

# YOLO-based Approach for Pavement Damage Detection

Xue Lin<sup>1</sup>, Huanxin Zhou<sup>1</sup>, He Zhang<sup>2</sup>, Mingjian Liu<sup>3</sup>, and Jiaqi Li<sup>1,\*</sup>

<sup>1</sup> School of Civil Engineering, University of Science and Technology Liaoning, Anshan 114051, China

<sup>2</sup> Anshan Iron & Steel Group Refractory Materials Co., Anshan 114000, China

<sup>3</sup> Angang Steel Group Mining limited, Anshan, China

\*lijiaqi@ustl.edu.cn

---

## Abstract

As the important role of highway in modern social and economic development becomes increasingly prominent, maintaining the good condition of road surface is crucial to ensure traffic safety and extend the service life of road. Although the traditional pavement damage detection method is effective, there are limitations in convenience and practicability, especially in the allocation of personnel and cost control. In view of this, a new pavement damage detection method based on computer vision is proposed in this paper. This study uses the latest version of YOLOv8, an object detection framework, and a homemade datasets that covers different types of pavement damage, including lateral cracks, longitudinal cracks, potholes, cracks, and damage around manhole covers. The datasets collected 5,965 images from an iPhone 7 Plus smartphone and shot under different lighting and backgrounds to enhance the model's generalization. The experimental results show that the mean Average Precision of the model reaches 0.98 in the task of target detection using YOLO v8, which has satisfactory detection effect. In addition, the study also uses YOLOv8 to carry out a case segmentation task, which can not only accurately locate the damage area, but also provide pixel-level segmentation results, which is helpful to evaluate the specific scope and severity of damage. The mean Average Precision of the instance segmentation model reaches 0.604. In conclusion, the method proposed in this study not only improves the efficiency and accuracy of pavement damage detection, but also provides strong support for subsequent road maintenance.

## Keywords

Pavement; Damage Detection; Object Detection; YOLO; Computer Vision.

---

## 1. Introduction

Nowadays, the highway is not only a modern transportation channel, but also plays an increasingly important role in the development of the national economy. As one of the most important links of urban infrastructure, the smooth, safe and durable road directly affects people's economic life. Among them, asphalt pavement structure has become the main form of highway at home and abroad with its advantages of short construction period, low maintenance cost, driving comfort and renewable utilization, and has been widely used. However, with the longer service time of the pavement, due to various reasons such as overload, heavy load, rain erosion and temperature and humidity changes, whether it is cement pavement or asphalt pavement, its surface will inevitably appear a variety of damage, deformation and other defects. Serious road damage is often formed by the gradual development of minor injuries, if the relevant departments ignore this problem, until major injuries

occur, leading to premature withdrawal of roads from service, not only will cause traffic congestion, but also bring serious safety risks.

A large number of experience facts show that effective road management system is very necessary. Therefore, many domestic and foreign scholars and engineers have carried out research on pavement damage detection, and put forward a variety of detection methods. For example, A. Haghghattalab et al. [1] proposed a method for assessing road injuries using high-resolution satellite imagery. In recent years, similar studies on the identification of post-disaster road damage through high-resolution images have been continuously improved, and the representative studies include Zhang et al. (2013) [2] and Wang et al. [3] For the detection of cracks on the road surface, some scholars also proposed a variety of methods. For example, Benedetto [4] used radar technology to detect pavement damage. Yamada [5] introduced a road crack detection method using a mobile robot equipped with 2D laser scanner. Compared with radar detection and laser scanning, direct image-based processing schemes can be more intuitive. Kil and Shin [6] report a two-step algorithm for road damage detection, and obtain good accuracy. With the progress and development of technology, image-based crack identification has become more efficient and fast, which is confirmed by the research of Chu [7], Huang & Zhang [8] and Jo&Ryu [9].

These early development of testing methods and equipment have promoted the development of this field and achieved better results, but most of the technology in the convenience and practicality are limited, because they often need to be equipped with professional testing personnel, which is not conducive to the optimal allocation of personnel, and some equipment is often expensive, and can not be economical and large-scale Mold production. Therefore, the traditional manual record inspection is still the main detection method used at present (as shown in Figure 1). With the rapid development of deep learning technology, many fields such as translation, medical treatment, finance, logistics and so on have gradually changed significantly because of artificial intelligence, and its application to the detection of pavement damage will certainly have great practical benefits.

