Proceedings of CONF-MLA 2024 Workshop: Neural Computing and Applications
DOI: 10.54254/2755-2721/105/2024TJ0060

Exploring the Effects of Attention Mechanisms on Road Crack Detection and Classification

Xingye Peng

School of Mechanical, Electrical and Information, Shandong University, Weihai, China

202000800359@mail.sdu.edu.cn

Abstract. Pavement cracks are a prevalent form of roadway distress that pose significant safety hazards, necessitating prompt detection and repair. Due to the extensive road network, traditional image processing-based crack detection methods exhibit limitations in recognition accuracy and speed. To address these challenges, this paper proposes an efficient pavement crack detection model based on YOLOv8, termed YOLOv8-ACD (YOLOv8 - Attention for Cracks Detecting), which integrates a global attention mechanism. YOLOv8-ACD enhances detection efficiency and accuracy by focusing on crack identification while filtering out most irrelevant information. We evaluated YOLOv8-ACD on the RDD2022 dataset, and experimental results demonstrate significant improvements in key performance metrics, such as F1-Score and mean Average Precision (mAP), compared to the original YOLOv8 and other mainstream models. The real-time processing capability of this model makes it suitable for practical road inspection and maintenance, effectively reducing the workload of maintenance personnel and enhancing roadway safety.

Keywords: Yolo, deep neural networks, road crack detection.

1. Introduction

Road crack detection and classification are critical tasks for maintaining and improving the safety, longevity, and efficiency of transportation infrastructure. Cracks, if not identified and addressed early, can lead to severe deterioration of road surface, resulting in higher maintenance costs and, more importantly, safety risks for vehicles and pedestrians. Recently, numerous methods have been developed to automate the detection and classification of road cracks. In traditional ways, Hansang et al. [1] used wavelet analysis as a method to detect structural damage by identifying changes in structural behavior and analyzing signal patterns. This method performs well in detecting some normal crack., but wavelet transform is sensitive to noise, and the structure is more complex, so it has a limited range of application. Artis et al. [2] evaluated several pothole detection algorithms based on accelerometer data, focusing on simple, resource-efficient methods like thresholding and standard deviation to identify potholes on urban roads. This method is simple to detect which can run-on small-scale platforms, but obviously, the accuracy is not enough to competent some complex environment.

In recent years, object detection based on deep learning has advanced swiftly. Qin et al.[3] used DeepCrack to automatically detect by learning the feature of crack based on fusing hierarchical convolutional stages. This method can represent not only more details in larger scale but also richer information in small scale, but when facing the multi-element environment, it may focus on most of the element which will retard the detecting and training process. Cao et al.[4] used attention-based network “ACNet” to improve a better crack detection network. This method can get a well performance on crack detecting and some relevant parameter is better than the traditional detection network.

The challenge of detecting cracks in real-world scenarios, where the environment may introduce noise, varying illumination, and complex backgrounds, remains. Object detection frameworks like SSD (Single Shot MultiBox Detector), and YOLO have been applied to address these issues, with YOLO standing out due to its real-time performance and high accuracy. To tackle the aforementioned issues, we suggest an integration with a Global Attention Mechanism (GAM). into the YOLOv8 framework to enhance its performance for road crack detection in complex environments. This work seeks to improve the resilience and accuracy of the detection model by minimizing information loss and maximizing global interactions. The experimental results show that the proposed integrated method significantly outperforms the simple model for detection in complex scenarios, which provides new ideas and technical support to enhance automated road crack detection.

2. Related works

The attention mechanism draws inspiration from the human visual system, utilizing selective focus to prioritize significant information. In deep learning, attention allocates varying weights to highlight significant components of the input, helping the system efficiently highlight relevant data. J et al.[5] produce a SE (Squeeze-and-excitation) to target channel relationships interdependencies by explicitly modeling and adaptively recalibrating channel responses. S et al.[6] create an attention module CBAM can inference on both channel and spatial dimensions of feature maps, which is lightweight and versatile.

For crack detection, M et al.[7] relies on the Gabor filter's ability to extract directional and frequency-specific features, which helps in distinguishing cracks from background noise or other surface irregularities. Gabor filters are used to extract texture and edge features that are characteristic of pavement cracks. The filter’s orientation and frequency sensitivity make it particularly adept at detecting cracks of different sizes and orientations. Q et al.[8] create a structure based on trees as CrackTree to automatically detect the cracks. This method focuses on enhancing the automation and efficiency of crack detection, reducing the need for manual inspection. Key components of the method include preprocessing techniques, feature extraction, and crack identification algorithms. By utilizing morphological operations and edge detection algorithms, the method identifies cracks as continuous or connected segments, improving the identification of cracks with varying shapes and sizes.

