, its mAP50 is 0.812 and recall is 0.747. Additionally, an integrated visual operation interface was developed to support functions such as image upload, automatic detection, segmentation visualization, and report generation, reducing the technical threshold for engineering application. This research provides an efficient and reliable technical solution for the quality detection of old residential buildings, supporting urban renewal and public safety governance.

Keywords: Old Residential Buildings; Defect Detection; YOLOv11-seg; Instance Segmentation; Deep Learning; Computer Vision; Dataset Construction; Model Optimization

1. INTRODUCTION

With the continuous advancement of urbanization, a large number of residential buildings have been constructed rapidly over the past few decades. Entering a new stage, some of the residential buildings built in the early days have gradually aged. According to statistics, the total area of existing buildings in China exceeds 60 billion square meters, among which old residential buildings account for a relatively large proportion. These buildings have been affected by natural environments (such as settlement, wind and rain erosion, temperature and humidity changes) and human factors (such as improper use and insufficient maintenance) for a long time, resulting in the gradual degradation of their structural performance and service functions. Common defects in old buildings include leakage, cracks, hollowing, spalling and structural loosening. These defects not only affect residential safety and comfort, damage the urban landscape, but also even threaten the lives and property safety of residents. Therefore, conducting scientific and systematic detection and evaluation has become a key link in urban renewal and public safety assurance. At present, building quality detection still mainly relies on manual inspection and empirical judgment, which has problems such as low efficiency, high cost, limited data volume and strong subjectivity, making it difficult to meet the actual needs of high-density and large-scale building detection. Moreover, its ability to identify hidden damage is insufficient, which easily leads to the omission of potential risks. However, with the development of technologies such as artificial intelligence, computer vision, unmanned aerial vehicles (UAVs) and the Internet of Things (IoT), building detection is accelerating its transformation towards intelligence and automation. Relying on deep learning and image recognition technologies, it is possible to achieve efficient, objective and repeatable identification and classification of building surface defects, significantly improving the accuracy and efficiency of detection. Therefore, exploring AI-based detection methods for defects in old residential buildings not only helps to break through the technical bottlenecks of traditional detection methods, but also provides new ideas and technical support for the renewal and transformation of old urban communities, building health monitoring and urban safety governance in China.

In recent years, deep learning technology has made remarkable progress in the field of building defect detection, greatly improving the recognition performance of computer vision. However, most of the current studies still focus on the detection of single-type defects such as cracks and spalling areas, and it is difficult to effectively deal with the actual situation where multiple defects with complex forms coexist and overlap in old buildings. Although multi-category defect detection methods have developed to a certain extent, they still face problems of missed detection and false detection in typical engineering scenarios such as small targets, blurred edges, uneven illumination and complex backgrounds, leading to a significant decline in detection accuracy. In addition, high-precision algorithms are usually accompanied by high computational costs, which makes it difficult to meet the real-time requirements of mobile devices and edge devices, limiting their practicality.

Traditionally, defect detection in old residential buildings has mainly relied on manual inspection. However, this method is not only time-consuming but also greatly affected by human factors, making it difficult to ensure the accuracy and consistency of detection results. To improve detection efficiency, some studies have integrated photoelectric detection technology for defect identification. For instance, Meng, X.B. [1] used eddy current sensors to evaluate the thickness of wall coatings; Wang, G. [2] adopted multi-frequency AC magnetic flux leakage testing to detect surface defects in building steel; Jing, X. [3] utilized infrared thermal imaging technology to detect defects in newly constructed buildings. Although these methods have certain advantages in specific scenarios, their implementation costs are high and their application scenarios are limited, making it difficult to promote them on a large scale.In modern times, deep learning has become the primary method for detecting defects in old residential buildings [4]. For example, Cheng, M. [5] proposed a baseline model called Deblur-DetNet for defect detection in ancient buildings; Yang, Z. [6] used the DeeplabV3+ model based on MobileNetV2 to detect building defects. While these methods have achieved good results in terms of accuracy, they consume significant computational resources, and further improvements in model lightweighting are still needed to meet the requirements of practical applications.As a leading single-stage object detection algorithm, the YOLO series has been widely used in building defect detection due to its efficient detection performance. In recent years, relevant studies have continuously made breakthroughs in expanding model functions and enhancing scenario adaptability. Ye, G. [7] used YOLOv7 to detect surface cracks in concrete; this method maintained excellent performance and robustness when dealing with cracks of different sizes and images affected by various noise levels and types, providing an effective solution for the automated identification of micro-cracks. Cai, P. et al. [8] proposed the YOLOv11-PC model, which optimized feature extraction through a cross-view contrastive learning strategy for concrete structure defect detection scenarios. This model significantly improved the recognition accuracy of multiple types of defects (such as cracks and spalling) under complex backgrounds, breaking the reliance of traditional models on single-type defect detection. Additionally, Wang, Q. et al. [9] developed the Mask-YOLO algorithm, which added instance segmentation functionality to the YOLO framework. This enables pixel-level localization and contour extraction of building defects, providing technical support for refined detection needs such as defect area quantification and morphological analysis. Despite the significant progress achieved, most existing studies still have limitations, which may hinder the wider practical application of such models.

