Road Damage Detection Algorithm Based on Multi-scale Feature Extraction

Zhixian Zhang, Wenhua Cui*, Ye Tao, and Tianwei Shi

***Abstract*—Deep learning is proliferating within the field of computer vision. Road damage detection technology, as its offshoot, already plays vital role in road maintenance and traffic safety. With road damage such as potholes and cracks, accurate and efficient detection results are essential for timely road safety repair and maintenance. Therefore, road damage detection algorithms based on deep learning have attracted wide attention. YOLOv5 is an advanced target detection algorithm known for its efficient detection speed and good accuracy. However, there is still room for further improvement in its performance for road damage detection. The ability of multi-scale damage detection and spatial structure capture is not perfect. Therefore, this paper proposes three improvement points to improve the accuracy of road damage detection based on YOLOv5. The first introduced module is the Non-linear Spatial Pyramidal Pooling-Fast (NSPPF) module. This module allows for better capture of detailed features of road damage areas. Non-linear transformation and fast pyramid operation improve the sensing ability and multi-scale damage detection ability. Secondly, a combination of the CoordConv and SK attention modules is constructed. The CoordConv module fuses coordinate information with features to provide a more spatially informed representation. The SK attention module also learns correlations between global and local features, enhancing the model's ability to detect damages at different scales. This paper can better capture road injuries' spatial structure and context information by combining these two modules. Finally, experimental results on the RDD2020 dataset demonstrate the effectiveness of our model. Compared to the baseline model, the proposed improvement algorithm increases the accuracy by 2.2%, resulting in a mAP of 58.2%. This demonstrates its effectiveness and feasibility.**

***Index Terms*—YOLOv5, NSPPF, CoordConv, SK Attention, RDD2020, Road Damage Detection**

Manuscript received June 13, 2023; revised November 6, 2023.This work was supported by Joint Fund Project of the National Natural Science Foundation of China (U1908218), the Natural Science Foundation project of Liaoning Province (2021-KF-12-06), and the Department of Education of Liaoning Province (LJKFZ20220197).

Zhixian Zhang is a Postgraduate of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China. (e-mail: 1849535553@qq.com).

Wenhua Cui* is a Professor of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China. (Corresponding author to provide phone: +86-133-0422-4928; e-mail: taibeijack@126.com).

Ye Tao is a Lecturer of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China. (e-mail: taibeijack@163.com).

Tianwei Shi is an Associate Professor of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China. (e-mail: tianweiabbcc@163.com).

I. Introduction

ROADS are an essential part of the modern transport system, carrying people on their journeys and transporting goods. Their good condition is vital in ensuring traffic safety, improving traffic efficiency, and promoting economic development. It is therefore important to know the extent of damage to the road and its subsequent maintenance [1]. However, over time and with frequent use, roads inevitably suffer from various types of damage, such as cracks and potholes. If these road damages are not detected and repaired in time, they do not only cause inconvenience to traffic. They can even lead to traffic accidents and vehicle damage, seriously threatening to traffic flow and safety. It will expand further and harm the overall quality and longevity of the road infrastructure. The development of Intelligent Transport Systems (ITS) and Advanced Driver Assistance Systems (ADAS) requires the use of road damage detection technology [2, 3]. Traditional road damage detection algorithms rely on manual inspection and empirical judgment, with problems such as high workload, low efficiency, and subjectivity. Therefore, there is a tremendous practical need to research and apply fast, efficient, and accurate methods for detecting road damage [4]. In current years, the speedy improvement of computer vision and deep learning technologies has opened new opportunities for automated road damage detection. This rise in computing power has played a pivotal role in propelling the progress of artificial intelligence [5]. In this context, the improved model based on YOLOv5 has become an important research direction in road damage detection. YOLOv5 is the fifth generation detection algorithm in the YOLO series [6]. It has excellent potential for improving detection accuracy and robustness. However, the traditional YOLOv5 model still faces some challenges in road damage detection. For example, these issues include the lack of universality for road damage of different shapes and sizes, as well as accuracy issues in complex backgrounds and adverse weather conditions.

In order to address these issues, this paper introduces three key improvement modules within the enhanced YOLOv5 model. Firstly, a Non-linear Spatial Pyramid Pooling-Fast (NSPPF) module is incorporated. By conducting multi-scale non-linear pooling operations on the input feature maps. This module is able to extract road damage features at different scales in an efficient manner. The module introduces non-linear operations compared to traditional Spatial Pyramid Pooling-Fast (SPPF). The model's ability to sense road damage is further enhanced. Secondly, this paper introduces the addition of the CoordConv module. This module fuses coordinate information with feature maps to enhance the model's understanding of road damage geometry and location information. The module uses coordinate information as an additional channel input. The improvement enables the model to better distinguish between different locations of road damage, improving the accuracy and robustness of the detection. Finally, this paper establishes the combination of the CoordConv and SK attention modules. The SK attention module enables finer-grained feature selection and weighting by learning correlations between feature map channels. The SK attention module utilizes the output of the CoordConv module as its input. As a result, the model becomes more capable of accurately capturing essential features of road damage areas.

Through the improvements above, this study aims to enhance the performance and practicality of the road damage detection system that utilizes the YOLOv5 model. To validate the effectiveness of the improved model, the mAP on the RDD2020 dataset in this paper reached 58.2%, an enhancement of 2.2% in contrast to the baseline model. This result indicates that the enhanced model in this work has enhanced robustness and detection accuracy in tasks involving the identification of road damage.