## 2. Related Works

Computer vision is an important application field of deep learning, which has been widely used in pavement damage recognition in recent years. Zhang et al. [10] collected 500 pavement crack pictures with smartphones, and then used Deep Convolutional Neural Network (DCNN) for training. Compared with SVM and Boosting and other traditional machine learning algorithms, The Precision has been significantly improved, up to 0.869. Zhang et al. [11] then proposed CrackNet, a CNN algorithm for automatically detecting pavement cracks. Different from traditional CNN, CrackNet removes the pooling layer, and the algorithm shows better recognition ability with a Precision of more than 0.9. Meada et al. [12] collected 9053 images containing multiple damage types in seven cities in Japan and trained the SSD-MobileNet model to achieve real-time detection with an Average Precision of over 0.75. Li et al. [13] designed four different CNNs to classify pavement cracks (Longitudinal crack, Transverse crack, Block crack and Alligator crack, respectively), and the classification accuracy reached 0.94. Li et al. [14] used YOLO v4 algorithm to detect pavement cracks, and the research results showed that after adding CBAM module to YOLO v4, the accuracy could be improved by about 3%. Jiang et al. [15] proposed to use Ghost module and CBAM attention mechanism to improve the ability of YOLO v4 to detect pavement cracks, and the results showed that the improved network model had improved detection accuracy and speed. Ni et al. [16] proposed a detection method based on YOLO v7 and Improved LaneNet, which can determine damage classes and lane localization. Li et al. [17] proposed a detection method combining CycleGAN and YOLO v5, and the mean Average Precision (mAP) reached 0.88.

In general, the use of computer vision to identify pavement damage will be a popular development direction in the field of road traffic operation and maintenance, but most studies only focus on pavement cracks, ignoring other injuries. Therefore, this paper expands the damage categories. Besides the conventional Transverse cracks and Longitudinal cracks, We also considered Pothole &

Separations, Alligator cracks, and Defects around the well cover in the data set. In terms of algorithms, we chose the latest YOLO v8 model in the field of object detection, and in addition to regular detection tasks, we also used YOLO v8 for pixel-level segmentation of road cracks and potholes.

### 3. Methodology

#### 3.1 Data Acquisition and Classification

In this paper, the authors collected 5965 photographs (see Figure 1) using a hand-held approach via an iPhone 7 Plus smartphone. Among the collected samples, different lighting conditions and background conditions were included, and in order to further improve the accuracy of the damage identification, photographs were taken from multiple angles for the same road damage. As a part of road damage, road damage around urban manhole covers is not considered in any of the available datasets. Usually, this type of damage occurs due to the well itself, construction, vehicle loading and long term service, which if not prevented can also lead to major accidents and economic losses, so it is necessary to take this type of damage into account. The dataset produced by the authors focuses more on the injuries themselves, which are defined in this paper in six categories: Lateral Crack (DMG1), Longitudinal Crack (DMG2), Pothole and separation (DMG3), Alligator Cracks (DMG4), Well Cover without Cracks (DMG4), and Alligator Cracks (DMG5). Well Cover without Cracks (DMG5), Cracks around Well Cover (DMG6). It should be noted that DMG5 is not a true injury, but to better distinguish it from DMG6, an additional category of "injury" was added, because when the injury around the well cover is not particularly obvious, it is easy to confuse the two during training if only one category is set. After determining the damage category, the Labellmg software loads the image to create a rectangular labelled box, and generates an xml file with the corresponding labelling information of the image. Figure 1 shows some of the damage images in the dataset.

