Eng Int Syst (2025) 1: 77–87
© 2025 CRL Publishing Ltd

Engineering
Intelligent Systems

Intelligent Monitoring Technology for Traffic Road Construction Quality under Deep Learning

Xi Chen and Limin Zhang*

Department of Civil Engineering, Shijiazhuang Tiedao University Sifang College, Shijiazhuang 051132, Hebei, China

Traffic road construction is an important component of urban development. In response to the low accuracy of crack detection in traffic road construction and the difficulty in adaptingto complexroad construction scenarios, in this study, the Mask R-CNN (Mask Region-based Convolutional Neural Network) model combined with the SCSE (Spatial Channel and Squeeze Excitation) attention mechanism was adopted to study the intelligent monitoringof the quality of traffic road construction. Firstly, drones equipped with high-definition cameras were used for real-time construction data collection, integrated RI datasets, and then Mask R-CNN models were used to capture crack features. The SCSE attention mechanism module was introduced before the output of the mask branch to help the model better utilize feature information and improve the detection of crack-related features for the monitoring of construction quality. Finally, the Mask R-CNN SCSE model was embedded into the monitoring system. The experimental results showed that the recognition accuracy of the Mask R-CNN SCSE model reached 98.85%, which was 1.08% higher than the Mask R-CNN model. The mAP (mean average precision) reached 94.38%, which improved the accuracy of crack detection in traffic construction. Moreover, it can adapt to different light intensities, weather conditions, and road conditions.

Keywords: traffic road, intelligent monitoring of construction quality, crack detection, deep learning, SCSE attention mechanism

1. INTRODUCTION

With the continuous development of urban transportation infrastructure, the monitoring of the construction quality of roads has become crucial. The consistent monitoring of road construction quality is important to ensure road safety, sustainable development, and improve the quality of life of urban residents. At present, this monitoring is facing increasingly complex challenges, and there is a problem of poor detection ac curacy when dealing with large road networks and complex construction scenarios. In particular, for s mall but critical construction issues, such as crack detection, it is difficult to provide real-time monitoring and timely alerts, leading to delays in resolving the problem. This study was conducted to overcome the traditional shortcomings of fault detection methods, improve the intelligence level of traffic road construction quality monitoring, and provide innovative solutions for the sustainable development of urban transportation infrastructure.

With the increasing attention paid to infrastructure engineering, many scholars have begun to study the intelligent monitoring of traffic road construction quality and have achieved a large number of promising research results. For instance, Ranyal introduced intelligent sensing and artificial intelligence technologies to detect the aging of road infrastructure structures in a timely manner, providing reference for future research [1–3]. Di Graziano adopted intelligent sensor networks for timely health detection of asphalt pavement structures in order to reduce the cost of road deterioration [4]. Shtayat developed a system to monitor the condition of paved and unpaved road surfaces, which can promptly determine the severity of road damage and the type of road distress [5]. Staniek used crowdsourcing based on data

*Address for correspondence: Limin Zhang, Department of Civil Engineering, Shijiazhuang Tiedao University Sifang College, Shijiazhuang 051132, Hebei, China, Email: zlmhxp@163.com

from smartphone users in transportation systems to identify and evaluate road surface defects by analyzing the dynamics of vehicle movement in the road network [6]. Braunfelds used FBG (fiber Bragg grating) sensors to monitor the health of road infrastructure structures and predict the structural life of the pavement in a timely manner [7]. To some extent, these scholars have improved the accuracy of road surface detection; however, the accuracy of their methods is still not high and it is difficult to adapt their solutions to complex road environments.

In recent years, deep learning technology has been widely applied in various industries, and many scholars have applied it to intelligent monitoring of road construction quality to solve practical problems. Wang et al. proposed the introduction of neural network models for monitoring underground pavement engineerin g to improve detection accuracy [8–9]. Yamaguchi et al. used the improved VGG19 and self-organizing map technology to automatically detect and classify road cracks, reducing cases of false detection[10] . Kulambayev et al. used the Mask R-CNN model for real-time detection of road surface losses, with a detection accuracy of over 90% [11]. Qu introduced attention mechanisms and combined them with multi-feature fusion to detect cracks in concrete pavement, improving the accuracy of detection and adaptability to certain complex situations [12–13]. Pratico et al. used CNN (Convolutional Neural Network) to solve the problem of detecting bottom-up cracks, and the results showed an accuracy of 95.6% [14]. The aforementioned scholars demonstrated that using the Mask R-CNN model combined with the SCSE attention mechanism for intelligent monitoring of traffic road construction quality is feasible and can solve real-world problems. In order to solve the problem of low accuracy of crack detection in traffic road construction and difficulty in adapting to complex road construction scenarios, this current study adopted the Mask R-CNN-SCSE model to study crack detection in traffic road construction. Construction site images were collected through drone mounted high-definition cameras and RI datasets were integrated for image filtering, grayscale processing, crack labeling, random cropping, and random rotation preprocessing. Then, the Mask R-CNN-SCSE model was applied to detect the cracked area, learn the crack features, , and verify the results in the construction quality monitoring system of traffic channels based on the model. The results show that the introduction of the SCSE attention module improves the accuracy of crack detection and has good adaptability to different construction scenarios.

2. DEEP LEARNING TECHNOLOGY IN INTELLIGENT MONITORING OF TRAFFIC ROAD CONSTRUCTION QUALITY

Deep learning technology has a wide range of applications in the intelligent monitoring of traffic road construction quality, mainly focusing on defect recognition, object detection, image segmentation, etc., to achieve comprehensive monitoring of the construction site. In regard to object detection, deep learning models are used to detect construction site equipment, personnel, etc., providing a foundation for overall monitoring of construction quality [15]. In terms of defect recognition, deep learning technology can accurately locate and identify construction defects on road traffic, such as potholes, cracks, etc., through image recognition, and timely discover and solve construction problems [16]. In image segmentation, deep learning techniques are used to segment the construction site into specific areas, which helps to analyze the construction quality more precisely, especially when there are very short cracks.

