Research on Building Crack Detection Based on YOLOv11-Seg

Xiaoying An*, Siqin Shi, Xiaofeng Wu, Chenguang Ma, Kezierkailedi Bahezhuoli
China Institute of Building Standard Design & Research, Beijing, China
**Corresponding Author*

Abstract: To investigate the performance of convolutional neural network models in the field of building crack detection, this study selected the lightweight and efficient YOLOv11-seg model for object detection experiments on building cracks. By using polygons to annotate cracks at the pixel level, the model calculated experimental results for both bounding boxes and masks. The results show that the accuracy and recall rate of object detection both exceed 70%, indicating that the YOLOv11-seg model performs well in detecting building cracks. When the IoU threshold is 0.5, the model achieves a mean average precision (mAP) of over 75%, demonstrating its strong recognition capability for building cracks under moderate precision requirements. However, when the IoU threshold ranges from 0.5 to 0.95, the average mAP significantly decreases, and the mAP for bounding box prediction is notably higher than that for mask prediction, indicating that the model's accuracy in predicting crack edges is insufficient. These experimental results suggest that the YOLOv11-seg model can rapidly locate building cracks but still requires improvements in the precise detection of crack edges.

Keywords: Yolov11-Seg; Cracks; Object Detection

1. Introduction

As a common damage in buildings, cracks not only affect the living experience of buildings, but also have an impact on their load-bearing capacity. At present, the main method for detecting cracks in buildings is manual detection, but manual detection has problems such as low detection efficiency and strong subjectivity, making it difficult for manual detection of high-rise and super high-rise buildings. Meanwhile, in recent years, with the continuous development of image recognition technology, many scholars have applied image recognition technology to the field of crack recognition to achieve automatic identification of building cracks.

Traditional crack recognition is based on image processing techniques, including several steps such as image processing, feature extraction, and detection classification. Early image processing techniques were mainly based on the linear features of cracks. Linear features based on cracks, such as Ayenu et al. [1], were used to identify road cracks using the edge detection algorithm Sobel. Li et al. [2] used algorithms to automatically identify the starting and ending points of cracks, and extracted features of road surface cracks based on their continuity. However, in complex backgrounds, there may be interference from other linear shapes in the background of crack images. Therefore, some researchers have implemented crack feature extraction based on the difference in grayscale values between crack pixels and background pixels. Gavil et al. [3] added gray values to crack areas based on the assumption that the crack area is darker than the surrounding area, in order to distinguish between crack and non crack areas. Yamaguchi et al. [4] achieved the goal of quickly identifying cracks by setting thresholds to binarize the cracks and background colors based on their differences. However, methods that rely on different grayscale values of cracks and background are not suitable for complex lighting conditions, and methods based on different grayscale values and linear shapes of cracks require pre-set thresholds or differential operators and variational theory, making calculations more complex.

With the gradual maturity of artificial intelligence technology, some scholars have applied this technology to the field of crack recognition. In terms of road and bridge crack recognition, Liang et al. [5] constructed a large-scale bridge crack dataset and proposed a bridge crack detection algorithm based on improved GooLeNet, which improved the accuracy of crack recognition and reduced the time required for crack recognition. The cracks were accurately located using a sliding window and the length and width of the cracks were calculated using a skeleton extraction algorithm. Wang et al. [6] preprocessed the images of cracks, expanded the dataset, and used semantic segmentation algorithms to identify road cracks. They used an improved UNet MobileNetV3 network model to reduce the parameter count of the classical UNet model and optimize the structure. Wu [7] proposed an improved YOLOv7 Tiny road defect detection algorithm, and aimed at the problem of low detection accuracy caused by dense road potholes and large differences in pothole shapes from different perspectives. An improved RSG-YOLO model was proposed, and both models showed a certain improvement in evaluation indicators compared to the original model. In terms of identifying building cracks, Miao8] developed a pixel level recognition method based on deep convolutional neural networks for surface cracks, concrete spalling, and exposed steel bars in concrete components. He also developed efficient post-processing techniques for crack damage, effectively improving the accuracy and precision of crack recognition. Furthermore, he established a degradation assessment model for the mechanical properties of earthquake damaged concrete components based on deep convolutional neural networks. Wang [9] constructed a crack recognition model called Cracknet based on U-Net and proposed an optimization algorithm for crack recognition using CrackNet output as a mask combined with Otsu. The inverse perspective error correction was applied to reduce the measurement error of crack width caused by the camera optical axis not being parallel to the component plane vector.

