Balanced Learning for Road Crack Segmentation

Subjects: Computer Science, Artificial Intelligence

Contributor: Yi Zhang, Junfu Fan, Mengzhen Zhang, Zongwen Shi, Rufei Liu, Bing Guo

Road crack segmentation based on high-resolution images is an important task in road service maintenance. The undamaged road surface area is much larger than the damaged area on a highway. This imbalanced situation yields poor road crack segmentation performance for convolutional neural networks.

Keywords: road crack segmentation ; convolutional neural networks ; balanced deep learning ; adaptive loss

1. Introduction

Road crack segmentation has important application value in road maintenance, especially for highways. It is challenging for road management departments to quickly determine the large-scale technical conditions of asphalt road pavement because of the rapid increase in road network mileage [1]. Scholars proposed various traditional computer vision-based road crack segmentation methods before convolutional neural networks became mainstream. Traditional digital image processing (DIP) methods, such as the Canny edge detector [2], wavelet transform [3], and crack index [4], have been carried out for crack segmentation. However, these algorithms focus on a small group of pixels and lack an understanding of the image content. Notably, DIP-based methods are very sensitive to noise [5], such as shadows and stains, and are unable to meet the needs of large-scale road maintenance. In the past five years, methods for extracting road cracks via convolutional neural networks have gradually replaced traditional methods and have become mainstream approaches in scientific research. Deep learning methods can better preserve the geometry of detected crack patterns, and their predictions (in terms of the pixels belonging to a crack) have ultimately become more accurate than those of the threshold method [6]. Therefore, the DIP method is not discussed in the following sections.

With the development of deep learning technology, convolutional neural networks such as ResNet [7] and DeepLabV3 [8] have achieved great success in computer vision and have been naturally introduced to the road crack segmentation task. Due to their powerful ability to learn high-level feature representations, road crack segmentation approaches using supervised convolutional neural networks have achieved good results. However, most studies have only applied advanced computer vision methods to the field of road damage segmentation and have not deeply considered the differences between road cracks and natural images. Compared with the semantic segmentation task involving natural images, the road crack segmentation task has unique characteristics:

- Small targets: According to the latest ″Highway Technical Condition Evaluation Standard JTG 5210-2018″, cracks with average widths greater than 1 mm should be extracted. Many road crack instances are only two or three pixels wide, making them much smaller than natural images in datasets such as PASCAL Visual Object Class (VOC), and Microsoft Common Objects in Context (MS-COCO), even in very high-resolution images.

- Imbalance at the sample level: The positive and negative data are imbalanced at the sample level. For example, in the DeepCrack dataset [9], the ratio of positive and negative samples is approximately 1:33, and in the HRRC dataset, the ratio greatly increases to 1:705 at the pixel level. The randomness and sparsity of the positive class are even more apparent in the engineered HRRC dataset.

- Imbalance at the feature level: Driven from samples, imbalance also occurs at the feature level. Deep high-level features in backbones have more semantic meanings, while shallow low-level features are more descriptive in terms of content [10]. The gradient generated by a small number of positive features, especially at a deep level, may be submerged by the gradient generated by a large number of negative features, and the utilized model will have difficulty achieving enough positive characteristics.

2. Network Architectures

The fully convolutional network (FCN) [11] architecture transforms fully connected layers into convolution layers, enabling a classification network to output a heatmap. A convolutional layer is used after the pooling layer to convert the fully connected layers in VGG16 [12] and up-sample the predictions back to pixels in a single step. The FCN architecture builds ″fully convolutional″ networks that take inputs of an arbitrary size and produce correspondingly sized outputs with efficient inference and learning processes.

With an FCN architecture, Chun [13] proposed a fully convolutional neural network-based road surface damage detection approach with semisupervised learning. The model is updated by using pseudolabeled images derived from semisupervised learning methods to improve the performance of road surface damage detection techniques.

The U-Net [14] architecture achieves very good performance in very different biomedical segmentation applications and has become one of the most commonly used network structures. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. Using skip connections between the two parts, U-Net can mix shallow, low-level feature maps obtained from the encoder and deep, semantic features derived from the decoder. Benefiting from skip connections, the network can better retain low-dimensional features with high resolution to achieve good results in medical segmentation tasks with rich details, such as boundaries and seams.