II. Related Work

A. Road Damage Detection Algorithm Based on Conventional Image Processing

In the evolution of road damage detection, many studies have used techniques based on traditional image processing algorithms. Some of these algorithms mainly rely on threshold segmentation and feature extraction techniques. In previous road breakage detection algorithms, researchers have often used threshold segmentation to segment the target from the background. For example, Akagic et al. [7] proposed a combined algorithm based on grey-scale histograms and OTSU thresholds. Pavement cracks are detected by segmenting the input image into sub-images and searching for pavement cracks. Sari et al. [8] used OTSU threshold algorithm and Gray-level Co-occurrence Matrix (GLCM) to detect and extract features of road cracks and used Support Vector Machine (SVM) for classification and statistical analysis. However, these algorithms tend to be more sensitive to changes in lighting and background complexity. These algorithms have limited performance in complex road environments. Some algorithms use edge detectors, such as Canny [9] and Sobel [10], to extract edge information from road damage. Maode et al. [11], for example, used a modified median filter and morphological filter to detect cracks. Although these algorithms can extract edge information better, they have limitations for complex road damage shapes and noise. In addition, some studies have used machine learning algorithms, such as Support Vector Machine (SVM), to apply this to classify and detect road damage. Hoang [12] proposed a supervised learning algorithm based on SVM, which establishes an automatic road pothole classification algorithm, as opposed to single-road pothole detection. Gao et al. [13] proposed a fast detection algorithm using a Library of Support Vector Machines (LIBSVM) machine learning model. This algorithm can distinguish between different types of road damage.

However, road damage detection algorithms based on traditional image processing algorithms have common shortcomings and limitations. Firstly, these algorithms usually rely on hand-designed feature extraction and threshold selection. These algorithms may not generalize well to different types and shapes of road damage. Secondly, conventional algorithms are sensitive to lighting conditions and background interference changes and are susceptible to noise and complex environments. These issues reduce the accuracy of the detection.

B. Road Damage Detection Algorithm Based on Deep Learning

With its powerful feature extraction capabilities, researchers are increasingly using deep learning-based models. Convolutional neural networks [14] can be used for various tasks, including image classification [15], object detection [16], and semantic segmentation [17]. Researchers commonly classify current object detection networks for road damage into two categories. One of these is a two-stage model based on a candidate region. Xu et al. [18] proposed a novel road damage detection algorithm based on Mask R-CNN. In order to enhance the network's accuracy, researchers incorporated a Path Augmentation Feature Pyramid Network (PAFPN). At the same time, they also integrated an edge detection branch. In recent years, many scholars have conducted experimental research on the Faster R-CNN, a two-stage target detection algorithm that utilizes convolutional neural network features [19]. Hacıefendioğlu et al. [20] used the two-stage network Faster R-CNN to detect cracks in concrete pavements. The effect of different lighting and weather conditions on the detection effectiveness of the model was also investigated. Maeda et al. [21] employed a network architecture for detecting cracked images and successfully applied this network for road crack detection on smartphones. The algorithm features road damage detection on mobile devices but may suffer from poor model generalization. Another kind of regression-based is the single-stage network. Naddaf-sh et al. [22] achieved the seventh position in the IEEE Big Data Challenge 2020 by utilizing the single-stage network EfficientDet-D7 [23] for the detection and classification of asphalt pavement images. Yang et al. [24] instead used a Fully Convolutional Network (FCN) to operate pixel-level road crack detection. Wang et al. [25] targeted road breakage with slender and tiny characteristics. Based on the YOLOv3 model, the detection accuracy was improved by integrating low-level and high-level features and optimizing the loss function. However, the model's accuracy is limited to transverse or longitudinal crack detection, and it lacks universality for the diverse types of road damage encountered.

In summary, existing deep learning-based road damage detection algorithms have their characteristics and limitations in different aspects. Some algorithms have achieved good results in accuracy or adaptation to complex environments. However, problems remain, such as a lack of generalization ability.

Therefore, this article proposes an improved road damage detection algorithm for multi-scale feature. To address the issues with deep learning-based road damage processing and conventional image processing, this paper introduces vital improvements, including the Non-linear Spatial Pyramid Pool Fast (NSPPF) module, CoordConv module, and SK attention module. This paper aims to improve the model's generalization capability to diverse shapes and sizes of road damage. The model's perception of road damage is improved, and in challenging weather and complex backgrounds, detection accuracy is increased.

III. Algorithm

A. Network Architecture Design

The modified YOLOv5 is divided into four parts: input, backbone, neck, and detector. Fig. 1 illustrates the network structure. This paper redesigns the SPPF module in the original YOLOv5 network, naming it NSPPF and replacing SPPF with NSPPF. In the end, this paper constructed a combination of the CoordConv and SK modules. This combination allows for better feature selection and weighting using location information. The model's ability to represent road damage and its detection performance is further enhanced.

B. NSPPF Module

In YOLOv5, this paper proposes a non-linear spatial pyramid module, which improves the SPPF. SPPF is a commonly used module for spatial pyramid pooling but has some drawbacks and shortcomings when dealing with multi-scale features.

Fig. 2 illustrates the SPPF structure. The problem of missing information in feature representations obtained at different levels of spatial pyramidal pooling. The reason for this is the exclusive use of the maximum pooling operation. In addition, the pooling operation in SPPF is linear and does not capture non-linear feature representations. This issue limits the model's ability to identify and locate complex road damage.

To address the shortcomings of SPPF, the NSPPF module is introduced to improve SPPF, as shown in Fig. 3. The NSPPF module adds a 1x1 convolutional layer between the CBS module of the SPPF and the first maximum pooling. This additional convolution layer allows the introduction of non-linear transformations. Non-linear operations of 1x1 convolutional layers can enhance the representation of features. It enables the model to better capture the complex non-linear features of road damage. Reducing the features' dimensionality can also improve the model's efficiency. This paper continues to preserve the CBS and three maximum pooling operations of the SPPF module and adds a 3x3 convolutional layer after these operations. The convolution operation with 3x3 convolution layers allows further feature extraction and drives better integration of features over space. This step's improvement helps to capture a broader range of contextual information about road damage and enhances the model's ability to perceive road damage. In addition, the Concat operation is performed after the 3x3 convolutional layer. This operation further preserves and integrates features from the previous pooling layer to better fuse feature information at multiple scales.