2. RESEARCH SCHEME

2.1 Scheme Design

At present, mainstream target detection methods can be roughly divided into three categories: two-stage detection algorithms (such as Faster R-CNN), one-stage detection algorithms (such as SSD and YOLO series) and end-to-end methods based on Transformer (such as DETR). Although two-stage methods perform well in terms of accuracy, they are not suitable for application scenarios with high real-time requirements due to their complex structure and slow inference speed. One-stage methods such as SSD have a fast inference speed, but they have certain defects in the detection of small targets and dense targets. The DETR model based on Transformer has global modeling ability, but it takes a long time to train and consumes a lot of computing resources, making it difficult to be widely deployed in embedded or edge devices.

Among the YOLO series, YOLOv5 has been widely recognized for its lightweight structure and high-speed detection. YOLOv8 has improved the overall accuracy by introducing anchor-free detection and deeper semantic modeling, but it still faces problems of blurred boundaries and insufficient feature fusion, which affects the accurate identification of complex and diverse defects. In addition, traditional single-defect detection models are difficult to cope with the complex environment where multiple types and multi-scale defects coexist in old residential buildings, and multi-category detection models also face challenges in stability and accuracy when there is strong noise and texture interference. In contrast, on the basis of inheriting the efficient and lightweight characteristics of the YOLO series, YOLOv11 has significantly enhanced the ability of fine-grained feature extraction and adaptation to complex scenarios. Through the optimized network structure and improved feature fusion mechanism, it has achieved more accurate boundary positioning and multi-scale target detection, effectively reducing the rate of missed detection and misjudgment. At the same time, YOLOv11 maintains excellent inference speed, making it suitable for deployment in industrial application environments with limited resources. Its advantages of high accuracy, high speed and model lightweight make it an ideal and reliable detection solution in practical industrial scenarios such as building surface defect detection and metal defect identification.

To sum up, this project selects YOLOv11-seg as the basic architecture, combines target detection with instance segmentation, and conducts targeted optimization for the defect detection task of old buildings. It realizes the refined contour extraction of defect areas, significantly improves the accuracy and stability of the model in the identification of multi-category and multi-scale defects, and enhances its practicality and robustness in complex environments. In addition, to improve the engineering availability and operational convenience of the model, we have designed and developed an integrated visual operation interface, enabling users to complete integrated operations such as image upload, automatic detection, segmentation visualization and detection report generation through this interface. This significantly reduces the technical threshold for algorithm deployment and use, and facilitates promotion and application in engineering practice. The overall block diagram of the scheme is shown in Figure 1.

