Journal of Computer Science and Frontier Technologies
ISSN Print: 3104-4204
ISSN Online: 3104-4212
CD ISSN: 3104-4220
https://doi.org/10.63313/JCSFT.9020

The improved model of YOLOV11 integrating Starnet

Dong You[1]. Liqiang Wang[1.2.*] and Yuezong Wang[2]
[1]School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin, 300222, China
[2]Tianjin Engineering Research Center of Fieldbus Control Technology, Tianjin 300222, China
Email: 1610715014@qq.com

How to cite this paper: Dong, Y., Wang, L. Q., & Wang, Y. Z. (2025). The improved model of YOLOV11 integrating Starnet. Journal of Computer Science and Frontier Technologies, 1(3), 16-22. ISSN Print: 3104-4204; ISSN Online: 3104-4212.
https://doi.org/10.63313/JCSFT.9020
Published: 2025-11-10

Abstract

Due to the small size, widespread distribution, and complex background of road cracks, traditional detection methods are prone to misdetection and missed detection. To improve the model's detection performance, we propose an im-proved YOLOv11 model for road crack detection. The model uses Starnet as the backbone network and dynamically adjusts its large spatial receptive field to more accurately capture the details of road cracks. Validation on the RDD2022 dataset shows that the improved model achieves an average precision (mAP) of 0.586, which is a 4.8 percentage point improvement over the traditional model. These improvements significantly enhance detection accuracy and efficiency, providing a more reliable and efficient solution for road crack detection.

Keywords

Road crack detection; RDD2022 dataset; Yolov11; C3k2-star backbone

1. Introduction

Urban roads are a key component of the urban transportation network, and their maintenance quality directly impacts the safety operation and traffic efficiency of the city. With the rapid pace of urbanization, the pressure on road infrastructure continues to increase. In areas with heavy traffic flow and variable climatic conditions, the durability of the road surface faces significant challenges. Cracks, as a typical form of road damage, if not identified and treated promptly, can accelerate the deterioration of road performance, increase maintenance costs, and even cause traffic accidents, negatively affecting traffic order and urban management. Current crack detection techniques mainly rely on manual inspections or traditional image analysis methods. Although these methods have basic recognition capabilities, manual approaches are inefficient and highly subjective, while conventional image algorithms lack stability in complex scenarios, making them difficult to meet the current demands for precision and timeliness in road maintenance. Thanks to

breakthroughs in deep learning, automated crack detection solutions using target recognition technologies such as YOLO have increasingly attracted attention in academia. The YOLO architecture, with its end-to-end training and real-time detection features, shows unique advantages in real-time monitoring applications and has been successfully applied in various fields.

However, existing YOLO models still have limitations when dealing with the diverse forms and rich details of road cracks, especially in complex backgrounds and small target detection. To improve detection accuracy, this paper improves the YOLOv11 model by introducing a crack direction attention module to enhance the model's ability to perceive cracks in different directions.

2. Related Work

2.1. Road Crack Detection Technology

In recent years, deep learning methods have made significant progress in road crack detection, especially in the application of convolutional neural networks (CNNs) and object detection algorithms. Traditional road crack detection methods typically rely on manual inspections or image processing techniques, but these methods often suffer from inefficiency and low accuracy, especially in complex environments where stability and precision are hard to maintain. With the continuous development of computer vision technology, deep learning models, especially the YOLO (You Only Look Once) series and other CNN architectures, have shown great potential in automated detection tasks.

In road crack detection applications, YOLO series algorithms have become one of the most commonly used object detection technologies due to their outstanding real-time performance and efficiency. Zhou et al. [1] proposed a road defect detection method based on EC-YOLO, which significantly improved detection accuracy by combining the C2f module with the EMA attention mechanism. The EMA attention mechanism enhances the model's focus on the road crack areas, enabling better handling of complex backgrounds and small targets, thereby reducing false positives and missed detections. This method made important optimizations to the traditional YOLO architecture, effectively improving the model’s performance in dynamic and complex environments.