C. CoordConv Module

The CoordConv [26] module is a special kind of convolution module. It introduces coordinate information to enhance the model's understanding of geometric shape and position information. In traditional convolution operations, the model does not have direct access to the absolute location information on the input feature map. This limitation may restrict the model's ability to perceive the road damage's location accurately.

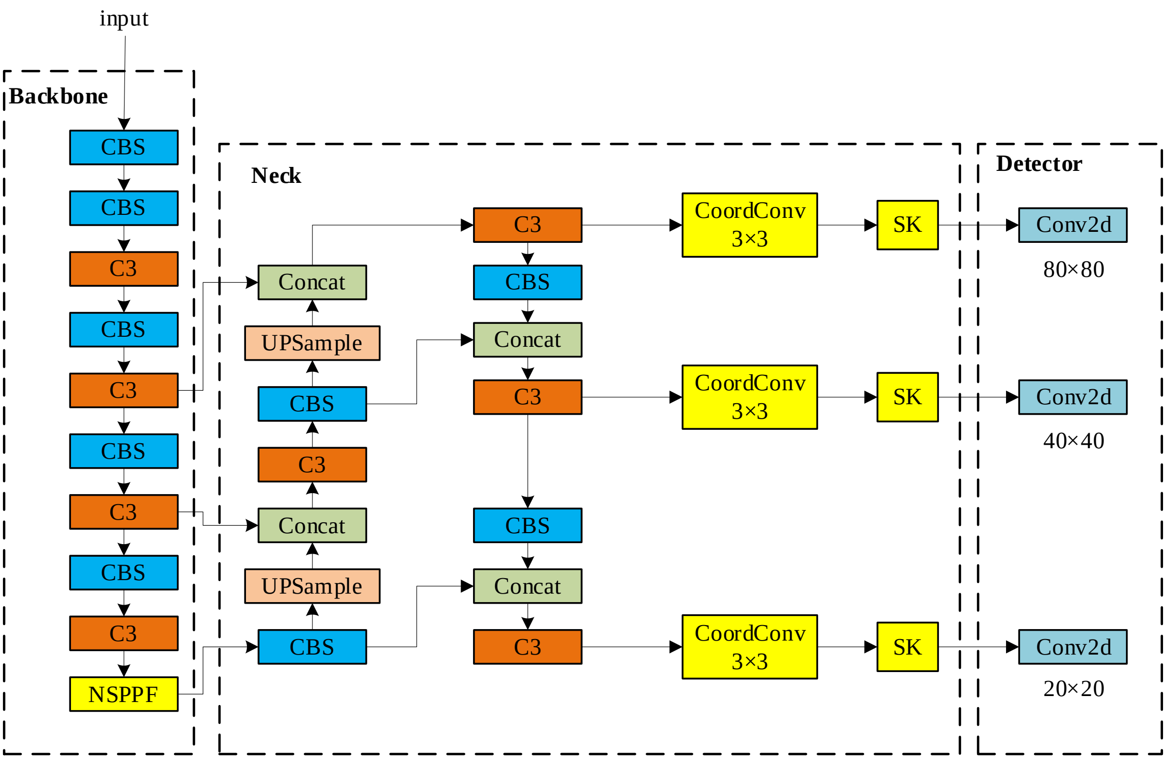


Fig. 1 Improved YOLOv5 network architecture

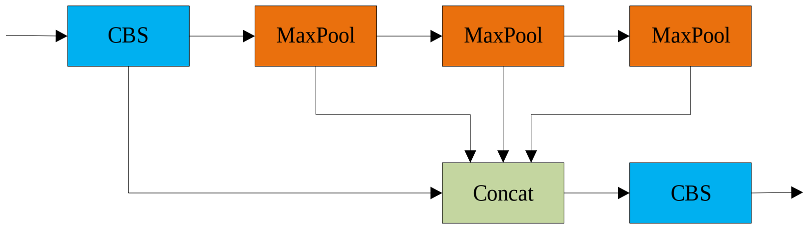


Fig. 2 Structure of the SPPF

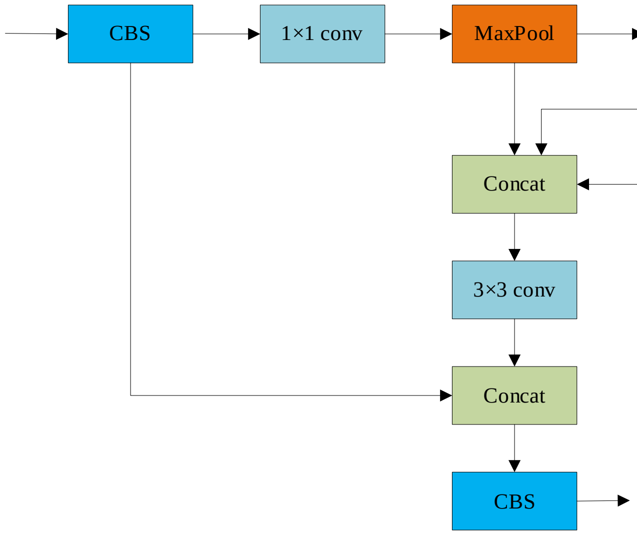
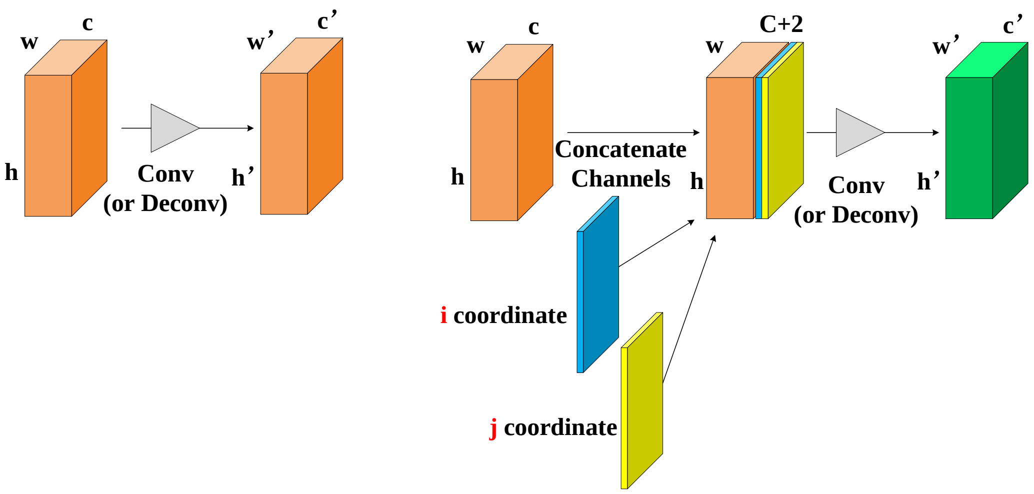