3. Proposed methodologies

In this paper, the traditional crack detection network is improved based on yolov8 model with added attention mechanism and named YOLOv8-ACD (YOLOv8-Attention for Cracks Detecting). YOLOv8-ACD enhances the efficacy of the YOLOv8 model for road crack identification with the incorporation of attention mechanisms, specifically the Global Attention Mechanism (GAM). Figure 1 illustrates how GAM is applied to the backbone network to enhance global context awareness.

DOI: 10.54254/2755-2721/105/2024TJ0060

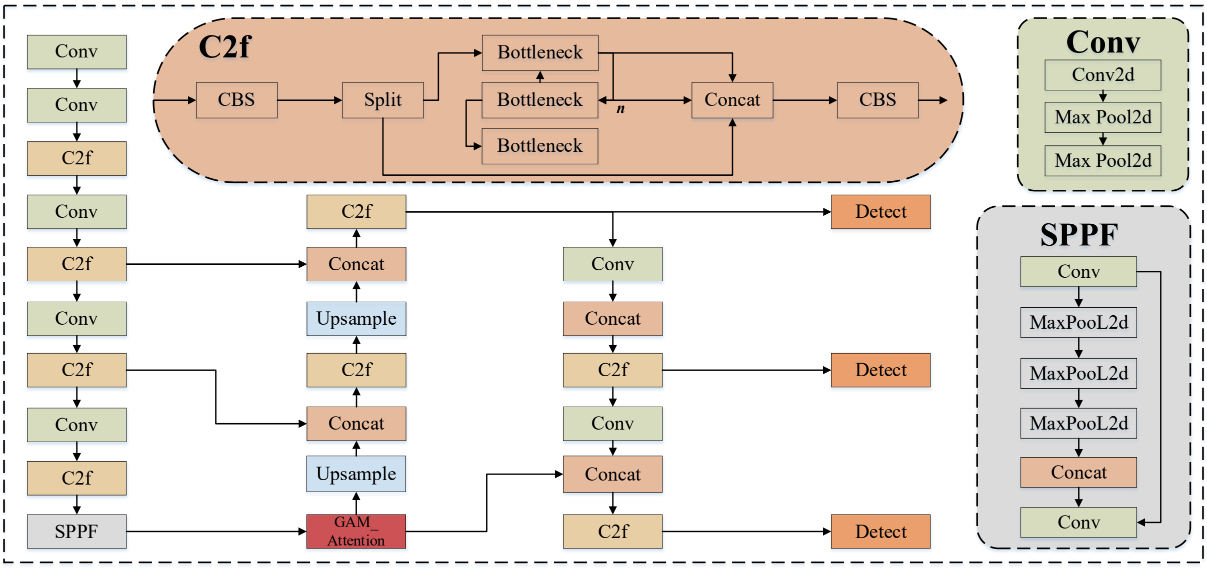

Figure 1. YOLOv8-ACD structure (Photo/Picture credit : Original).

3.1. Global attention mechanism (GAM)

Road cracks are often thin and elongated, blending into their surroundings, making it difficult for conventional convolutional layers to differentiate them accurately. In the context of YOLOv8-ACD, The integration of GAM into the network enhances the model's capacity to grasp global context, which is crucial for distinguishing cracks from the background. Global interaction features are amplified by the GAM mechanism while information dispersion is minimized. The authors incorporate a sequential channel-spatial attention mechanism into their design of the CBAM submodule [9]. The whole process is shown in Figure 2 and in equations (1) and (2).

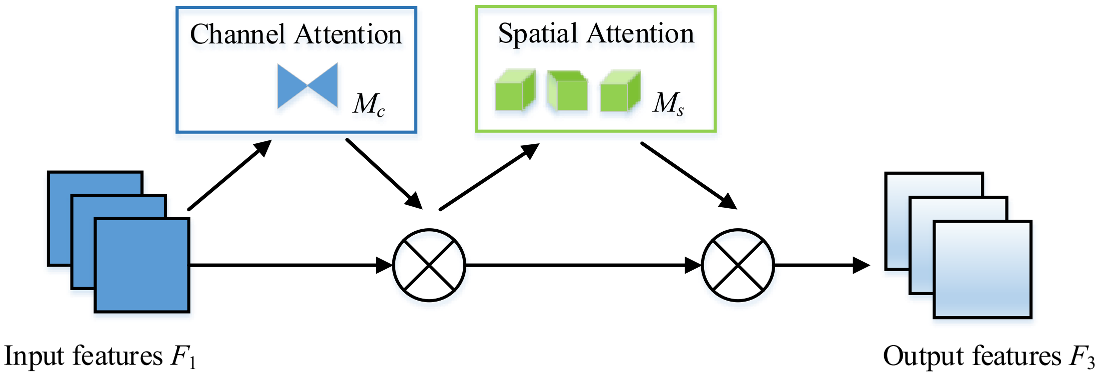

Figure 2. GAM mechanism structure (Photo/Picture credit : Original).