With the U-Net architecture, Rodrigo [15] proposed a novel synthetic dataset and a weakly supervised learning method to overcome the inaccurate annotation problem and improved the results by up to 12%. Hong [16] proposed an improved identification technique based on the U-Net architecture that was enhanced with a convolutional block attention module, an improved encoder, and the strategy of fusing long- and short-skip connections. This method could effectively predict highway cracks in unmanned aerial vehicle (UAV) images.

UNet++ [17] alleviates the unknown network depth problem with an efficient ensemble of U-Nets with varying depths, which partially share an encoder and simultaneously perform co-learning using deep supervision. Furthermore, the skip connections are redesigned to aggregate features of varying semantic scales at the decoder. These network structures effectively preserve shallow features and are widely used in medical image segmentation.

With the UNet++ architecture, Yang [18] proposed a method composed of two stages, crack recognition and crack semantic segmentation, making it easy to meet the needs of efficient and reliable detection for large-scale collected images. The fine-tuned VGG16 model in the first stage can accurately identify crack images and avoid the computing costs incurred by the further processing of non-crack images, and the UNet++ model in the second stage provides pixel-level semantic segmentation for crack images.

The pyramid scene parsing network (PSPNet) [19] was proposed to embed difficult scenery context features in an FCN-based pixel prediction framework. The pyramid pooling module fuses features under four different pyramid scales, aggregates multilevel context to global context information, and provides global-scene-level information to the developed model. This architecture exploits the capability of global context information derived from different-region-based context aggregation through a pyramid pooling module and a suitable strategy to utilize global scene category clues.

A feature pyramid network (FPN) [20] naturally leverages the pyramidal shape of a ConvNet’s feature hierarchy while creating a feature pyramid that has strong semantics at all scales. An FPN relies on an architecture that combines low-resolution, semantically strong features with high-resolution, semantically weak features via a top-down pathway and lateral connections, thereby developing a top-down architecture to build high-level semantic feature maps at all scales.

With a similar network structure, DeepCrack [9] resizes, concatenates, and fuses these multiscale feature maps at all scales, and a guide filter [21] is used to perform denoising at the feature level before prediction. APLCNet [22] combines a channel attention mechanism and a spatial attention mechanism during FPN feature fusion, highlights the attribute information and location information of cracks, and achieves good results on road crack instance segmentation tasks.

DeepLabV3+ [23] improves the encoder–decoder structure of other approaches. The encoder module encodes multiscale contextual information by applying atrous convolution on multiple scales, while the simple and efficient decoder module refines the segmentation results along the target boundary. The Xception model is further explored for the segmentation task. Deep separable convolution is applied to the ASPP and decoder modules to produce faster and stronger encoder–decoder networks.

A high-resolution network (HRNet) [24] maintains high-resolution representations by connecting high-to-low resolution convolutions in parallel and strengthens high-resolution representations by repeatedly performing multiscale fusions across parallel convolutions. HRNet has achieved great success in key point detection, attitude estimation, and multiperson attitude estimation tasks and exhibits enormous potential for scientific research and applications.

Based on HRNet, Chen [25] proposed an enhanced version called HRNete by removing the down-sampling operation in the initial stage, reducing the number of high-resolution representation layers, using dilated convolution, and introducing hierarchical feature integration. Using HRNet as the backbone, Bai [26] proposed the robust mask R-CNN, an end-to-end deep neural network for crack detection and segmentation on structures or their components that may be damaged during extreme events, such as earthquakes.

3. Imbalance Problem

Libra R-CNN [27] utilizes a simple but effective framework for balanced learning. It integrates IOU-balanced sampling by splitting the sampling interval into K bins according to their IOU measures and uniformly selects samples from these bins. It obtains a balanced feature pyramid by resizing the multilevel features in an FPN, averaging them to obtain an integrate feature, making refinements with convolutions, and resizing the features to their original sizes. A balanced L1 loss is set by separating inliers from outliers and clipping the large gradients produced by outliers with a maximum value of 1.0.

A two-stage convolutional neural network [28] was proposed for road crack detection and segmentation in images at the pixel level. The first stage serves to remove noise or artifacts and isolate the potential cracks in a small area via a classification network that is composed of five-layer convolutional neural networks and two fully connected (FC) layers. The second stage is a U-Net-structured segmentation network that can learn the context of cracks in the detected area. In this two-step framework, the first network filters out pure negative sample images, and images containing positive samples are allowed to proceed to the second stage; this approach can reduce the proportion of negative samples.