In behavior recognition and violation detection, deep learning models can be used to detect non-standard behavior (e.g. not wearing safety helmets) and the incorrect operation of equipment on construction sites, thereby improving the safety management on road construction sites , and improve the level of safety management on road construction sites. In terms of real-time monitoring and early warning, deep learning technology can be used to monitor road construction sites in real time, and the early warning system can give timely detection of abnormal road construction. Strict quality monitoring of traffic road construction makes it more intelligent and precise, which helps to improve construction efficiency, reduce risks, and make significant contributions to the sustainable development of urban transportation infrastructure.

3. MASK R-CNN MODEL AND ATTENTION MECHANISM

3.1 Mask R-CNN Model

Mask R-CNN is an improved model based on Faster R-CNN (Faster Region-based Convolutional Neural Network). It modifies the RoI Pooling (Region of Interest Pooling) with nearest neighbor interpolation used in Faster R-CNN to RoI Align (Region of Interest Align) with bilinear interpolation. This solves the problem of region mismatch caused by quantization operations, ensures the accuracy of feature pixels, and adds mask prediction techniques [17–18]. In addition, on Faster R-CNN, FCN (Fully Convolutional Network) is used for the semantic segmentation of each suggestion box [19].

In the Mask R-CNN model, the image feature extraction part adopts a backbone network structure that combines ResNet (Residual Network) and FPN (Feature Pyramid Network). Multiple fusion of low-level and high-level features can be utilized to compensate for localization errors, enhance semantic information of features, and further improve crack detection performance [20]. Region Proposal Network (RPN) [21–22] is used in Mask R-CNN to generate refined region candidate boxes on the features output by the feature extraction network with FPN added. It generates marker points within each candidate box, which can calibrate the detection area and ultimately output the real construction road crack area. RoI Align eliminates quantization operations and calculates pixel values through bilinear interpolation, solving the problem of difficult boundary matching. Ro I Align outputs the overall pixel value for pixels that are relatively

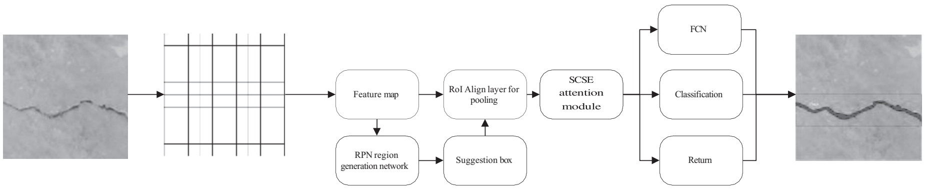


Figure 1 Mask R-CNN model network structure diagram.

close on the feature map, and finally outputs the region of interest. The result is maximum pooling of sampling points within all cells and, through backpropagation calculation, the image of the crack segmentation area is output.

The network structure diagram of the Mask R-CNN model is shown in Figure 1. Firstly, for the image to be detected, ResNet residual network is used to extract features, while FPN is used to fuse multi -scale feature information to obtain the image feature map. Then it integrates the feature map into RPN, outputs a suggestion box for road cracks, and introduces the SCSE attention mechanism to capture detailed features. Finally, the feature images in the suggestion box are classified, regressed, and segmented.

3.2 Loss Function

The total loss of Mask R-CNN for all regions of interest is calculated with formula (1) [23–24].

$$F = F_d + F_k + F_n \tag{1}$$

where F represents the total loss of the model; F_d is the classification loss; F_k represents regression loss, and F_n is the mask loss.

The classification loss is calculated by the logarithm of the softmax loss function, calculated with formula (2).

$$F_d(q, v) = -log_2(q_v) \tag{2}$$

Among them, $q = (q_0, q_1, \ldots q_i)$.

The regression loss function is calculated using the loss function, F_k calculated with formula (3).

$$F_k(s^v, u) = \Sigma_{s\epsilon\{x,y,a,b\}} smooth_{F_k}(s_s^v - u_j) \tag{3}$$

where $u = u_x, u_y, u_a, u_b$ represents the coordinates of the boundary box of the crack, and $s_v = (s_x^v, s_y^v, s_a^v, s_b^v)$ represents the coordinate correction of the boundary box of the crack.

Formula (4) is used to calculate the smooth$_{F_k}$ loss function.

$$smooth_{F_k}(x) = \begin{cases} \frac{1}{2}x^2 & |x| < 1 \\ |x| < \frac{1}{2} & \text{other} \end{cases} \tag{4}$$

3.3 SCSE Attention Mechanism

The SCSE attention module is a spatial squeezing and channel excitation attention module that combines channel attention mechanisms, as well as a channel squeezing and spatial excitation attention module that combines spatial attention mechanisms. This helps the model to better capture and understand important channel information related to cracks, facilitating the localization and detection of cracks [25–27]. The Channel Squeeze Excitation (cSE) module is calculated as follows [28]:

(1) Firstly, the input feature mapping $E = [e_1, e_2, \ldots, e_n]$ is used to generate a vector Q through global average pooling.

(2) Then, The Channel Squeeze Excitation (cSE) module can be achieved by passing through a fully connected layer, followed by a ReLU (Corrected Linear Unit) layer.

(3) Finally, the sigmoid layer can be used to output activation values, focus ing more on important paths and ignoring the unimportant ones.

The calculation formula for the cSE module is shown in formula (5) [29].

$$\hat{E}_{cSE} = E_{cSE}(E) = [\delta(\hat{Q}_1)e_1\delta(\hat{Q}_2)e_2, \ldots, \delta(\hat{Q}_n)e_n] \tag{5}$$

Spatial Squeeze Excitation (sSE) module is calculated as follows:

(1) Firstly, the inputtable feature map $E = [e_{1,1}, e_{1,2}, \ldots, e_{i,j}, \ldots, e_{A,B}]$ is used, and then the o vector is output through 1 * 1 convolution.