#### 3.2 Convolutional Neural Networks



Figure 1. Road pavement damage images

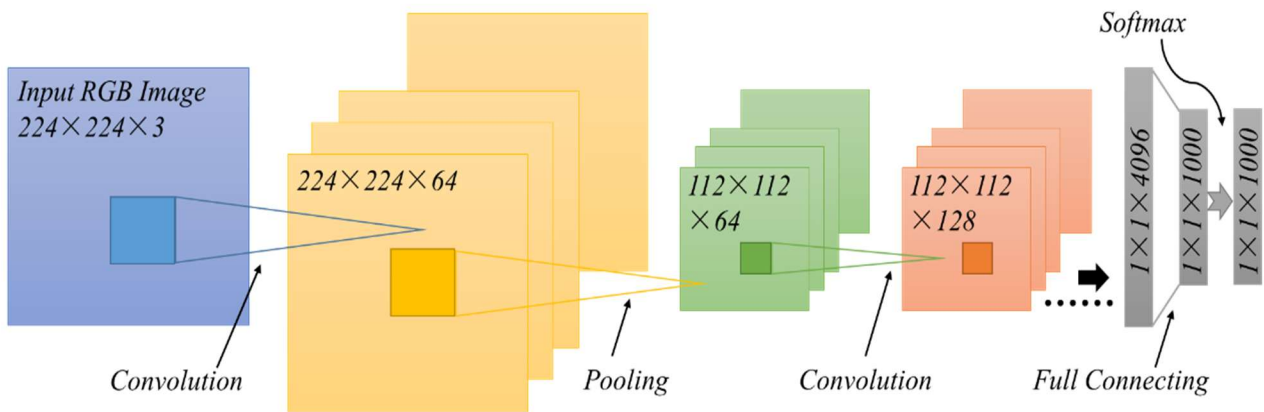


Figure 2. Basic structure of a CNN

Convolutional Neural Networks (CNNs) are deep learning models [18,19] that are particularly suited for processing data with a grid structure, such as images and videos. CNNs automatically detect spatially hierarchical features in the input data through mechanisms such as local receptive fields, weight sharing, and pooling. Partial receptive fields mean that each neuron responds to only a small region of the input; weight sharing ensures that features of the same type are consistently identified throughout the input space; and pooling is used to reduce the spatial dimensionality of the data, which reduces the amount of computation and prevents overfitting. These features allow CNNs to efficiently learn hierarchical representations from raw data without the need for manual feature engineering.

Compared to traditional machine learning algorithms, CNNs are not only able to automatically extract features, but also learn more complex nonlinear relationships. Traditional machine learning methods usually require manual feature engineering and have limited model complexity to capture subtle patterns in high-dimensional data. In addition, CNNs tend to outperform traditional methods on large datasets, especially in computer vision tasks such as image classification, target detection, and semantic segmentation. Figure 2 shows the basic structure of a CNN with the VGG16 network as an example.

### 3.3 YOLO v8 Model

YOLOv8 (You Only Look Once version 8) is a state-of-the-art real-time object detection algorithm [20], which improves the model architecture to increase the detection accuracy and speed with respect to its predecessor YOLO v5. It employs efficient multi-scale feature fusion and lightweight design, and integrates advanced data enhancement techniques such as Mosaic data enhancement. YOLOv8 achieves faster inference by optimising the computational process and utilising hardware acceleration, and provides easy-to-use API interfaces and pre-trained models suitable for a wide range of application scenarios.

#### 3.3.1 Backbone

Figure 3 shows the network structure of YOLO v8. The backbone network of YOLO v8 follows the CSPDarknet53 that has been used since YOLO v4, which mainly applies the Cross Stage Partial Network (CSPNet). In CSP, the input features will be split into two parts, one of which is output by the residual module and then spliced with the other part to form a new output. The use of the CSP module enhances the learning ability of the CNN and speeds up the computation. Different from the previous version YOLO v5, YOLO v8 uses the C2f convolution module, which can obtain more information about the gradient flow compared to the C3 module. YOLO v8 is divided into five types according to the size of the network parameters, and the main difference lies in the width value  $w$  and the depth value  $d$  in Figure 3, and the specific values are shown in Table 1.