Focal loss [29] addresses the class imbalance problem by reshaping the standard cross-entropy loss such that it down-weights the losses assigned to well-classified examples. Focal loss focuses training on a sparse set of hard examples and prevents the vast number of easy negatives from overwhelming the detector during training. Two preset hyperparameters are employed; one is used to reduce the loss contributions provided by easy examples and extend their range, and the other addresses class imbalance.

Dice loss [30] considers that predictions are strongly biased toward the background if the area of interest occupies only a very small region of an image, and a loss function based on the Dice coefficient (related to precision and recall) can ameliorate this situation. The F1 score considers both the precision and recall of classification models and can comprehensively describe the performance of a model. The Dice loss is defined as 1 minus the F1 score, which involves maximizing the value of the F1 score as the optimization condition.

RAO-UNet [31] proposed a foreground perception optimization method. This method calculates the ratio of the sum of the probabilities of all pixels predicted as the background to the sum of the probabilities of all pixels predicted as the foreground in the predicted probability map. This ratio is used as an adaptive parameter for weighted binary cross entropy (BCE) loss in each batch to improve the performance damage caused by data imbalance.

Generative adversarial networks (GANs) [32] can extend the data by generating virtual data to solve the data imbalance problem. This approach has been used to deal with the problem of unbalanced landslide remote sensing data [33]. In Ref. [34], crack detection was enhanced by a generative adversarial network. CrackGAN [35] uses the generator as a segmentation network and adds a new constraint, generative adversarial loss, to regularize the objective function, which makes the network always generate a crack-GT detection result. These studies demonstrate the effectiveness of GANs in combating sample imbalance.

The technique of randomly under-sampling the majority class (RUMC) [36] involves randomly selecting examples from the majority class and removing them for the training dataset. The majority-class instances are discarded at random until a balanced class distribution in the training set is reached.

4. Dataset Situation of Road Crake Segmentation

At present, existing road crack segmentation datasets, such as DeepCrack [9], CrackForest [37], and 2StageCrack [28], are used by researchers to conduct experiments. The DeepCrack dataset contains 537 images with dimensions of 544 × 384 pixels. The CrackForest dataset contains 118 labeled images of 544 × 384 pixels. The 2StageCrack dataset, the newest open-source dataset, contains 1276 images for training and 354 for testing, each with dimensions of 96 × 96 pixels.

5. The Proposed Method

The main idea of researchers' method is inspired by the second law of thermodynamics: heat can be spontaneously transferred from a hotter body to a cooler body. Researchers hope to establish a similar mechanism during the convolutional neural network training process to regard precision and recall as body temperature and make them naturally flow from high to low. Therefore, researchers need to solve three problems: how to determine the direction of the flow, how to determine the strength of the flow, and how to make the precision and recall flow spontaneously.

To overcome these problems, researchers' proposed method is as follows. (1) Researchers define an adaptive parameter called PRF, which is associated with precision and recall, to evaluate the gap between precision and recall at a defined interval. (2) The flow direction is determined according to the positive and negative values of the PRF, and the flow intensity is determined according to the absolute value of the PRF. (3) To bridge precision and recall at the sample level and feature level, the sampling method and loss weight are repeatedly adjusted during the whole training process. Finally, spontaneous flow and a dynamic balance between precision and recall are achieved to narrow the gap between them and obtain a better trained model. The overall design of the recurrent adaptive network framework is shown in **Figure 1**.

