



# YOLOv8-CAACA: A Context-Aware Adaptive Confidence Adjustment and Target Fusion Algorithm for Pavement Crack Recognition in Complex Scenarios

Muhan Bai

Robotics College, Beijing Union University, Beijing, China  
2024250391001@buu.edu.cn

**Abstract.** To address the issues in current crack detection from road images—such as inaccurate identification of long cracks, misdetection and missed detection of cracks in complex environments, and splitting of a single crack into multiple segments—this paper proposes an improved YOLOv8 algorithm (YOLOv8s-CAACA) incorporating context-based adaptive confidence adjustment and crack clustering-merging. The proposed method first integrates a context-based adaptive confidence adjustment module into the YOLOv8s algorithm, enabling the model to more accurately identify road cracks and reduce isolated false detections caused by road potholes, protrusions, or similar conditions. Secondly, a crack clustering module is added to merge cracks within a certain threshold into a single instance, allowing the model to directly recognize entire long cracks without splitting them into fragmented segments that render detection results unreadable. In tests on a dataset of 5,000 images under simple environments, the optimized YOLOv8 model achieved a slight improvement in the evaluation. In tests under complex conditions, however, the optimized algorithm showed significant enhancements and greatly improved the visibility of detection results. Experimental results demonstrate that YOLOv8s-CAACA achieves good accuracy in crack detection under complex environments. Meanwhile, the model maintains lightweight properties, indicating its application value and prospects in road maintenance and upkeep.

**Keywords:** YOLOv8, Road Crack Detection, Context-aware Adaptive Confidence Adjustment, Clustering-Based Merging.

## 1 Introduction

Road infrastructure constitutes the cornerstone of modern transportation, facilitating economic activities and daily commutes [1, 2, 3]. However, prolonged service and environmental stressors inevitably lead to pavement deterioration, with cracks being the most prevalent and critical defect [4]. Unaddressed cracks can escalate into potholes and structural failures, increasing maintenance costs by up to 80% and posing severe safety risks [5, 6, 7]. Traditional crack detection methods, such as manual inspection

and conventional image processing, suffer from inefficiency, subjectivity, and poor robustness under complex conditions, such as varying (illumination) and background clutter [8].

The advent of deep learning has revolutionized automated crack detection, with single-stage object detectors, such as the You Only Look Once (YOLO) series gaining prominence for their speed and accuracy [9, 10]. YOLOv3-based methods, such as crack-YOLOv3 conducted by Nie et al, have enhanced feature extraction through multi-scale information fusion and attention mechanisms, improving detection precision in dense crack scenarios [11]. YOLOv8, meanwhile, has been optimized for lightweight deployment, with studies, such as proposing a streamlined YOLOv8n model that balances efficiency and performance, making it suitable for mobile inspection devices [12].

Despite these advancements, challenges persist. Complex pavement backgrounds, diverse crack morphologies, and the need for real-time processing remain unresolved [13]. For instance, small and irregular cracks are often misdetections due to insufficient feature representation [14]. To address this, recent studies have integrated innovative techniques: Yao et al fused dynamic snake convolution with YOLOv8 to better capture tubular crack structures [15]; Pei et al employed deep feature transfer and data augmentation to mitigate domain shifts between diverse pavement datasets [16]; and semi-supervised learning approaches, reduced reliance on large labeled datasets [17].

YOLOv11, the latest iteration, has further improved feature extraction with modules, such as C3k2 and C2PSA, enhancing adaptability to small targets and complex backgrounds [18]. However, optimizing its performance for crack detection—particularly in balancing speed, accuracy, and computational efficiency—remains an active area of research [19].

This paper focuses on improving the YOLOv8 algorithm for road crack recognition, integrating features such as lightweight design, real-time performance, strong anti-noise capability, and high visibility. It innovatively develops a context-based adaptive confidence adjustment module and a crack clustering algorithm. The innovative algorithms in this paper are designed to be applied when using the YOLO algorithm to detect multiple targets of the same type in the same image. The model can adaptively adjust the confidence based on context content, significantly enhancing its ability to filter noise. Meanwhile, it clusters and merges adjacent detected targets to improve the visualization of detection results. After optimization in this paper, the YOLOv8 algorithm can accurately identify various types of cracks, effectively filter environmental noise, and obtain recognition results that are convenient for road inspectors to observe.