As the most effective deep learning model for image processing in the current field of computer vision, Convolutional Neural Networks (CNNs) have the ability to automatically learn and recognize image features. When identifying cracks, the shape of the cracks usually presents a slender and irregular distribution, with a small proportion in the image, which belongs to the problem of small object detection. In addition, the lighting conditions of the building surface background are often variable, which makes it difficult to significantly improve the accuracy of crack detection. The width and length of cracks in some key structural components can have an impact on structural safety, so when identifying cracks, attention should also be paid to the accuracy of crack edge recognition. Therefore, this study aims to train a convolutional neural network model using a pixel level segmented dataset of building cracks, so that the model can extract image features of cracks through automatic learning of annotated data, and obtain the recognition accuracy of cracks and the recognition accuracy of crack edges.

2. Overview of Relevant Theories

Convolutional neural networks are mainly composed of input layers, convolutional layers, pooling layers, fully connected layers, and output layers. Convolutional layers extract local features through sliding windows and use learnable filters to map features of the input image; The pooling layer (usually max pooling or average pooling) is used to reduce the spatial dimension of feature maps and enhance the translational invariance of the model; The fully connected layer integrates the extracted high-level features and outputs the final prediction result. Compared with traditional image processing methods, CNN has significant advantages in identifying building cracks. Firstly, CNN can automatically learn discriminative features of cracks without the need for manually designed complex feature extraction algorithms; Secondly, CNN has strong robustness to changes in lighting and background interference, and can adapt to crack images under different shooting conditions; Furthermore, through data augmentation and transfer learning, CNN can achieve good performance on relatively small annotated datasets. For scenes where cracks are slender, irregular structures in the building background and have low contrast with the background, CNN can effectively capture this special texture pattern through a local sensory weight sharing mechanism, while a multi-level network structure can integrate contextual information of different scales to improve the accuracy of crack detection.

The commonly used object detection models include one-stage models and two-stage models, each with its own advantages in the field of object detection [10-13]. One stage models such as YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector) are known for their high efficiency, directly predicting the category and location of targets in the network; Two stage models such as Faster R-CNN and Mask R-CNN achieve higher detection accuracy through a two-stage process of “region proposal + classification regression”. In the task of identifying cracks on exterior walls, both types of models have been widely used, but they need to have their own advantages according to specific scene requirements. Two-stage models include Fast R-CNN and Mask R-CNN, which consists of two parts: Regional Proposal Network (RPN) and Detection Network. RPN first generates candidate regions (Region Proposals) that may contain targets, and then the detection network classifies and performs bounding box regression on these proposed regions. This two-stage process, although computationally intensive, typically achieves higher positioning and recognition accuracy. One-stage models mainly include YOLO, SSD, etc., which integrate target localization and classification into a single forward propagation during computation, thus achieving higher computational efficiency. Compared to traditional two-stage models, the YOLO (You Only Look Once) series models have the advantages of high efficiency and lightweight, making them more suitable for rapid engineering application detection. The workload of building crack detection is usually large, and it requires rapid screening at the construction site. The YOLO model's fast and efficient processing speed can meet the real-time requirements of crack detection. At the same time, in order to determine the impact of cracks on structural safety, it is necessary to accurately determine the morphological information such as the length and width of cracks. The YOLOv11 seg model introduces instance segmentation function into the object detection framework, achieving the unity of object detection and instance segmentation. Therefore, this article chooses to use the YOLOv11 seg model to identify building cracks.

