Pavement Crack Detection Method of Street View Images Based on Deep Learning

Zekai Shu*, &, Zhaoyu Yan &, Xihang Xu &

*These authors contributed equally to this work and should be considered co-first authors.

School of Chang’an Dublin International College of Transportation, Chang’an University, Xi’an 710064, China

*Corresponding author e-mail: 2018905835@chd.edu.cn

Abstract. Pavement crack detection is a challenging task for carrying out pavement maintenance works. Deep learning method is regarded as an effective and accurate way to detect pavement cracks. However, this often requires a large dataset composed of different crack images. This paper introduces a convenient and low-cost method to collect pavement images by using street view images. 400 images from 5 cities are collected and labeled to form the dataset. Then, it is applied to train a target detection network YOLOv5, which is the latest version of YOLO network. The result shows that this network can effectively detect crack with mAP of over 70% and detection time of 152ms, which are all better than another classical method YOLOv3. Considering the easiness of collecting images, this method can be a suitable way to evaluate the pavements.

1. Introduction

The maintenance of transportation infrastructure is important to the working of many activities. Delayed repairing and maintenance can result in significantly increased cost. In the road maintenance work, the pavement cracks detection is important. So the detection of cracks at its early developing stage can be essential. Generally, the widely adopted method is by manual detection. However, it still has some defects such as low efficiency, large error and are sensitive to subjectivity [1]. To solve these problems, some other methods such as Gabor filter [2] and 3D laser scanning technology [3] is proposed to do the detection. In recent years, deep-learning techniques are applied to this region and a series of researches demonstrate that it can quickly find cracks with high accuracy. Zhang et al. [4] put forward CrackNet to detect 3D asphalt pavement cracks. Fan et al. [5] used the deep convolution neural network (CNN) on images and has achieved an accuracy of 99.92% for classification. Kaseko et al. [6] suggested a methodology which combines neutral network and traditional method to identify the cracks on pavement. A new network named FPHBN was proposed in [7] to detect the cracks. Mandal et al. [9] applied a target detection model YOLOv2 and images from mobile cameras to do automatic pavement distress identification works. Eisenbach et al. [8] used GAPs pavement distress dataset and convolutional neutral networks (CNN) to automatically detect pavement distress.

Machine learning often requires large dataset for training and testing. For pavement cracks detection, images are mainly obtained by devices such as Charge coupled device (CCD) [10] and Ground
However, it often requires a lot of time and generate extra costs. In recent years, as a new type of picture data provided to the public, street view pictures can be used as a convenient source of images for many researchers. Internet companies such as Google and Baidu have collected and built their own street view images database and open it to the public [12]. These data have been widely used in street safety assessment [13] and evaluating the environment of neighborhood [14]. Compared with traditional dataset, street view pictures have a large amount of accessible data and has covered many different places around the world [15]. Google and Baidu have also provided API for researchers to obtain images with specified location, customized shooting angle and picture size. From the view of the timeliness, to ensure the effectiveness, internet companies are willing to update their database periodically, this enables researchers to evaluate pavement changes through time. Therefore, it is a suitable way for researchers to collect data, which can help to reduce the time and cost spent.

To effectively apply street view images to the work of automatic pavement detection, we propose a new target detection model based on YOLOv5 network to identify road cracks in street view images.

2. Pavement crack detection model

2.1. Algorithm flow
Firstly, get the street view pictures with cracks to build the dataset. Then using open-source tool to create labels for each image. The dataset is then divided into two parts and use the training part to train YOLOv5 network. The performance of network on validation dataset is evaluated after each epoch. Saving the best weight after finishing training part and after this the model can be applied to detect the unknown pictures. The overall flow diagram is shown in Fig. 1.

![Fig. 1 Flow chart](image)

2.2. Dataset acquisition
We get 400 street view images from five cities in China (Xi’an, Jilin, Shenyang, Harbin and Shiyan). The cracks are divided into three main types, the number of each type of cracks are shown in Fig. 2. In the figure, 0,1,2 represent transverse, longitudinal and alligator cracks respectively. The size of image is set to $1024 \times 512$ and then cut into $512 \times 512$ to satisfy the input requirement of network. The labeling is operated by LabelImg, which is an open-source image labeling software. After labeling, xml files are formed containing the crack classes and coordinates of the boxes and then converted to txt file which can be recognized by YOLOv5. Finally, the dataset is divided into training and validation part by 8:2.

The diversity of dataset can be easily achieved by using street view images. For traditional images acquisition methods, researchers have to drive to the destination to take photos of pavement cracks, which may restrict the scope of image sources as the pavement conditions are similar for each city and
even each region. This may make the algorithm not universally applicable to other regions. However, through the use of street view images, this could be solved. The images can be easily captured from different regions in China without using professional equipment. So, this paper chooses images with different pavement conditions, weathers, lighting conditions and so on to ensure that the algorithm can be applied to detection work on different conditions.