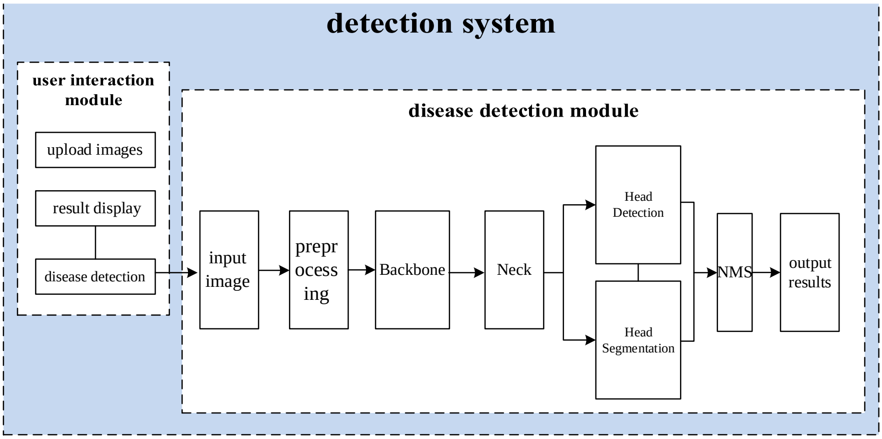


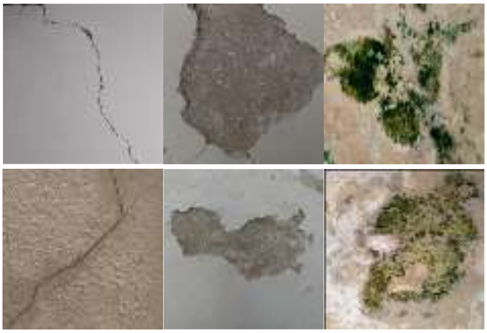
Figure 1 Overall Block Diagram of the Scheme

2.2 Dataset

The dataset used in this study is independently constructed by the research team, mainly covering three common types of defects on the building surface, including cracks, spalling and algae. It contains a total of 11,702 color images, with a unified image resolution of 640×640 pixels, all stored in JPG format.

To achieve high-quality supervised learning, each image in the dataset is labeled manually one by one. The labeling information is provided in the form of text files, including the accurate bounding box coordinates and category labels of defects, realizing refined and multi-category target labeling. Statistical data show that the dataset contains 4,545 crack images, 3,752 spalling images and 3,405 algae images. This detailed labeling information provides a solid foundation for training multi-category recognition models.

To ensure the reliability of the experiment and the universality of the model performance, the dataset is randomly divided into a training set, a validation set and a test set, which contain 9,437, 1,665 and 600 images respectively. Random division effectively avoids data sorting bias and ensures the balance of sample distribution in each subset, thereby improving the robustness of the model training process and the generalization ability in the test phase. Examples of each type of defect in the dataset are shown in Figure 2.



(a) (b) (c)
Figure 2 (a) Cracks (b) Spalling (c) Algae

2.3 Model Architecture Analysis

2.3.1 YOLOv11-seg Target Detection Algorithm

YOLOv11 is a new generation of single-stage target detection algorithm, designed based on a deep convolutional neural network architecture, with the advantages of lightweight structure, fast detection speed and high accuracy. This model inherits the core idea of the YOLO series, converting the target detection task into a unified regression problem, and can complete both target localization and classification in a single forward propagation, significantly improving the inference efficiency. YOLOv11-seg is the instance segmentation version of YOLOv11 and one of the latest generations of multi-task visual models in the YOLO series. It not only has the high-precision target detection capability of YOLOv11 (outputting target categories and bounding boxes), but also adds a pixel-level target contour segmentation function on the basis of detection, which can complete localization, classification and accurate shape extraction at the same time. Compared with previous-generation models, YOLOv11-seg has been systematically optimized in terms of feature expression ability, multi-scale fusion strategy and regression loss function. It shows stronger robustness and accuracy especially in the detection tasks of small targets, targets with blurred edges and under complex backgrounds. Therefore, YOLOv11-seg is very suitable for deployment in application scenarios with high requirements for real-time performance and accuracy, such as building defect detection, metal defect identification and unmanned inspection. The network architecture diagram of the YOLOv11-seg algorithm is shown in Figure 3.

