Research on Baidu Street View Road Crack Information Extraction Based on Deep Learning Method

Min Zhang¹, Yang Liu¹,²,³, Shaohua Luo¹, Siyan Gao¹

¹School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing, 102616, China
²Beijing Advanced Innovation Center for Future Urban Design, Beijing University of Civil Engineering and Architecture, Beijing, 100044
³Beijing Key Laboratory of Urban Spatial Information Engineering, Beijing 100038, China

Corresponding author’s e-mail: 2634039766@.com

Abstract. Rapid and effective identification and extraction of road cracks has always been a major difficulty in road detection and maintenance. This paper applies the Deeplabv3+ model to road crack extraction, and proposes a new joint identification method for road cracks based on open network data and open source convolutional neural network. Based on the semantic segmentation theory and Deeplabv3+ neural network, this method uses baidu street view map as the training data set. By adjusting the proportion weight of road cracks and background, the training network model can quickly and accurately identify road cracks. Experimental results show that this method has achieved a good effect on crack segmentation: Mean Intersection over Union (MIOU) of this method is more than 70%, and the processing speed is 1 second/sheet, which is better than FCN algorithm. The results of road crack information extraction are similar to those of manual interpretation, which indicates that this method is feasible for quality evaluation of municipal roads.

1. Introduction

Crack diseases are one of the main road surface diseases [1]. Traditionally, road crack detection research adopts the method of manual on-site detection, which is not only time-consuming and labor-intensive, but also has hidden safety hazards and is not objective enough to be difficult on a large scale and refined scale on the application [2]. Under the condition of maintaining normal traffic and saving maintenance costs, how to carry out efficient and accurate crack detection has important research and practical value.

At present, the continuous development of high-precision Internet street view images has provided new data sources and technical ideas for the identification of urban characteristics on a large scale of space [3]. Compared with traditional on-site image data collection, Internet street view data has the characteristics of wide coverage and large amount of data. In addition, Internet map providers provide free download services for street view images, which makes image data collection efficient and easy to operate. Internet street view images have been applied to road traffic sign detection [4], community environment [5], urban security [6], income prediction and research on architectural characteristics [7]. Street view images have become an important data source in urban environmental assessment research, and also provide new research ideas for the detection of urban road conditions.
In recent years, deep learning has made rapid progress in the field of image recognition, promoting its research and application in road crack detection. Young-Jin Cha et al. [8] proposed a fully automated convolutional neural network (CNN) to extract and classify road cracks in order to solve the problem of road images affected by the environment. Lei Zhang et al. [9] trained classifier for pavement crack identification is better than manual identification. Zhang et al. [10] proposed CrackNet to perform automated pavement crack detection on asphalt pavement.

In summary, the training data set used in the existing research is orthophoto data taken parallel to the road surface. The image contains only road surface and crack information, and cracks account for a relatively large proportion in the image and are easy to identify. The Internet Street View data is taken from the non-orthogonal perspective of the human eye, and does not focus on road surface diseases, which increases the difficulty of road disease identification. Aiming at the current difficulty of manual detection of pavement cracks, this paper proposes an automated pavement crack identification method. Using Baidu Street View data, using Deeplabv3+ network model, by adjusting the parameters in the network structure, segmenting road cracks, exploring open data sources and open source convolutional neural network support, the practical application of road crack recognition is possible. The research significance of the thesis is that the deep learning neural network combined with the Internet street scene image can greatly improve the recognition accuracy and efficiency of pavement cracks, save the detection cost, and provide road managers with fast, accurate and efficient pavement crack information.

2. Road Crack Segmentation Algorithm

2.1. Algorithm Flow

The method of road crack segmentation in this paper is mainly based on DeepLabv3+ network. Firstly, the street view map with fixed angle was downloaded through baidu panorama static image API, and the downloaded image was cut and screened to make the data set of road crack image. LabelMe was used to make sample labels for road cracks, and the prepared training samples were divided into training set, test set and verification set. The training set is used to train the DeepLabv3+ network model. During the training, the pixel values of road crack and background data were counted and weighted according to the calculation results. When the loss value converged to a certain degree, the training was stopped. Then, the trained network model is used to segment the validation set to verify the accuracy of the training model and the segmentation results of the visual model of the test set. The overall flow chart is shown in figure 1.

