ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul856

Identification and Detection of Road Damages Based on Improved YOLOv5

Moslema Chowdhuray Momi[1]; Lin Bai[2]; Muhammad Arslan Ghaffar[3]

School of Electronics and Control Engineering[1;2]; School of Automobile[3]
Chang'an University Xi'an 710064, China

Publication Date: 2026/07/27

Abstract: Road damage significantly affects transportation safety, vehicle condition, and road maintenance efficiency. Traditional inspection methods are often costly and time-consuming, leading researchers to explore image processing and deep learning techniques for automated road damage detection. This work aims to investigate the application of an object detection approach for damage identification and detection on road surfaces. This work proposed an improved YOLOv5-based model for accurate road damage identification. The proposed approach enhances YOLOv5s by introducing three key improvements: (1) a P2 detection head to improve the detection of small and distant road damages, (2) the CBAM attention mechanism to reduce background interference and highlight important damage features, and (3) a BiFPN-inspired feature fusion structure to strengthen multi-scale feature integration and improve information flow across network layers. Experimental results on the GRDD2020 dataset demonstrate the effectiveness of the proposed YOLOv5s-P2-CBAM-BiFPN model. Compared with the original YOLOv5s, the improved model achieves higher detection performance, increasing Precision by 3.9%, Recall by 7.7%, mAP@0.50 by 8.4%, and mAP@0.50:0.95 by 7.3%. These results confirm that the proposed method provides a reliable and accurate solution for automated road damage detection.

Keywords: *Road Damage Detection; Deep Learning; YOLOv5s; Multi-Scale Object Detection; CBAM Attention Mechanism; BiFPN.*

How to Cite: Moslema Chowdhuray Momi; Lin Bai; Muhammad Arslan Ghaffar (2026) Identification and Detection of Road Damages Based on Improved YOLOv5. *International Journal of Innovative Science and Research Technology*, 11(7), 1212-1221. https://doi.org/10.38124/ijisrt/26jul856

I. INTRODUCTION

Road damage, such as cracks, potholes, and surface distress, poses significant challenges to transportation safety and infrastructure maintenance. Traditional manual inspection methods are often slow, labor-intensive, and prone to errors, making timely maintenance difficult. In recent years, with the development of deep convolutional neural networks and deep learning algorithm, object detection have become the mainstream approach in computer vision.

➢ *Research Challenges*

The low resolution of road damage images and various factors such as noise, complex backgrounds, and other sources of interference, which pose challenges to accurate detection. To address these issues, the research focuses on the following aspects to enhance the road damage detection capabilities of YOLOv5s:

- Addressing the low-resolution and insufficient information of road damages: two solutions are proposed. Firstly, employing a larger feature map for road damage detection can mitigate the issue of low resolution. By using a larger feature map, the local receptive field is reduced, enabling the network to detect more damages even with lower resolution. Secondly, incorporating multi-scale feature fusion aims to enrich the network with detailed information. Because the deep neural network lost a lot of characteristic information after several down sampling operations, the problem of insufficient information of road damage was aggravated. Shallow feature maps contain more fine-grained information such as color, texture, and edges, so in shallow feature maps, the detailed information of road damage is more abundant. To improve the useful information of minor road damage, we can thus think about refining the YOLOv5 feature fusion structure to fully combine more shallow features with deep features.
- Mitigating the interference from image noise and background: A method to increase attention mechanism is proposed. By leveraging attention mechanisms to amplify key features and suppress noise interference, the model's focus on road damage detection is enhanced, enabling better identification of occluded damage.

➢ *Research Contribution*

This work presents a novel deep learning-based method for detecting road damage, with a particular focus on the YOLOv5 specially, the lightweight YOLOv5s model. We enhanced the models road damage detection capability by refining its network structure and validating the improved

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul856

model on the GRDD2020 dataset. Our main contributions to the field of study are:

- To mitigate the issue of missing road damage detection, we introduced a P2 detection head to enable the network to detect smaller damages. The YOLOv5s-p2 model integrates a 4-fold sub-sampling feature map with rich location information, facilitating the transfer of shallow features to deep features. This enhancement improves the network's ability for multi-scale target detection.
- To suppress background interference, noise, and other factors, and enhance the attention to road damages, we incorporated several CBAM modules into the neck structure of YOLOv5s. The neck network can now efficiently handle a high volume of feature data by prioritizing important damage features and lessening the influence of irrelevant information with to the addition of the CBAM attention mechanism. This enhancement boosts the feature information related to road damages and enhances detection accuracy.
- To improve feature fusion and enhance the network's feature processing capability, we drew inspiration from the bi-directional cross-scale connection concept of the BiFPN network to optimize the feature fusion structure of YOLOv5s.

In the enhanced network, shallow and deep features are fused bi-directionally and across scales. This fusion method strengthens the relationship between feature maps, enriching the detected feature map with more information, especially low-level geometric details. It aids the network in focusing more on the road damage location area, thus enhancing the road damage recall rate.

II. BACKGROUND AND LITERATURE CONTEXT

The 1980's marked the beginning of the development and deployment of a system that detects road damage [1]. In the past, early methods for road damage detection relied on image recognition techniques like local binary patterns and filters [2]. However, these approaches faced challenges in distinguishing features within images, accurately classifying cracked and non-cracked images, and were considerably impacted by outside variables like road roughness and sunlight. Scholars have increasingly used deep learning approaches for road damage identification as a result of their rise in domains such as image recognition and classification. Road damage detection techniques that use deep learning can be broadly divided into three categories: classification annotation[3], semantic segmentation[4], and object detection.

