Road Crack Detection Using Deep Neural Network with Receptive Field Block

Jing Yang, Qin Fu and Mingxin Nie

School of Wuhan University of Technology, Wuhan, China

*Corresponding author e-mail: 624968304@qq.com, 1468158086@qq.com, fuqin@whut.edu.cn

Abstract. Cracks are common pavement diseases that affect pavement performance. To maintain the road in good condition, localizing and fixing the cracks is a vital responsibility for transportation maintenance department. However, traditional manual detection methods are considerably tedious and require domain expertise. Therefore, the research on automatic detection and identification of pavement crack is of great significance for ensuring traffic safety and pavement maintenance decisions. In this paper, we propose an automatic pavement crack detection network based on the Single Shot MultiBox Detector (SSD) deep learning framework, and introduce the receptive field module to enhance the feature extraction capability of the network, which ensures real-time crack detection and also improves the performance of accuracy in pavement crack detection.

1. Introduction

With the rapid economic development and the improvement of people's living standards, the carrying capacity of highways has greatly increased. Once various diseases occur on the road, the safety of the road surface will decrease and the road traffic safety will be affected. Therefore, the road surface cracks need to be detected in real time. Traditional road crack detection is mostly performed manually. However, as the road network and traffic flow continue to increase, this time-consuming method is no longer practical, and manual detection is also seriously affected by the subjectivity of decision-making.

In contrast, semi-automated or automated pavement inspection systems rigorously designed can quickly detect pavement cracks with high accuracy and eliminate subjective effects. At present, pavement crack detection is still a very challenging task. In recent years, people have become more and more interested in this field. Many deep learning algorithms have been applied to the research of automatic detection of pavement cracks. In this paper, we use deep learning-based methods to detect pavement cracks, which will help people to perform automated pavement crack detection without manual execution.

Several researchers in the past have proposed their work on detecting road cracks using computer vision technology. Currently, most pavement crack detection methods are based on manual features and patch-based detection. Many types of these features have been used for crack detection. For example, a road surface crack detection method based on the Gabor filter is proposed in [2], while [5] proposed a pavement crack detection operator based on local binary mode (LBP). Liu et al. [1] and Kaseko et al. [3] make the assumption that real crack pixels are always darker than around the crack, and use a...
threshold-based method to detect pavement crack areas. These methods use local patterns for coding, but lack a global view of cracks. In order to perform crack detection from a global perspective, some studies [4, 12, 13] combined the photometric and geometric features of pavement crack images to automatically detect pavement cracks. These methods eliminate noise to a certain extent and enhance the continuity of crack detection. But its detection performance is not superior, mainly due to the lack of robust feature representation.

In order to overcome the above shortcomings, [7] proposed a method for automatic detection of pavement cracks based on crack trees; crackForest [6] used structural information in crack patches to fuse complementary features from multiple levels to characterize cracks. However, these methods are based on manual features for crack detection, and their recognition ability is insufficient to distinguish cracks from complex backgrounds and low-level clues.

Recently, deep learning has been widely used in the field of computer vision with its good expression ability. Zhang et al. Developed a crack detection method that used convolutional neural networks to learn discriminative features from original image blocks. Similarly, [8] proposed a deep CNN structure to detect concrete cracks. A complete crack detection and characterization model is proposed in [10], and a complete image processing algorithm for detecting and characterizing pavement crack damage is described in [11]. In this paper, an model based on SSD network is proposed to automatically detect and identify pavement cracks according to the effectiveness of deep learning in the field of object detection, and a receptive field module is introduced to enhance the network's feature extraction capabilities to improve the accuracy of pavement crack detection.

2. Methodology

2.1. SSD

The SSD algorithm is one of object detection algorithms [14] that directly predicts the coordinates and categories of the bounding box. The entire network of SSD adopts the idea of one stage to improve the detection speed. The network extracts features hierarchically and calculates bounding box regression and classification operations in order, so that it can adapt to training and detection tasks of multiple scale targets.