**Table 1.** Values of w and d for various YOLO v8 models

Model	d	w
YOLO v8n	0.33	0.25
YOLO v8s	0.33	0.50
YOLO v8m	0.67	0.75
YOLO v8l	1.00	1.00
YOLO v8x	1.00	1.25

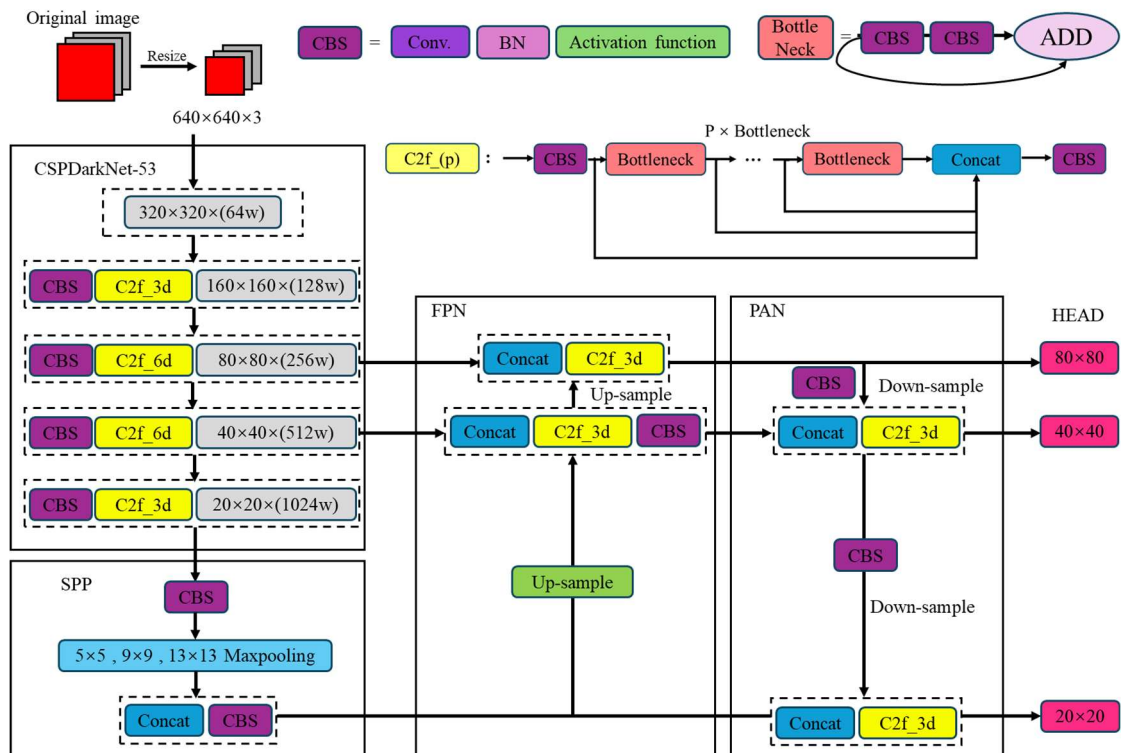
### 3.3.2 Feature Fusion and Prediction Section

After the image features are extracted by the backbone network, the feature information will be further enriched by feature fusion. The feature fusion part mainly consists of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN). Through FPN, low-level feature maps can be up-sampled and spliced with high-level feature maps to better fuse the features before and after the network. Subsequently PAN can pass the high-level information back to the bottom to make the feature fusion more adequate.

Through top and bottom sampling and feature fusion, YOLO v8 can achieve target prediction from three scales, corresponding to different scales of receptive fields, where small-size feature maps can be used to predict large-size targets, and vice versa, large-size feature maps can be used to predict small-size targets.