The first type of image classification annotation is to label images for different damage categories or levels, dividing overlapping areas into two categories (with or without damages) and multiple categories (with different damage levels). The main research includes: A Cubero Fernandez, Fco J. After Rodriguez Lozano [5] captures the image, several processes are applied to extract the main features to emphasize cracks. Zhang et al. [6] used convolutional neural networks (CNNs) to reduce road surface images and classified crack and crack free road images by outputting probabilities. Li B et al. used DCNN [7] technology to classify cracks in 3D road surface images. Road cracks were labeled into 5 categories, and they used four original CNN models to classify the images. These CNNs have different receptive fields, and different models have almost the same overall accuracy. A method for crack identification and classification based on deep convolutional neural networks was proposed by N A M Yusof et al. [8]. This method successfully detected cracks in images with recall, accuracy, and precision rates of 98%, 99%, and 99%, respectively. They use two different filter sizes (3×3 and 5×5) Train DCNN. However, compared to larger filter sizes, smaller filter sizes require more processing and training time. Overall, this method successfully achieved a classification accuracy of 94.5%.

The second type of semantic segmentation is to annotate road damaged using pixel level annotation, and then use semantic segmentation forms such as U-Net to classify images at the pixel level. Its advantage is that it can more accurately express the shape of the damages. The defect is that the labeling cost is too high. For example, Chunde Yang, Mingyue Geng pointed out that road crack images are affected by a large amount of complex noise, and the main information of crack edges is easily lost. Therefore, he proposed a road crack detection algorithm based on edge information. Experiments have shown that this method can accurately detect road cracks and preserve edge information.[9]

To make road crack texture learning easier, Fan et al. [10] developed a supervised deep-learning neural network model that uses crack data for image analysis. Likewise, road damage identification was achieved by Jenkins et al. using the U-net [11] semantic segmentation method. To classify fractures in road images, Zou et al.[12] proposed an end-to-end deep-learning network approach. To identify several distinct fractures, this model makes use of several expansion modules that may perform convolutional expansion at different rates.

Maeda et al. pioneered the development of a software program for road damage detection in Japan. In 2018, they compiled the first road damage detection dataset using a cellphone camera. They employed the SSD-MobileNet algorithm, known for its speed, which can be executed on a smartphone for detection purposes. Their research achieved up to 77% accuracy in classifying 8 classes of road damage. An advantage of using images instead of vibrations for classifying road damage is the ability to detect damage to the paint of traffic signs on the road, such as crosswalks. However, a drawback is that shadows can lead to false classifications, and rainy conditions may impair the phone's visibility of the road.[13]

Many researchers have started using faster models, such as YOLO, Faster R-CNN, and other faster neural network models, for road damage identification to increase detection speed. Yanbo et al. employed the SSD and Faster-RCNN to identify the damages to the roads. Researchers were able to improve the accuracy of their detection model by using the

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul856

Faster-RCNN, adjusting hyper-parameters like the anchor box ratio, and executing several data augmentation strategies (such as varying brightness and Gaussian blur).[14]

The YOLOv3[15] was used along with the DarkNet53 by Alfarrarjeh et al.[16] to detect the road damages. These authors also tried to improve the models detection performance by replenishing the data for classes with inadequate images. Also, Kluger et al.[17] employed the Faster-RCNN, the RetinaNet[18], and the random forest algorithm to deal with the data scarcity problem.

Luo Hui et al.[19] proposed an improved multi-scale road damage detection algorithm based on YOLOv4, which uses CSPDarknet-53 backbone network instead of ordinary convolutional, optimizes the loss function in YOLOv4 using Focal loss, and expands the dataset using data augmentation methods such as cropping, flipping, adjusting brightness, and noise disturbance. Zhang Ning suggested using data augmentation techniques like transfer learning in addition to the Faster R-CNN (Faster Region Based Convolutional Neural Networks) model for identifying road deterioration.[20]

By improving the YOLOv5s technique, Wan et al.[21] presented a lightweight road damage recognition model. Their primary focus is on enhancing the detection accuracy of the lightweight ShuffleNetV2 model by adding an ECA attention module. Additionally, they aim to extract more rich feature information by utilizing the BiFPN structure rather than the feature pyramid structure. Additionally, create a higher quality anchor frame in order to rectify the sample imbalance. They employed Focal-EIOU, which offers better performance in terms of efficiency and accuracy, as a positioning loss.

The impact of different backbone models, including CSPDarknet53, Hourglass-104, and EfficientNet, on classification performance was examined by Vishal Mandal et al. Simultaneously training deep network models using over 20000 images obtained on different roads in Japan, Czech Republic, and India. Finally, the performance of damage classification for each model was evaluated.[22]

III. PROPOSED METHODOLOGY

The YOLOv5s model was selected due to its lightweight nature and advantageous accuracy for detecting road damages. Additionally, YOLO incorporates an automatic anchor box selection process, which enables it to learn and utilize the most suitable anchor box for a given dataset without manual intervention. These anchor boxes serve as predefined boxes that best fit the objects of interest. Notably, YOLOv5 demonstrates superior accuracy compared to YOLOv4 and YOLOv3.[23]

Due to the frequent exposure of roads to different environmental conditions, such as repeated traffic loads, temperature fluctuations, and changes in moisture, it is easy to lead to road defects. Cracks and potholes are the most serious defects in road. Road damage is classified according to its size, shape, and depth. This study mainly studies 7 types of road damages, namely alligator cracking, longitudinal cracking, linear cracking, zig-zag cracking, potholes, reveling, and upheaval shown in figure 1.

Fig 1 Road Damage Types with Samples for Detection

- *CSP Module*

To address gradient repeated calculation issues during network optimization and reduce computational load during inference, CSP module introduced based on the concept of CSPNet[24]. This module enhances network learning capacity, ensuring accuracy while minimizing calculations and memory consumption. YOLOv5 introduces the CSP module into the neck network. YOLOv5 introduces two CSP modules, figure 2 illustrates the architecture, with CSP1_X serving as the backbone network and CSP2_X serving as the neck network. Depending on the channel, CSP1_X divides the feature map into two sections. The first section is subjected to standard convolutional processes, while the second section includes residual components that are modeled after residual networks. These parts are then combined to yield a new feature map, avoiding gradient value repetition and improving

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul856

model inference speed. Furthermore, the residual component CSP structure improves gradient values during backpropagation, reducing gradient vanishing problems and enhancing the network's ability to extract features. In contrast, CSP2_X splits the input feature map into two halves, which are then fused using convolutional layers to preserve more picture information.