![Fig.1 Flow chart](image)

DeepLabv3+ combines the advantages of the spatial pyramid module and the encoder-decoder structure [11] [12]. As shown in Figure 2, DeepLabv3 is used as the encoder architecture, and an effective decoder module is added for refinement on this basis. Segment the results and use dilated
convolution to control the resolution of the feature under the specified computing resources. In the decoder module, 1x1 convolution is used to reduce the number of channels from low-level features, and 3x3 convolution is used to gradually obtain the segmentation results. The overall architecture of Deeplabv3 + network for road crack identification in this paper is shown in Figure 3.

![Fig.2 Deeplabv3+ structure](image)

**2.2. Unbalanced Data Set**

In machine learning, unbalanced data set refers to the unbalanced distribution of samples of different categories, that is, samples of one category are far more than samples of another category [13]. Since Internet street view images are not collected for road cracks, the background is relatively complex, and the road cracks present a slender shape. The proportion of background and cracks is extremely unbalanced.

In order to solve the problem of data imbalance, this study weighted the categories of road cracks in the data set of Internet street view to reduce the impact of the unbalanced distribution of a few categories and improve the accuracy of network model identification. The specific method is to calculate the weight of the number of pixels in each category of the statistical training picture, and set
the weight coefficient of each category according to the calculation result, and multiply the corresponding category by the corresponding weight coefficient.

First, as shown in formula (1), the image pixels in this study are divided into two categories: background and crack, and the calculated pixel ratio of crack and background is calculated:

$$\frac{p_x(Bg)}{p_x(Cr)} = \frac{\sum_{i=0}^k p_x(i)}{\sum_{j=0}^m p_x(j)}$$

(1)

Where px represents the number of pixels, Bg and Cr represent the background and road cracks, respectively. In this study, the statistical result for Baidu Street View is 1/25. Next, modify the weight coefficient of loss, as shown in formula (2), where weights represents the weight of the overall category, px (Cr) represents the weight of the background, and px (Bg) cracks.

$$weights = Bg \times px(Cr) + Cr \times px(Bg)$$

(2)

3. Experiment and Analysis

3.1. Dataset Production
Since there is no publicly available street view image road crack data set, the image downloaded by Baidu Street View Map is used to manually interpret the data set.

In order to collect more information about road cracks, the downloaded street view image is the maximum downloaded image with a size of 1024x512. Each baidu street view image contains information such as location point unique identifier, latitude and longitude, horizontal angle and vertical angle of view. First, based on the geographic coordinates of the street view data sampling points and the line of sight (including horizontal perspective, vertical perspective and horizontal range), the street view image of each sample point can be obtained. In the study, baidu street view API is called in the form of HTTP URL to obtain massive street view data. The relevant parameter Settings of street view image downloading are shown in table 1. The experiment shows that the street view image downloaded from the horizontal angle of 0 ° and the vertical Angle of 40 ° is the perspective with the highest proportion of road information downloaded from the street view image. Figure 4 shows the street view image downloaded.

| Table 1. The camera parameters |
|-----------------------------|
| parameters | width | height | heading | pitch | fov |
| value       | 1024  | 512    | 0       | 40    | 90  |

![Fig.4 Download the street view image](image)
In order to facilitate the network to read data and improve the processing speed, this study used the segmentation raster tool in Arcmap software to segment the image, and the size of the image after segmentation was 512x512. Then, the image annotation software LabelMe was used to mark the road cracks. The original image and the annotation image were unified as 512x512x3. The format of the annotated image is 24bit. Due to the 8bit format of the label map required by DeeplabV3+ network, the program converts the final image into a grayscale image containing only the pixel values of 0 and 1. This is shown in figure 5 below.

3.2. Experiment and Analysis

In this study, a data set containing 400 baidu street view was used for network model training. The initial learning rate was 0.01 and the learning momentum parameter was 0.9. The global training step length was set to 15,000 steps. During the training, the current training model was saved once for every 150 steps, and then tested once. After Deeplabv3+ network training, the indicators have reached a good level.

