033

深層学習に基づく道路ひび割れの検出法

DEVELOPMENT OF ROAD CRACK DETECTION METHOD BASED ON DEEP LEARNING

21WM1333 陸　海翔　　指導教員 丸山 喜久
Lu Haixiang　　劉　ウェン

SYNOPSIS

This study tries to develop a method that can automatically detect cracks in concrete structure images by using a convolutional neural network, and it is applies to detect road cracks caused by the 2016 Kumamoto earthquake. We adopt the deep learning method YOLOv5 for the detection and modify it to overcome the weaknesses of the YOLOv5 model. The used road images are taken by vehicle-mounted cameras after the earthquake on April 17 and 20, 2016. After the image preprocessing, we create two datasets for training by the original YOLOv5 model. Then the dataset with better results were applied to the modified model. The test results show that the modified model has better detection ability for distant targets and smaller targets, it also can detect nearby targets better. However, the confidence level of the modified model generally drops by 5-20% and some near cracks were omitted. The modified model is larger by adding more structures, the training time became double of the original YOLO model.

1. Introduction

As the investment in infrastructure such as highways and bridges in various countries increases year by year and natural disasters (such as earthquakes) become more frequent, structural health issues have always been the top priority of traffic safety. Meanwhile, road cracks are one of the main forms of road damage. At present, crack detection is mainly performed manually in practice. This method has high work costs, high labor intensity, and low detection efficiency. Therefore, it is an urgent engineering practice problem to grasp the road surface information quickly and timely.

In recent years, object detection algorithms have made significant breakthroughs. However, most methods use the R-CNN algorithm based on Region Proposal (R-CNN, Fast R-CNN, Faster R-CNN), which is called a two-stage method. Its accuracy rate is high, but speed is slow, which is not suitable for real-time detection. Therefore, this study tries to use the faster one-stage methods[1)] and improve accuracy weakness[2)].

Considering the purpose of detecting road damage in real-time through onboard cameras, the fastest YOLOv5 model[3)] is selected. In additional, we add a new feature fusion layer[4)], a new prediction head, and a Coordinate Attention mechanism[5)] to the model to enhance the detection ability of small objects.

2. Dataset and Image Processing

In this study, the used road images were taken by onboard cameras after the Kumamoto earthquake on April 17 and 20, 2016. Since road surfaces deteriorate due to the daily use, we distinguish the cracks caused by the earthquake from the cracks due to the daily use manually. The definition of the cracks caused by the earthquake is according to three principles:1) cracks that forming height differences and potholes, which affect safe passage; 2) large-scale breakage; 3) deep cracks, forming faults with the surrounding.

In this experiment, the cracks are classified into four types: transverse crack (TC), longitudinal crack (LC), alligator crack (AC), and road pit (RP) as shown as in Figure 1. There are a total of 73,500 cracks. Because the road images contain the road, the sides of the road, the sky, etc., we cropped the road part from the whole image. In this research, we built the dataset twice.

Figure 1 Samples of four types of the road damage

In the first-made dataset, the lower 1/4 part of each image was cut off. Only the road and both sides were kept as shown as the red box in Figure 2. In this case, the road will be included in the image regardless of whether the vehicle is going straight or turning. After the manual screening, 4,676 images with damage and 68,824 images without damage were selected. Among them, 500 image with damage and 500 images without damage were randomly selected as the test set, and the rest were used as the training set and validation set.

After several times of training using the first dataset., it was found that there were serious overfitting problems. After research, we decided to carry out the second dataset construction. This time, images were re-cropped that only contained the main road,

as shown as the green box in Figure 2. In additional, the numbers of each class of cracks were counted. In the second-made dataset, 500 images with cracks and 500 images without cracks are still divided as the test set. The training set and the validation set are divided by 3:1, which includes 3132 and 1044 images with cracks respectively. Also, four data enhancement methods are used to increase the number of samples, which namely horizontal flip, vertical flip, Gaussian noise, and mosaic data enhancement.

Figure 2 The used road image, where the red box shows the road area in the first crop and the green box is the road area after the second crop.

3. YOLOv5 Model and Modification

YOLOv5 is a one-stage target detection algorithm with region-free method[1]. It integrates more tricks on the basis of YOLOv4[3], and the weight file size is only about 10%, also the speed is almost up to twice. In this experiment, the latest YOLOv5 5.0 version was used.

The network structure is shown in Figure 3, and the whole structure has four parts: Input, Backbone, Neck, and Prediction. Those parts consist of the following: 1) Input: adaptive anchor box calculation[6] and adaptive image scaling; 2) Backbone: focus module and C3 module; 3) Neck: FPN+PAN[7] structure; 4) Prediction: CIOU_Loss[8]. The main difference between YOLOv5 and YOLOv4 is the use of FOCUS module and C3 module.