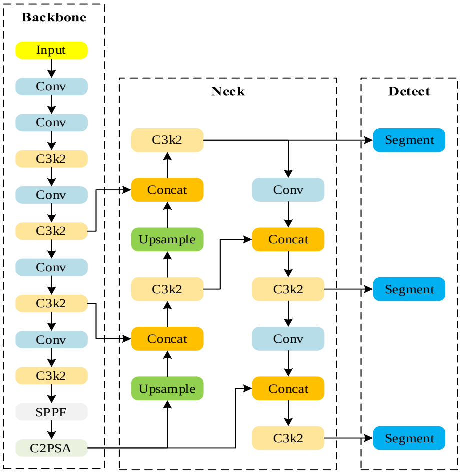


Figure 3 Network Architecture Diagram of the YOLOv11-seg Algorithm

2.3.2 Structure Overview

YOLOv11-seg inherits the modular design idea consistently adopted by the YOLO series. Its overall structure is composed of three functional modules: Backbone (main network), Neck (neck network) and Head (detection head), forming an end-to-end and highly integrated target detection system. Each module in YOLOv11-seg has been customized to achieve the best balance between detection accuracy, inference speed and model lightweight.

(1) Backbone

As the core module for feature extraction, the Backbone is responsible for extracting multi-scale and multi-level semantic features from the original image. In this part, YOLOv11-seg introduces an efficient C3K2 module structure, which integrates shallow detailed features and deep semantic information. By stacking multiple small convolution blocks with 3×3 convolution kernels, combined with residual connections and lightweight bottleneck structures, the nonlinear expression ability of the network is effectively enhanced, and the number of model parameters and computational costs are significantly reduced. This design not only maintains the integrity and discriminability of features but also improves the modeling ability for small targets and complex backgrounds.

(2) Neck

The Neck is used to further fuse feature maps of different scales from the Backbone, so as to improve the model's ability to perceive multi-scale targets. YOLOv11-seg draws on the idea of the Path Aggregation Network (PAN) structure and adopts a design strategy that combines the Feature Pyramid Network (FPN) with a path enhancement mechanism to realize the information transmission and feature fusion from bottom to top and top to bottom. In this process, the spatial information in the low-level features and the semantic information in the high-level features are fully integrated, thereby enhancing the model's robustness to the shape, size and position of targets.

(3) Head

The Head is responsible for directly predicting the target's position (bounding box), category probability and confidence score based on the fused feature maps. YOLOv11-seg inherits the single-stage regression design idea of the YOLO series and performs parallel detection of targets of different sizes through multiple scale branches. Each detection branch is composed of a set of lightweight convolution layers to reduce the delay during inference and improve the real-time detection performance. This Head effectively controls the false positive rate while maintaining a high recall rate, and has a good speed-accuracy trade-off characteristic, making it suitable for target detection tasks in a variety of complex scenarios.

To sum up, YOLOv11-seg integrates a few lightweight and performance-enhancing strategies in its structural design. It can significantly improve the adaptability to complex environments, small-sized targets and multi-scale scenarios while ensuring the detection speed, providing a stable and reliable technical foundation for high-accuracy and high-efficiency target detection tasks.

2.3.3 Workflow of the YOLOv11-seg Algorithm

YOLOv11-seg is a single-stage, end-to-end target detection and instance segmentation algorithm. Its overall workflow includes four stages: input preprocessing, feature extraction, feature fusion, and target detection and segmentation prediction.

(1) Input Preprocessing YOLOv11-seg receives input images of any size and standardizes the images to a fixed size (such as 640×640) through a series of preprocessing operations (such as scaling, normalization and color enhancement) to meet the input requirements of the network structure. This step ensures the stability and consistency of the subsequent feature extraction process.

(2) Feature Extraction (Backbone)The preprocessed images are sent to the Backbone for initial feature extraction. The Backbone part of YOLOv11-seg adopts an improved C3K2 module as the core structure. This module combines multi-layer 3×3 convolution, small residual connections and lightweight bottleneck structures, which can effectively extract multi-scale semantic information. While maintaining low computational overhead, it enhances the feature expression ability and the convergence performance of the model.

(3) Feature Fusion (Neck)In the feature fusion stage, YOLOv11-seg adopts a structure combining FPN and PAN to fully integrate the spatial detailed features of the low levels and the semantic abstract features of the high levels, thereby improving the model's robustness in small target recognition and complex background scenarios.

(4) Detection and Segmentation Prediction (Head)The fused multi-scale features are transmitted to the segmentation detection Head module to realize the simultaneous prediction of target position, category and mask. YOLOv11-seg adopts the Anchor-Free strategy in the detection branch to avoid the limitation of prior anchors; in the segmentation branch, it realizes instance segmentation by generating pixel-level binary masks. The three-scale detection Heads can predict the bounding boxes, category probabilities and corresponding masks of targets of different sizes in parallel. Finally, the detection results are processed by the Non-Maximum Suppression (NMS) strategy to remove redundant boxes, and the target boundaries and segmentation regions with high confidence are output.