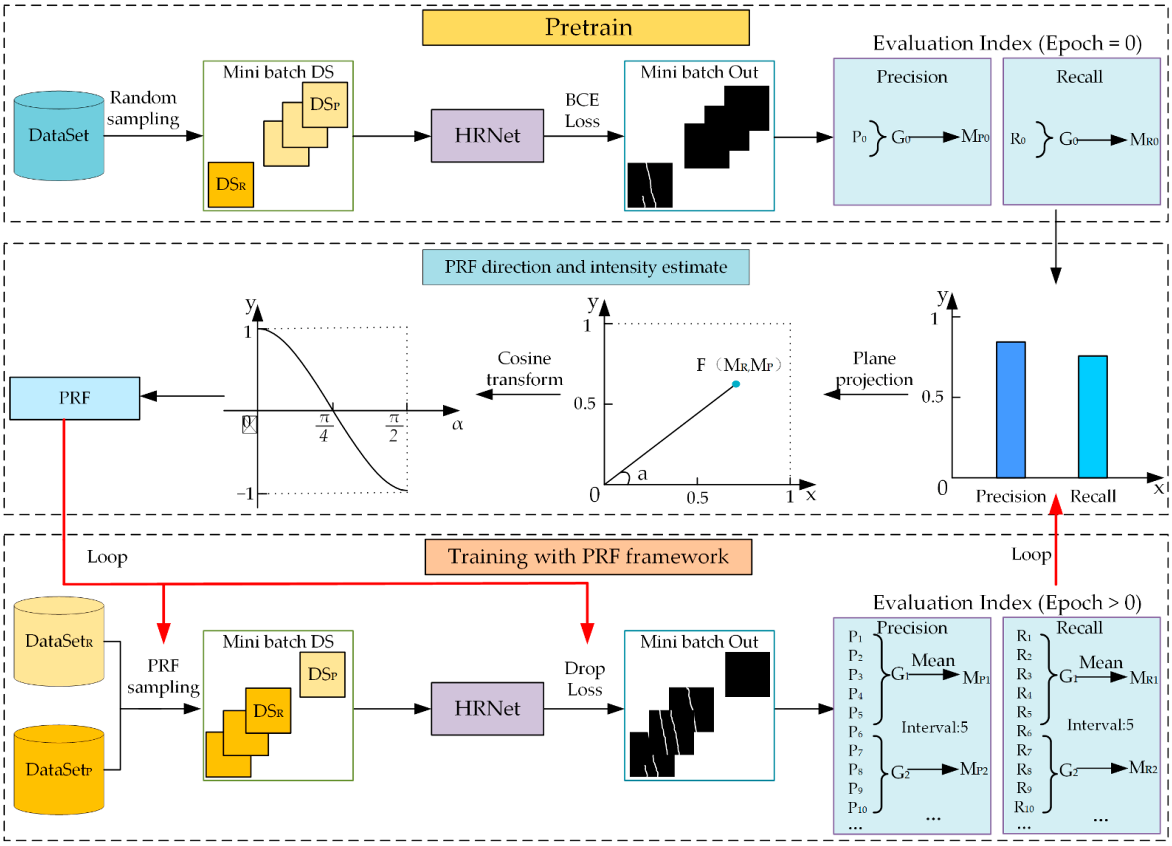

Figure 1. The overall design of the recurrent adaptive network framework.

References

1. Pan, Y.; Zhang, X.; Cervone, G.; Yang, L. Detection of Asphalt Pavement Potholes and Cracks Based on the Unmanned Aerial Vehicle Multispectral Imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2018, 11, 3701–3712.

2. Xu, H.; Tian, Y.; Lin, S.; Wang, S. Research of image segmentation algorithm applied to concrete bridge cracks. In Proceedings of the IEEE Third International Conference on Information Science and Technology (ICIST), Yangzhou, China, 23–25 March 2013; pp. 1637–1640.

3. Zhou, X.; Xu, L.; Wang, J. Road crack edge detection based on wavelet transform. IOP Conf. Series: Earth Environ. Sci. 2019, 237, 032132.