(2) Then, the sigmoid layer can be used to output the activation value.

Formula (6) is used for the calculation of the spatial incentive attention module:

$$\widehat{E_{sSE}} = E_{sSE}(E) = [\delta(o_{1,1})e_{1,1}, \delta(o_{2,2})e_{2,2}, \ldots, \delta(o_{A,B})e_{A,B}] \tag{6}$$

3.4 Fusion Methods

For the SCSE attention mechanism and Mask R-CNN model, in this study, a SCSE attention module is inserted between the output of the backbone and the output of the mask branch based on the Mask R-CNN model. This ensures that the SCSE attention module introduces spatial and channel attention mechanisms while extracting features. The SCSE attention mechanism helps the model better utilize feature information and improve the identification of task-related features. At the same time, the SCSE module improves the model's ability to resist noise, making it more robust in practical scenarios. The various interferences and noises that may exist in road construction sites have better adaptability.

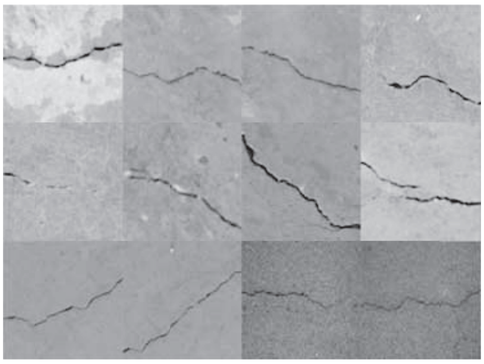


Figure 2 Photos of some original road surface cracks in the RI dataset.

Figure 3 Photos taken at the construction site.

4. EXPERIMENTAL DATA ON CRACK DETECTION IN TRAFFIC ROAD CONSTRUCTION QUALITY

4.1 Experimental Dataset

The dataset used for the experiment was sourced from the publicly available RI dataset and images collected from several construction sites. The RI dataset contains a total of 500 images of construction cracks, with 651 on-site images, including images of constructions with and without cracks. In this paper, 322 images were randomly selected from the public dataset, and 651 images were selected from the self-collected on-site images as the self-built dataset for this experiment, totaling 973 images. The experimental data partitioning method adopts the ten-fold cross validation method to partition the self-built dataset, with 30% as the test set and 70% as the training set. The experiments were conducted alternately and the average of the results was taken as the final experimental result. The original images of cracks are shown in Figures 2 and 3.

4.2 Data Preprocessing

(1) Grayscale processing In order to reduce feature dimensions, avoid color interference, and reduce computational complexity, in this study, grayscale preprocessing was performed on crack images. Firstly, a color crack image is loaded, and then the grayscale averaging method [30] is used to calculate the corresponding grayscale values based on the pixel values of the three channels in the RGB color space. Finally, the color values of each pixel in the road crack image are converted into corresponding grayscale values. The processed grayscale images are shown in Figure 4A and 4B.

(2) Image crack annotation In this experiment, the labeling program was used to label the crack image dataset. After labeling, a target area was created based on the contour of the image containing the crack, and the crack area was labeled. The annotation of the crack image is shown in Figure 5.

(3) Image enhancement In order to balance the dataset, random rotation and random cropping enhancement operations were performed on the images [31-32]. To enhance the crack images, some were uniformly rotated by 20 degrees and others were randomly rotated, as shown in Figure 6A and 6B.

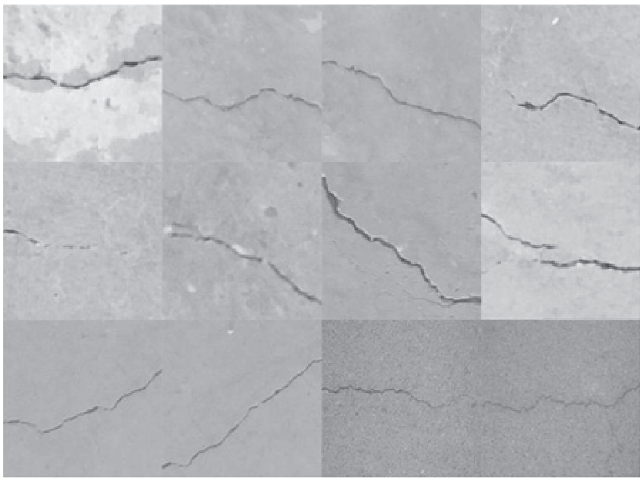
5. DETECTION AND QUALITY MONITORING EXPERIMENT OF TRAFFIC ROAD CONSTRUCTION CRACKS

5.1 Experimental Environment

This experiment was conducted using the Windows 10 operating system, equipped with 64GB of memory, Inter Core i9-10900X CPU (Central Processing Unit), 32GB (Gigabyte) graphics memory, and NVIDIA Quadro RTX5000 GPU (Graphics Processing Unit).

5.2 Experimental Process

The experimental field data was collected from multiple construction sites using drones equipped with high-definition



(A)

(B)

Figure 4 Image after grayscale processing.