In this paper, the SSD network is divided into 6 stages. Each stage can learn a feature map, and then perform bounding box regression and classification. The backbone of the SSD uses the basic network structure of VGG-16. The first five layers of the convolutional network in VGG16 are used as the first stage. Then the two fully connected layers fc6 and fc7 are converted into two convolutions layers conv6 and conv7, and serve as the second and third stages of the network. Based on this, the SSD network adds four layers of conv8, conv9, conv10, and conv11 at the end to extract higher-level semantic information. In each stage, the default boxes are generated according to the size of the feature map and according to the fixed scale and radio. In each feature map cell, we predict the offset from the default box shape in the cell and the scores for each class. Specifically, for each square other than k at a given position, we compute a class c score and 4 offsets from the original default square shape, which will result in a \((c+4)\times k\times m\times n\) output of the m*n feature map.

The overall object loss function is the weighted sum of the confidence loss(conf) and the positioning loss (loc):

\[
L(x, c, l, g) = \frac{1}{N} (L_{conf} (x, c) + \alpha L_{loc} (x, l, g))
\]

Where N is the default number of matching boxes. If N = 0, we set the loss to 0. The positioning loss is a smooth L1 loss between the prediction box(l) and the ground truth box(g) parameters. We regress the center(cx, cy) of the default bounding box(d) and its width(w) and height(h) offsets.

The definition of the positioning loss function is as follows:
Confidence loss refers to the softmax loss(c) of the confidence between multiple classes. The definition of the confidence loss function is as follows:

$$L_{conf}(x,c) = - \sum_{i \in Pos} x_i^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0)$$  \hspace{1cm} (3)$$

In formula (3), $\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_{p} \exp(c_i^p)}$, and the sum weight $\alpha$ is set to 1 through cross-validation.

Suppose we use $m$ feature maps to make predictions. For each feature map, the scale of its default box is calculated according to the following formula:

$$s_k = s_{min} + \frac{s_{max} - s_{min}}{m-1} (k-1), k \in [1,m]$$  \hspace{1cm} (4)$$

As for the aspect ratio, use $a_r \in \{1, 2, 3, \frac{1}{2}, \frac{1}{3}\}$. So we can calculate the width $w_k^a = s_k \sqrt{a_r}$ and height $h_k^a = s_k / \sqrt{a_r}$ of each default box. In addition, when the aspect ratio is 1, we also add a scale default box:

$$s_k^1 = \sqrt{s_k \cdot s_{k+1}}$$  \hspace{1cm} (5)$$

2.2. Receptive Field Block (RFB)

Inspired by the structure of Receptive Fields (RFs) in human visual systems, a novel RF Block (RFB) module is proposed in [9]. RFB is a multi-branch convolution block. Its internal structure can be divided into two parts: a multi-branch convolutional layer with different kernels and a tail expansion pooling or convolutional layer. The former part simulates PRFs of various sizes and the latter part reproduces the relationship between PRF size and eccentricity in the human visual system.

Specifically, first, we use a bottleneck structure in each branch, using a 1*1 convolution layer to reduce the number of channels in the feature map, followed by an n*n convolution layer. In order to reduce the amount of parameters in the network, we replaced the 5*5 convolution layer with two stacked 3*3 convolution layers, and replaced the original n*n convolution layer with a 1*n and a n*1 convolution layer. Finally, we applied the shortcut design from ResNet.

The design of the dilated pooling or convolutional layer is to generate higher resolution feature maps and obtain a larger receptive field with the same amount of calculation. Figure 1 shows a receptive field block structure composed of a multi-branch convolutional layer and a dilated convolutional layer. At each branch, the convolutional layer of a specific kernel size is followed by the corresponding dilated pooling or convolutional layer, and finally the feature maps of all branches are fused to achieve the effect of the receptive field in the visual cortex.
2.3. Model and Dataset

The model is mainly an SSD network structure based on crack detection. The SSD network framework is mainly divided into three parts, including the basic network (we use the VGG-16 network as the basis), the multi-scale feature maps for detection, and the convolutional predictors for detection. In this paper, we embed a receptive field block in the SSD network to improve the features extracted from the lightweight backbone network. The overall network architecture is shown in Figure 2.