The training results of the experimental network are shown in figure 6 below, which shows that the loss value changes with the number of iterations, and the loss value gradually becomes stable with the increase of step size. After the training of 15,000 step length, the loss value has converged to around 0.2, and the visualization effect is shown in figure 6. The change of MIoU in the experiment with the increase of step size is shown in figure 7. It can be seen that the value of MIoU gradually increases. When the training step size is 15000, the value of MIoU exceeds 70%.
To verify the effectiveness of the network model, the DeepLabv3+ model and FCN network training model to carry out the contrast experiment using the same road cracks tag data set as the training data. The comparison results of the algorithm are shown in Table 2. In terms of the time taken by MIOU and a single test, Deeplabv3+ network parameter model trained after adjusting parameters in this paper is superior to FCN.

| Algorithm               | MIOU | Time (single test time) |
|-------------------------|------|-------------------------|
| FCN                     | 65%  | 3s                      |
| The model in this paper | 72%  | 1s                      |

Figure 8 for training the model after the segmentation result, street view can see baidu's background is more complex, including trees, sky, lane line, part of the baidu cars, pedestrians and vehicles, etc.. The predict images using the network model of the training division of road cracks can segment the main part of the road cracks and close to the human eye recognition effect.

4. Conclusion
In this paper, a new joint recognition method for road cracks based on the combination of open network datasets and open source convolutional neural network is proposed for the difficulties of road crack identification. The conclusions are obtained as below:

(1) This method uses baidu street view map as the data set and adopts the combination method of open network data set and open source convolutional neural network, which is low in cost and high in efficiency and can solve the problem of large difficulty in data source acquisition and artificial identification. It has practical application value in road maintenance.
(2) Based on the semantic segmentation of the network Deeplabv3+, it is verified that the Deeplabv3+ network model can solve the data set imbalance and complexity in the data set with complex and unbalanced background information.

(3) Baidu street view map was used as the data set to display the models of 400 data sets after 15,000 steps, and the road crack segmentation was carried out. After the assessment of the evaluation criteria, Mean Intersection over Union (MIOU) reached 72%.

Acknowledgments
This work was financially Supported by the National Key Research and Development Program of China (No. 2018YFC0706003); Funded by Beijing Key Laboratory of Urban Spatial Information Engineering (NO. 2020216).

References
[1] Seongdeok Bang, Somin Park, Hongjo Kim, Hyoungkwan Kim. Encoder–decoder network for pixel-level road crack detection in black-box images[J]. Computer-Aided Civil and Infrastructure Engineering, 2019, 34(8).
[2] Oliveira H, Correia P L. Supervised crack detection and classification in images of road pavement flexible surfaces[M]// Recent advances in signal processing. In Tech, 2009.
[3] LONG Y, LIU L. How Green are the Streets. An Analysis for Central Areas of Chinese Cities Using Tencent Street View[J]. PIOS One, 2017, 12(2): e0171110.
[4] Campbell Andrew, Both Alan, Sun Qian (Chayn). Detecting and mapping traffic signs from Google Street View images using deep learning and GIS[J]. Campbell Andrew; Both Alan; Sun Qian (Chayn), 2019, 77.
[5] Rundle A G, Bader M D M, Richards C A, et al. Using google street view to audit neighborhood environments[J]. American Journal of Preventive Medicine, 2011, 40(1): 94-100.
[6] Ponzi L, Rota Bulo S, Epri B, et al. Predicting and understanding urban perception with convolutional neural networks[C]. Proceedings of the 23nd ACM international conference on Multimedia. ACM, 2015: 139-148.
[7] Glaeser E L, Kominers S D, Luca M, et al. Big data and big cities: The promises and limitations of improved measures of urban life[J]. Economic Inquiry, 2018, 56(1): 114-137.
[8] Cha Y J, Choi W, Buyukozturk O. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks[J]. Computer-aided Civil & Infrastructure Engineering, 2017, 32(5): 361-378.
[9] Zhang L, Yang F, Zhang Y D, et al. Road crack detection using deep convolutional neural network[C]// IEEE International Conference on Image Processing. IEEE, 2016.
[10] ZHANG A, WANG K C P, Li Bao-xian, et al. Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network[J]. Computer-aided Civil and Infrastructure Engineering, 2017, 32(10): 805-819.
[11] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 801-818.
[12] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. arXiv:1606.00915v2.
[13] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (12): 2481-2495.