In addition, Meng et al. [2] improved the YOLOv8 model and optimized it for road detection tasks, achieving a 92.7% mAP50, demonstrating the strong capabilities of the YOLO series in road crack detection. After introducing multi-scale feature fusion and an improved network architecture, YOLOv8 not only improved detection accuracy but also optimized its ability to recognize small targets and cracks in various shapes. These improvements enable YOLOv8 to achieve higher detection efficiency in complex road environments, making it an important choice in the field of road crack detection.

Beyond the YOLO framework, other deep learning architectures offer complemen-

tary approaches. Hu et al. [3] explored two paths: an optimized YOLOv5s model using EIoU loss for improved localization, achieving over 74% crack classification accuracy, and a Res2Unet-CBAM network that employs a channel attention mechanism for precise pixel-level crack segmentation. In a similar vein, Hong et al. [4] developed EB-UNet++, an enhanced segmentation network that combines EfficientNet-B2 and UNet++ with a Boundary Extraction Module (BEM), showing superior performance in segmenting irregular, low-contrast cracks.

In summary, deep learning, particularly through the evolution of YOLO series algorithms and specialized segmentation networks, has significantly boosted the automation, intelligence, accuracy, and efficiency of road crack detection.

2.2. YOLO Series Algorithms

The YOLO series is a leading single-stage object detection algorithm renowned for its end-to-end training capability and high-efficiency real-time performance. According to a comprehensive review by Sapkota and Karkee [5], the evolution from YOLOv5 to YOLO26 has introduced key innovations such as native NMS-free inference, Progressive Loss Balancing, and Small-Target-Aware Label Assignment. Building upon this developmental trajectory, YOLOv11 inherits foundations established by YOLOv8, including decoupled detection heads and anchor-free predictions, laying a solid groundwork for efficient multi-task visual capabilities.

YOLOv11 introduces several critical improvements for complex detection tasks. Its core innovations include adopting the C3k2 module to replace the C2f module from YOLOv8, which captures richer local and global features through larger convolutional kernels, significantly enhancing detection accuracy and robustness. Additionally, the C2PSA module added after the SPPF layer utilizes a self-attention mechanism, enabling the model to automatically focus on key regions in the feature maps. This effectively enhances the detection capability for small targets like cracks in complex backgrounds while reducing noise interference.

The practical effectiveness of this architecture has been validated in multiple crack detection applications. One study on sewer pipe crack recognition [6] enhanced YOLOv11n by incorporating a coordinate attention mechanism and a BiFPN module, raising its performance metric to 87.3%. Another study [7] developed an improved YOLOv11-EfficientHead system, optimizing detection performance using a specialized dataset. Its applicability even extends to the field of medical imaging, where an optimized YOLOv11-TDSP model [8] achieved a precision of 94.5% and a mAP of 95.8% in dental anomaly detection by integrating a Single-Head Self-Attention mechanism and a small-target detection layer. These improvements make YOLOv11 an ideal choice for road crack detection, providing robust technical support for intelligent transportation systems and road maintenance.

3. Model Optimisation

3.1. Starnet Module

This paper introduces the lightweight StarNet network [5], which replaces the original backbone in the model. The architecture of StarNet is depicted in Figure 1. Designed with a hierarchical approach, the StarNet network is structured into four stages. At each stage, the initial feature map is downsampled using convolutional layers (Conv), which reduces its size and extracts preliminary features. These are followed by the Star Blocks layer, where efficient feature extraction and fusion are achieved through a unique star-shaped operation.

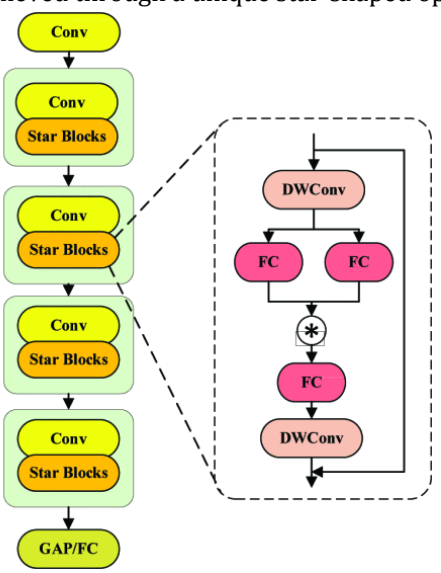


Fig 1. The structure of StarNet

Depth-separable convolutions are incorporated both before and after the Star Blocks layer, effectively doubling the channel count at each stage. This modification enhances the network's feature representation capabilities. In this study, a channel expansion factor of 4 is used, ensuring that the network's width doubles at each stage. To further improve feature extraction efficiency, Batch Normalization is employed instead of the original Layer Normalization, optimizing the network's overall performance.