On the original SSD, the basic network structure is followed by a series of waterfall-like convolutional layers to produce a feature map that decreases in spatial resolution and increases in receptive field. We retain the same waterfall structure as SSD, but replace the convolutional layer on the large-resolution feature map with an RFB module. In addition, we remove all dropout layers and fc8 layer. The RFB module is used as a prototype for simulating the effect of eccentricity. Then, considering the proportion of pRF and the ratio of eccentricity in different visual maps, RFB-s is used to simulate the shallow pRFs in the human retina and placed in the conv4_3 convolution layer. The last few convolutional layers are retained because the feature map is too small to be processed with a 5 *5 kernel.

In this paper, stochastic gradient descent (SGD) is used to optimize the model, and the learning rate update strategy adopts the exponential decay method. In the initial stage of training, a large learning rate is set to quickly reach the vicinity of the optimal solution, and then the learning rate is gradually reduced to avoid fierce oscillations due to the large learning rate.

The training of the network model is based on a total of 4,000 pictures in different complex weather, different time periods and other environments. In addition, it contains 4 types of data sets with a total of 1200 pictures with different damage conditions for fine-tuning learning parameters, of which 800 pictures are used as training sets and 400 pictures are used as test sets. According to the pavement diseases, the pre-treated disease pictures are artificially divided into cracks (longitudinal cracks and transverse cracks), construction joint part, loose deformations (pits, ruts and wave surge package) and other corruption (crosswalk blur and white line blur), with D00, D01, D22, and D44 as the corresponding category labels.
3. Experimental Results

In our experiments, model training and testing are done under the PyTorch framework. Hardware environment: Intel(R) Xeon(R) CPU E5-1620 v4 @ 3.50GH, TITAN X (Pascal) GPU. Software environment: CUDA Toolkit 9.0.176, CUDNN V7.0, Python 3.6.3, PyTorch 0.4.0, Ubuntu 16.04.5, 64-bit operating system.

Model training and testing are accelerated by GPU. Taking the hardware performance and training time into consideration, the number of batch samples for testing and training is set to 64. The total number of training epochs is 250. The initial learning rate for our model is 0.01. Besides, we utilize a weight decay of 0.0005 and a momentum of 0.9.

For evaluating crack detections, a prediction box that captures over 50% Intersection over Union (IoU) in the area with the ground truth box is termed a successful match. In our study, evaluation metric focused on comparing the ground truth Mean F-1 score as shown in equation (6). We compute $n_{tp}$, $n_{fp}$, $n_{fn}$ and $n_{tn}$, which represent the number of testing images that are true positive, false positive, false negative and true negative, respectively.

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

The precision and recall can be computed using the following equations:

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

The partial test results of the model are shown in the Figure 3 below. The results displayed in the upper-left corner of the bounding box indicate the road damage category and category score (probability values belonging to this category). It can be seen from results in the following images that the model in this paper can automatically extract pavement features, has excellent crack detection performance and good classification performance.