Fig. 3 Structure of the NSPPF

To address this problem, this paper introduces the 3x3 CoordConv module. This module can introduce coordinate information into the feature map as an additional channel compared to conventional convolution. It can improve model perception while enhancing the understanding of road damage of different sizes and shapes. Fig. 4 shows the structure of CoordConv. The figure shows (a) a conventional convolutional layer and (b) a CoordConv layer.

CoordConv compares with the traditional convolution module by adding two additional channels, *i* and *j*, to the input feature map. These two channels, *i* and *j* represent each pixel point's horizontal and vertical coordinates. At the same time, CoordConv retains the advantages of fewer parameters and efficient computation found in traditional convolution. Nevertheless, it allows the network to learn to keep or discard the translation invariance, as is needed for the learning task. In this way, the model can obtain information on the absolute coordinates of each position on the input feature map through



(a) Convolution Layer

(b) CoordConv Layer

Fig. 4 Structure of CoordConv

a convolution operation. During training, the model can automatically learn and adjust the coordinate information to adapt to the location distribution of different road damage. By introducing the CoordConv module, the model in this paper can better understand the geometric and positional information of road breaks. This enhancement improves the accuracy and robustness of the detection.

D. SK Attention Module

In order to enhance the model's ability to focus on road damage areas and effectively identify various sizes of road damage, further improving the CoordConv module's erformance. This paper constructs a combination of the CoordConv module and the SK [27] attention module. The SK module is a channel attention module. It adaptively adjusts the receptive field and dynamically reorganizes features. This ability can enable it to understand global information better and, thus, better adapt to different scales of road damage. Fig. 5 shows the SK module.

The SK module comprises three operators: Split, Fuse, and Select. The Split operator creates several pathways with diverse kernel sizes, aligning with the distinct Receptive Field (RF) sizes of the neurons. The Fuse operator combines and aggregates information from multiple paths to attain a global and combined representation of selection weights. The Select operator combines feature maps of various kernel sizes by utilizing selection weights for aggregation.

Split: The feature map $X \in R^{H'*W'*C'}$ is obtained after the CoordConv module. By default, two transformations are performed first $F1: X \to \hat{U} \in R^{H'*W'*C'}$ and $F2: X \to \hat{U} \in R^{H'*W'*C'}$ first, with convolution kernel sizes of 3 and 5, respectively. To further improve efficiency, a 3×3 convolution kernel is used in conversion *F2* to decrease the complexity of the model through replacing the 5×5 convolution kernel with a dilation convolution of dilation size 2.

Fuse: Integrating the $\tilde{U}$ and $\hat{U}$ information gives U as in equation (1). A feature vector S of the shape $C \times 1 \times 1$ is obtained for the integrated information by the global average pooling operation F_{gp}, as shown in equation (2). where S_c denotes the *c-th* element of S and U_c denotes the *c-th* element of the message U.

$$U = \tilde{U} + \hat{U} \tag{1}$$

$$S_c = F_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U_c(i,j) \tag{2}$$

The globally averaged pooled feature vector S is then fed into a fully connected layer (F_{fc}) for linear transformation to obtain a shape $1 \times 1 \times (C/r)$ feature Z as in equation (3). Where r is the scale factor that controls the size of dimension d of Z. As shown in equation (4), the computational workload of the model is reduced by dimensionality reduction. Meanwhile, L stands for the lowest dimension of Z.

$$\begin{gathered} Z = F_{fc}(s) = \delta(B(W_s)) \\ W \in R^{d*C} \ , \ Z \in R^{d*1} \end{gathered} \tag{3}$$

$$d = \max(\frac{C}{r}, L) \ , \quad L = 32 \tag{4}$$

where W denotes the fully connected operation on feature S at value d. B is batch normalized to optimize the fully connected layer's output and avoid gradient vanishing. The *RELU* activation function is chosen for δ better to capture the non-linear relationship of the feature S.

Select: Here, two weight matrices, a and b, are used to weight the matrices $\tilde{U}$ and $\hat{U}$, and then the final output vector V is obtained by summation.

Soft attention across channels is first generated, then information from the adaptation is used to pick distinct spatial scales. Then, we obtain from Z the attention weights a_c for $\tilde{U}$ and $\hat{U}$ as in equation (5) and b_c as in equation (6).

$$a_c = \frac{e^{A_c * z}}{e^{A_c * z} + e^{B_c * z}} \tag{5}$$

$$b_c = \frac{e^{B_c * z}}{e^{A_c * z} + e^{B_c * z}} \tag{6}$$

Where $A, B \in R^{C*d}$ $A, B \in R^{C*d}$, a, b denote the soft attention of $\tilde{U}$ and $\hat{U}$, $A_c \in R^{1*d}$ $A_c \in R^{1*d}$ is the *c-th* row of A, and a_c is the *c-th* element of a. In the case of two branches, since $ac+bc=1$, the matrix B is redundant. The final mapping V, as in equation (7), is obtained from the attention weights of the different convolution kernels. Where several branches generate several attention weights. Here, the attention weight a_c controls the $\tilde{U}$ branch, and the attention weight b_c controls the $\hat{U}$ branch.