Additionally, to streamline the feature extraction process and avoid complex operations, StarNet uses star operations to map the input into a higher-dimensional space. Feature fusion is then performed via pointwise multiplication of matrix elements, enabling highly efficient feature extraction.

4. Experimental Result and Analysis

4.1. Datasets

The RDD2022 dataset is a public resource designed for road damage detection, supporting advancements in intelligent transportation and automated road maintenance. It includes high-quality images from various countries, covering dam-

age types such as cracks, potholes, and surface degradation. These images, captured in diverse conditions like lighting and weather, are annotated with damage type, location, and boundary box coordinates, making the dataset ideal for object detection and semantic segmentation tasks. This research uses 10,000 images from the dataset to explore deep learning algorithms for accurate road damage detection, aiming to enhance intelligent transportation and road management systems.

4.2. Experimental Configuration

All experiments were conducted using the same hardware configuration to ensure consistency. The specific hardware setup for this study is outlined as follows.

Tab 1. Experimental Configuration

Hardware	Version or model
Operating system	Windows 10
GPU	NVIDIA RTX A4000
Python	3.8.0
Pytorch	2.4.0
CUDA	11.8

4.3. Evaluation Metrics

To assess the performance of the proposed model, this study utilizes commonly adopted evaluation metrics in object detection, including Precision (P), Recall (R), Average Precision (AP), and mean Average Precision (mAP). The mathematical definitions for these metrics are presented in Equations (1) through (4).

$$P = \frac{TP}{TP+FP} \times 100\% \quad (1)$$

$$R = \frac{TP}{TP+FN} \times 100\% \quad (2)$$

$$AP = \int_0^1 P(R)dR \quad (3)$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \quad (4)$$

Specifically, P represents the proportion of positive samples that were correctly predicted by the model among those predicted as positive. TP refers to the number of instances that the model correctly identified as positive samples. FP indicates the number of instances where the model incorrectly predicted a negative sample as positive. R reflects the proportion of actual positive samples that were correctly predicted by the model. FN denotes the number of instances where the model mistakenly predicted a positive sample as negative. AP is the area under the precision-recall curve for a given category, reflecting the model's performance for that category. mAP is the average of the AP values across multiple categories, serving as an indicator of the model's overall performance across all categories. N represents the number of categories in the dataset.

4.4. Comparative Experiments

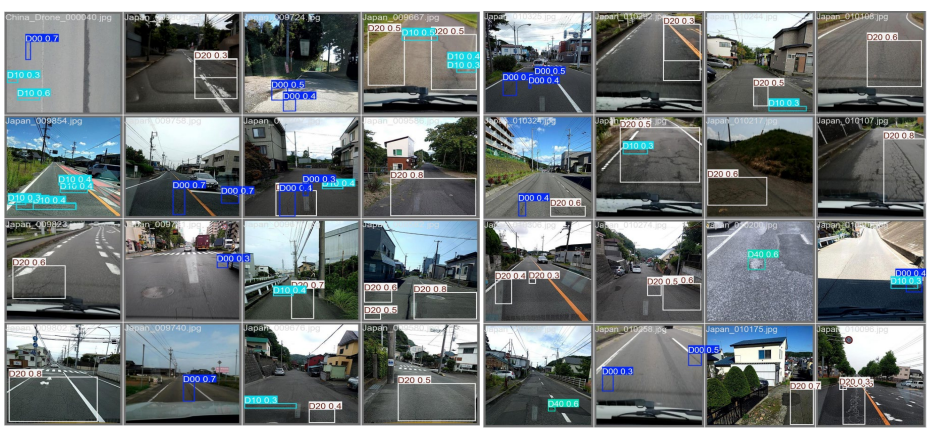
In this study, the Starnet module, which possesses efficient feature extraction capabilities, was employed to replace the backbone network of the original model. Its performance was compared against five mainstream architectures: Yolov11n, ConvNeXt, MobileNet V3, PVT, and ShuffleNet V2. As shown in Table 2, the improved model integrating the Starnet module achieved a precision of 71.5%, a recall of 56.4%, and an mAP@0.5 of 58.6%. Compared to the baseline model Yolov11n, Starnet improved the mAP@0.5 by 2.6 per-centage points. It also outperformed the MobileNet V3 architecture by 4.1 percentage points in mAP@0.5. Although MobileNet V3 showed a slightly higher recall (57.6%), Starnet demonstrated significant advantages in both comprehensive detection preci-sion and mean average precision, exhibiting more balanced and superior detection performance.