The detection principle of the YOLO model is to treat object detection as a single-stage regression problem. By dividing the image into grids and directly predicting bounding boxes and class probabilities for each grid cell, the detection speed is significantly improved. YOLOv11 was released by the Ultralytics team in September 2024, and its overall architecture continues the classic three-stage design of the YOLO series, which includes the backbone network, neck network, Neck network, and detection head. The network structure is shown in Figure 1.

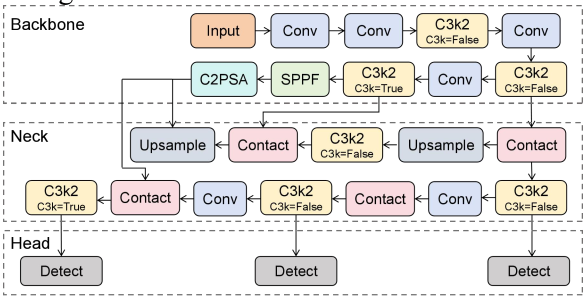

Figure 1. Yolov11 Network Structure

Compared with the previous generation YOLOv8 model, the main improvements in the overall architecture of YOLOv11 include the use of C3K2 module instead of C2F module in the backbone network. The C3K2 module reduces computational complexity and enhances the detection ability of small targets through parallel convolution and flexible parameter configuration, while ensuring feature extraction capability; Added C2PSA module after sppf layer, introduced parallel spatial attention mechanism, enhanced attention to key regions, and improved detection accuracy; In the detection head, depthwise separable convolution is used instead of conventional convolution. Depthwise separable convolution decomposes standard convolution into depthwise convolution and pointwise convolution, reducing computational and parameter complexity while preserving model accuracy. In addition, YOLOv11 has improved the convergence speed and classification accuracy by enhancing the loss function and dynamic weight allocation strategy through EIoU. The improvement of the above modules has comprehensively improved YOLOv11's accuracy while retaining detection speed, and enhanced its ability to detect small targets in complex scenes.

Starting from YOLOv8, the YOLO model has further added instance segmentation models. By using a single model to simultaneously output bounding boxes and masks, the model has the ability to perform both object detection and pixel level segmentation, making it more suitable for detecting small targets such as cracks. At the same time, on-site detection of building cracks can be divided into two steps.

The first step is to quickly screen for cracks and determine the location of cracks on large areas such as walls through on-site photography and other methods; Secondly, for structural cracks, in some scenarios, it is necessary to measure the length and width of the cracks to determine their impact on structural safety. In addition to locating the cracks, it is also necessary to accurately identify the edges of the cracks. To verify the model's ability to quickly identify and locate building cracks, as well as its accuracy in detecting crack edges, this paper chooses to use the instance segmentation model (YOLOv11 seg) for object detection of building cracks.

3. Experimental Analysis

3.1 Data Preparation

The dataset used in this article includes 800 original images of building cracks. The collected images of building cracks were uniformly modified to a size of 500 pixels by 500 pixels. The Labelme polygon annotation tool was used to annotate the cracks in the images at the pixel level, and the annotation labels were all "crack". The images before and after annotation are shown in Figure 2.

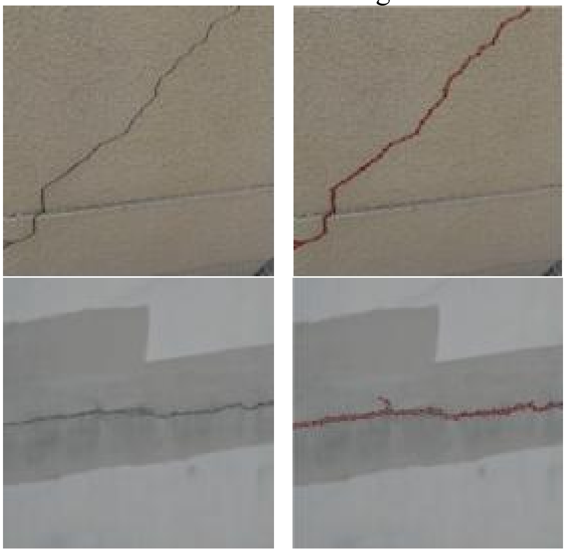

Figure 2. Image and Annotation of Building Cracks

Randomly divide the annotated building crack images into training set, validation set, and testing set in a ratio of 7:2:1, which will be used for model training, validation during the training process, and final testing and evaluation of the model. In addition, this article uses horizontal flipping, brightness increase, saturation increase and other methods to enhance the training set data to improve the robustness of the model.