![Figure 3. Partial test results of the model.](image-url)
The precision, recall and F1 score are calculated as shown in Table 1. From Table 1, it can be observed that the F1 value of the crack category (D00) is the highest among the four categories. However, D01 and D44 obtain lower F1 values than D00 and D22. The cause of this result may be that most of the pavement pictures in our dataset are cracks-type pavement images, while insufficient number of construction joint part and other corruption type road images in the training database. Another reason may be that under complex conditions, construction joint parts are easily misclassified at low resolution. The main reason for the inaccurate crack detection is that the cracks are not labeled correctly in the image, and the discrimination between cracks and the background image and shadow formation is low.

| Crack Class Name | Precision (%) | Recall (%) | F1 (%) |
|------------------|---------------|------------|--------|
| D00              | 89.53         | 86.23      | 87.85  |
| D01              | 77.75         | 69.89      | 73.61  |
| D22              | 83.00         | 85.75      | 84.35  |
| D44              | 64.50         | 74.75      | 69.25  |

4. Conclusion
In this paper, an SSD deep learning framework network embedded with the receptive field module is applied to the automatic detection and classification of pavement cracks. The F1 score for classification includes crack class was 0.7787. The model shows good performance in pavement crack detection and classification. It was observed that the distress analyzer so developed is more accurate in evaluating cracks but struggles with other corruption. This may be caused by the insufficient number of other corruption images in the training database.

At present, pavement crack detection models are limited to images with a simple background and a single type of damage. In reality, there may be many characteristics of damage types on the road surface, and applications in the field may require real-time pavement detection. In the future, we should further enrich the image dataset of pavement cracks. At the same time, in terms of optimization, we can make full use of multi-scale features and consider optimizing models from the aspect of feature fusion to further improve the model's recognition rate and practicability.

Acknowledgments
This work has been supported by Project of Hubei Provincial Highway Bureau under Grant Numbers 20141h0288.

References
[1] F. Liu, G. Xu, Y. Yang, X. Niu, and Y. Pan, “Novel approach to pavement cracking automatic detection based on segment extending,” in International Symposium on Knowledge Acquisition and Modeling, 2008, pp. 610–614.
[2] M. Salman, S. Mathavan, K. Kamal, and M. Rahman, “Pavement crack detection using the gabor filter,” in Proceedings of IEEE International Conference on Intelligent Transportation Systems, Oct. 2013, pp. 2039–2044.
[3] M. S. Kaseko and S. G. Ritchie, “A neural network-based methodology for pavement crack detection and classification,” Transportation Research Part C: Emerging Technologies, vol. 1, no. 4, pp. 275–291, 1993.
[4] Q. Zou, Y. Cao, Q. Li, Q. Mao, and S. Wang, “Cracktree: Automatic crack detection from pavement images,” Pattern Recognition Letters, vol. 33, no. 3, pp. 227–238, 2012.
[5] Y. Hu and C. X. Zhao, “A novel LBP based methods for pavement crack detection,” Journal of pattern Recognition research, 5 (1), pp. 140–147, 2010.
[6] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, “Automatic road crack detection using random structured forests,” IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 12, pp. 3434–3445, 2016.
[7] Q. Zou, Y. Cao, Q. Li, Q. Mao and S. Wang, “CrackTree: Automatic crack detection from pavement images,” Pattern Recognition Letters, 33 (3), pp. 227-238, 2012.

[8] Y. J. Cha, W. Choi, and O. Büyüköztürk, “Deep learning - based crack damage detection using convolutional neural networks,” Computer - Aided Civil and Infrastructure Engineering, 32 (5), pp. 361-378, 2017.

[9] Liu S, Huang D, Wang Y. Receptive Field Block Net for Accurate and Fast Object Detection [J]. 2017.

[10] H. Oliveira and P. L. Correia, “Automatic road crack detection and characterization,” IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 1, pp. 155–168, 2013.

[11] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, “Automatic road crack detection using random structured forests,” IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 12, pp. 3434–3445, 2016.

[12] R. Amhaz, S. Chambon, J. Idier, and V. Baltazart, “Automatic crack detection on two-dimensional pavement images: An algorithm based on minimal path selection,” IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 10, pp. 2718–2729, 2016.

[13] K. Fernandes and L. Ciobanu, “Pavement pathologies classification using graph-based features,” in ICIP, 2014, pp. 793–797.

[14] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector [J]. 2015.