Based on the input feature map $F1 \in R^{C \times H \times W}$, the intermediate states and outputs are defined as:

$$F_2 = M_C(F_1) \otimes F_1 \tag{1}$$

$$F_3 = M_S(F_2) \otimes F_2 \tag{2}$$

The channel attention map and spatial attention map are represented by M_C and M_S , while the element-wise multiplication operation is represented by $\otimes$.

DOI: 10.54254/2755-2721/105/2024TJ0060

By integrating global information, GAM enhances the model's precision in crack detection by reducing false positives. CNNs, while effective at identifying localized patterns, struggle to capture broader contextual cues needed for detecting small, irregular objects like cracks. GAM addresses this limitation by offering a more comprehensive view, enabling the model to better distinguish cracks from background textures. It focuses attention on key areas likely to contain cracks, filtering out irrelevant regions. This improves detection accuracy and makes the model more robust in complex environments with varied lighting and road textures.

4. Road damage detection

A set of experiments was conducted to verify the effectiveness of the recommended YOLOv8-ACD model in detecting road cracks. The purpose of these experiments was to assess the performance of YOLOv8-ACD compared to the YOLOv8 model's baseline and other advanced crack detection techniques. A comprehensive assessment of both accuracy and efficiency was achieved by analyzing the evaluation metrics of precision, recall, F1-score, and inference time.

4.1. Dataset

RDD2022: The Multi-National Road Damage Dataset 2022 [10]: there are more than 40K road images in the dataset that come from six countries: Japan, India, the Czech Republic, Norway, the United States, and China (Table 1). The dataset captures five types of road damage, which include longitudinal crack, transverse crack, alligator crack, pothole, and repair. Due to the size of the dataset, the China_Drone sub-dataset is the sole training method for the YOLOv8-ACD model. The China_Drone sub-dataset contains eight classes; Meanwhile, the format of this dataset is Pascal VOC, which conflicts with the general format of YOLO, so we need to convert the PASCAL VOC format to a format supported by YOLO.

Table 1. Details in China_Drone sub-dataset.

Types of road damage	Class name	Amount
Longitudinal Crack	**D00**	1426
Transverse Crack	**D10**	1263
Alligator Crack	**D20**	293
Pothole	**D40**	86
Repair	**Repair**	769

4.2. Implementation details

The model can be trained using the dataset with an initial learning rate of 0.01, a batch size of 64 and we stopped training after 300 epochs. The NVIDIA 3090 GPU is used for all experiments with PyTorch.

4.3. Evaluation criteria

We utilized different performance indicators, including F1-Score and mAP50/mAP50-95, to conduct a quantitative analysis of the experimental results.

In statistics, the F1 score is a metric that evaluates the accuracy of binary classification models by analyzing both the precision and recall of the model. The detail is shown in equation (3).

$$F1\text{-}Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{3}$$

To assess object detection algorithms, Mean Average Precision (mAP) is a crucial measure. It's calculated by aggregating the Average Precision (AP) across all detected categories. The detail is shown

DOI: 10.54254/2755-2721/105/2024TJ0060

in Equation 5. The Precision-Recall (PR) curve is used to average precision values at different recall levels to determine the AP for each category, which is shown in equation (4).

$$AP = \int_0^1 P(r)dr = \sum_{k=1}^{N} P(k)\Delta r(k) \tag{4}$$

$$mAP = \frac{\sum_{i=1}^{c} AP_i}{C} \tag{5}$$

The average precision of the model when the IoU threshold is 0.5 is measured by the mAP50, which is the mean average precision at intersection over union threshold 0.5. IoU is a metric that gauges the degree of overlap between the anticipated bounding box and the actual bounding box, mAP50-95 calculates the average accuracy of the model in the range from IoU 0.5 to 0.95.

4.4. *Qualitative and quantitative evaluation*

A quantitative analysis experiment was carried out to confirm the reliability of the YOLOv8-ACD improved in this paper on a test set that contained 2401 road crack images. Table 2 displays the results, and the optimal results are highlighted in bold.

Table 2. Comparison of F1-Score in different networks.