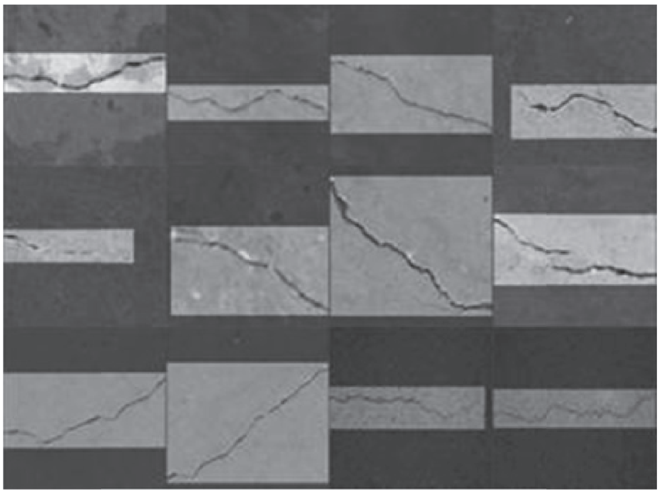


Figure 5 Crack image annotation.

cameras, and monitored in the background based on the real-time interface of the construction site. If cracks appear during construction, an alarm can be issued to alert construction personnel to take timely response measures. Firstly, for this experiment, the original crack dataset was collected and preprocessing steps were performed: grayscale, crack image annotation, random rotation, and cropping. Then, the training set was trained 240 times and the number of batches was set to 4. To avoid model fitting, the learning rate was set to 0.004 and the weight screening attenuation coefficient to 0.0001. Finally, the performance of the crack detection method was determined using the test set. The experimental flowchart is shown in Figure 7.

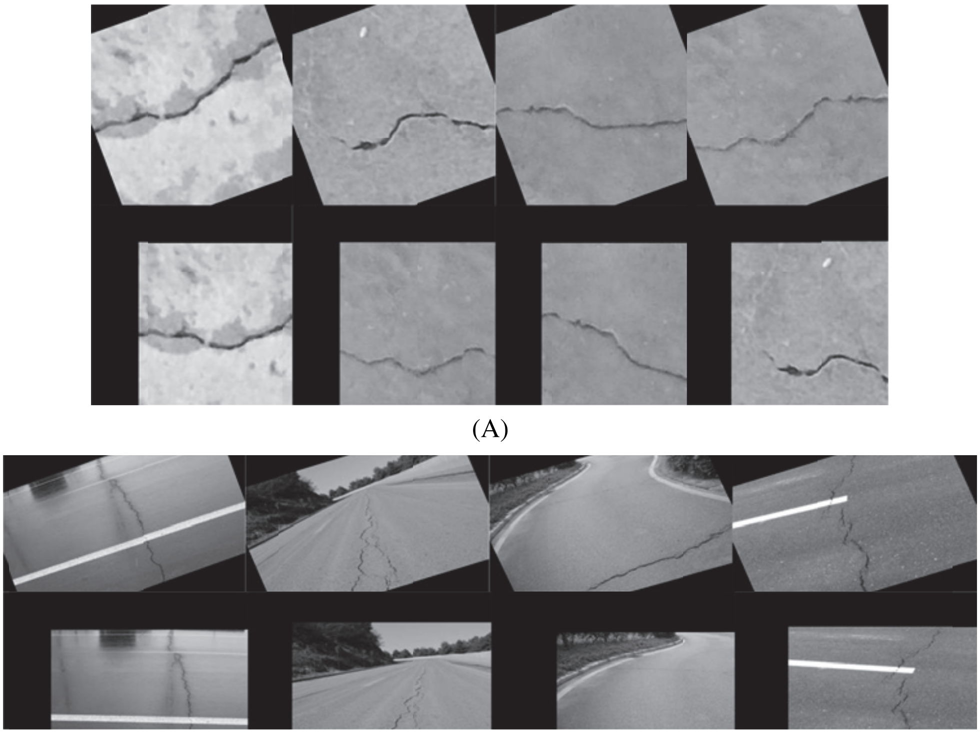
5.3 Experimental Results

Evaluating indicator
Accuracy: the proportion of all correctly judged results in the classification model to the total.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TNP} + \text{N}}$$

Precision: the proportion of truly correct predictions to all positive predictions.

$$\text{precision} = \frac{\text{TP}}{\text{TP+P}} \tag{7}$$



(A)

(B)

Figure 6 Image enhancement.

Recall rate: the proportion of true positives to all actual positives.

$$\text{recall} = \frac{\text{TP}}{\text{TP+N}} \tag{8}$$

f1 score: the f1 value is the arithmetic mean divided by the geometric mean. The f1 value is weighted for both accuracy and recall, and the f1 score belongs to 0-1.

$$\text{f1} = \frac{\text{2TP}}{\text{2TP+P+N}} \tag{9}$$

The results of crack identification obtained by the experiment are shown in Table 1.

6. EXPERIMENTAL DISCUSSION ON CRACK DETECTION ON ROAD CONSTRUCTION QUALITY

6.1 Results of Crack Identification in Road Construction

Table 1 shows the results of crack identification in the construction of traffic roads. Overall, both wider and finer cracks can be detected well, and instance segmentation is successful, achieving good results. The area of cracks can be estimated while detecting. Number 6 and number 11 are normal construction site images, and there are no cracks on the road's surface. The remaining images are of construction scenes with cracks.

The overall recognition of traffic road construction cracks is good, but there is a certain degree of error for some small and discontinuous cracks. For instance, in number 5, there are some pits and cracks that have not been segmented and detected. Images 9 and 12 show that the fractured cracks can also be accurately detected, indicating that the experimental model is able to detect fractured cracks.

6.2 Crack Detection Performance of Different Models

Based on the above experiments, in order to verify the performance of Mask R-CNN-SCSE, different models were applied to this experiment under the same experimental environment. Then, the performance of the proposed method was compared with Mask R-CNN, Yo lov3 (You Only Look Once version 3), SSD (Single Shot MultiBox Detector), and HRNet (High Resolution Network) models, as shown in Figure 8. In Figure 8, orange represents the accuracy of crack recognition, green indicates the accuracy of recognition, and purple represents the recall rate of recognition.