2.4 Improvement and Analysis of Experimental Methods

To further improve the feature expression and learning ability of the YOLOv11-seg model in complex scenarios, this study introduces two key enhancement modules on the basis of the original network architecture: the DCA module and the DS_C3K2 module, which systematically optimize the spatial feature perception ability and channel selective expression ability respectively.

(1) DCA Module: Deformable Convolution and Channel Attention Fusion Structure The DCA module is composed of Deformable Convolution and the Efficient Channel Attention (ECA) mechanism. By introducing a spatially variable

convolution kernel offset mechanism, this module enables the model to have stronger flexibility and perception ability when dealing with unstructured deformed targets and complex background images. Specifically, Deformable Convolution can automatically adjust the sampling position according to the change of the target shape in the image, thereby enhancing the receptive field coverage and feature response of the model to irregular defect areas (such as cracks, erosion, rust, etc.). At the same time, the ECA attention mechanism weights the channel information in a lightweight convolution way, strengthens the feature response to key regions, and effectively improves the robustness and detection accuracy of the model. The structure diagram of the DCA module is shown in Figure 4.

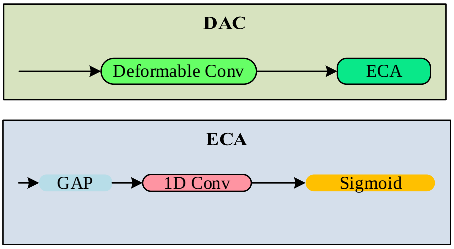


Figure 4 Structure Diagram of the DCA Module

(2) DS_C3K2 Module: Optimization of the Backbone Structure by Integrating Dynamic Convolution and the SE Attention Mechanism The DS_C3K2 module introduces Dynamic Convolution, which can dynamically generate convolution kernel combination weights according to different input images, enabling the model to have stronger input adaptability and nonlinear feature expression ability. Then, by integrating the Squeeze-and-Excitation (SE) attention mechanism, the channel responses of the convolution output are recalibrated, further improving the selective expression ability of feature channels. Without significantly increasing the number of model parameters, the DS_C3K2 module effectively improves the modeling accuracy of the network for multi-scale features and heterogeneous textures. The structure diagram of the DS_C3K2 module is shown in Figure 5.

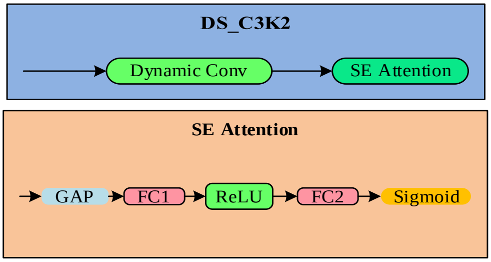


Figure 5 Structure Diagram of the DS_C3K2 Module

The network structure of YOLOv11-seg after integrating the two modules (DCA and DS_C3K2) is shown in Figure 6.

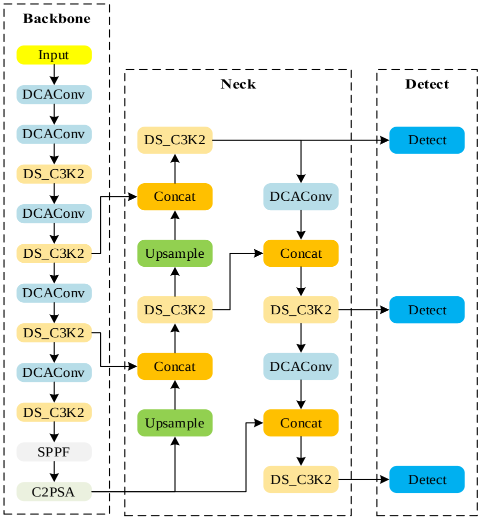


Figure 6 Network Architecture Diagram of the Improved YOLOv11-seg Algorithm

2.5 Introduction to Evaluation Indicators

(1) Precision

Precision measures the proportion of correctly detected positive samples among the results predicted as positive samples by the model, reflecting the reliability of the detection results. Its calculation formula is shown in Equation (1). Among them, True Positive (TP) represents the number of samples correctly classified as positive, that is, the number of instances that are actually positive and classified as positive by the classifier; False Positive (FP) represents the number of samples incorrectly classified as positive, that is, the number of instances that are actually negative but classified as positive by the classifier. A higher precision indicates a lower false positive rate of the model.