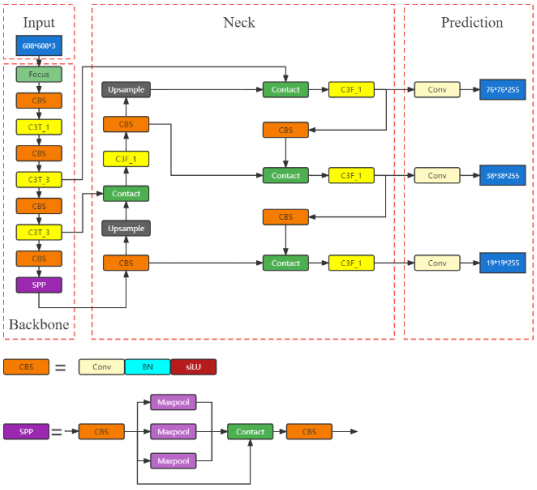

Figure 3 The structure of the original YOLOv5 model

FOCUS module reduces the number of layers, the number of parameters, and computation. Thus, the required memory usage of CUDA decreases, and the speed of inference and gradient back-propagation are improved. The C3 module is born out of the CSP[9] (Cross Stage Partial Networks) module, which is the cross-stage local network. It can simplify the network structure, reduce the amount of computation, and reduce the model inference time.

The learning and detection speed of YOLOv5 is very fast, which can detect up to 45-155 images per second, much higher than other algorithms. But the accuracy of the result is low, and the detection effect for small objects is poor, especially dense small objects. However, due to the position and the angle of the car camera, the images and videos are taken parallel to the ground, resulting in the cracks that were originally at a certain distance from each other becoming denser in the images. To overcome the weakness of small objects, a new feature fusion layer, a new prediction head, and the Coordinate Attention mechanism[5] are added to the model. The structure of the modified model is shown in Figure 4.

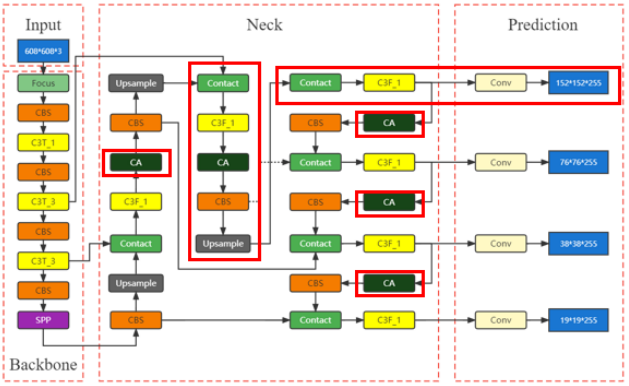

Figure 4 The structure of the modified YOLOv5 structure by adding a new feature fusion layer, a new prediction head, and a Coordinate Attention mechanism.

Coordinate Attention[5] is a lightweight attention module proposed in 2021. The benefit is that long-range dependencies can be captured along one spatial direction, and precise location information can be preserved along the other. Then, the generated feature maps are separately encoded to form a pair of feature maps, which can be applied to the input feature maps to enhance the representation of objects of interest.

In the new added fusion layer[4], the C3 module and CBS module consistent with the original model are used. At the same time, the fused feature map is further upsampled and spliced with the one from the backbone part of the original network to generate a new fused feature map. When the feature information from the Backbone part is brought into the feature fusion layer, the new fusion layer is also brought in. These connections can enhance the back-propagation of the gradient, avoid gradient decay and reduce the loss of feature information of small objects.

4. Model Training

We used the original YOLOv5 for seven times of trainings. The first-made dataset was used for three times, and the second-made dataset was used for four time by the original YOLOv5. The best results for two datasets are shown in Figure 5.

Precision is calculated as the ratio between the number of positive samples correctly classified to the total number of samples classified as positive (either correctly or incorrectly). It measures the model's accuracy in classifying a sample as positive.

Recall is calculated as the ratio between the number of positive samples correctly classified as positive to the total number of positive samples. It measures the model's ability to detect positive samples.

Average Precision (AP) is a way to summarize the Precision-Recall curve into a single value representing the average of all precisions. Each class can calculate its Precision and Recall to get a curve, and the area under it is the value of AP. mAP_0.5 means mean Average Precision when IOU (intersection over union) is set as 0.5, and mAP_0.5:0.95 means mean Average Precision over different IOU (from 0.5 to 0.95 in steps of 0.05).

The loss function is used to measure the degree of inconsistency between the predicted value of the model and the real value. The smaller it is, the better the robustness.