Tab 2. Backbone Architecture Experiment

Network	P/%	R/%	mAP@0.5/%
Yolov11n	68.9	54.3	56
ConvNeXt	68.8	54.8	55.3
MobileNet V3	69.9	57.6	54.5
PVT	63.9	50.8	53.3
ShuffleNet V2	69.3	56.1	55.8
Starnet	71.5	56.4	58.6

4.5. Visualization experiment

This is a comparative analysis chart of road crack detection model tests, divided into two parts: (a) Chinese Road Surface Detection Image and (b) Japanese Road Surface Detection Image. It shows the detection results in different road scenarios of the two countries, with bounding boxes presenting the recognition of targets such as road cracks. It can be used to compare and analyze the performance of the detection model in different road environments of China and Japan, providing a visual experimental reference for optimizing road crack detection technology and improving road maintenance efficiency.



(a) Chinese Road Surface Detection Image (b) Japanese Road Surface Detection Image

Fig 3. Model Test Comparative Analysis Chart

5. Conclusion

To address the challenges of road crack detection such as small target size, widespread distribution, and complex backgrounds, this paper proposes an improved YOLOv11 model. It adopts Starnet as the backbone network and dynamically adjusts the large spatial receptive field to accurately capture crack details. Validated on the RDD2022 dataset, the model achieves a mAP of 0.586, a 4.8 percentage point improvement over the traditional model. Comparative experiments show it outperforms mainstream architectures like Yolov11n and MobileNet V3. Visualization results from Chinese and Japanese road surface images also demonstrate its effectiveness in different road environments. This improvement enhances detection accuracy and efficiency, providing a reliable solution for road crack detection and contributing to intelligent transportation and road maintenance.

Acknowledgements

We acknowledge the support of the following projects: Tianjin Science and Technology Program (Project No. 25KPKJRC00050), Tianjin Jinnan District Science and Technology Program (Project No. 20230107).

References

[1] Zhou C, Yang Y, Lin L, et al. Research on pavement crack detection technology based on improved YOLO11[C]//2025 40th Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE, 2025: 2220-2223.

[2] Meng A, Zhang X, Yu X, et al. Investigation on lightweight identification method for pavement cracks[J]. Construction and Building Materials, 2024, 447: 138017.

[3] Hu G X, Hu B L, Yang Z, et al. Pavement crack detection method based on deep learning models[J]. Wireless Communications and Mobile Computing, 2021, 2021(1): 5573590.

[4] Hong P T H, Hang T T T, Van Giap D, et al. EB-UNet++: An enhanced crack segmentation network combining EfficientNet-B2 and UNet++ with boundary extraction module[J]. Journal of Military Science and Technology, 2025 (IITE): 148-159.

[5] Sapkota R, Karkee M. Ultralytics YOLO Evolution: An Overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 Object Detectors for Computer Vision and Pattern Recognition[J]. arXiv preprint arXiv:2510.09653, 2025.

[6] Wu D. YOLO-Based Enhanced Crack Detection Algorithm for Underground Pipeline Scenarios[J]. Traitement du Signal, 2025, 42(2): 787.

[7] Żarski M, Wójcik B, Miszczak J A. KrakN: Transfer learning framework and dataset for infrastructure thin crack detection[J]. SoftwareX, 2021, 16: 100893.

[8] YOLOv11-TDSP: Lightweight and high-precision abnormal tooth detection model for oral panoramic films[J]. Journal of Southern Medical University, 2025, 45(8): 1791.