$$Precision = \frac{TP}{TP + FP} \qquad (1)$$

(2) Recall

Recall measures the model's ability to identify positive samples, reflecting the coverage of the detection results. Its calculation formula is shown in Equation (2). Among them, False Negative (FN) represents the number of samples incorrectly classified as negative, that is, the number of instances that are actually positive but classified as negative by the classifier. A higher recall indicates a lower missed detection rate than the model.

$$Recall = \frac{TP}{TP + FN} \qquad (2)$$

(3) Intersection over Union (IoU)

IoU is used to evaluate the overlap between two bounding boxes. It requires a real bounding box and a predicted bounding box, and its calculation method is the area of intersection of the two detection boxes divided by the area of union of the two detection boxes, as shown in Figure 7. Through IoU, it can be judged whether the detection is valid or invalid.

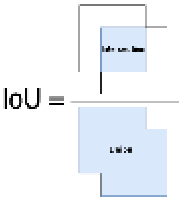


Figure 7 Schematic Diagram of Intersection over Union (IoU)

(4) Mean Average Precision (mAP)

mAP refers to the area enclosed by the precision-recall function curve and the coordinate axis under a certain IoU. A higher mAP means that the model can detect target objects more accurately. Among them, mAP50 refers to the area under the precision-recall function curve when the detection boxes with IoU greater than 0.5 are regarded as positive examples, which is usually used as the basic indicator for general target detection projects; mAP50-95 refers to the average of the areas under the precision-recall function curve when IoU takes a value every 0.05 step from 0.5 to 0.95, and detection boxes greater than the IoU of each step are regarded as positive examples. It is usually used to evaluate the model's performance under different overlap degrees.

3. EXPERIMENTAL PART

3.1 Experimental Setup

All experiments in this study are carried out in the following software environment: Python 3.8, PyTorch 1.8.1 and CUDA 11.1. All training images are uniformly scaled to 640×640 pixels, aiming to balance image resolution and model computing efficiency. This resolution can retain sufficient image details for the model to extract effectively, while avoiding introducing excessive computing costs.

The total number of training epochs of the model is set to 100, and the batch size is 8. This configuration can ensure the stable training of the model while making rational use of computing resources. The number of iterations is sufficient to ensure that the model fully learns from the training data, and the setting of the batch size helps to improve the stability of model convergence and the overall training efficiency.

In terms of optimization, the Stochastic Gradient Descent (SGD) algorithm is used in the training process, with an initial learning rate of 0.01 and a momentum factor of 0.937. The setting of the above hyperparameters is based on empirical rules and the best practices of previous studies, aiming to achieve a good balance between training speed and model convergence. All hyperparameters have been verified through systematic experiments, and the final configuration performs best on the validation set.

3.2 Experimental Results

The operation results of the YOLOv11-seg model used in this project in the instance segmentation mode are shown in Figure 8. The eight graphs on the left show the change trends of various losses in the training and validation stages respectively. Among them, box_loss represents the bounding box regression loss, and a lower value indicates higher accuracy of the model in target position

Table 1 shows the performance comparison of different models on the same validation set. The evaluation indicators include Precision (P), Recall (R), mAP50 and mAP50-95, which are for the bounding box (Box) detection and mask (Mask) segmentation tasks respectively.

In the bounding box detection, the mAP50 of YOLOv11-seg (ours) reaches 0.845, which is 0.23 and 0.14 higher than that of YOLOv5-seg and YOLOv8-seg respectively, and slightly 0.031 higher than that of the original YOLOv11-seg; the mAP50-95 is 0.669, which is 0.242 and 0.152

prediction; dfl_loss also belongs to the bounding box regression loss, and as a more advanced loss function, it can further optimize the bounding box positioning effect for hard-to-detect targets; cls_loss represents the category classification loss, and a lower value indicates that the model has higher accuracy in distinguishing target categories; seg_loss is the segmentation loss function, which is used to measure the difference between the model's segmentation results and the real labels. The eight graphs on the right show multiple evaluation indicators in the model inference stage, and the relevant values are detailed in the subsequent tables of this chapter. higher than that of YOLOv5-seg and YOLOv8-seg respectively; the precision is the highest, reaching 0.872.