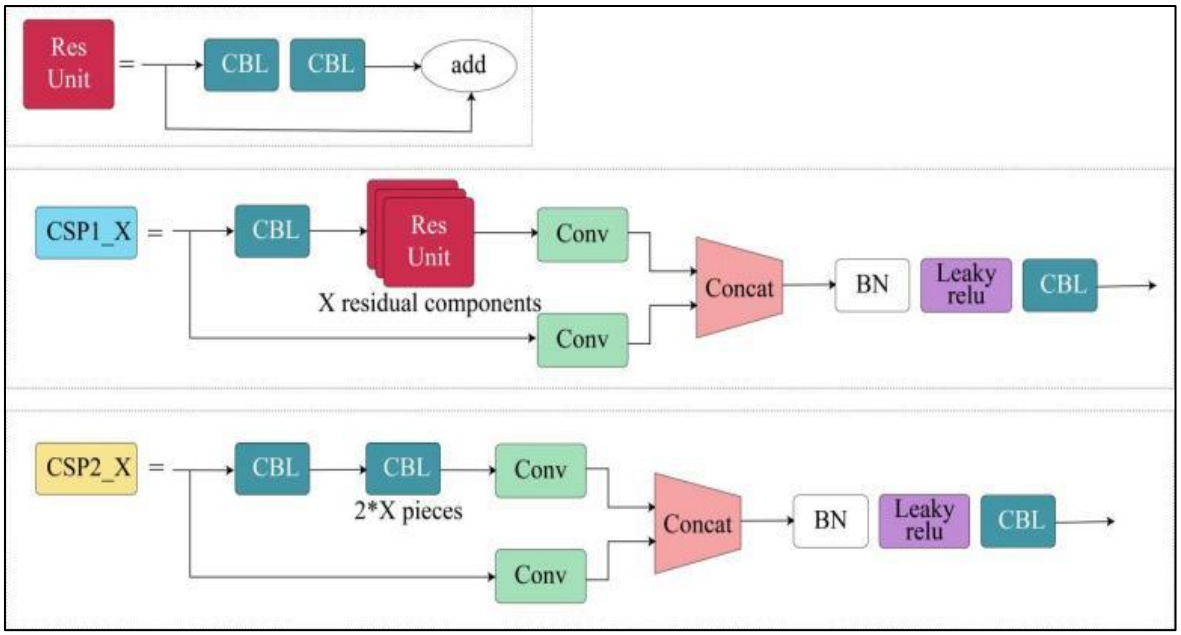


Fig 2 CSP Module Structure

➢ *CBAM Network Structure*

The neck structure is the most important component of feature processing in the YOLOv5 network. The neck structure acts as a link between the previous and the subsequent since it is situated between the output layer and the backbone network. The features that were taken out of the backbone network are fully aggregated and sent to the output layer for prediction. Thus, the key to influencing the algorithms success is the neck network's processing capacity for the features. To better utilize processing power and retrieve more relevant feature data for the damage identification task, we added the CBAM attention mechanism to the YOLOv5s neck network. By adding CBAM, we can learn more effective features and improve network performance without increasing network depth, width and input image resolution.

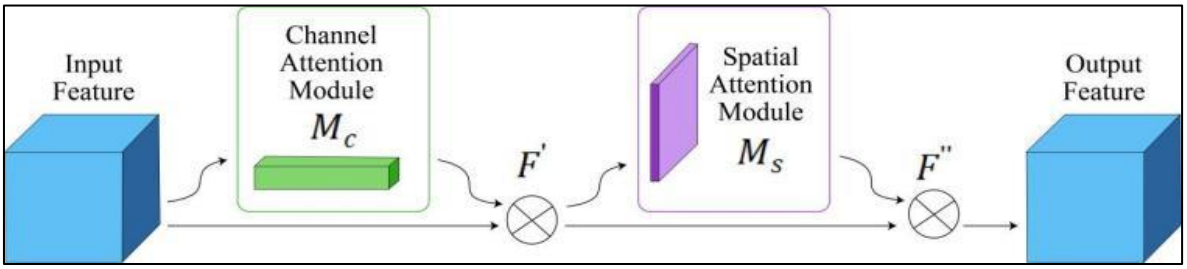


Fig 3 CBAM Network Structure

➢ *Model Improvement Based on BiFPN*

The improved model is recorded as YOLOv5s-p2-CBAM-BiFPN, which has multi-layer feature fusion, and the fusion method is shown in figure 4. The deep features are always up sampled and fused with the shallow features. After the up sampling is completed, the shallow features are always down sampled and fused with the deep features. Furthermore, to integrate the features gathered from the backbone network into the feature map that needs to be identified, the network extends the original FPN+PAN structure by two horizontal link paths. With only minimal computation, this updated method can enhance the expression capacity of the feature map, provide more feature information to the feature map to be detected, and boost the detector's efficiency.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul856