## 4. Model Training and Detection Results

### 4.1 Object Detection Models



**Figure 3.** Basic structure of YOLO v8

80% of the images in the dataset were used to train and fit the models, and the remaining 20% of the images were used to evaluate the performance of the models. Epoch was set to 200 and the batch size was 8. Table 2 lists the Average Precision (AP) and Mean Average Precision (mAP) for each of the

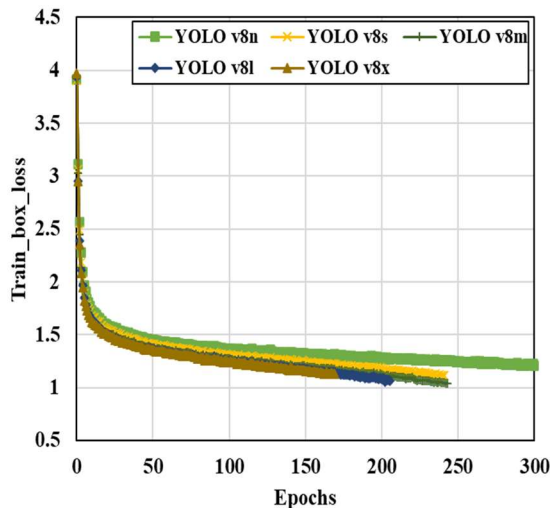
models trained in this paper. Precision, mAP). It can be seen that the YOLO v8 model has a satisfactory precision of 0.98. Compared with SSD MobileNet, which is also a one-stage model, the mAP is higher than 0.1. Compared to the classical two-stage algorithm Faster R-CNN, the mAP is 0.017 higher.

Loss value indicates the degree of deviation of the model's prediction from the true value, and the smaller the value, the better the detection ability of the trained model. Figure 4 shows the curve of the loss value with the training epoch, which shows that as the number of network parameters increases, the model convergence speed will be accelerated. From the loss value of the validation set, the model starts to converge when the training reaches 100 epochs, and YOLO v8x and YOLO v8m show some overfitting phenomenon.

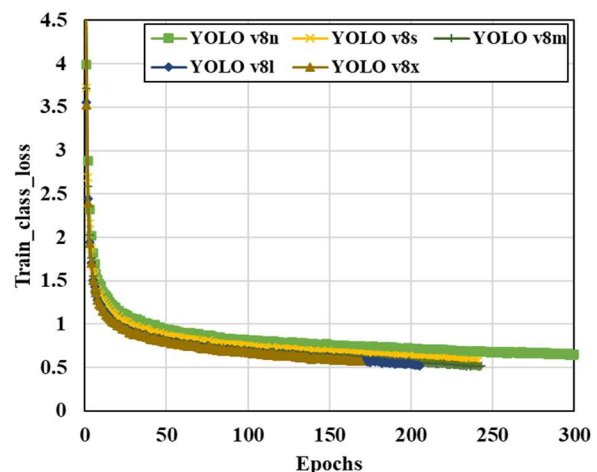
**Table 2.** AP and mAP of various YOLO v8 models

	DMG1	DMG2	DMG3	DMG4	DMG5	DMG6	mAP
YOLO v8n	0.962	0.971	0.994	0.966	0.988	0.994	0.979
YOLO v8s	0.967	0.974	0.994	0.964	0.987	0.994	0.980
YOLO v8m	0.967	0.973	0.993	0.965	0.988	0.995	0.980
YOLO v8l	0.968	0.971	0.995	0.967	0.986	0.994	0.980
YOLO v8x	0.963	0.979	0.995	0.961	0.983	0.995	0.979
SSD-MobileNet-1	0.936	0.798	0.799	0.889	0.794	0.911	0.855
SSD-MobileNet-2	0.944	0.814	0.839	0.904	0.832	0.919	0.876
Faster R-CNN	0.971	0.971	0.888	0.970	0.988	0.991	0.963

In terms of training time, YOLO v8n takes the shortest time, with one epoch consuming 19 s. YOLO v8x takes the longest time, with 70 s. YOLO v8s, YOLO v8m, and YOLO v8l consume 22 s, 32 s, and 47 s, respectively, to train one epoch. Subsequently, we select the one that has a better balance of mAP and detection speed, the YOLOv8m for practical testing on the validation set images. Figure 5 shows some of the test results, which show that the trained YOLO v8m model can accurately locate the road damage in the images, and demonstrates high stability even in the face of different light and viewpoint changes.