## 2 Dataset Source and Description

The dataset used in this study contains 5,466 road crack images in JPG format (Figure 1). Referring to the recommendations from pavement defect detection research, the complexity of real-world scenarios is constructed by integrating diverse environmental conditions. The images cover variations in illumination (strong sunlight, overcast, and low light), weather (dry, rainy, and snowy road surfaces), shooting angles ( $0^\circ$  - $60^\circ$ )

from the horizontal), and crack types (longitudinal, transverse, and irregular reticular cracks). This diversity ensures the robustness of the model in practical scenarios.

Each image is individually annotated, with specific crack locations marked by rectangular boxes in the images, as shown in Figure 2. These image annotations are converted into TXT format that can be used for YOLO model training. The TXT format annotation files, as presented in Table 1, include class labels (crack) and normalized bounding box coordinates (x\_center, y\_center, width, height).



Fig. 1. Original data image. (Picture credit: Original)

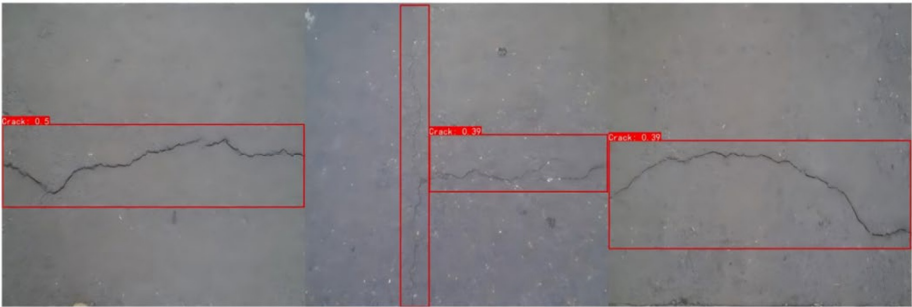


Fig. 2. Annotation of Image Format. (Picture credit: Original)

Table 1. Image annotation converted to txt formatt.

Image1	0 , 0.5 , 0.51953125 , 0.990234375 , 0.333984375
Image2	0 , 0.361328125 , 0.541015625 , 0.0654296875 , 0.916015625 0 , 0.7021484375 , 0.568359375 , 0.591796875 , 0.1123046875
Image 3	0 , 0.4990234375 , 0.630859375 , 0.998046875 , 0.412109375
...	...

### 3 Introduction to Core Algorithms

#### 3.1 Overall Workflow and Model Structure of YOLOv8-CAACA

First, the image is input into the model and undergoes basic detection by the YOLOv8 model. It then enters the first core innovative algorithm, namely the context-based adaptive confidence adjustment module, which adjusts confidence to filter out noise. Next, it proceeds to the second core innovative algorithm, the crack clustering and merging algorithm, which is used to cluster and merge cracks within a certain threshold to improve the visibility of detection results. Finally, after post-processing (including adding crack labels and enhancing display), the annotated image after recognition is output. The overall workflow is shown in Figure 3.

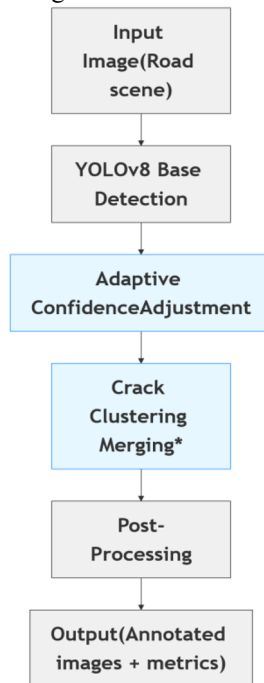


Fig. 3. Overall algorithm flow chart of YOLOv8-CAACA. (Picture credit: Original)

#### 3.2 Context-Aware Adaptive Confidence Adjustment

Traditional object detection models use a fixed confidence threshold across the entire image, which often leads to missed detections in low-contrast regions and false positives in noisy areas. To address this, this paper proposes a novel context-aware adaptive confidence adjustment mechanism that dynamically adjusts the confidence threshold based on local crack density.