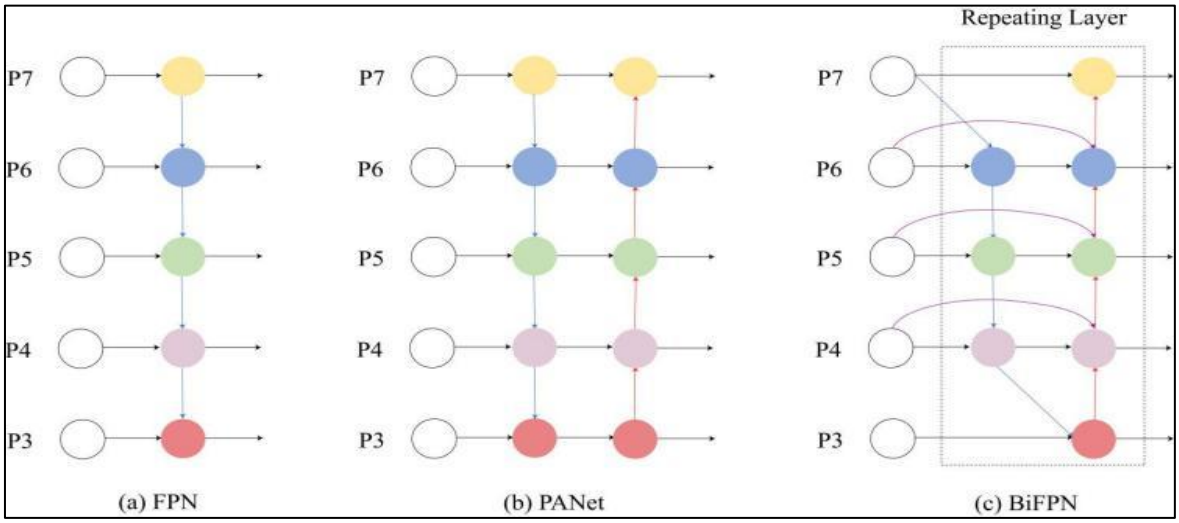


Fig 4 Variant of Multi-Scale Feature Fusion Network

➢ *Overall Model Architecture*

In order to improve small object detection problem, a 4-fold down sampling detector head is added and named YOLOv5s-p2. The minimum object resolution that YOLOv5s-p2 can detect is 4×4, which is more powerful than the original network. The YOLOv5s-p2 model's feature fusion configuration is shown in figure 3. The feature maps, that were retrieved from the backbone network and subsampled by 4, 8, 16, and 32 fold, respectively, are represented by the letters C2, C3, C4, and C5, in this image. The feature fusion layers are denoted by F3 and F4, and the detection layers by P2, P3, P4, and P5. The network can detect four different scale feature maps with the inclusion of the P2 detection layer. P3 detection layer becomes F3 feature fusion layer through the upgraded network. Upsampling the feature map and fusing the four times subsampled feature map that the backbone network produced come next, following F3. Following this, the fused feature map is used to identify road damages, and the P2 detection layer finally outputs the detection findings. Furthermore, a feature output is added and linked to the neck end of the backbone network following the first C3 module. At the same scale, this feature fusion takes place. Predictions are made using the P2 detection layer after the C3 module and convolutional procedures. More shallow feature information can be transferred to the deep feature map during the downsampling process because the P2 detector head includes both shallow and high-resolution feature maps. As a result, the P2 detection head improves the algorithms road damage detection performance by making the network more sensitive to minute defects.

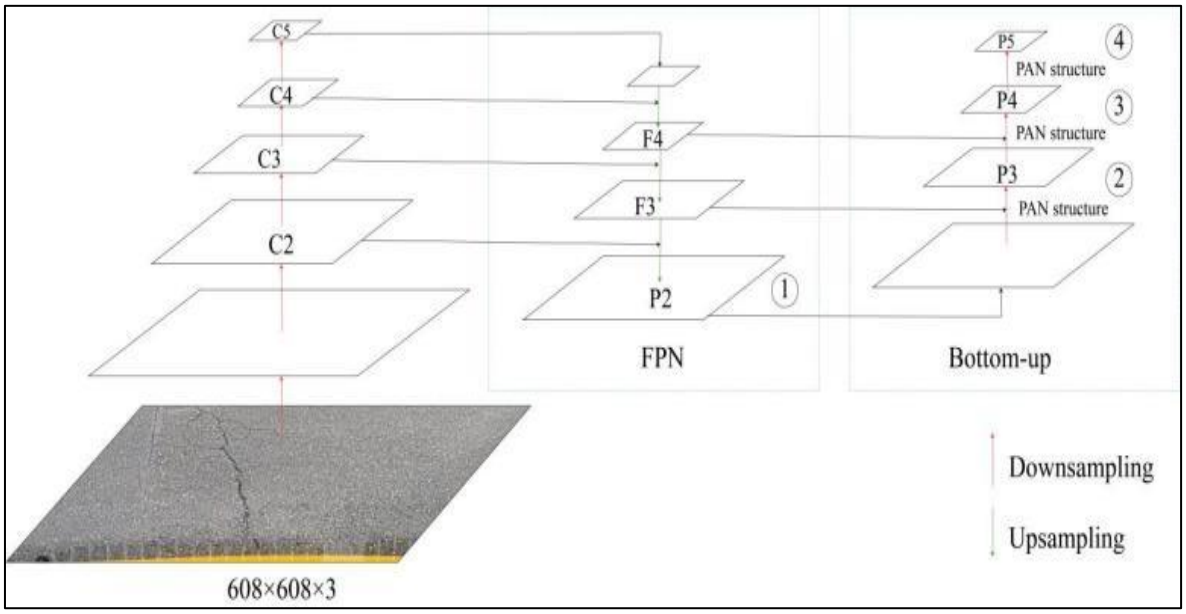


Fig 5 YOLOv5s-p2 Feature Fusion Structure

The deep feature map is subjected to sequential upsampling within the FPN structure in order to integrate it with shallow feature maps of a greater scale. In order to minimize network computations, a C3 operation is started after every fusion. The CBAM attention mechanism is used in the PAN structure to update the feature map after the previous fusion, improving the network's focus on important features and incorporating more valuable shallow geometric

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul856

information into the deep features. The feature map is then downscale and combined with the low-resolution feature map. The CBAM module improves feature fusion at the neck end and strengthens the model's analytical capacity for complicated scenarios. This allows the model to concentrate more on pertinent areas by automatically filtering out background interference and other distractions. As a result, making better use of the limited computational resources makes it possible to obtain more useful data and improves damage detection accuracy.