(a) Train box loss



(b) Train class loss

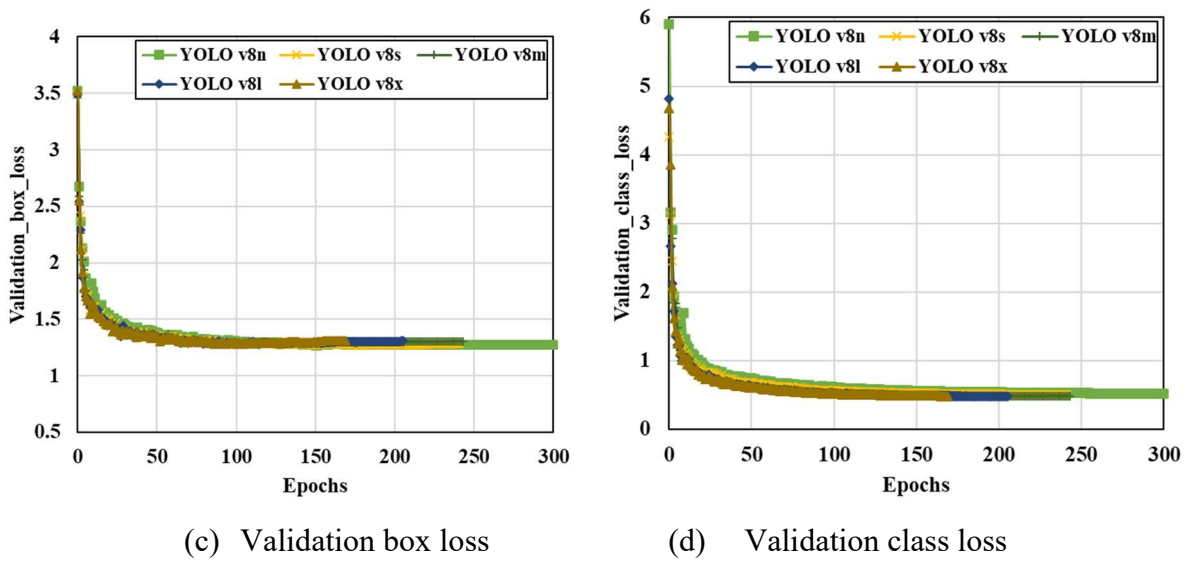


Figure 4. Loss values

### 4.2 Instance Segmentation Model

YOLOv8 instance segmentation consists of identifying and outlining individual objects in an image, providing a detailed understanding of the spatial distribution. Unlike semantic segmentation, it uniquely labels and precisely divides each object. For instance segmentation, YOLO v8 uses a Mask R-CNN-like scheme to accurately segment each object in an image.

Applying instance segmentation algorithms for pavement damage identification has several significant advantages over target detection algorithms: firstly, instance segmentation not only accurately locates the damage region, but also provides segmentation results at the pixel level, which means that it can accurately depict the boundaries and shapes of the damages; secondly, when faced with a complex scene that contains multiple types of damages, instance segmentation is able to differentiate and label different instances for each damage, which provides an opportunity for subsequent damage assessment. instances, which provides more details for subsequent damage assessment; moreover, the rich information provided by instance segmentation can support more in-depth analyses, such as calculating the damage area and evaluating the damage severity, and help to achieve more refined data management and model training.

From the target detection dataset, we selected 1200 images and labelled them with instance segmentation using Labelme software, retaining only the two common types of damage, Crack and Pothole. Subsequently, the pre-training model based on YOLO v8l-seg completes the model training, the epoch is 300, the batch size is 4. Figure 6 shows the curve of the loss value, which shows that the loss value tends to be stable at the end, and the model reaches the convergence state. the average precision (AP) of Crack and Pothole recognition is 0.522 and 0.686, respectively, and the mAP is 0.604. Figure 7 shows the example segmentation results of some images, Crack and Pothole can be accurately segmented, with the improvement of the dataset and the model algorithm, there is still much room for improvement in the future segmentation effect.