For each detected crack candidate, the algorithm calculates an adaptive confidence threshold using the following steps, first is Context Window Definition: A square window of size  $(W \times W)$  pixels centered at the candidate crack is defined. The window

size  $W$  is a hyperparameter determined through empirical evaluation. Second is Crack Density Estimation: The number of crack candidates  $N$  within the context window is counted. This serves as a measure of local crack density. Third is Threshold Adjustment: The confidence threshold ( $T_{adaptive}$ ) for the candidate is computed as:

$$T_{adaptive} = \begin{cases} T_{base} - \Delta T, & \text{if } N \geq C \\ T_{base} + \Delta T, & \text{otherwise} \end{cases} \quad (1)$$

Where  $T_{base}$  is the base confidence threshold,  $\Delta T$  is the adjustment magnitude, and  $C$  is the crack density threshold. This formula ensures that in regions with high crack density (likely actual cracks), the threshold is lowered to detect subtle cracks, while in sparse regions (likely noise), the threshold is raised to reduce false positives.

### 3.3 Clustering-Based Crack Merging

Road cracks often appear as fragmented segments due to occlusion, shadowing, or model limitations. To address this, we introduce a clustering-based crack merging algorithm that groups adjacent crack segments into continuous cracks.

The distance between two crack bounding boxes  $B_1$  and  $B_2$  is defined as:

$$d(B_1, B_2) = \sqrt{(center_x(B_1) - center_x(B_2))^2 + (center_y(B_1) - center_y(B_2))^2} \quad (2)$$

Where  $center_x$  and  $center_y$  denote the center coordinates of the bounding boxes.

This paper uses a greedy clustering approach where each crack segment is assigned to the nearest cluster if its distance is within a predefined threshold  $D$ .

### 3.4 YOLOv8 Detection Framework

The YOLOv8 model is adopted for road crack detection, leveraging its advantages in real-time object detection; its architecture, as shown in Figure 4, consists of three core components: Backbone, which utilizes a CSPDarknet structure with C2f modules to extract multi-scale features, enhancing the capture of fine crack details (such as edges) and high-level semantic information (such as crack patterns) [12, 13]; Neck, which employs a Path Aggregation Network (PAN) for feature fusion, integrating low-level high-resolution features with high-level semantic features to improve the detection capability for cracks of different sizes [12]; and Head, an anchor-free detection head that directly predicts bounding boxes and class probabilities, simplifying the process and improving efficiency compared to anchor-based methods.

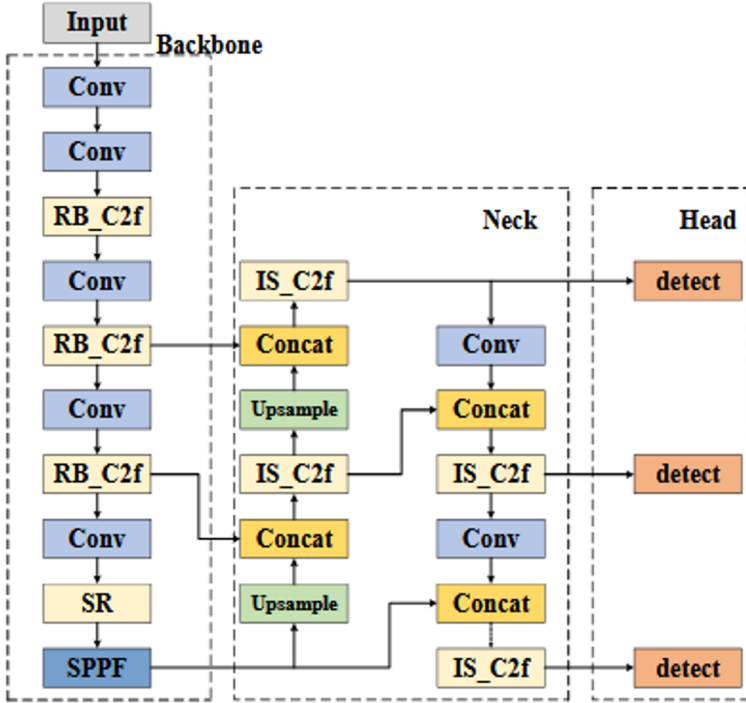


Fig. 4. YOLOv8 architecture diagram (cited from reference [20]).