In the mask segmentation, the mAP50 of this model is 0.812, which is 0.157, 0.125 and 0.005 higher than that of YOLOv5-seg, YOLOv8-seg and the original YOLOv11-seg respectively; the mAP50-95 is 0.508, with the increase ranges of 0.104, 0.031 and 0.004 respectively. Both the precision and recall (0.855, 0.747) are higher than those of the comparison models.

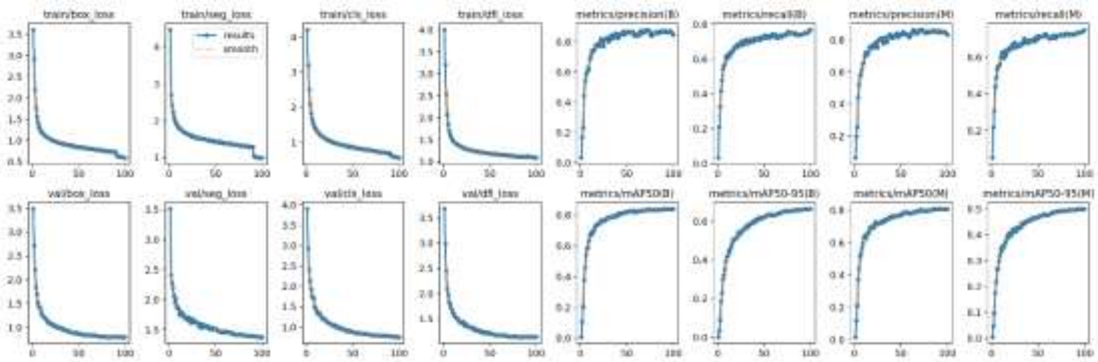


Figure 8 Operation Results of the YOLOv11-seg Model

Table 1 Performance Results of Different Models

	Box				Mask			
Model Type	P	R	mAP50	mAP50-95	P	R	mAP50	mAP50-95

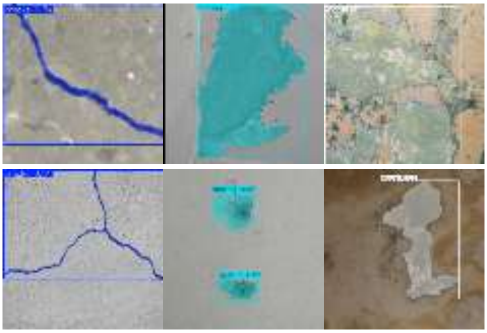
YOLOv5-seg	0.646	0.582	0.612	0.427	0.638	0.611	0.655	0.404
YOLOv8-seg	0.727	0.611	0.707	0.517	0.708	0.630	0.687	0.477
YOLOv11-seg	0.818	0.785	0.814	0.672	0.829	0.737	0.807	0.504
YOLOv11-seg(ours)	**0.872**	**0.771**	**0.845**	**0.669**	**0.855**	**0.747**	**0.812**	**0.508**

Table 2 compares the performance of YOLOv5-seg, YOLOv8-seg, YOLOv11-seg and the improved model YOLOv11-seg (ours) in the detection tasks of three types of defects (cracks, spalling and algae). The results show that YOLOv11-seg (ours) is superior to other models in all evaluation indicators. Especially in crack detection and spalling detection, its precision reaches 0.901 and 0.890 respectively. In addition, the YOLOv11-seg used in our study also shows obvious advantages in the target detection and segmentation performance of cracks and algae. In conclusion, the YOLOv11-seg model used in this project performs more accurately and stably in the detection of multiple defects, which proves its effectiveness and practical value in the identification of complex building defects.