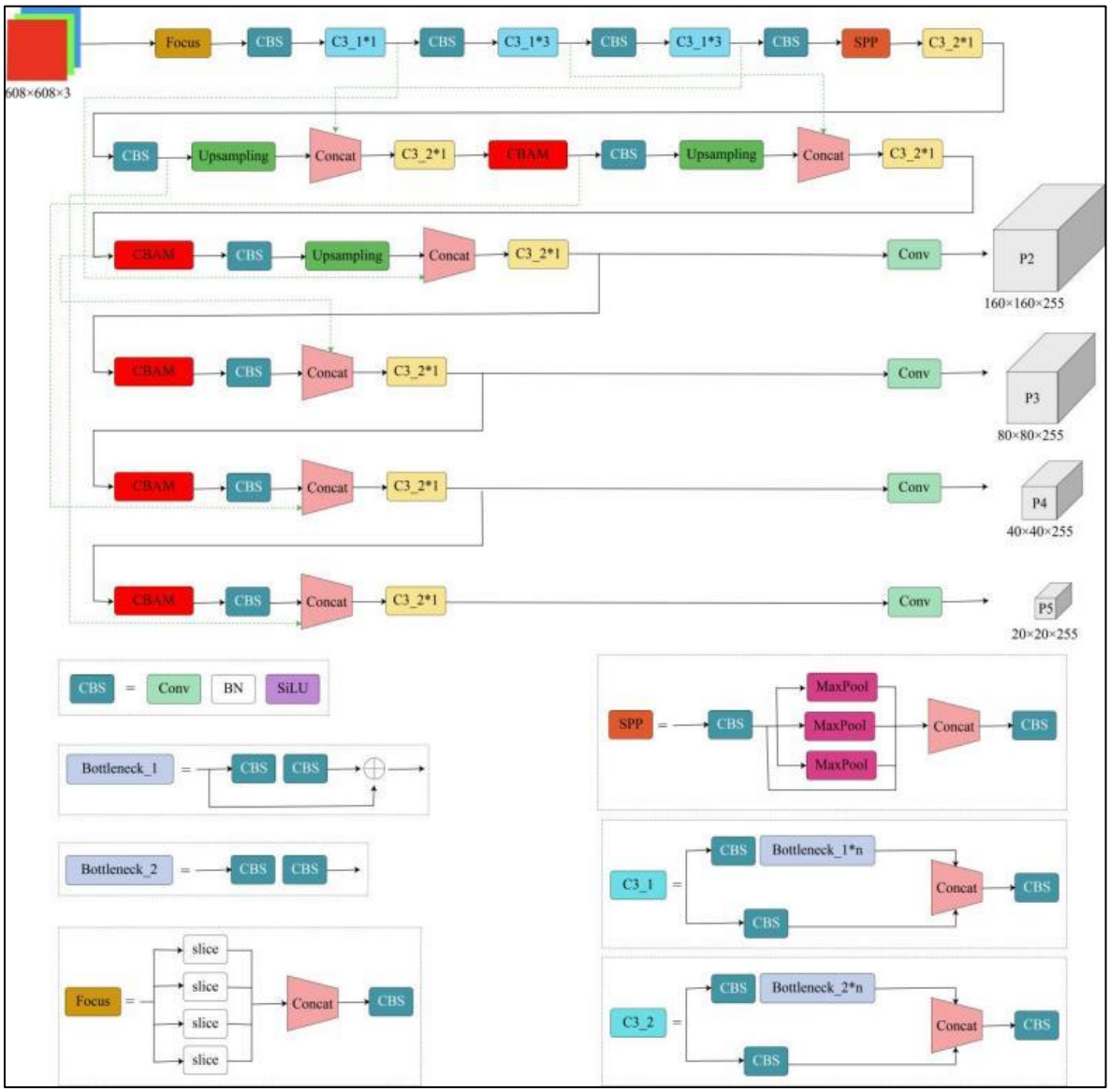


Fig 6 Overall Model Architecture

The overall network architecture is shown in Figure 6. The 8-fold and 16-fold down-sampling characteristics are represented by the second and third C3 modules of the backbone network in this image, respectively. In the figure 6, the red dotted line represents the improvement approach. The second C3 module's feature map is initially attached to the neck end and fused with the feature map at the P3 detection layer's location on the same scale. Similarly, the feature map of the P4 detection layer and the feature map produced by the third C3 module are fused. A feature fusion path is integrated into the detection layer of the YOLOv5s-p2-CBAM-BiFPN model. In essence, for the two P3 and P4 scales, the features that need to be identified are combined from three sources: features from various network tiers on the same scale, such as the feature map from the upper network layer's lower sampling, the feature map from the neck end's upper sampling fusion, and the feature map that was taken from the backbone network.The YOLOv5s feature fusion structure is optimized, drawing inspiration from the BiFPN network. By increasing the connectivity of information between feature maps with varying scales, this optimization seeks to lessen the negative consequences of feature information loss caused by upsampling and downsampling. As a result, the feature fusion performance of the original YOLOv5 is much enhanced. It is important to emphasize that this study specifically concentrates on enhancing the efficacy of road damage detection with YOLOv5, and validates the refined model using road damage dataset. The results show that weighting the features involved in the fusion will reduce the performance of the algorithm.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul856

IV. EXPERIMENTAL DETAILS

➢ *Experimental Datasets and Evaluation Metrics*

The experiments in this article utilize on the GRDD2020[98] (Global Road Damage Detection) dataset in order to evaluate the upgraded model based on YOLOv5s detection capabilities for road damage. The dataset contains of 26,336 road pictures from sources such as India, Japan, and the Czech Republic. Remarkably, 121 teams from various nations took part in this championship. The answers that were submitted were assessed using two different dataset, test1 and test2, which consisted of 2,631 and 2,664 pictures, respectively. 1,610 labeled images are in the dataset; 816 images are in the test set and 794 images are in the training set.

The evaluation metrics used in the experiments include intersection over union (IOU), F1-score, Error rate, Recall, Accuracy, and Precision. Additionally, Mean average precision (mAP) were analyzed to evaluate network complexity.

➢ *Experimental Details*

The effectiveness of the expriment is verified through experiments conducted on an NVIDIA T4 TENSOR CORE GPU with Intel Core i7-11800h operating system. The code is implemented via Python. Our proposed method employs stochastic gradient descent (SGD) for network optimization and adopts a cosine linear learning strategy, utilizing a batch size of four images and training for 300 epochs. The initial learning rate is set to 0.01, whereas the weight decay and momentum are set to 0.0005 and 0.937, respectively.