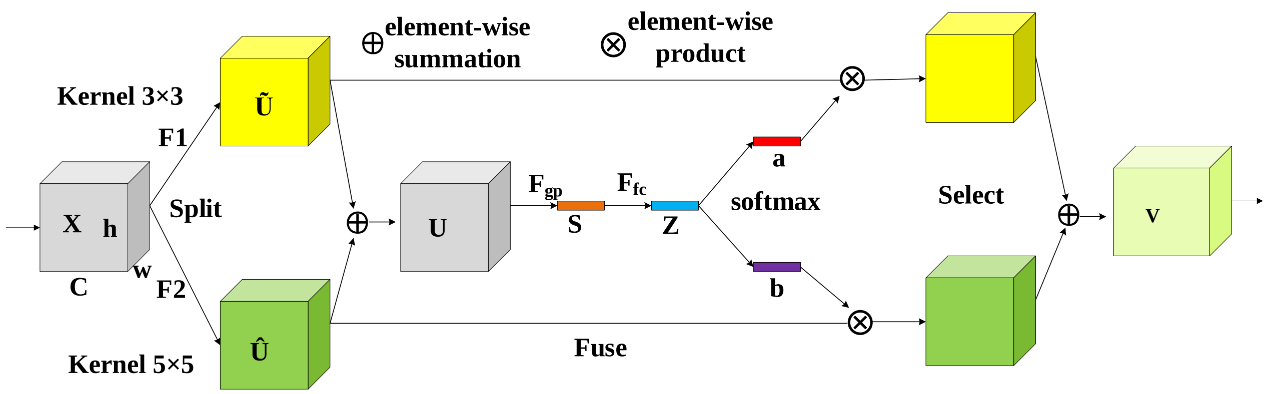


Fig. 5 Structure of the SK module

$$V_c = a_c \cdot \tilde{U} + b_c \cdot \hat{U}, a_c + b_c = 1 \tag{7}$$

Where $V = [V_1, V_2, V_3, ..., V_c,]$, $V_c \in R^{H*W}$, and this paper employs a three-branch convolution kernel of 3, 5, and 7 in the SK attention module. This paper provides a formula for the two-branch case, and the three-branch principle and formula are like the above.

IV. Experiments

The experimental setup used in this work is equipped with an Intel(R) Xeon(R) Silver 4210R CPU running at 2.40GHz and a GPU called the RTX 3090. Software configured for Windows 10 and Cuda 11.3. Deep learning framework platforms: Pytorch 1.10.0, Python 3.8. All photos were resized to 640×640 pixels to comply with the model's input specifications. The computer hardware-corresponding batch size was set to 32. The network was made more efficient with the use of an SGD optimizer. The model utilizes a migration learning approach to reduce training time and begins training by loading officially provided pre-training weights.

A. *Experimental Dataset*

The Road Damage Detection 2020 (RDD2020) dataset was used to assess the road damage detection network suggested in this paper. The dataset utilised in this paper includes 21041 photos with damage annotations, sourced from Japan, India, and the Czech Republic. The road damage information consists of the coordinates of the bounding boxes and labels that describe the type of damage associated with the bounding boxes. The training and validation sets were divided in 8:2 by random distribution of 16833 and 4208 images. Table I shows the specific types in this dataset and their definitions.

B. *Evaluation Indicators*

To evaluate the experimental results objectively, the model's performance is measured using two widely used metrics: Precision (P) and Recall (R). Precision (P) represents the probability of accurately predicting a positive sample out of all the samples predicted as positive. Recall (R) denotes the probability of correctly predicting a positive sample out of all the actual positive samples. Equations (8) and (9) provide formulas for Precision (P) and Recall (R).

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

True Positives (*TP*): The model correctly predicted the positive cases. False Positive (*FP*): The model incorrectly predicts negative cases as positive. False Negative (*FN*): The model incorrectly predicts positive cases as negative cases. True Negative (*TN*): The model correctly predicted the negative cases.

This paper uses the composite assessment criteria *F1-Score* to assess the model holistically and the Average Precision (AP) to characterize detection accuracy. The *F1-Score* aims to balance the impact of precision and recall, a reconciled average of precision and recall. A higher value of *F1-Score* indicates a higher quality model. Increased network accuracy is implied by higher AP and F1-Score values. The average accuracy across all categories is shown by Mean Average Precision (mAP). The formulas for *AP*, *F1-Score*, and *mAP* are in equations (10), (11), and (12).

$$AP = \int_0^1 P(R)dR \tag{10}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{m} AP^i \tag{11}$$

TABLE I
EXPERIMENTAL DATA

Class Name	Type Detail	Number of training Samples	Number of validation Samples
D00	Longitudinal Crack, Tire indentation	5230	1362
D01	Longitudinal Spile Crack, Construction joint	133	46
D10	Transverse Crack, Equal interval	3562	884
D11	Transverse Spile Crack, Construction joint	32	13
D20	Alligator Crack	6714	1667
D40	Rutting, bump, pothole, separation	4506	1121
D43	Crosswalk Blur	642	151
D44	Lane Line Blur	4071	986
D50	Manhole Cover	2842	739

$$F1-Score = 2 \times \frac{P \times R}{P + R} \quad (12)$$

C. *Experimental Results*

Tests were carried out on the RDD2020 dataset, processed by way of the technique described in this paper, to verify the effectiveness of the model improvements. Fig. 6 and Fig. 7 show the changes in *mAP* and *F1-Score* during the proposed and baseline model training process. These two figures clearly demonstrate that the enhanced model in this paper consistently surpasses the baseline model in accuracy and demonstrates more excellent stability. These outcomes validate the proposed model's efficacy in road damage detection. Compared to the baseline model, the model in this research performs better in road damage identification and validates the improvement's efficacy.