Table 2 Detection Results of Different Models on Three Types of Building Defects

		Box				Mask			
Model Type	class	P	R	mAP50	mAP50-95	P	R	mAP50	mAP50-95
YOLOv5-seg	Crack	0.638	0.611	0.623	0.417	0.607	0.593	0.581	0.19
YOLOv8-seg		0.727	0.611	0.690	0.487	0.727	0.611	0.690	0.27
YOLOv11-seg		0.865	0.78	0.827	0.663	0.796	0.714	0.724	0.28
YOLOv11-seg(ours)		**0.901**	**0.79**	**0.847**	**0.688**	**0.837**	**0.729**	**0.754**	**0.3**
YOLOv5-seg	Spalling	0.611	0.570	0.592	0.403	0.603	0.567	0.577	0.423
YOLOv8-seg		0.732	0.643	0.679	0.589	0.705	0.614	0.662	0.605
YOLOv11-seg		0.872	0.829	0.892	0.722	0.886	0.836	0.897	0.7
YOLOv11-seg(ours)		**0.89**	**0.818**	**0.892**	**0.719**	**0.9**	**0.82**	**0.892**	**0.704**
YOLOv5-seg	Algae	0.587	0.460	0.502	0.387	0.568	0.423	0.497	0.32
YOLOv8-seg		0.667	0.587	0.648	0.523	0.638	0.601	0.611	0.487
YOLOv11-seg		0.806	0.698	0.801	0.611	0.821	0.702	0.797	0.519
YOLOv11-seg(ours)		**0.825**	**0.705**	**0.796**	**0.601**	**0.829**	**0.694**	**0.79**	**0.521**

The detection and segmentation results are shown in Figure 9. It can be seen that the model has high detection accuracy for various defects. Especially, it shows clear performance in the edge segmentation of small cracks and achieves complete coverage of fragmented targets in spalling areas. This is due to the fact that the YOLOv11-seg model used in this project integrates the two modules (DCA and DS_C3K2). While maintaining low computational complexity, it can still maintain high accuracy and robustness in scenarios where targets are small or have significant shape changes.



(a) (b) (c)

Figure 9 (a) Crack Detection Results (b) Spalling Detection Results (c) Algae Detection Results

To sum up, the YOLOv11-seg used in this project is superior to some YOLO series models in indicators such as mAP50, recall and precision, and has good deployment efficiency and engineering practicality.

References

[1] Meng, X. B., Lu, M. Y., Yin, W. L., Bennecer, A., and Kirk, K. J. 2021. Evaluation of Coating Thickness Using Lift-Off Insensitivity of Eddy Current Sensor. Sensors, 21(2), 419.

[2] Wang, G., Xiao, Q., Gao, Z. H., Li, W., Jia, L., Liang, C., and Yu, X. 2022. Multifrequency AC Magnetic Flux Leakage Testing for the Detection of Surface and Backside Defects in Thick Steel Plates. IEEE Magn. Lett., 13, 8102105.

[3] Jing, X., Yang, X.-Y., Xu, C.-H., Chen, G., and Ge, S. 2012. Infrared thermal images detecting surface defect of steel specimen based on morphological algorithm. J. China Univ. Pet., 36, 146–150.

[4] Zhu, Z., and Al-Qadi, I. L. 2023. Crack Detection of Asphalt Concrete Using Combined Fracture Mechanics and Digital Image Correlation. J. Transp. Eng. Part B Pavements, 149, 04023012.

[5] Cheng, M., Zhang, X., Xia, L., et al. 2025. Visual defect detection for historical building Preservation. Expert Systems with Applications, 291, 128376.

[6] Yang, Z. 2024. Consider Computer Vision for Building Component Recognition and Defect Detection. In 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA) (pp. 457–460).

[7] Ye, G., Qu, J., Tao, J., Dai, W., Mao, Y., and Jin, Q. 2023. Autonomous surface crack identification of concrete structures based on the YOLOv7 algorithm. J. Build. Eng., 73, 106688.

[8] Cai, P., and Zhang, L. 2025. PConv: Pinwheel convolution for enhanced feature representation. Measurement Science and Technology, 36(8), 085401. https://doi.org/10.1088/1361-6501/ad123c

[9] Wang, Q., Chen, Z., and Liu, H. 2025. Unified real-time instance segmentation with Mask-YOLO. IEEE Transactions on Image Processing, 34, 1120–1135. https://doi.org/10.1109/TIP.2025.123456