Metrics	Algorithms				
	CNN	UNet	YOLOv5s	YOLOv8n	YOLOv8-ACD
Recall	0.957	**0.975**	0.914	0.950	0.961
Precision	0.833	0.779	0.924	0.941	**0.959**
F1-Score	0.858	0.813	0.919	0.945	**0.960**

As can be seen from the table 2, the Recall, Precision and F1-Score of YOLOv8-ACD are 0.961, 0.959, 0.96 and respectively, which are 1.1%, 1.8%, 1.5% higher than YOLOv8n; 0.4%, 12.6%, 10.2% higher than CNN. Although YOLOv8-ACD did not reach the highest rate of Recall, the gap between the best rate in Recall, UNet, and YOLOv8-ACD is acceptable, because YOLOv8-ACD can maintain high rate of Recall while maintaining high rate of Precision, which means it can reach the highest rate of F1-Score.

Table 3. Comparison of mAP50/mAP50-95 in different networks.

Metrics	Algorithms		
	YOLOv5s	YOLOv8n	YOLOv8-ACD
mAP50	0.965	0.979	**0.985**
mAP50-95	0.678	0.783	**0.802**

The table 3 shows that among three different YOLO algorithms. Notably, YOLOv8-ACD demonstrates superior efficacy, achieving the highest mAP50 value of 0.985, outperforming both YOLOv5s (0.965) and YOLOv8n (0.979), which means YOLOv8-ACD maintain the higher accuracy and completeness. Therefore, from the overall experimental results, due to integrating the GAM-Attention, By reducing information loss and amplifying global interaction representations, we created an optimized model that is better than other comparison algorithms.

DOI: 10.54254/2755-2721/105/2024TJ0060

5. Conclusion

To enhance the efficiency of pavement crack identification and classification, this paper introduces a YOLOv8 network model augmented with an embedded GAM, termed YOLOv8-ACD. This innovative model leverages the YOLOv8n architecture, which is known for its high accuracy and speed, and integrates the attention mechanism to further improve performance. The Global Attention Mechanism allows the model to prioritize relevant features associated with pavement cracks while effectively minimizing the influence of extraneous background information. To evaluate the recognition and classification efficacy of YOLOv8-ACD, the model was trained on a dataset consisting of 3,837 pavement crack images, which encompasses a variety of crack types, sizes, and lighting conditions, ensuring a comprehensive training process. The model was then tested on a separate validation set comprising 2,401 images that were not included in the training phase. This approach ensures a rigorous assessment of the model's generalization capabilities. Experimental results indicate that YOLOv8-ACD achieves superior evaluation metrics compared to baseline algorithms, with significant improvements in both F1-Score and mAP. These results reflect the model's capacity to accurately identify and classify pavement crack images under diverse background conditions, highlighting its practical applicability in real-world scenarios. The high computational efficiency of YOLOv8-ACD makes it suitable for real-time applications in road maintenance, thereby providing a valuable tool for enhancing roadway safety and reducing maintenance costs.

References

[1] Kim, H., & Melhem, H. (2004). Damage detection of structures by wavelet analysis. *Engineering Structures,* 26(3), 347-362.
[2] Mednis, A., Strazdins, G., Zviedris, R., et al. (2011). Real time pothole detection using android smartphones with accelerometers. *In 2011 International Conference on Distributed Computing in Sensor Systems and Workshops* (DCOSS) (pp. 1-6). .
[3] Zou, Q., Zhang, Z., Li, Q., et al. (2018). Deepcrack: Learning hierarchical convolutional features for crack detection. *IEEE Transactions on Image Processing*, 28(3), 1498-1512.
[4] Cao, J., Yang, G., & Yang, X. (2020). Pavement crack detection with deep learning based on attention mechanism. *Journal of Computer-Aided Design & Computer Graphics*, 32(8), 1324-1333.
[5] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7132-7141).
[6] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. *In Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3-19).
[7] Salman, M., Mathavan, S., Kamal, K., et al. (2013). Pavement crack detection using the Gabor filter. *In 16th International IEEE Conference on Intelligent Transportation Systems* (ITSC 2013) (pp. 2039-2044). .
[8] Zou, Q., Cao, Y., Li, Q., et al. (2012). CrackTree: Automatic crack detection from pavement images. *Pattern Recognition Letters*, 33(3), 227-238.
[9] Liu, Y., Shao, Z., & Hoffmann, N. (2021). Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv preprint arXiv:2112.05561.*
[10] Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., & Sekimoto, Y. (2022). RDD2022: A multi-national image dataset for automatic Road Damage Detection. *arXiv preprint arXiv:2209.08538*