3.2 Experimental Environment

The experimental environment in this article is the Ubuntu 22.04 operating system, Cuda12.1, Python3.8, The deep learning framework is Pytorch 2.4.1, and the GPU is NVIDIA Geforce RTX4080 (24GB).

The experimental parameters are 50 training epochs, a learning rate of 0.01, SGD dynamic parameter of 0.937, and a weight decay rate of 0.0005.

4. Analysis of Experimental Results

This article uses precision (P), recall (R), and mean accuracy (mAP) as evaluation indicators for identifying building cracks. Among them, TP represents pixels with real labels as cracks and predicted results as cracks, FP represents pixels that misclassify the background as cracks, FN represents pixels that misclassify cracks as background, and TN represents pixels that correctly identify the background. The precision and recall can be expressed as:

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

The intersection to union ratio IoU represents the ratio of intersection to union between the predicted result and the true label, mAP@0.5 Represents the detection accuracy when IoU is 0.5, mAP@0.5 0.95 represents the average accuracy before IoU of 0.5~0.95. The accuracy of crack area localization and pixel level recognition accuracy of crack edges are evaluated by calculating the calculation indicators of bounding box and mask separately. The calculation results are shown in Table 1.

Table 1. Experimental Result

Calculated metrics	Box	Mask
P	74.6%	76.8%
R	70.5%	71.3%
mAP@0.5	76.4%	76.5%
mAP@0.5:0.95	60.5%	33.8%

According to the experimental results, the accuracy and recall of the model bounding box and mask are both greater than 70%, indicating that the model has fewer false alarms for cracks and a lower rate of missed detections. Therefore, it is considered that the YOLOv11 seg model is relatively reliable for detecting building cracks. Average accuracy index mAP@0.5. All values are greater than 75%, indicating that the

YOLOv11 seg model has stable performance in identifying building cracks at a moderate level of accuracy. mAP@0.5: 0.95 (box) and mAP@0.5 0.95 (mask) is significantly lower than mAP@0.5 As the positioning accuracy requirements increase, the performance of the model gradually decreases, but mAP@0.5 0.95 (box) significantly higher than mAP@0.5 0.95 indicates a sharp decline in the performance of the model when the requirement for segmentation boundary accuracy is increased. Based on the above experimental results, this article believes that the YOLOv11 seg model performs well in identifying building cracks, but has some shortcomings in accurately predicting the boundaries of cracks.

The visualization analysis of the validation set images is shown in Figure 3. Through the visualization analysis results, it is intuitively demonstrated that the YOLOv11 seg model can achieve the recognition of building cracks.