For the evaluation indicators Precision, Recall and AP, the higher values show better results, whereas the lower value for the loss function shows better results. From Figure 5, it can be confirmed that the loss of the model using the first-made dataset made was overfitting at step 50. Its results were not as good as the results using the second-made dataset. Then we applied second-made dataset to our modified YOLOv5 model.

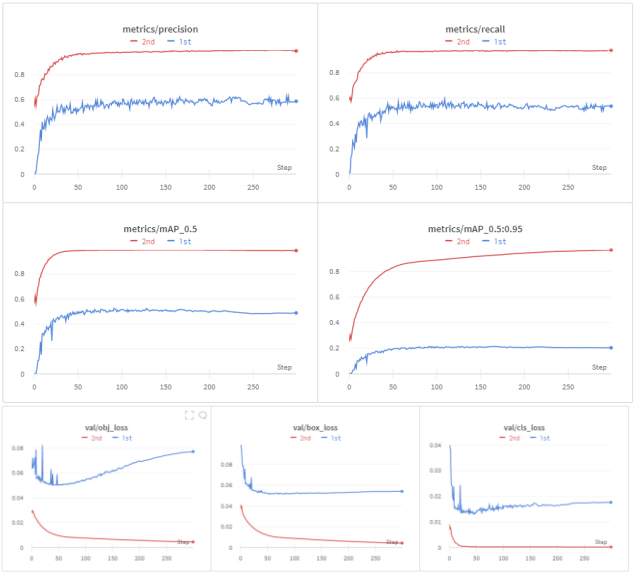

Figure 5 Comparison of the training results of the original YOLOv5 with two different datasets

The comparison of the results using the original YOLOv5 and the modified models with the second datasets is shown in Figure 6. We can see differences of the performance evaluation indicators are not large. The modified version shows poor performance than the original version, where the indicators are 0.02~0.05 lower.

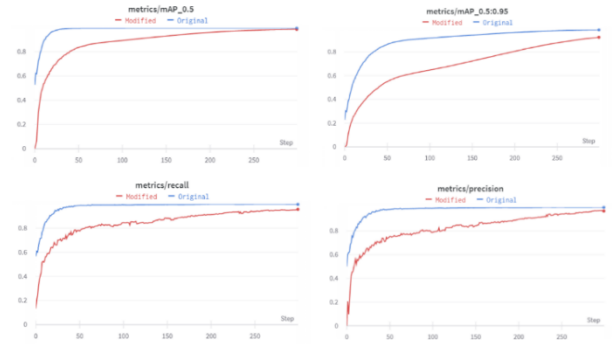

Figure 6 Comparison of the training results of the original and modified YOLOv5 using the second-made dataset

5. Detection of Road Damage

We performed two detectors, using the original and the modified YOLOv5 with the second-made datasets. Considering the application scenario, our envisage is real-time detection using in-vehicle cameras, which means all the images without cropping. So firstly, we try to use the uncropped images for detection. In the first detection, there are 1000 images, including 500 images with cracks and 500 images without. They are detected by the original and modified YOLOv5 models respectively. The samples of the images with cracks are shown in Figure7, which were detected successfully. The recalls of the two models are shown in the Table 1.

Errors are mainly concentrated in three types: undetected (fail to identify cracks), off-road objects (identify non-ground targets as cracks), and wrong objects (identify targets on the road or roadside as cracks). The samples are shown in Figure 8, 9 and 10. Undetected is the most common, followed by off-road objects and the least false objects.

Table 1. Comparison of the detected results using the uncropped images by the original and modified models that were trained by the second-made dataset

	Images with cracks (500)	**Images without cracks (500)**
Original	97.6% (488)	74.4% (372)
Modified	96.4% (482)	72.2% (361)

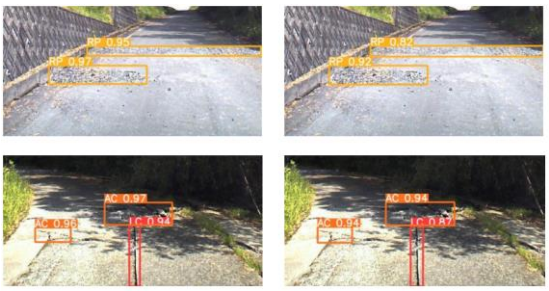

Figure 7 Both the original and the modified YOLOv5 can correctly identify the cracks from the uncropped road images.

Figure 8 An example of the undetected error, where the longitudinal crack could not be identified.