## 4 Result and Discussion

### 4.1 Experimental Process and Results

To comprehensively assess the model's performance, four key metrics were selected based on standard practices in road defect detection: Precision (P): The ratio of true positives (TP) to all detected instances (TP + false positives, FP), measuring the model's ability to avoid false alarms. Recall (R): The ratio of TP to all actual cracks (TP + false negatives, FN), reflecting the model's capability to capture existing defects. mAP@0.5: Mean average precision at an intersection-over-union (IoU) threshold of 0.5, a critical metric for evaluating detection accuracy in crack localization.

For the experimental data foundation of the proposed model in this paper, 4000 road crack images similar to the dataset in this study were used. Each crack image contains at least 1 crack and at most 3 cracks. Although there are no independent annotations for individual cracks, statistical information on the number of cracks in each image is available.

Regarding model selection, this paper tested a conventional CNN, three differently sized YOLOv8 models (YOLOv8n, YOLOv8s, and YOLOv8m), as well as a modified

YOLOv8 integrated with adaptive confidence adjustment and crack clustering algorithms (YOLO v8n-CAACA, YOLO v8s-CAACA, and YOLO v8m-CAACA). The CNN model was trained for 200 epochs, while the remaining models were trained for 150 epochs. The Precision (P), Recall (R), and mAP@0.5 metrics of these algorithms are presented in Table 2.

**Table 2.** Evaluation results of each model.

Model	Precision (P)/%	Recall (R)/%	mAP@0.5/%	mAP@0.5:0.9/%
CNN	52.7	61.3	41.4	22.3
YOLOv8n	80.4	70.1	75.5	54.8
YOLOv8s	81.1	80.9	76.2	59.7
YOLOv8m	78.6	79.7	72.8	53.3
YOLOv8n- CAACA	83.8	81.2	76.3	59.9
YOLOv8s- CAACA	84.3	82.6	76.8	61.2
YOLOv8m- CAACA	79.2	80.5	73.1	55.1

As can be seen from Table 2, the performance metrics generally show an increasing trend with the rise in model complexity. However, a decline in metrics is observed when moving from YOLOv8s to YOLOv8m. The study attributes this phenomenon to the insufficient scale of the dataset used in the experiment, which limited the performance of the larger model.

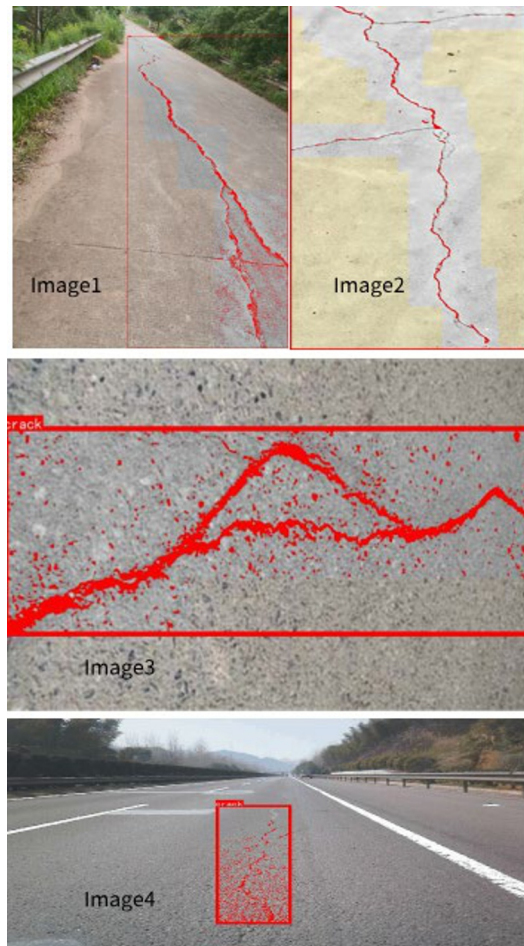
From another perspective, all YOLO models achieved improved metrics after integrating the adaptive confidence adjustment and crack clustering algorithms. The research suggests that the complexity of the initial test set was insufficient—for instance, all test samples were small images containing only cracks and asphalt pavement, which were overly simple for the models and failed to fully demonstrate the advantages of the core algorithms proposed in this paper, namely the adaptive confidence adjustment and crack clustering algorithms.

To address this, the study conducted a second test using a more complex dataset consisting of real-world wide-view road images, which included lane markings, vegetation around the road, and soil environments. However, due to the small sample size of this supplementary dataset, no further quantitative analysis of metrics was performed. Instead, the paper presents the direct visualization of the detection results.