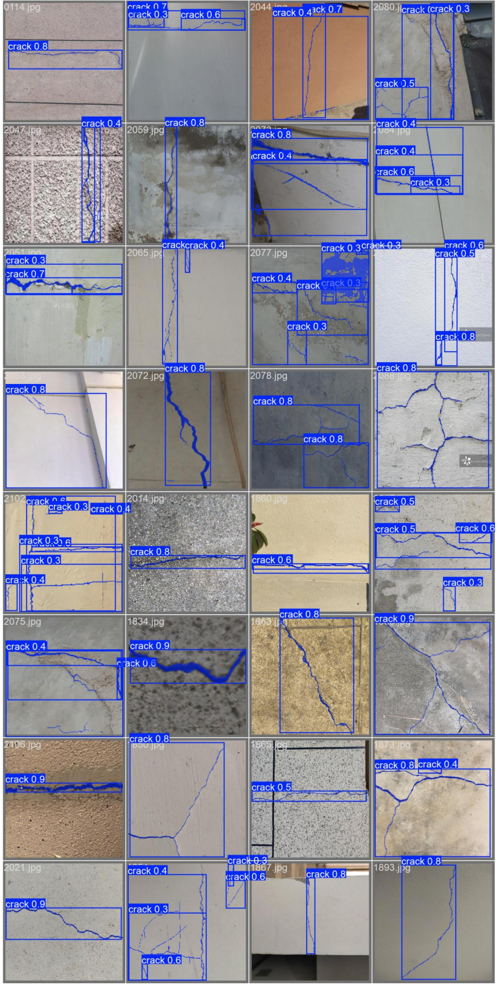

Figure 3. Visualized Detection Results

5. Conclusion

This article uses the YOLOv11 seg model to detect building cracks. The Labelme tool is used to annotate building cracks at the pixel level during detection to verify the model's recognition accuracy of crack edges. Data augmentation is applied to the building crack dataset to improve the model's versatility. The accuracy, recall, mAP, and other calculation metrics of bounding boxes and masks were obtained by using the YOLOv8 seg model for dataset detection. According to the calculation results, the accuracy and recall of bounding boxes and masks mAP@0.5 All have reached a high level, but the mask mAP@0.5 The 0.95 index is significantly lower than the bounding box, so this article believes that the YOLOv11 seg model performs well in the rapid screening and localization of building cracks. However, the prediction accuracy of the model for crack edges is slightly insufficient, and further improvements are needed to accurately measure the length and width of building cracks.

Acknowledgments

This paper is supported by China Architecture Design & Research Group Youth Fund Project (No. Z2024Q20).

References

[1] Ayenu-Prah A, Attoh-Okine N. Evaluating pavement cracks with bidimensional empirical mode decomposition. EURASIP Journal on Advances in Signal Processing, 2008, 2008: 1-7.
[2] Li Q, Zou Q, Zhang D, et al. FoSA: F* seed-growing approach for crack-line detection from pavement images. Image and Vision Computing, 2011, 29(12): 861-872.
[3] Gavilán M, Balcones D, Marcos O, et al. Adaptive road crack detection system by pavement classification. Sensors, 2011, 11(10): 9628-9657.
[4] Yamaguchi T, Hashimoto S. Fast crack detection method for large-size concrete surface images using percolation-based image processing. Machine Vision and Applications, 2010, 21: 797-809.
[5] Liang X H, Cheng Y Z, Zhang R J, etc Bridge crack identification and measurement method based on convolutional neural network. Computer Applications, 2020, 40 (04): 1056-1061
[6] Wang S Y, Liu J. Research on Road Crack Recognition Method Based on

Convolutional Neural Network. Smart City, 2023, 9 (12): 18-22. DOI: 10.19301/j.cnki. zncs. 20223.12.005

[7] Wu T Y. Research on Road Defect Recognition Based on Improved YOLO Algorithm. Dalian Jiaotong University, 2025.DOI:10.26990/d.cnki.gsltc.2025.0007 32

[8] Miao Z H. Research on Seismic Damage Identification and Performance Evaluation of RC Components Based on Computer Vision. Tsinghua University, 2022. DOI:10.27266/d.cnki.gqhau.202200023

[9] Wang W B. Research on post earthquake safety emergency assessment method for visual based framework structures. Institute of Engineering Mechanics, China Earthquake Administration, 2023. DOI: 10.27490/d.cnki.ggjgy.2023.000032

[10]Zhou X, Luo L H, Zhang Ruijun, etc Identification of Wall Cracks in Old Buildings Based on SSD Algorithm. Technological Innovation and Application, 2023, 13 (28): 36-39. DOI: 10.19981/j. CN23-1581/G3.2023.28.009

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

[12]Liao Y N, Dou D Y Design and Research of Bridge Crack Detection Method Based on Mask RCNN. Applied Optics, 2022, 43 (01): 100-105+118

[13]Zhou S X, Yang D, Pan Y, etc YOLOv5 pavement crack detection and recognition based on attention mechanism. Journal of East China Jiaotong University, 2024, 41 (02): 56-63. DOI: 10.16749/j.cnki. jecjtu. 20220307.002