This paper utilizes YOLOv5 as the baseline and incorporates additional improvement modules in a stepwise manner to demonstrate their validity and necessity through ablation experiments. *AP*, *mAP*, and *F1-Score* are used as assessment metrics. From the ablation experiments, the NSPPF, CoordConv, and SK added in this paper can increase model's accuracy. This paper's model chooses better approaches according to the properties of road damage. First, the SPPF in the original YOLOv5 is replaced with NSPPF, enhancing the model's non-linear transformation capability. This modification results in a 1.1% increase in the model's *mAP* and a 1.2% advancement in the *F1-Score*. Secondly, this paper constructs a combination of CoordConv and SK attention mechanism modules. This change allows the model to better adapt to changes in location and scale in road detection tasks. Furthermore, it enhances the model's ability

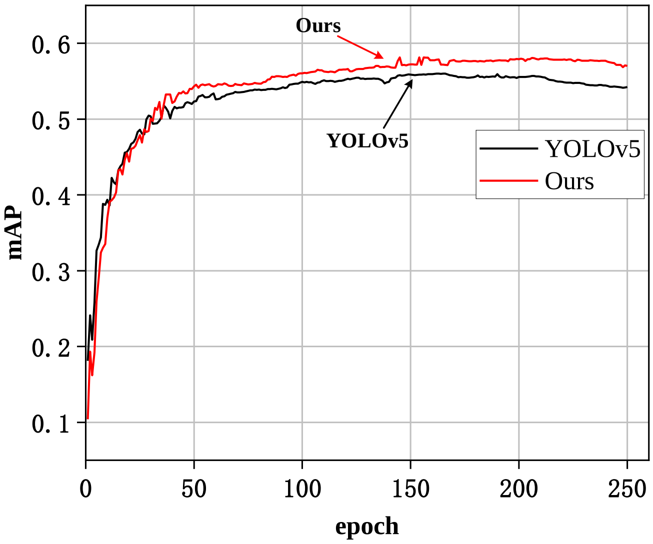


Fig. 6 Comparison chart of experimental results mAP

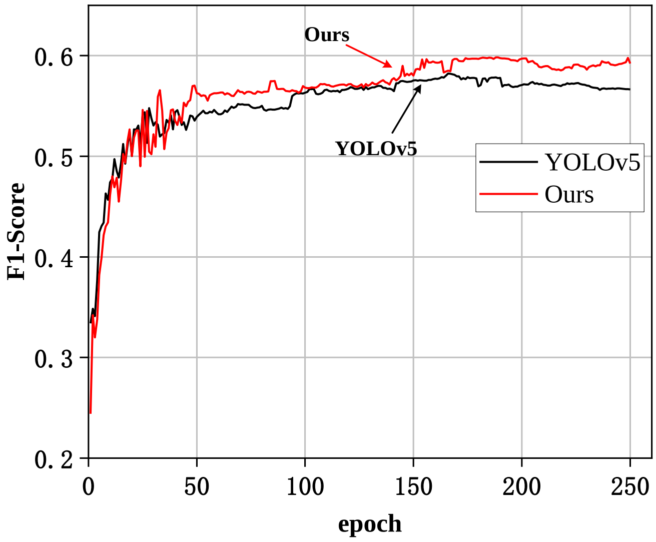


Fig. 7 Comparison chart of experimental results F1-Score

TABLE II
RESULT OF ABLATION EXPERIMENTS

Evaluating Indicator	Class Name	YOLOv5	YOLOv5+ NSPPF	YOLOv5+ CoordConv	YOLOv5+ SK	Ours
AP (%)	D00	50%	49%	46.7%	49.1%	50.4%
	D01	36%	33.1%	41.3%	36.7%	43%
	D10	42.5%	45.7%	42.6%	44.2%	44.9%
	D11	23.5%	31.9%	24.1%	30.1%	29.1%
	D20	66.3%	64.7%	64.8%	66.9%	65.3%
	D40	51.9%	52.7%	53.8%	52.8%	53.8%
	D43	75.9%	78.9%	77.9%	76.3%	78.4%
	D44	68.5%	69%	68.7%	69.7%	69.5%
	D50	89%	88.9%	88.1%	88.4%	89.2%
mAP (%)	All	56%	57.1%	56.4%	57.1%	58.2%
F1-Score(%)	All	58.2%	59.4%	58.6%	58.6%	60.4%

to capture road damage targets. Adding the CoordConv and SK attention mechanisms improved the model by 0.4% and 1.1% *mAP*, respectively, and both *F1-Score* by 0.6%. Also, in the model of this paper, each of the nine categories of the dataset RDD2020 processed by the method in the text is tested in the improved model. The experiment's results are displayed in Table II.

From Table II, the *mAP* of the improved YOLOv5 model improved by 2.2%, and the *F1-Score* improved by 2.2%. The *AP* of the dataset's nine damage categories, D00, D01, D10, D11, D20, D40, D43, D44 and D50 has improved. This result demonstrates that the optimized YOLOv5 model can better identify and detect various forms of road damage. Moreover, it can learn a more generalized feature representation.