In terms of accuracy, the Mask R-CNN achieved an accuracy of 97.77%, while the Mask R-CNN SCSE model achieved an accuracy of 98.85%, an improvement of 1.08% compared to the Mask R-CNN model. It can be seen that the introduction of the SCSE attention mechanism has a good improvement effect on the recognition accuracy of the model. The recognition accuracy of Yolov3 reached 92.52%, a decrease of 6.33% compared to the Mask R-CNN-SCSE model, indicating a significant decrease. In addition, the SSD model achieved 88.75%, while the HRNet model achieved 78.30%. In terms of accuracy, the Mask R-CNN SCSE model achieved 97.20%, an improvement of 1.64% compared to the Mask R-CNN model, and the Yo lov3 model achieved 90.80%.

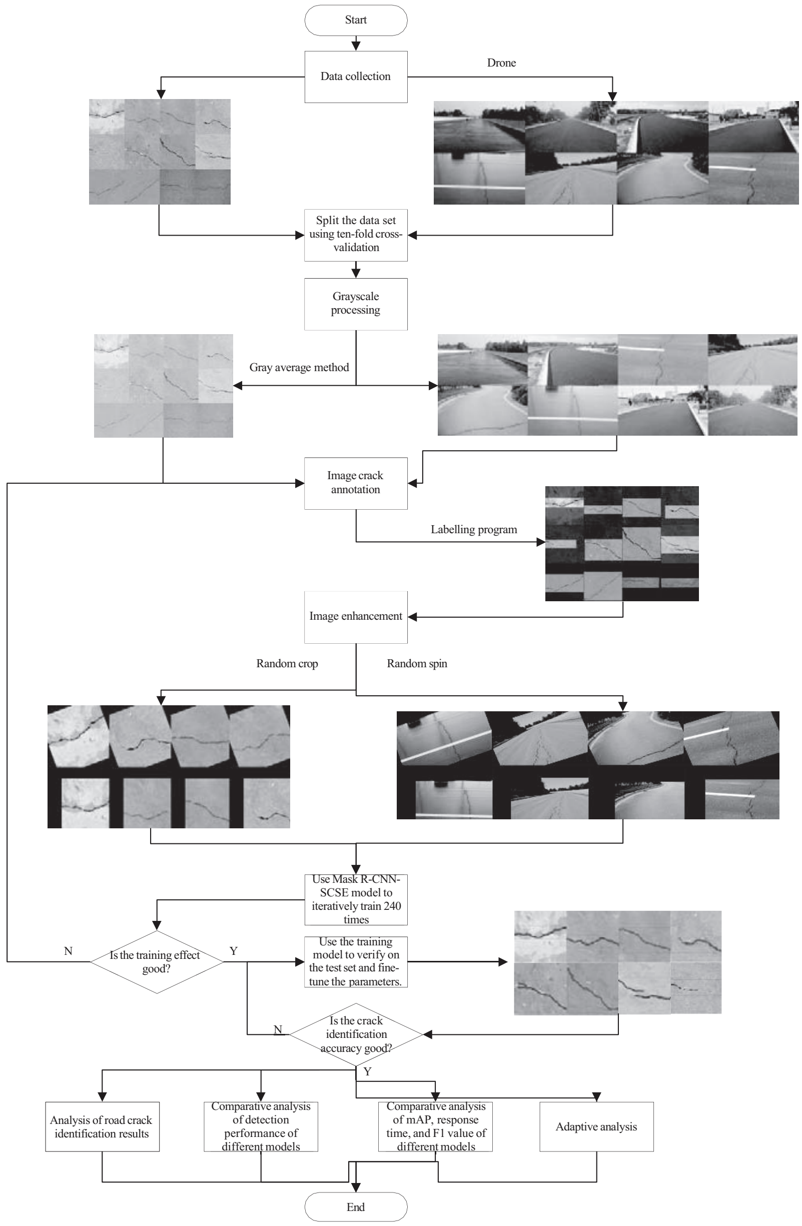


Figure 7 Experimental flowchart.

The SSD model and HRNet model are relatively poor, with only 84.21% and 81.49%, respectively. In terms of recall rate, the lowest is the HRNet model, which is only 79.10%, and the highest is the Mask R-CNN-SCSE model, achieving up to 99.59%. The Mask R-CNN model and Yolov3 model both achieved over 93.0% accuracy, with the SSD model being the worst performer at only 87.46%.

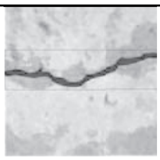
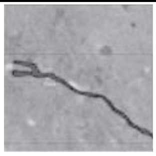
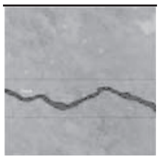
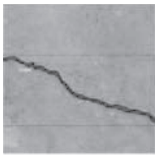
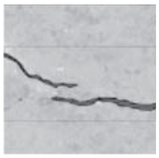
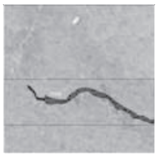
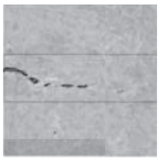
Overall, the performance of the Mask R-CNN was improved by the addition of the SCSE attention mechanism. The Mask R-CNN-SCSE model achieved the best performance in crack detection accuracy, precision, and recall.

6.3 mAP, Response Time, and F1 Values Among Different Models

In order to further verify the performance of the Mask R-CNN SCSE model, the mAP, response time, and F1 values of different models were analyzed and compared, as shown in Figure 9. In Figure 9, the blue line represents mAP, the red line represents response time, and the green line represents F1 value.