Due to the lens distortion, the inclination stage of cracks on adjacent lanes are changed. Especially for longitudinal cracks, this could make them more like transverse cracks, which could be an obstacle for the detection accuracy of the neutral network. To solve this problem, this paper tries to restrict the width of images to filter the cracks on adjacent lanes.

Fig. 2 Classes of cracks and number of images

2.3. Model Introduction
To achieve the target of this paper on classifying the cracks and location them on images, this paper proposes a convolutional neutral network called YOLO. You only see once (YOLO) is a classic one-stage target detection algorithm, which can be applied to get the classes and also the location of the target at the same time [16,17]. YOLO are able to acquire bounding boxes and the classification of targets on images with fast speed, which also enables it to do the detection on videos. Due to these benefits, it has been widely used for the detection of steel defects [18], pedestrian [19] and license plate [20]. YOLOv5 is the latest version of YOLO, it has four version with different applications, YOLOv5s has the simplest network structure but also has quickest speed. YOLOv5x has the deepest network but slowest speed. This paper chooses YOLOv5l with second deepest network with higher accuracy than YOLOv5s and YOLOv5m.

YOLOv5 is integrated with many advanced techniques to improve its detection ability [21]. For the input images, YOLOv5 uses mosaic method to increase the number of images by randomly cutting multiple images and splicing them. In the backbone of YOLOv5, the “Focus” module is used to extract the features of image, which can extend the number of channels without losing image information. Through the use of this module, the computation can be significantly decreased for the following training operations. CSP2 structure is used in the Neck of the network, which can help to mix the network characteristics. For the output part YOLOv5 selects GLOU_Loss as the loss function and also applies weighted NMS to remove the redundant overlapping frame.

Transfer learning is used to save time on training the neutral networks. This is to apply the previously fully trained model on a new task. YOLO has a pretrained weights based on COCO, which is a widely used dataset for the performance evaluation of models for recognition, segmentation works. This dataset includes over 300,000 images divided into 91 categories and so that contains many features. Before training on this paper’s dataset, loading COCO weights can help to save training time.
3. Experimental Results and Analysis

mAP (mean average precision) is used to evaluate the accuracy of the model. This is a widely used index to evaluate the performance of deep learning models [22]. This metric is closely related to precision and recall rate. According the classification result, the images can be divided into four groups. Precision is the ratio of TP over (TP+FP). And recall is the ratio of TP over (TP+FN). By adjusting the judging threshold, precision and recall will change at the same time. So, to evaluate the performance of neutral network more objectively, mAP is used. For each class of cracks, an array of precision-recall values is calculated to draw the precision-recall curve, AP (average precision) is used to represent the area of the graph composed of curve, x-axis and y-axis. Finally, the mean value of AP of all classes is calculated to get mAP. The higher value of mAP, the better performance for the model.

In order to verify the advantages of crack detection model using YOLOv5 model, a comparative experiment was carried out on YOLOv5 and YOLOv3 under the condition of using the same dataset and 100 epochs of training. The results are shown in Table 1. The result of mAP test demonstrates that YOLOv5 is 0.15 higher than YOLOv3. In the perspective of detection time for one image with the size of $512 \times 512$, the model in this paper is 22ms faster than YOLOv3 model and has better practical effect.

| Algorithm      | mAP@0.5 | detection time (ms) |
|----------------|---------|---------------------|
| YOLOv3         | 0.648   | 174                 |
| This paper’s model | 0.733   | 152                 |

Fig. 3 demonstrates the detection result on four random images after training. It shows that cracks are successfully located and classified on images with confidence level of over 0.55. It can be seen from these four images that after training on images from different pavement conditions, the network can effectively recognize cracks from dark and light environment. The longitudinal crack is inclined with a small angle, and this may have an impact on the classification of cracks.

![Fig. 3 Prediction result on six random images](image)
To compare the detection ability of two kinds of networks on the same pavement diseases, same images are detected through two networks. The results are shown in Fig. 4, where the upper three figures are by YOLOv5 network and the other three figures under them is from YOLOv3 network. It can be seen that the IOU value of YOLOv3 is lower than 0.3, but for YOLOv5 are much higher and Some parts of the crack are not being detected by YOLOv3. This shows that YOLOv5 can effectively detect the location of cracks. As to the other parts of alligator cracks outside the box, this may be because in the labeling stage some small branches are not included in the ground truth box, this require making the labels more precisely if needed. It can also be seen from the last group of images that YOLOv3 treat the shadow of streetlamp as a short crack, which present that YOLOv5 have better effect on detecting smaller target. However, for the last group of images, YOLOv5 only detected parts of the longitudinal crack, this may because the shadow breaks the crack into two parts and makes an obstacle to the detection of cracks, so this requires more training images with the same condition.