Figure 5. Detection results

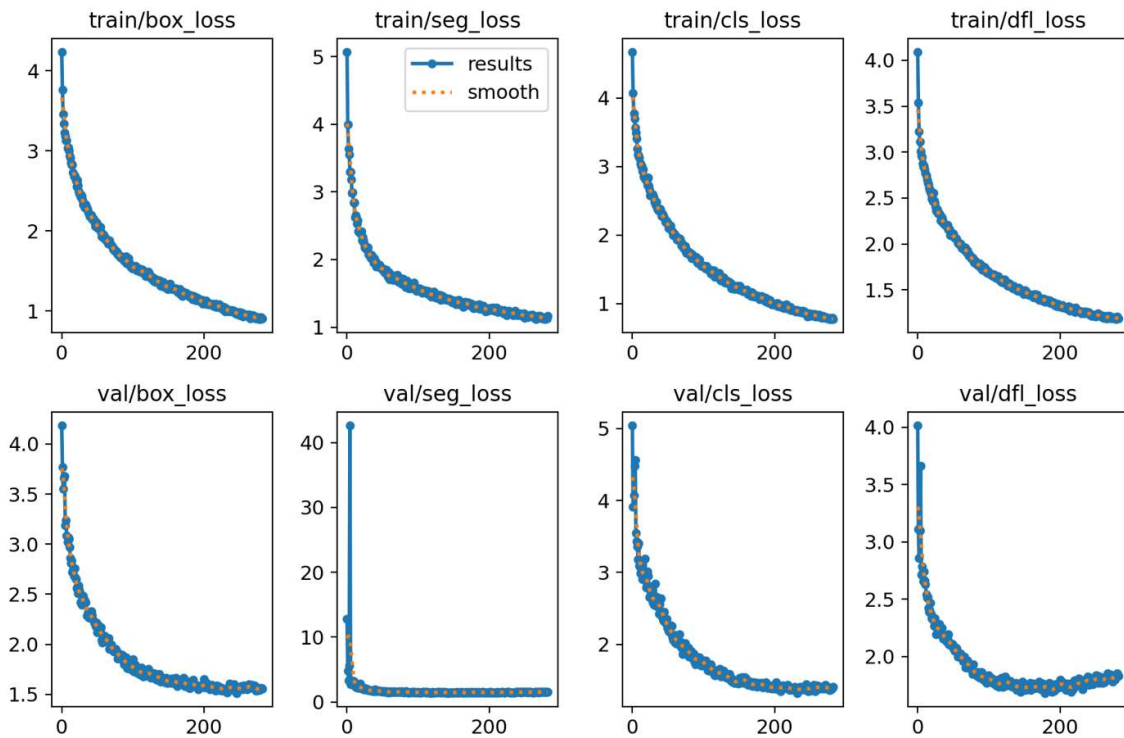


Figure 6. Loss values for segmentation model



Figure 7. Segmentation results

## 5. Conclusion

This paper proposes a road damage detection method based on YOLO v8. Based on self-made datasets, this research trains the target detection algorithm and case segmentation algorithm. The mAP of the target detection algorithm can reach 0.98, which has higher detection accuracy and Faster detection speed than the previous SSD and Faster R-CNN algorithm. The case segmentation model

can accurately segment and classify the pavement damage in the image, which provides more details for the subsequent damage assessment.

## Acknowledgments

This work was funded by the following programs:

Innovation and Entrepreneurship Training Project of University of Science and Technology Liaoning (Research on Intelligent Detection and Evaluation Methods for Pavement Damage Based on Pixel-Level Segmentation).