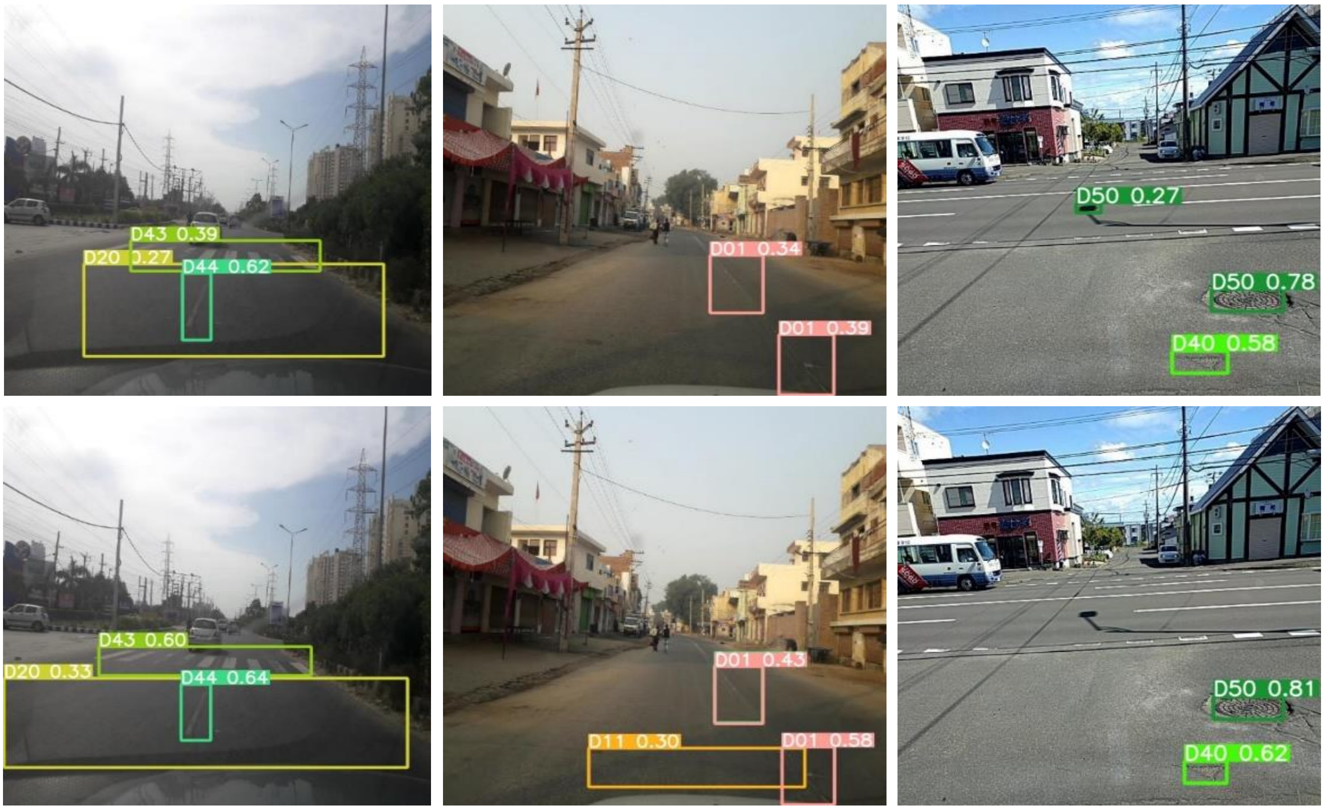
Fig. 8 shows three representative sets of images to illustrate different aspects of the model's improvement. The top image in each group demonstrates the experimental outcomes for the baseline model, and the bottom image demonstrates the experimental outcomes for the improved model. In group (a), both models identify the same category and number of road damage. However, as can be seen from the figure, the accuracy of the improved model has improved in each category. It shows that the improvements in this model have improved the model's accuracy for road damage detection. In group (b), the model in this paper has a greater variety of detections and improved accuracy compared to the experimental results of the baseline model. This paper's improved model has improved the perception of multi-scale damage, as demonstrated. In group (c), the baseline model misidentifies the shadow of a streetlamp in sunlight as a D50 class dataset. The improved model has shown advancements in this aspect, enhancing accuracy in detecting real road damage. The experimental outcomes demonstrate that the optimized model in this paper can scientifically detect road damage under complex weather conditions. The ability of the model to focus on road damage is further validated, enabling the model to capture important features of road damage areas greater accurately.

The algorithm is further compared with other target detection algorithms to validate the model's effectiveness in this paper. This includes YOLOv5, Faster R-CNN, EfficientDet, and two other literatures. The above algorithms are trained and validated on the RDD2020 dataset processed by the method in the paper. Two metrics, *mAP* and *F1-Score*, are selected to evaluate the algorithms. Table III shows the experimental results of different algorithms for road damage detection on the RDD2020 dataset.

TABLE III
PERFORMANCE COMPARISON OF VARIOUS ALGORITHMS

Model	mAP (%)	F1-Score (%)
YOLOv5	56%	58.2%
Faster R-CNN	51.2%	51.4%
EfficientDet	56.9%	57.2%
Ref. [28]	57%	58.6%
Ref. [29]	57.6%	58.7%
Ours	**58.2%**	**60.4%**

According to the experiments, the algorithm suggested in this paper achieves 58.2% and 60.4% for *mAP* and *F1-Score*, respectively, in the road damage detection task. Compared with the baseline model YOLOv5, the improvements are 2.2% and 2.2%, respectively. Compared to Faster R-CNN, the improvements are 7% and 9%, respectively. Compared with EfficientDet, the improvements are 1.3% and 3.2%



(a) (b) (c)

Fig. 8 Comparison chart of experimental real-world effects respectively. Compared to Ref. [28], the improvements are 1.2% and 1.8% respectively. Compared to Ref. [29], the improvements are 0.6% and 1.7% respectively. These results demonstrate the effectiveness of the model improvement in this paper.

V. Conclusion

This paper proposes three improvement points in the road damage detection task based on multi-scale feature extraction. Firstly, introducing the Non-linear Spatial Pyramid Pooling-Fast (NSPPF) module. Better capturing the detailed features of road damage areas improves the perception of multi-scale damage. Secondly, by building a combination of the CoordConv and SK attention modules. The ability to fuse coordinate information and learn correlations between global and local features enhances the model's capacity to identify damage at various sizes. These improvements enable our algorithm to achieve an accuracy of 58.2% on the RDD2020 dataset, These improvements enable our algorithm to achieve an accuracy of 58.2% on the RDD2020 dataset, our algorithm is able to outperform the baseline model by 2.2%.