Fig. 4 The comparison between two models

4. Conclusion
Transportation infrastructure plays an important role in the development of the society, so its maintenance is important for ensure its service life and high performance. Delayed maintenance works may increase the cost. To detect the pavement cracks at its early stage and so that implement timely maintenance works, this paper proposed a pavement crack recognition model which combines the street view image data source and YOLOv5 target detection network. Compared with the traditional model, this method can reduce the cost on collecting dataset and increase the diversity of dataset by extracting the images with different conditions. Using the latest version of YOLO to detect the cracks and has achieved over 70% of mAP and 0.15s of detection time. However, to better increase the robustness and accuracy of this system, a larger and diverse dataset is required.

References
[1] Gavilán, M., Balcones, D., Marcos, O., Llorca, D. F., Sotelo, M. A., Parra, I., ... & Amirola, A. (2011). Adaptive road crack detection system by pavement classification. Sensors, 11(10), pp. 9628-9657.
[2] Salman, M., Mathavan, S., Kamal, K., & Rahman, M. (2013). Pavement crack detection using the Gabor filter. In 16th international IEEE conference on intelligent transportation systems (ITSC 2013), pp. 2039-2044.
[3] Zhang, D., Zou, Q., Lin, H., Xu, X., He, L., Gui, R., & Li, Q. (2018). Automatic pavement defect detection using 3D laser profiling technology. Automation in Construction, 96, pp. 350-365.
[4] Zhang, A., Wang, K. C., Li, B., Yang, E., Dai, X., Peng, Y., ... & Chen, C. (2017). Automated pixel - level pavement crack detection on 3D asphalt surfaces using a deep - learning network.
Fan, R., Bocus, M. J., Zhu, Y., Jiao, J., Wang, L., Ma, F., ... & Liu, M. (2019). Road crack detection using deep convolutional neural network and adaptive thresholding. In 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 474-479.

Kaseko, M. S., & Ritchie, S. G. (1993). A neural network-based methodology for pavement crack detection and classification. Transportation Research Part C: Emerging Technologies, 1(4), pp. 275-291.

Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X., & Ling, H. (2019). Feature pyramid and hierarchical boosting network for pavement crack detection. IEEE Transactions on Intelligent Transportation Systems, 21(4), pp. 1525-1535.

Eisenbach, M., Stricker, R., Seichter, D., Amende, K., Debes, K., Sesselmann, M., ... & Gross, H. M. (2017). How to get pavement distress detection ready for deep learning? A systematic approach. In 2017 international joint conference on neural networks (IJCNN), pp. 2039-2047.

Mandal, V., Uong, L., & Adu-Gyamfi, Y. (2018). Automated road crack detection using deep convolutional neural networks. In 2018 IEEE International Conference on Big Data (Big Data), pp. 5212-5215.

Wang, W., Wang, M., Li, H., Zhao, H., Wang, K., He, C., ... & Chen, J. (2019). Pavement crack image acquisition methods and crack extraction algorithms: A review. Journal of Traffic and Transportation Engineering (English Edition), 6(6), pp. 535-556.

Gao, J., Yuan, D., Tong, Z., Yang, J., & Yu, D. (2020). Autonomous pavement distress detection using ground penetrating radar and region-based deep learning. Measurement, 164, 108077.

Long, Y., & Liu, L. (2017). How green are the streets? An analysis for central areas of Chinese cities using Tencent Street View. PloS one, 12(2), e0171110.

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore-predicting the perceived safety of one million streetscapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 779-785.

Rundle, A. G., Bader, M. D., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google Street View to audit neighborhood environments. American journal of preventive medicine, 40(1), pp. 94-100.

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., ... & Weaver, J. (2010). Google street view: Capturing the world at street level. Computer, 43(6), pp. 32-38.

Laroca, R., Severo, E., Zanlorensi, L. A., Oliveira, L. S., Gonçalves, G. R., Schwartz, W. R., & Menotti, D. (2018). A robust real-time automatic license plate recognition based on the YOLO detector. In 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1-10

Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: unified, realtime object detection. In: IEEE conference on Computer Vision and Pattern Recognition, pp. 779–788.

Li, J., Su, Z., Geng, J., & Yin, Y. (2018). Real-time detection of steel strip surface defects based on improved yolo detection network. IFAC-PapersOnLine, 51(21), pp. 76-81.

Lan, W., Dang, J., Wang, Y., & Wang, S. (2018). Pedestrian detection based on YOLO network model. In 2018 IEEE international conference on mechatronics and automation (ICMA), pp. 1547-1551.

Chen, R. C. (2019). Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning. Image and Vision Computing, 87, pp. 47-56.

Li, S., Gu, X., Xu, X., Xu, D., Zhang, T., Liu, Z., & Dong, Q. (2021). Detection of concealed cracks from ground penetrating radar images based on deep learning algorithm. Construction and Building Materials, 273, 121949.

Yue, Y., Finley, T., Radlinski, F., & Joachims, T. (2007), A support vector method for optimizing average precision. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 271-278.