Since this model is designed for real-time road crack detection, it is necessary to balance model accuracy and inference speed. Therefore, the YOLOv8s model was ultimately selected as the final model, and all subsequent experimental data were generated using the YOLOv8s model.

For the adaptive confidence adjustment and crack clustering algorithms proposed in this paper, there are several key parameters as follows: Base confidence threshold: The initial confidence threshold of the model for cracks; a higher value results in stricter detection, while a lower value leads to more lenient detection. Context window size: A user-defined local region used to calculate the presence of other cracks in the surrounding area. Crack density threshold: The critical value for determining whether the current region is a crack-dense area. Confidence adjustment magnitude: The extent to which the threshold is lowered in dense regions and raised in sparse regions. Crack clustering range: Cracks within this range are grouped into one cluster.

In this experiment, the base confidence threshold was set to 0.1, the context window size to 200 pixels, the crack density threshold to 4, the confidence adjustment magnitude to 0.2, and the crack clustering range to 50 pixels. Additionally, the display of crack regions was enhanced.



**Fig. 5.** Detection results based on the YOLOv8s-CAACA. (Picture credit: Original)

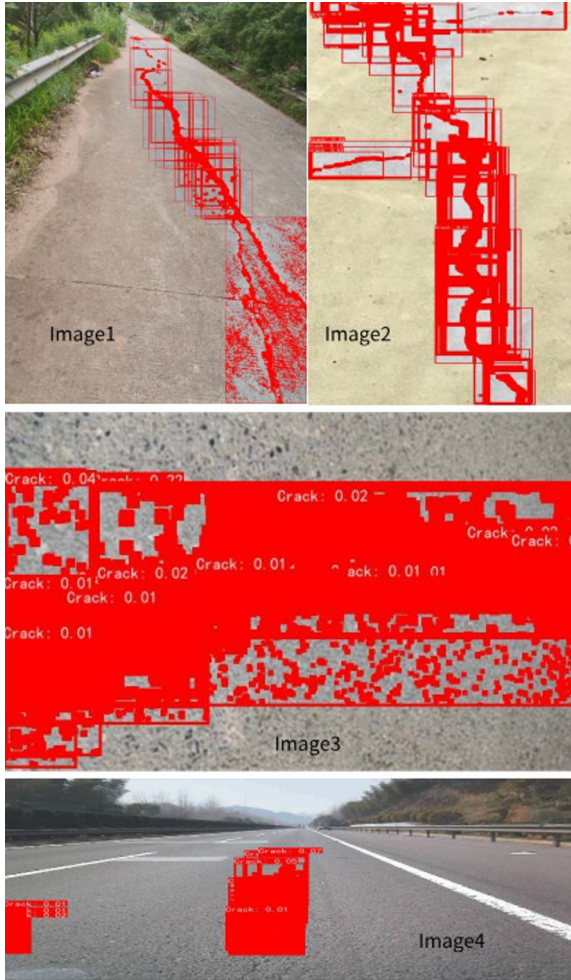


Fig. 6. Detection results based on the original YOLOv8s algorithm. (Picture credit: Original)

As shown in Figure 5 and Figure 6, Figure 5 presents the detection results of YOLOv8s-CAACA, while Figure 6 shows the detection results of the conventional YOLOv8s algorithm. Four most representative complex road crack images were selected, specifically: long cracks with multiple noises and interferences under a wide-angle view, long cracks extending in different directions, a single crack with noises caused by road conditions such as surrounding stones and potholes, and indistinct road cracks captured by a driving recorder.

It is evident that the cracks in Figure 5 are more distinguishable, with their contours perfectly marked. In complex environments, the adaptive confidence adjustment and crack clustering algorithms play a significant role in isolating noise to a large extent. Particularly for the third scenario involving a single crack mixed with potholes, the conventional YOLO model misidentified all surrounding potholes as cracks, rendering

the detection results completely unreadable. In contrast, YOLOv8s-CAACA can distinguish most potholes and ultimately outline the main contour of the crack. Figure 7 shows the evaluation results of the two models on complex scenarios.