## References

- [1] A. Haghighattalab, A. Mohammadzadeh, M. J. V. Zoej, et al.: Post-earthquake road damage assessment using region-based algorithms from high-resolution satellite images, Proc. Image and Signal Processing for Remote Sensing XVI (SPIE, 2010), Vol. 7830, p.439-447. DOI:10.1117/12.864538.
- [2] X. Zhang, Y. Chen, M. Jia, et al.: The study of road damage detection based on high-resolution SAR image, Proc. 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS (2013), p.2633-2636. DOI:10.1109/IGARSS.2013.6723363.
- [3] J.Wang,Q.Qin,J.Zhao, et al.:A knowledge-based method for road damage detection using high-resolution remote sensing image, Proc. 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)(2015),p.3564-3567.DOI:10.1109/IGARSS.2015.7326591.
- [4] A. Benedetto, F. Benedetto, M. R. D. Blasiis, et al.: Reliability of Radar Inspection for Detection of Pavement Damage, Road Materials and Pavement Design (2004).DOI:10.1080/14680629.2004.9689964
- [5] T. Yamada, T. Ito, A. Ohya: Detection of road surface damage using mobile robot equipped with 2D laser scanner, Proc. 2013 IEEE/SICE International Symposium on System Integration (2013), p.250-256. DOI:10.1109/SII.2013.6776679.
- [6] D. H. Kil, f. b. Shin: Automatic road-distress classification and identification using a combination of hierarchical classifiers and expert systems-subimage and object processing, Proc. International Conference on Image Processing(1997), Vol. 2, p.414-417. DOI:10.1109/ICIP.1997.638795.
- [7] C. Yanli: Algorithm Study on Thinning and Keeping Connectivity of Bituminous Pavement Crack Images, Proc. 2010 International Conference of Information Science and Management Engineering (2010), Vol. 1, p.550-553. DOI:10.1109/ISME.2010.265.
- [8] W. Huang, N.Zhang: A novel road crack detection and identification method using digital image processing techniques, Proc. 2012 7th International Conference on Computing and Convergence Technology (ICCT) (2012), P.397-400. DOI:10.1109/ICCCT.2012.6530365.
- [9] Y. Jo, S. Ryu: Pothole Detection System Using a Black-box Camera, Vol. 15 (2015) No.11, p.29316-29331. DOI:10.3390/s151129316.
- [10]L. Zhang, F. Yang, Y. Daniel Zhang, et al.: Road crack detection using deep convolutional neural network, Proc. 2016 IEEE International Conference on Image Processing (ICIP) (2016), p.3708-3712. DOI:10.1109/ICIP.2016.7533052.
- [11]A. Zhang, K. C. P. Wang, Y. Fei, et al: Deep Learning-Based Fully Automated Pavement Crack Detection on 3D Asphalt Surfaces with an Improved CrackNet, Journal of Computing in Civil Engineering, Vol. 32 (2018) No.5, p.04018041.DOI:10.1061/(ASCE)CP.1943-5487.0000775.
- [12]H. Maeda, Y. Sekimoto, T. Seto, et al.: Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images: Road damage detection and classification, Computer-Aided Civil and Infrastructure Engineering, Vol. 33 (2018). DOI:10.1111/mice.12387.
- [13]B. Li, K. Wang, A. Zhang, et al.: Automatic classification of pavement crack using deep convolutional neural network,International Journal of Pavement Engineering, Vol. 21 (2018), p.1-7. DOI:10.1080/10298436.2018.1485917.
- [14]L. Li, B. Fang, J. Zhu: Performance Analysis of the YOLOv4 Algorithm for Pavement Damage Detection with Different Embedding Positions of CBAM Modules, Applied Sciences, Vol. 12 (2022), p.10180 DOI:10.3390/app121910180.

- [15]Y. Jiang, D. Pang, C. Li, et al.: Two-step deep learning approach for pavement crack damage detection and segmentation, International Journal of Pavement Engineering, Vol. 24 (2022), p.1-14. DOI:10.1080/10298436.2022.2065488.
- [16]Y. Ni, J. Mao, Y. Fu, et al.:Damage Detection and Localization of Bridge Deck Pavement Based on Deep Learning, Sensors, Vol. 23 (2023), p.5138. DOI:10.3390/s23115138.
- [17]Z. Li, X. Xiao, J. Xie, et al.: Cycle-YOLO: A Efficient and Robust Framework for Damage Detection, 2024.
- [18]Y. LeCun, L. Bottou, Y. Bengio, et al.:Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, Vol. 86 (1998), p.2278-2324. DOI:10.1109/5.726791.
- [19]Y. LeCun, L. Bottou, Y. Bengio, G. Hinton: Deep learning, Nature, Vol. 521 (2015) No.7553, p436-444. DOI:10.1038/nature14539.
- [20]G. Jocher, A. Chaurasia, J. Qiu: Ultralytics YOLO, (2023-01). <https://github.com/ultralytics/ultralytics>.