➢ *Ablation Experiment*

In six sets of ablation tests, we evaluated the performance gains attained by combining the three upgraded techniques of the P2 detector head, CBAM module, and BiFPN network. Table 1 provides an overview of the outcomes on the GRDD2020 test set. The initial three rows indicate that when either the CBAM module or BiFPN is individually added to the YOLOv5s model, the mAP@.5 improved by approximately 2%, suggesting a modest enhancement in network performance. Introducing the P2 detection head to YOLOv5s led to a slight reduction in accuracy but a notable increase in recall rate by 2.7% and mAP@.5 by 1.6%, effectively addressing object missed detection. The latter three rows demonstrate that incorporating both the CBAM module and BiFPN network onto YOLOv5s-p2 resulted in respective improvements in model accuracy and recall rate. Compared to YOLOv5s-p2, YOLOv5s-p2-CBAM-BiFPN exhibited a notable increase in mAP@.5 by 5.8%, highlighting the positive impact of the CBAM module and BiFPN network on model performance. Overall, the most significant enhancement was observed when integrating the P2 detector head, CBAM module, and BiFPN into YOLOv5s. Compared to the original YOLOv5s, YOLOv5s-p2-CBAM-BiFPN demonstrated a remarkable 7.7% improvement in recall rate while maintaining accuracy, with mAP@.5 increasing by 8.4% and mAP@.5:.95 increasing by 7.3%.

➢ *Comparative Experiment*

Figure 7 shows the P-R curves for the YOLOv5s and enhanced YOLOv5s models. Better model performance is implied by the fact that the mAP value rise as the region between the P-R curve and the coordinate axis grows. For the first class “alligator cracking”, the original YOLOv5s has an accuracy of 53.0%, whereas the improved YOLOv5s model has improved its accuracy up to 99.5%. Similarly, the other classes such as zig-zag cracking, linear cracking, upheaval and pothole with YOLOV5s, has accuracies of 99.5%, 54.5%, 51.1% and 82.0%, respectively. The improved YOLOv5s model has improved the accuracies for all classes, the accuracy of zig-zag cracking, linear cracking, upheaval and pothole has improved to 99.5%, 94.0%, 99.5% and 99.5%, respectively. The P-R curve and mAP for each item in the original and improved YOLOv5s models, respectively, are shown in figures 7(a) and (b), which use the GRDD2020 dataset.

The experimental findings demonstrate that the original YOLOv5s model has less optimal detection performance in object detection when compared to the YOLOv5s enhanced model, which reduces the overall mAP. On average, the YOLOv5s-p2-CBAM-BiFPN model performs well in detecting each type of object, even though it does not have the greatest AP value for each class. Overall, the YOLOv5s-p2-CBAM-BiFPN model exhibits the best performance.

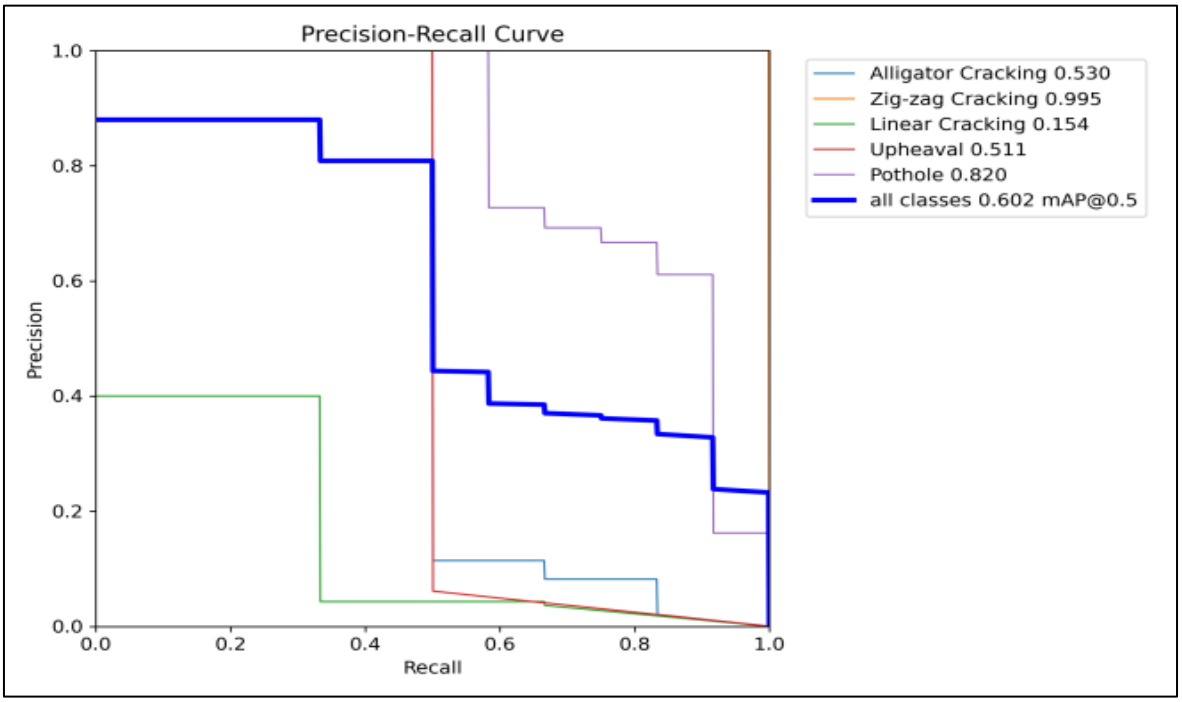
➢ *Experimental Result*

The proposed model in this article can effectively detect the all these appearing categories and can also detect twice appearing classes successfully, signifying that YOLOv5s-p2-CBAM-BiFPN has enhanced the overall detection accuracy. There are total five categories appearing in these three frames. Some detection results of the YOLOv5s-p2-CBAM-BiFPN model on the GRDD2020 dataset is shown in figures 8.