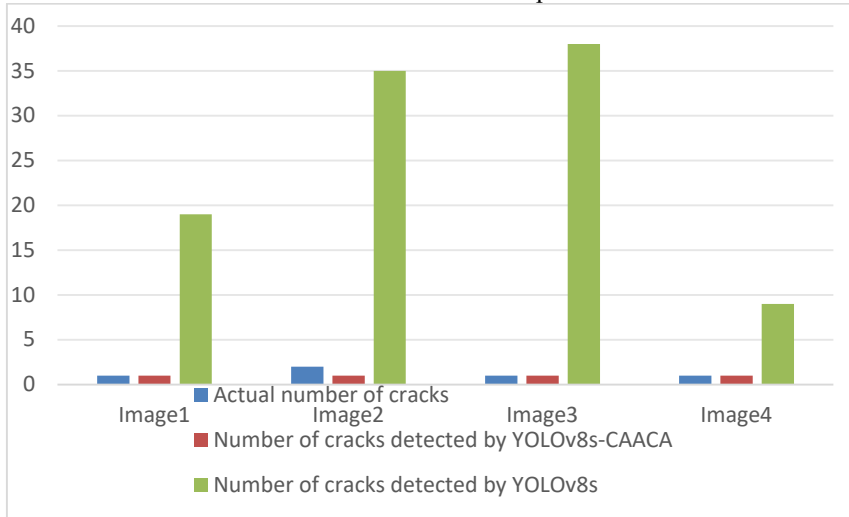
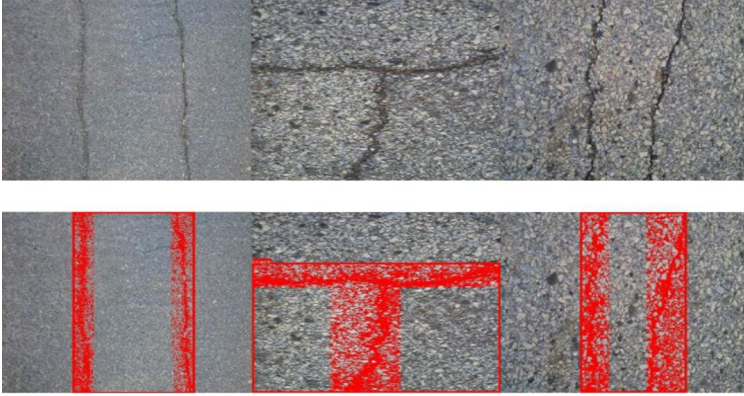


Fig. 7. Evaluation results of the two models on complex scenarios. (Picture credit: Original)

## 4.2 Discussion

It can be seen from the detection results that the conventional YOLOv8s model achieves very poor accuracy in detecting long cracks and in complex scenarios. For long cracks, the conventional YOLOv8s model tends to split them into multiple short cracks similar in length to those in the training set, resulting in a single crack being identified as a dozen or even dozens of short cracks. Meanwhile, due to the absence of a denoising algorithm in the model, it directly identifies many isolated potholes, stone protrusions, and other road conditions as cracks, which greatly reduces the accuracy and visibility of the detection results.

In contrast, the improved YOLOv8s-CAACA algorithm in this paper achieves quite excellent accuracy in long crack detection. Thanks to the crack clustering algorithm, all long cracks within the preset range are merged into a single crack, which significantly enhances the visualization of crack detection. Additionally, with the adaptive confidence adjustment, many isolated road protrusions, potholes, or stones can be accurately identified without being misclassified as cracks.



**Fig. 8.** Detection results of YOLOv8s-CAACA for multiple cracks in a small area. (Picture credit: Original)

However, in cases where there are multiple cracks within a small area of the same image, the crack clustering algorithm tends to merge them into a single crack, which affects detection accuracy. For instance, in the three typical examples shown in Figure 8—where there are either two parallel cracks or two perpendicularly intersecting cracks within a small range—all are identified as a single crack by the proposed YOLOv8s-CAACA algorithm.

Meanwhile, due to the presence of the adaptive confidence adjustment algorithm, when there are obvious dark-colored stones or large granular pavement around the cracks, the model is prone to misidentifying these normal road conditions as cracks. Such issues are all exemplified in Figure 8.