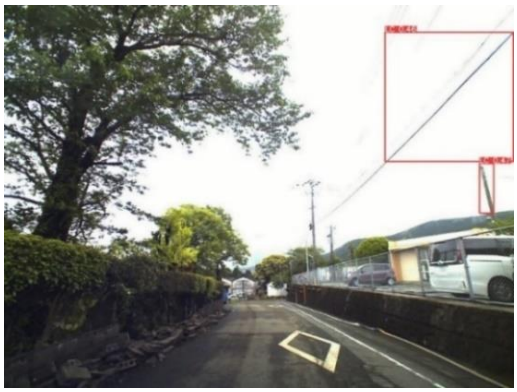

Figure 9 An example of the off-road objects error. The wires and poles are identified as longitudinal cracks

Figure 10 An example of the wrong objects error. Soil beside the road is identified as a road pit.

After confirming the images manually, it can be found that many errors were concentrated in both off-road objects error and wrong objects error. If we use the cropped images in the second-made dataset, those errors would be reduced. Therefore, we try to use the second-made dataset with only road content for the detection. including 500 images with cracks and 500 images without. The recalls of the two models are shown in the Table 2.

Table 2. Comparison of the detected results using the cropped images by the original and modified models that were trained by the second-made dataset

	Images with cracks (500)	**Images without cracks (500)**
Original	98.6% (493)	97.8% (489)
Modified	97.8% (489)	95.4% (477)

The recalls increase for both model after applying to the cropped images. The recalls of the original version are higher than that of the modified model, but the gap is small. The modified YOLOv5 is more sensitive than the original, which can detect more small cracks and far cracks. The comparison is shown in Figure 11. The modified model identified two transverse cracks, which were omitted by the original model. However, due to its sensitivity, some areas without cracks were recognized as cracks, such as the example shown in Figure 12. The training time of the modified model is almost doubled of the original model. Thus, the lightweight advantage of the YOLOv5 no longer exists.

Original Modified

Figure 11 The modified YOLOv5 can detect more cracks than the original one, especially small or long-range targets.

Original Modified

Figure 12 Due to the sensitivity of the modified version, the junction of the sidewalk was identified as a longitudinal crack due to the strong contrast.

6. Conclusion

In this research, the detectors for road damage were generated using the road images by deep learning methods. The road images were taken by the onboard cameras of ordinary vehicles after the 2016 Kumamoto earthquake. Four different damage types were classified manually. Then the YOLOv5 model with fastest calculation speed was adopted for training and testing. Considering the low accuracy of the original YOLOv5 model for small targets, we improved the network. A new feature fusion layer and a new prediction head are added in YOLOv5, and the Coordinate Attention mechanism is added to enhance the detection ability of small objects. Then we trained the original and modified models using the cropped road images.

When the two detectors were applied to the uncropped images, both models showed capability on detecting cracks and pits. However, the recalls are less than 75%. When the two detectors were applied to the cropped images, the accuracy increased significantly to more than 95%. The modified model has better detection ability for distant targets, smaller targets, and also nearby targets. However, the overall accuracy of the modified model is lower than the original model. Meanwhile, the modified model has more layer which increased the training time.

In the future, we plan to continue to optimize the model. On the one hand, we will improve sensitivity and increase the correct rate while maintaining detection rates. We plan to prune the program and improve the network structure to make the model more concise and reduce the training time. When better attention modules or network structures become available, we will try to incorporate them into the model as well.

REFERENCE

1) J Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi: You Only Look Once: Unified, Real-Time Object Detection, IEEE Conference on CVPR, pp. 779-788, 2016.
2) Junming Liu, Weihua Meng: Review on Single-Stage Object Detection Algorithm Based on Deep Learning, Aero Weaponry, 27(3): 44-53, 2020.
3) Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao: YOLOv4: Optimal Speed and Accuracy of Object Detection, ArXiv, abs/2004.10934, 2004.
4) Linlin Zhu, Xun Geng, Zheng Li, and Chun Liu: Improving YOLOv5 with Attention Mechanism for Detecting Boulders from Planetary Images, Remote Sensing 13, no. 18 (2021): -. doi: 10.3390/rs13183776, 2021.
5) Qibin Hou, Daquan Zhou, Jiashi Feng: Coordinate Attention for Efficient Mobile Network Design, IEEE Conference on CVPR, pp. 13713-13722, 2021.
6) Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, Silvio Savarese: Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression, IEEE/CVF CVPR, pp. 658-666, 2019
7) Golnaz Ghiasi, Tsung-Yi Lin, Quoc V. Le: NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection, IEEE/CVF CVPR, pp. 7036-7045, 2019.
8) Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, Dongwei Ren: Distance-IOU Loss: Faster and Better Learning for Bounding Box Regression. AAAI Conference on Artificial Intelligence, 34(07), pp.12993-13000, 2020.
9) Chien-Yao Wang, Hong-Yuan Mark Liao, I-Hau Yeh, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh: CSPNet: A New Backbone that can Enhance Learning Capability of CNN, IEEE/CVF Conference on CVPRW, pp. 1571-1580, 2020.