In regard to the mAP, both the Mask R-CNN-SCSE model and Mask R-CNN model achieved over 91.0% accuracy, and the Mask R-CNN-SCSE model improved by 2.43% compared to the Mask R-CNN model, followed by the Yolov3 model, which reached 84.36%. The SSD model and HRNet model are relatively poor, both below 80.0%, at 78.01% and 73.39%, respectively. In terms of response time, the Mask R-CNN SCSE model reached 0.9 seconds. Compared to the Mask R-CNN model, it requires an additional 0.3 seconds, but it consumes 0.6 seconds less than the HRNet model. This is because the addition of the SCSE attention module has increased the complexity of the model to a certain extent, resulting in a longer response time, while the Yo lov3 model and SSD model have lighter structures and shorter response

Table 1 Results of crack identification.

Serial number	Picture	Actual type	Forecast type	Serial number	Picture	Actual type	Forecast type
1		Crack	Crack	7		Crack	Crack
2		Crack	Crack	8		Crack	Crack
3		Crack	Crack	9		Crack	Crack
4		Crack	Crack	10		Crack	Crack
5		Crack	Crack	11		Normal	Normal
6		Normal	Normal	12		Crack	Crack

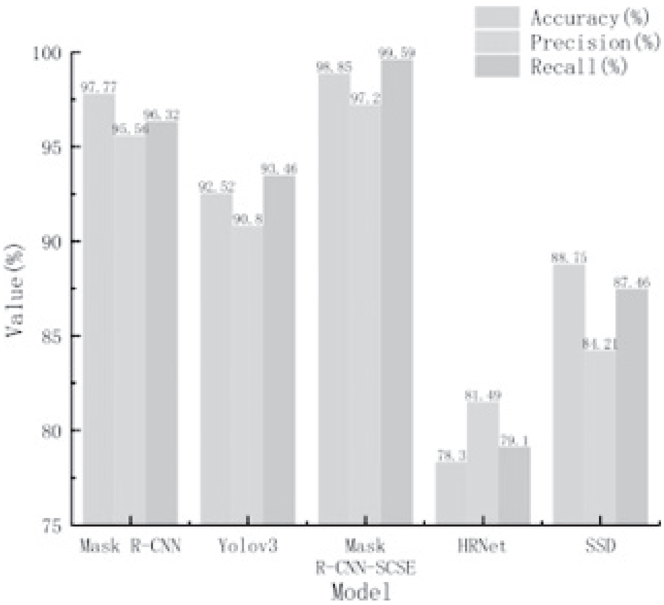


Figure 8 Crack detection performance of different models.

times. The response time for detecting cracks in the Yo lov3 model is 0.5 seconds, with the shortest being the SSD model, which only takes 0.3 seconds. From the F1 value, the Mask R-CNN SCSE model reached 0.96, the Mask R-CNN model reached 0.94, a decrease of 0.02 compared to the Mask R-CNN SCSE model, and the Yolov3 model reached 0.91. The SSD model and HRNet model have lower F1 values, with a decrease of 0.11,0.16 compared to the Mask R-CNN-SCSE model.

Overall, it can be seen that the Mask R-CNN-SCSE model has achieved good results in identifying cracks in traffic road

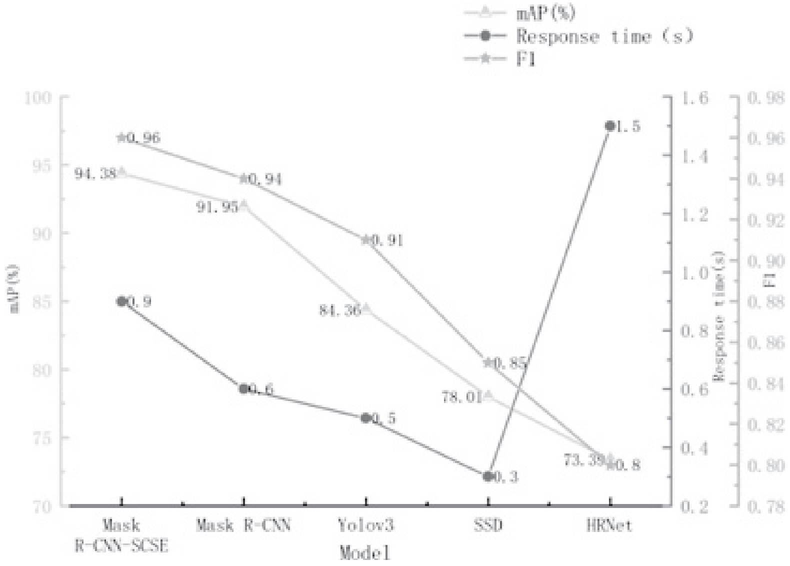


Figure 9 Comparison of mAP, response time, and F1 values for different models.

Table 2 Model adaptability.

Different environments	Light intensity			Weather conditions			Road surface material	
	Light intensity 24%	Light intensity 50%	Light intensity 92%	Sunny day	Cloudy day	Rainyday	Asphaltroad	Cementroad
Accuracy(%)	90.21	95.53	80.30	98.85	88.37	79.22	92.73	94.50
AUC	0.89	0.94	0.82	0.91	0.87	0.78	0.89	0.92
Precision(%)	87.04	94.23	80.52	97.20	90.81	79.93	90.55	92.83
F1	0.86	0.93	0.80	0.96	0.90	0.76	0.92	0.94

construction in terms of mAP and F1. Due to its high complexity, the response time is not ideal, but the actual demand for load is not ideal.

6.4 Adaptability

The construction site of transportation roads inevitably faces conditions such as light intensity, different weather conditions, and different pavement materials. In order to further explore the adaptability of the Mask R-CNN SCSE model, this article compares and analyzes the above three models in terms of accuracy, precision, AUC (Area Under Curve), and F1 value, as shown in Table 2. In Table 2, the light intensity in the experiment is divided into three component experiments, namely 24%, 50%, and 92% light intensity, represent ing low, medium, and high levels of light intensity respectively. The weather conditions are divided into three types: sunny, cloudy, and rainy. The road surface materials are divided into asphalt road and cement road.