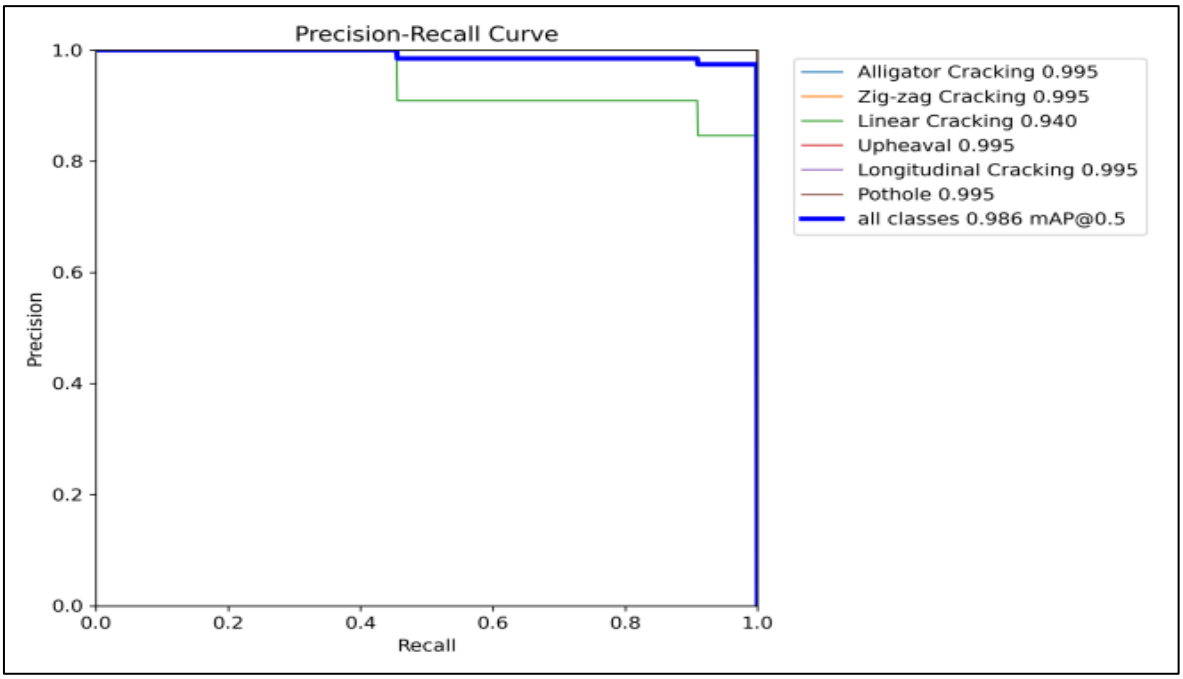
Table 1 Ablation Results of GRDD2020 Test Set

Methods				**P(%)**	**R(%)**	**mAP@.50**	**mAP@.50:.95**
YOLOv5s	P2	CBAM	BiFPN				
✓				87	90.8	90.2	71.5
✓		✓		87.3	92.1	91.5	72.1
✓			✓	87.6	91.7	91.3	72.3
✓	✓			86.8	93.5	92.8	72.6
✓	✓	✓		88.1	93.9	94.4	74.9
✓	✓	✓	✓	91.9	98.5	98.6	78.8