## 5 Conclusion

Through this research, the context-based adaptive confidence adjustment algorithm can be extended to various detection tasks involving multiple identical targets in the same image, such as dense detection targets like lane lines. This algorithm can effectively filter various environmental noises while avoiding missed detections in high-density areas, improving the model's recognition accuracy without requiring additional optimization during the model training phase, and meeting lightweight requirements. The crack clustering algorithm, essentially a greedy clustering strategy, can also be widely applied in image recognition, such as pollutant detection in environmental protection zones. This algorithm can merge multiple identified small targets into a single large target, effectively enhancing the visualization of image recognition results.

This paper analyzes the application of the YOLOv8 algorithm in road crack detection and incorporates context-based adaptive confidence adjustment and crack clustering algorithms. Through these optimizations, the model has achieved significant improvements in detecting long cracks and cracks in complex environments, enhancing the accuracy and visibility of the model in practical applications.

## References

1. Guo, F.: Analysis of the impact of road transport on regional economic development. *China Shipping Weekly* (11), 87–89 (2025)
2. Wu, B.: Discussion on the development of road traffic economic research. *Jiangsu Business Review* (04), 48–51 (2025).
3. Qin, H.: Research on urban sustainable development and municipal road planning and design. *Engineering Construction and Design* (04), 61–63 (2025).
4. Zhang, Y.: Research on urban road disease analysis and maintenance decision optimization. *Engineering Technology Research* **9**(24), 151–153 (2024).
5. Tong, G.: Research on the repair technology of asphalt pavement cracks in municipal road maintenance. *Jiangxi Building Materials* (03), 387–388, 393 (2023)
6. Zheng, H.: Analysis and prevention of causes of asphalt pavement cracks. *Rural Economy and Science and Technology* **31**(10), 52, 83 (2020)
7. Wang, S.: A brief discussion on the treatment of road diseases in urban road maintenance. *Urban Roads, Bridges and Flood Control* (10), 128–129, 143, 16 (2016).
8. Oliveira, H., Correia, P.L.: Automatic road crack detection and characterization. *IEEE Transactions on Intelligent Transportation Systems* **14**(1), 155–168 (2012)
9. Zhang, L., Yang, F., Zhang, Y.D., et al.: Road crack detection using deep convolutional neural network. In: *Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3708–3712. IEEE, Phoenix (2016)
10. Jiang, P., Ergu, D., Liu, F., et al.: A review of YOLO algorithm developments. *Procedia Computer Science* **199**, 1066–1073 (2022)
11. Nie, M., Wang, C.: Pavement crack detection based on YOLO v3. In: *Proceedings of the 2019 2nd International Conference on Safety Produce Informatization (IICSPI)*, pp. 327–330. IEEE, Chongqing (2019)
12. Yanfei, P., Yue, J.: Road crack detection algorithm based on improved YOLOv8. In: *Proceedings of the 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pp. 28–32. IEEE, Dalian (2023)
13. Wang, H., Han, X., Song, X., et al.: Research on automatic pavement crack identification based on improved YOLOv8. *International Journal on Interactive Design and Manufacturing (IJIDeM)* **18**(6), 3773–3783 (2024)
14. Zeng, K., Fan, R., Tang, X.: Efficient and accurate road crack detection technology based on YOLOv8-ES. *Autonomous Intelligent Systems* **5**(1), 4 (2025)
15. Yao, Q., Wang, Y., Feng, Y., et al.: Road crack detection based on YOLOv8 integrated with dynamic snake convolution. *Science, Technology and Engineering* **25**(12), 5083–5092 (2025)
16. Pei, L., Wang, P., Wang, G., et al.: Pavement crack detection method based on deep feature migration and data enhancement. *Journal of Highway and Transportation Technology* **41**(02), 39–46 (2025).
17. Guo, W., Zhong, L., Zhang, D., et al.: Pavement crack detection using fractal dimension and semi-supervised learning. *Fractal and Fractional* **8**(8), 468 (2024)
18. Khanam, R., Hussain, M.: YOLOv11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725* (2024)
19. El-Din Hemdan, E., Al-Atroush, M.E.: A review study of intelligent road crack detection: Algorithms and systems. *International Journal of Pavement Research and Technology*, 1–31 (2025)
20. Wang, Z., Chen, C., Dong, F.: Road crack detection algorithm based on improved YOLOv8n. *Modern Electronic Technology*, 1–9 (2025)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