In terms of different light intensities, the accuracy of 24% light intensity reached 90.21%, and the AUC reached 0.89. The accuracy of 50% light intensity reached 95.53%, and the AUC reached 0.94. The accuracy of 92% light intensity reached 80.30%, and the AUC reached 0.82, indicating that th e experimental model has good adaptability to 50% light intensity. It has the highest accuracy and AUC, and the best performance. At the same time, it can be found that the model is more appropriate for relatively weak light conditions compared to strong light. In addition, the precision of 24% light intensity reached 87.04%, the F1 value reached 0.86, and the precision reached 94.23% under 50% light intensity. The F1 value reached 0.93, the precision reached 80.52% under 92% light intensity, and the F1 value reached 0.80, which can also be concluded as above.

In regard to the different weather conditions, the accuracy under sunny conditions reached 98.85%, and the AUC reached 0.91. The accuracy under cloudy conditions reached 88.37%, AUC reached 0.87, and accuracy under rainy conditions reached 79.22%. The AUC reached 0.78, indicating the best performance in road construction quality under sunny conditions. Under sunny conditions, the precision reached 97.20%, F1 reached 0.96, and under cloudy conditions, the precision reached 90.81%, F1 reached 0.90. Under rainy conditions, the precision reached 79.93%, and F1 reached 0.76, indicating that the crack recognition performance was the worst under rainy conditions, while under cloudy conditions, the crack recognition performance was better than under rainy conditions. This is precisely because rainy days not only cause blurred vision, but also lead to a large amount of debris on the road surface that covers up the original cracks, resulting in detection errors.

Regarding the road surface materials, the accuracy rate of crack identification in asphalt pavement s reached 92.73%, and the AUC reached 0.89. The accuracy rate of crack identification in the construction cement pavement reached 94.50%, and the AUC has reached 0.92. It can be seen that crack detection is better in cement road, since the precision of crack detection on asphalt roads was 90.55%, with an F1 value of 0.92. The precision for cement roads reached 92.83%, and the F1 value reached 0.94. It can also be concluded that cracks can be more easily detected in cement roads.. This is because there is a certain color difference in images of asphalt roads captured by a camera, captured, but the difference between the two in terms of detection is not significant.

In summary, it can be seen that the Mask R-CNN-SCSE model exhibits good performance under different lighting intensities, conditions, and materials, and can adapt to certain complex environments.

7. CONCLUSIONS

In this study, the Mask R-CNN model combined with the SCSE attention mechanism was used to identify cracks in traffic road construction quality, and drones equipped with high-definition cameras were deployed for data collection. This study introduced the SCSE attention mechanism and then embedded the Mask R-CNN-SCSE model into the traffic road construction quality monitoring system. In fact, displaying the area of cracks in real time improves detection accuracy and has good adaptability to different conditions. There are still some shortcomings in this study, such as insufficient adaptability to different scenarios, especially in regard to the detection of cracks in road surfaces when the light intensity is strong, there are rainy conditions, and the road surface is asphalt. Moreover, this model is highly complex, resulting in unsatisfactory response time. In the future, the model's performance could be improved by making it more lightweight.

REFERENCES

1. Ranyal E, Sadhu A, Jain K. Road condition monitoring using smart sensing and artificial intelligence: A review[J]. Sensors, 2022, 22(8): 3044–3070. https://doi.org/10.3390/s22083044
2. Hou Y, Li Q, Zhang C, Lu G,Ye Z,Chen Y,et al. The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analys is[J]. Engineering, 2021, 7(6): 845–856. https://doi.org/10.1016/j.eng.2020.07.030
3. Oladimeji D, Gupta K, Kose N A, Gundogan K,Ge L,Liang F. Smart transportation: an overview of technologies and applications[J]. Sensors, 2023, 23(8): 3880–3911. https://doi.org/10.3390/s23083880
4. Di Graziano A, Marchetta V, Cafiso S. Structural health monitoring of asphalt pavements using smart sensor networks: A comprehensive review[J]. Journal of Traffic and Transportation Engineering (English Edition), 2020, 7(5): 639–651. https://doi.org/10.1016/j.jtte.2020.08.001
5. Shtayat A, Moridpour S, Best B, Shroff A,Raol D. A review of monitoring systems of pavement condition in paved and unpaved roads[J]. Journal of Traffic and Transportation Engineering (English Edition), 2020, 7(5): 629–638. https://doi.org/10.1016/j.jtte.2020.03.004
6. Staniek M. Road pavement condition diagnostics using smartphone-based data crowdsourcing in smart cities[J]. Journal of traffic and transportation engineering (English edition), 2021, 8(4): 554–567. https://doi.org/10.1016/j.jtte.2020.09.004
7. Braunfelds J, Senkans U, Skels P, Janeliukstis R,Salgals T,Redka D,et al. FBG-based sensing for structural health monitoring of road infrastructure[J]. Journalof Sensors, 2021, 2021(1): 1–11. https://doi.org/10.1155/2021/8850368
8. Wang L, Xu S, Qiu J, Wang K,Ma E,Li C,et al. Automatic monitoring system in underground engineering construction: review and prospect[J]. Advances in Civil Engineering, 2020, 2020: 1 -16. https://doi.org/10.1155/2020/3697253
9. Roberts R, Giancontieri G, Inzerillo L, Mino G D. Towards low-cost pavement condition health monitoring and analysis using deep learning[J]. Applied Sciences, 2020, 10(1): 319–340. https://doi.org/10.3390/app10010319
10. Yamaguchi T, Mizutani T. Road crack detection interpreting background images by convolutional neural networks and a self- organizing map. Computer-Aided Civil and Infrastructure Engineering, 2024, 39(11): 1616–1640. https://doi.org/10.1111/mice.13132
11. Kulambayev B, Nurlybek M, Astaubayeva G, Tleuberdiyeva G, Zholdasbayev S. Real-Time Road Surface Damage Detection Framework based on Mask R-CNN Model[J]. International Journal of Advanced Computer Science and Applications, 2023, 14(9):757–765. DOI:10.14569/IJACSA.2023.0140979
12. Qu Z, Chen W, Wang S Y, Yi T N,Liu L. A crack detection algorithm for concrete pavement based on attention mechanism and multi-features fusion[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(8): 11710–11719. https://doi.org/10.1109/TITS.2021.3106647
13. Sun X, Xie Y, Jiang L, Cao Y,Liu B. DMA-Net: DeepLab with multi-scale attention for pavement crack segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 18392–18403. https://doi.org/10.1109/TITS.2022.3158670
14. Pratico F G, Fedele R, Naumov V, Sauer T. Detection and monitoring of bottom-up cracks in road pavement using a machine-learning approach[J]. Algorithms, 2020, 13(4): 81–96. https://doi.org/10.3390/a13040081
15. Pan H. Application of 5Gwireless communication and BIM technology in management of construction projects. Engineering Intelligent Systems, 2024, 3:191–201.
16. Wang J. Design of intelligent traffic sign image recognition system based on machine learning algorithms. Engineering Intelligent Systems, 2024, 4: 1–8.
17. Zhang Y, Chu J, Leng L,Miao J. Mask-refined R-CNN: A network for refining object details in instance segmentation[J]. Sensors, 2020, 20(4): 1010–1025. https://doi.org/10.3390/s20041010
18. Xu X, Zhao M, Shi P, Ren R,He X,Wei X,et al. Crack detection and comparison study based on faster R-CNN and mask R-CNN[J]. Sensors, 2022, 22(3): 1215–1231. https://doi.org/10.3390/s22031215
19. Zhu Lei, Li Dongbiao, Yan Xingzhi, Liu Xiangyang, Shen Cai. Intelligent tunnel crack detection method based on improved Mask R-CNN deep learning algorithm [J]. Journalof Graphics, 2023, 44(1): 177–183. https://doi.org/10.11996/JG.j.2095-302X.2023010177
20. Liu Qingsong, Wen Xuguang, Luo Wenbin, Wang Yijun, Hui Chenliang.A method for identifying cracks in railway tunnels based on MASK R-CNN [J].Scientific and technological innovation,2023,5(03):5–9.