These improvements are of great importance for road maintenance and traffic safety. Accurate detection of damage on roads helps to repair and maintain them promptly, improving traffic safety, reducing accidents, and improving the driving experience for drivers. Future research could further explore and optimize road damage detection algorithms to improve accuracy and robustness. At the same time, combining other advanced deep learning techniques and computer vision algorithms can expand the application area of road damage detection. For example, this technology enables real-time monitoring and prediction of road damage. Various benefits can be achieved by integrating road damage detection into autonomous driving technology. These further studies and applications will advance the field of road maintenance and traffic safety. They will lead to a safer and more efficient road damage detection network for society.

References

[1] D. Myungsik, and H. Jung, "Enhancing road network resilience by considering the performance loss and asset value," *Sustainability*, vol. 10, no. 11, pp. 4188, 2018.

[2] V. Matthew, and M. Moussa, "Deep learning for intelligent transportation systems: A survey of emerging trends," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3152-3168, 2019.

[3] L. Han, and S. Seo, "Lightweight deep learning for road environment recognition," *Applied Sciences*, vol. 12, no. 6, pp. 3168, 2022.

[4] G. Guo, and Z. Zhang, "Road damage detection algorithm for improved YOLOv5," *Scientific Reports*, vol. 12, no. 1, pp. 15523, 2022.

[5] N. Tran, H. Nguyen, H. Luong, M. Nguyen, K. Luong, and H. Tran, "Recognition of Student Behavior through Actions in the Classroom," *IAENG International Journal of Computer Science*, vol. 50, no. 3, pp1031-1041, 2023.

[6] P. Zhang, S. Xu, and W. Wang, "Belt Deviation Detection System Based on Deep Learning under Complex Working Conditions," *IAENG International Journal of Applied Mathematics*, vol. 53, no. 3, pp863-868, 2023.

[7] A. Akagic, E. Buza, S. Omanovic, and A. Karabegovic, "Pavement crack detection using Otsu thresholding for image segmentation," *Proceedings of the 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics*, pp. 1092-1097, 2018.

[8] Y. Sari, P. B. Prakoso, and A. R. Baskara, "Road crack detection using support vector machine (SVM) and OTSU algorithm," *Proceedings of the 2019 6th International Conference on Electric Vehicular Technology*, pp. 349-354, 2019.

[9] H. Zhao, G. Qin, and X. Wang, "Improvement of canny algorithm based on pavement edge detection," *Proceedings of the 2010 3rd International Congress on Image and Signal*, vol. 2, pp. 964-967, 2010.

[10] N. Attoh-Okine, and A. Ayenu-Prah, "Evaluating pavement cracks with bidimensional empirical mode decomposition," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1-7, 2008.

[11] Y. Maode, B. Shaobo, X. Kun, and H. Yuyao, "Pavement crack detection and analysis for high-grade highway," *Proceedings of the 2007 8th International Conference on Electronic Measurement and Instruments*, IEEE, pp. 548-552, 2007.

[12] N. D. Hoang, "An artificial intelligence method for asphalt pavement pothole detection using least squares support vector machine and neural network with steerable filter-based feature extraction," *Advances in Civil Engineering*, vol. 2018, 2018.

[13] M. Gao, X. Wang, S. Zhu, and P. Guan, "Detection and segmentation of cement concrete pavement pothole based on image processing technology," *Mathematical Problems in Engineering*, vol. 2020, pp. 1-13, 2020.

[14] N. Aloysius, and M. Geetha, "A review on deep convolutional neural networks," *Proceedings of the 2017 International Conference on Communication and Signal*, IEEE, pp. 0588-0592, 2017.

[15] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3156-3164, 2017.

[16] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212-3232, 2019.

[17] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, "Understanding convolution for semantic segmentation," *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision*, pp. 1451-1460, 2018.

[18] Y. Xu, D. Li, Q. Xie, Q. Wu, and J. Wang, "Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN," *Measurement*, vol. 178, pp. 109316, 2021.

[19] Y. Liu, and Y. Tian, "DCMS-YOLOv5: A Dual-Channel and Multi-Scale Vertical Expansion Helmet Detection Model Based on YOLOv5," *Engineering Letters*, vol. 31, no. 1, pp373-379, 2023.

[20] K. Hacıefendioğlu, and H. B. Başağa, "Concrete road crack detection using deep learning-based faster R-CNN method," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, pp. 1-13, 2022.

[21] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127-1141, 2018.

[22] S. Naddaf-Sh, M. M. Naddaf-Sh, A. R. Kashani, and H. Zargarzadeh, "An efficient and scalable deep learning approach for road damage detection," *Proceedings of the 2020 IEEE International Conference on Big Data*, pp. 5602-5608, 2020.

[23] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10781-10790, 2020.

[24] X. Yang, H. Li, Y. Yu, X. Luo, T. Huang, and X. Yang, "Automatic pixel - level crack detection and measurement using fully convolutional network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1090-1109, 2018.

[25] Q. Wang, J. Mao, X. Zhai, J. Gui, W. Shen, and Y. Liu, "Improvements of YoloV3 for road damage detection," *Proceedings of the Journal of Physics: Conference Series*, vol. 1903, no. 1, pp. 012008, 2021.

[26] R. Liu, J. Lehman, P. Molino, F. Petroski Such, E. Frank, A. Sergeev, and J. Yosinski, "An intriguing failing of convolutional neural networks and the CoordConv solution," *Advances in Neural Information Processing Systems*, pp. 31, 2018.

[27] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 510-519, 2019.

[28] H. Liang, S. C. Lee, and S. Seo, "Automatic recognition of road damage based on lightweight attentional convolutional neural network," *Sensors*, vol. 22, no. 24, pp. 9599, 2022.

[29] F. Wan, C. Sun, H. He, G. Lei, L. Xu, and T. Xiao, "YOLO-LRDD: a lightweight method for road damage detection based on improved YOLOv5s," *EURASIP Journal on Advances in Signal Processing*, vol. 2022, no. 1, pp. 98, 2022.