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul856



(a) Original YOLOv5



(b) Enhanced YOLOv5

Fig 7 Comparison Between the Optimized and Original Model

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul856

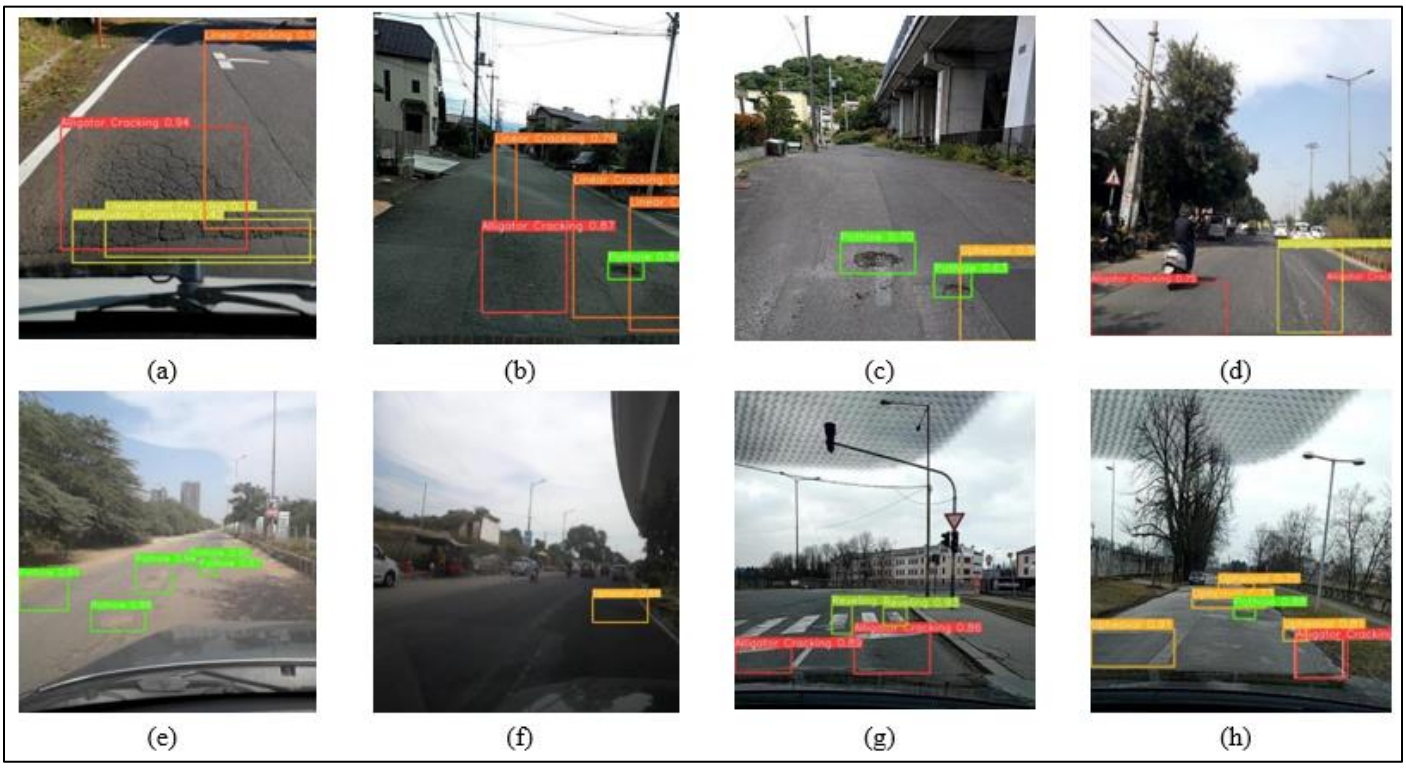


Fig 8 Visualization Results of Example of GRDD2020 Dataset

V. CONCLUSION

In this letter, we proposed YOLOv5s-p2-CBAM-BiFPN. The proposed enhancements to YOLOv5s, incorporating the CBAM attention mechanism and the BiFPN-based feature fusion architecture, effectively improve the model's capability for road damage detection. By enhancing feature representation and multi-scale feature fusion, the proposed method mitigates missed detections and false positives while improving localization accuracy. Experimental evaluation on the GRDD2020 dataset demonstrates consistent performance gains over the baseline YOLOv5s, validating the effectiveness of the proposed architecture for accurate and robust road damage detection in complex real-world scenarios.

Future research may focus on developing more robust detection frameworks through larger and more diverse datasets, advanced network architectures, and enhanced feature extraction techniques. In addition, real-time road damage detection from video streams and the integration of detection models with Geographic Information Systems (GIS) for accurate damage localization and condition assessment represent promising directions for practical intelligent road maintenance systems.

REFERENCES

[1]. Caroff, G.; Joubert, P.; Prudhomme, F.; Soussain, G. Classification of pavement distresses by image processing (MACADAM SYSTEM). In Proceedings of the 1st International Conference on Applications of Advanced Technologies in Transportation Engineering ASCE, San Diego, CA, USA, February 1989; pp. 46–51.

[2]. Xu Jian. Hand drawn sketch depth retrieval network based on Gabor filters [J]. Information Technology and Informatization, 2022 (07): 122-125.

[3]. Han Haihang, Zhang Li, Wu Haotian, Chen Haizhu, Hu Di. Application of Image Data Processing Mechanisms in the Training of Asphalt Road Crack damage Identification Algorithms [J]. Science and Technology Innovation and Application, 2021 (09): 29-34.

[4]. Zhang Zhihua, Wen Yanan, Mu Haowei, Du Xiaoping. Road crack detection using dual attention mechanism [J]. Chinese Journal of Image and Graphics, 2022, 27 (07): 2240-2250G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955.

[5]. Cubero Fernandez A, Rodriguez Lozano F, Villatoro R, et al. Efficient pavement crack detection and classification [J] EURASIP Journal on Image and Video Processing, 2017, 2017 (1): 1-11.

[6]. Zhang L, Yang F, Zhang Y D, et al. Road crack detection using deep convolutional neural network [C] 2016 IEEE international conference on image processing (ICIP) IEEE, 2016: 3708-3712.

[7]. Li B, Wang K C P, Zhang A, et al. Automatic classification of payment crack using deep convolutional neural network [J] International Journal of Payment Engineering, 2020, 21 (4): 457-463.

[8]. Yusof N A M, Ibrahim A, Noor M H M, et al. Deep convolutional neural network for crack detection on asphalt pavement [C] Journal of Physics: Conference Series IOP Publishing, 2019, 1349 (1): 012020.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul856

[9]. Yang C, Geng M. The crack detection algorithm of pavement image based on edge information [C] AIP Conference Proceedings AIP Publishing LLC, 2018, 1967 (1): 040023.
[10]. Fan Z, Wu Y, Lu J, et al. Automatic payment crack detection based on structured prediction with the convolutional neural network [J] ArXiv print arXiv: 1802.02208, 2018.
[11]. Jenkins M D, Carr T A, Iglesias M I, et al. A deep convolutional neural network for semantic pixel wise segmentation of road and pavement surface cracks [C] 2018 26th European Signal Processing Conference (EUSIPCO) IEEE, 2018: 2120-2124.
[12]. Zou Q, Zhang Z, Li Q, et al. Deepcrack: Learning hierarchical convolutional features for crack detection [J] IEEE Transactions on Image Processing, 2018, 28 (3): 1498-1512.
[13]. 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, 33(12):1127-1141, 2018.
[14]. Wang, Y.J., Ding, M., Kan, S., Zhang, S., Lu, C.: Deep proposal and detection networks for road damage detection and classification. In: IEEE International Conference on Big Data (Big Data), pp. 5224–5227. IEEE, Piscataway, NJ (2018).
[15]. Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
[16]. Alfarrarjeh, A., Trivedi, D., Kim, S.H., Shahabi, C.: A deep learning approach for road damage detection from smartphone images. In: IEEE International Conference on Big Data (Big Data), pp. 5201–5204. IEEE, Piscataway, NJ (2018).
[17]. Kluger, F., Reinders, C., Raetz, K., Schelske, P., Wandt, B., Ackermann, H., Rosenhahn, B.: Region-based cycle-consistent data augmentation for object detection. In: IEEE International Conference on Big Data (Big Data), pp. 5205–5211. IEEE, Piscataway, NJ (2018).
[18]. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988. IEEE, Piscataway, NJ (2017).
[19]. Luo Hui, Jia Chen, Li Jian Highway pavement damage detection algorithm based on improved YOLOv4 [J] Progress in Laser and Optoelectronics, 2021, 58 (14): 141-175.
[20]. Zhang Ning Research on Highway Pavement damage Detection Algorithm Based on Faster R-CNN [D] Nanchang: East China Jiaotong University, 2019.
[21]. Wan, F., Sun, C., He, H., et al. (2022). YOLO-LRDD: a lightweight method for road damage detection based on improved YOLOv5s. EURASIP Journal on Advances in Signal Processing, 2022(1), 98.
[22]. Mandal V, Mussah A R, Adu Gamfi Y. Deep learning frameworks for payment stress classification: A comparative analysis [C] 2020 IEEE International Conference on Big Data (Big Data) IEEE, 2020: 5577-5583.
[23]. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding Yolo Series in 2021, Computer Vision and Pattern Recognition. arXiv 2021, arXiv:2107.08430.
[24]. Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.