21. Liu Zhiyong. Research on intelligent identification of cracks in concrete Roads based on Mask R-CNN and migration learning [J]. Guangdong Architecture Civil Engineering, 2021, 28(8):5–43. DOI:10.19731/j.gdtmyjz.2021.08.002
22. Sun Z, Geng H, Lu Z, Scherer R,Wozniak M. Review of road segmentation for SAR images[J]. Remote Sensing, 2021, 13(5): 1011–1024. https://doi.org/10.3390/rs13051011
23. Yu Jiayong, Li Feng, Xue Xiankai, Zhu Ping, Wu Xinyun, Lu Peisheng. Intelligent identification of cracks in bridge structures based on drones and Mask R-CNN [J]. China Highway Journal, 2021, 34(12): 80–90.
24. Chen Q, Gan X, Huang W, Feng J,Shim H. Road damage detection and classification using mask R-CNN with DenseNet backbone[J]. Comput Mater Contin, 2020, 65(3): 2201–2215. doi:10.32604/cmc.2020.011191
25. Zhang Ruiyan.Highway crack detection combining multi-scale characteristics and attention mechanism [J].Modern electronic technology,2023,46(3):100–104. DOI:10.16652/j.issn.1004-373x.2023.03.019
26. Jia Rui, Zhao Hongyan. Research on crack detection based on channel attention mechanism [J].Value engineering,2022,41(32):120–122. doi:10.3969/j.issn.1006-4311.2022.32.037
27. Zhang Zhihua, Wen Yanan, Mu Zhiwei, Du Xiaop-ing.Road crack detection combined with dual attention mechanism [J].Chinese Journal of Graphics and Graphics, 2022,27(07):2240–2250. DOI:10.11834/jig. 200758
28. Geng P, Lu J, Ma H, Yang G. Crack Segmentation Based on Fusing Multi-Scale Wavelet and Spatial-Channel Attention[J]. Structural Durability & Health Monitoring (SDHM), 2023, 17(1):1–22. DOI: 10.32604/sdhm.2022.018632
29. Bakr E M, El-Sallab A, Rashwan M. EMCA: Efficient Multiscale Channel Attention Module[J]. IEEE Access, 2022, 10: 103447–103461. https://doi.org/10.1109/ACCESS.2022.3205602
30. Kaur H, Koundal D, Kadyan V. Image fusion techniques: a survey[J]. Archives of computational methods in Engineering, 2021, 28(1): 4425–4447. https://doi.org/10.1007/s11831-021-09540-7
31. Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques[J]. Global Transitions Proceedings, 2022, 3(1): 91–99. https://doi.org/10.1016/j.gltp.2022.04.020.
32. Wang Z, Yang J, Jiang H, Fan X. CNN training with twenty samples for crack detection via data augmentation[J]. Sensors, 2020, 20(17): 4849–4865. https://doi.org/10.3390/s20174849

Xi Chen was born in Cangzhou, Hebei. P.R. China, in 1987. She received the master's degree from Chongqing Jiaotong University, P.R. China. Currently, she works at Sifang College of Shijiazhuang Railway University. Her research interest is road and railway engineering.
E-mail: 18032059001@163.com

Limin Zhang was born in Shijiazhuang, Hebei. P.R. China, in 1980. She received the master's degree from GuangXi University for Nationalities, P.R. China. Currently, she works at Sifang College of Shijiazhuang Railway University. Her research direction is road and railway engineering.
E-mail: zlmhxp@163.com