4. Abdellatif, M.; Peel, H.; Cohn, A.G.; Fuentes, R. Pavement Crack Detection from Hyperspectral Images Using a Novel Asphalt Crack Index. Remote. Sens. 2020, 12, 3084.
5. Guo, B.; Wei, C.; Yu, Y.; Liu, Y.; Li, J.; Meng, C.; Cai, Y. The Dominant Influencing Factors of Desertification Changes in the Source Region of Yellow River: Climate Change or Human Activity? Sci. Total Environ. 2022, 813, 152512.
6. Rezaie, A.; Achanta, R.; Godio, M.; Beyer, K. Comparison of crack segmentation using digital image correlation measurements and deep learning. Constr. Build. Mater. 2020, 261, 120474.
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
8. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. arXiv 2017, arXiv:1706.05587.
9. Liu, Y.; Yao, J.; Lu, X.; Xie, R.; Li, L. DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. Neurocomputing 2019, 338, 139–153.
10. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 818–833.
11. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 640–651.
12. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
13. Chun, C.; Ryu, S.-K. Road Surface Damage Detection Using Fully Convolutional Neural Networks and Semi-Supervised Learning. Sensors 2019, 19, 5501.
14. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
15. Rill-García, R.; Dokladalova, E.; Dokládal, P. Pixel-accurate road crack detection in presence of inaccurate annotations. Neurocomputing 2022, 480, 1–13.
16. Hong, Z.; Yang, F.; Pan, H.; Zhou, R.; Zhang, Y.; Han, Y.; Wang, J.; Yang, S.; Chen, P.; Tong, X.; et al. Highway Crack Segmentation from Unmanned Aerial Vehicle Images Using Deep Learning. IEEE Geosci. Remote Sens. Lett. 2021, 19, 6503405.
17. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. IEEE Trans. Med. Imaging 2020, 39, 1856–1867.
18. Yang, Q.; Ji, X. Automatic Pixel-Level Crack Detection for Civil Infrastructure Using Unet++ and Deep Transfer Learning. IEEE Sensors J. 2021, 21, 19165–19175.
19. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
20. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
21. He, K.; Sun, J.; Tang, X. Guided Image Filtering. IEEE Trans. Pattern Anal. Mach. Intell. 2013, 35, 1397–1409.
22. Zhang, Y.; Chen, B.; Wang, J.; Li, J.; Sun, X. APLCNet: Automatic Pixel-Level Crack Detection Network Based on Instance Segmentation. IEEE Access 2020, 8, 199159–199170.
23. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the Computer Vision—ECCV 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 833–851.
24. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep High-Resolution Representation Learning for Human Pose Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 5686–5696.
25. Chen, H.; Su, Y.; He, W. Automatic crack segmentation using deep high-resolution representation learning. Appl. Opt. 2021, 60, 6080.
26. Bai, Y.; Sezen, H.; Yilmaz, A. End-to-end Deep Learning Methods for Automated Damage Detection in Extreme Events at Various Scales. In Proceedings of the 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 6640–6647.

27. Pang, J.; Chen, K.; Shi, J.; Feng, H.; Ouyang, W.; Lin, D. Libra R-CNN: Towards Balanced Learning for Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 821–830.

28. Nguyen, N.H.T.; Perry, S.; Bone, D.; Le, H.T.; Nguyen, T.T. Two-stage convolutional neural network for road crack detection and segmentation. Expert Syst. Appl. 2021, 186, 115718.

29. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. 2020, 42, 318–327.

30. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.

31. Fan, L.; Zhao, H.; Li, Y.; Li, S.; Zhou, R.; Chu, W. RAO-UNet: A residual attention and octave UNet for road crack detection via balance loss. IET Intell. Transp. Syst. 2021, 16, 332–343.

32. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. Advances in Neural Information Processing Systems. arXiv 2014, arXiv:1406.2661.

33. Al-Najjar, H.A.H.; Pradhan, B.; Sarkar, R.; Beydoun, G.; Alamri, A. A New Integrated Approach for Landslide Data Balancing and Spatial Prediction Based on Generative Adversarial Networks (GAN). Remote Sens. 2021, 13, 4011.

34. Duan, L.; Geng, H.; Pang, J.; Zeng, J. Unsupervised Pixel-level Crack Detection Based on Generative Adversarial Network. In Proceedings of the 5th International Conference on Multimedia Systems and Signal Processing, Chengdu, China, 8–10 May 2020; pp. 6–10.

35. Zhang, K.; Zhang, Y.; Cheng, H.-D. CrackGAN: Pavement Crack Detection Using Partially Accurate Ground Truths Based on Generative Adversarial Learning. IEEE Trans. Intell. Transp. Syst. 2020, 22, 1306–1319.

36. Fiorentini, N.; Losa, M. Handling Imbalanced Data in Road Crash Severity Prediction by Machine Learning Algorithms. Infrastructures 2020, 5, 61.

37. Shi, Y.; Cui, L.; Qi, Z.; Meng, F.; Chen, Z. Automatic Road Crack Detection Using Random Structured Forests. IEEE Trans. Intell. Transp. Syst. 2016, 17, 3434–3445.