

Diffusion-Based Method for Pavement Crack Detection

Subjects: Computer Science, Artificial Intelligence

Contributor: Haoyuan Zhang, Ning Chen, Mei Li, Shanjun Mao

Pavement crack detection is of significant importance in ensuring road safety and smooth traffic flow. However, pavement cracks come in various shapes and forms which exhibit spatial continuity, and algorithms need to adapt to different types of cracks while preserving their continuity. Some studies have already applied the feature learning capability of generative models to crack detection.

Keywords: pavement crack detection ; diffusion model

1. Introduction

Pavement crack detection plays a crucial role in the maintenance and management of road infrastructure worldwide. Cracks on road surfaces are common indicators of pavement distress, often resulting from various factors such as traffic load, environmental conditions, and material deterioration. These cracks, if left unattended, can lead to a multitude of issues, including compromised road safety, increased risk of accidents, and hindered traffic flow. Traditional methods of crack detection primarily rely on manual visual inspection, where engineers or inspectors visually examine the surfaces of structures or materials to identify cracks and defects. This approach relies on human expertise, demanding significant time and labor efforts.

In recent years, numerous research efforts have been dedicated to developing automated pavement crack detection algorithms and systems, utilizing various imaging modalities such as visible light photography ^[1], infrared thermography ^[2], and LiDAR scanning ^[3]. These technologies offer the potential for faster, more accurate, and cost-effective crack detection compared to manual inspection methods. Methods utilizing visible light images for pavement crack detection offer the advantages of cost-effectiveness, enhanced detection speed, and automation. Image-based crack detection methods mainly include threshold segmentation ^[4], edge detection ^{[5][6][7]}, traditional machine learning ^{[8][9][10]}, and deep learning techniques ^{[11][12][13][14][15][16][17][18]}. Deep learning models, particularly convolutional neural networks (CNNs) ^[19], have exhibited significant promise in the field of crack detection because they can automatically learn image features, leading to highly accurate crack detection. Influenced by materials and environment, pavement cracks exhibit various shapes and features, but they typically present spatial continuity. However, limited by the receptive field of the kernel of CNNs, deep learning methods based on CNNs often produce fragmented and discontinuous detection results. There are mainly two reasons for producing such results: 1. Cracks exhibit continuity in space, while CNN-based methods are unable to learn this underlying spatial relationship between pixels. 2. Cracks exhibit significant scale differences between length and width dimensions. Moreover, annotating ground truth for cracks can be challenging as it is difficult for humans to accurately delineate crack contours, which results in a decline in CNN performance.

In recent years, generative models have garnered substantial popularity in the domain of deep learning. Deep generative models are essentially designed to seek and express the probability distribution of (multivariate) data in some way. These models, by capturing the data generation process, possess a certain level of robustness and probabilistic inference capability. They can handle missing and unlabeled data, learn high-level feature representations, and, as a result, find wide application in tasks such as visual object recognition, information retrieval, classification, and regression. Particularly, diffusion models ^[20] have emerged as a cutting-edge generative model, surpassing traditional models like Generative Adversarial Networks (GANs) ^[21] and Variational Autoencoders (VAEs) ^[22] in image generation tasks. Diffusion models learn through the processes of adding noise and sampling, and they possess two major advantages in feature representation. Firstly, diffusion models can predict pixel-level noise at each diffusion step, thereby characterizing the joint probability distribution of all pixels. This implies that they can learn broader spatial relationships, capturing more details and structural information during the image generation process. Secondly, the sampling process of diffusion models progresses from noise to image and from coarse to fine, showcasing their strong learning ability for both shallow and deep image features. This sampling process helps the model gradually understand the structure and content of images, leading to the generation of higher-quality images.

However, currently, diffusion models are primarily used in the field of image generation. Unlike discriminative models, which can easily compute the correlation between predicted results and ground truth, there are only a few generative tasks that have well-defined ground truth, such as image super-resolution [23]. Currently, there is no research on incorporating the learning capability of diffusion models into crack detection.

CrackDiff uses the framework of the diffusion denoising probability model (DDPM) [20], where the crack image is embedded and input into a multi-task UNet [24] network alongside the noise image. In this network, one branch is responsible for predicting the image's segmentation result, while the other branch predicts the noise at the current step, as illustrated in **Figure 1**. With this design, CrackDiff is capable of accurately learning the distribution and spatial relationships of pavement cracks. During the initial phases of the sampling process, CrackDiff can learn the approximate outline of the cracks. In the later stages of sampling, CrackDiff is capable of generating detailed contours of the cracks. Finally, it can generate highly confident segmentation results.

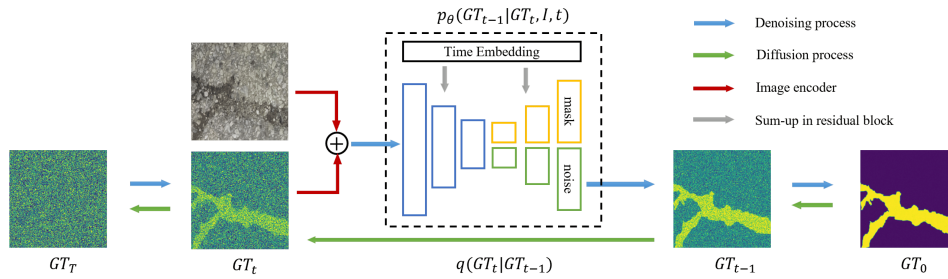


Figure 1. Overall framework of CrackDiff. The framework consists of a forward diffusion process and a backward denoising process. At any timestep t in the denoising process, a multi-task UNet network is employed for the simultaneous prediction of mask and noise. The network takes the sum of the embedding of the current estimation GT_t and the crack image as inputs and embeds timestep t into every residual block. Based on the predicted noise, the sampled result GT_{t-1} is obtained.

2. Diffusion-Based Method for Pavement Crack Detection

2.1. Pavement Crack Detection

Image-based pavement crack detection methods mainly include threshold segmentation, edge detection, traditional machine learning, and deep learning techniques. Threshold segmentation methods classify pixels in a given image into object and background classes based on pixel thresholds [4]. These methods are computationally efficient and fast, but they are highly sensitive to noise. The edge is the main feature of the cracks in the image. Edge detection methods use edge detection operators to identify these edges, such as Sobel, Canny, Prewitt, etc. [5][6][7]. However, the choice of parameters in these algorithms can significantly impact the detection of cracks, and finding the optimal parameters for different images, especially those with complex backgrounds, can be challenging. Traditional machine learning methods involve manually extracting crack features from images and then using techniques like support vector machines [8][9] or random forests [10] for feature classification. These methods offer better accuracy compared to traditional image processing techniques but rely on manually extracted crack features.

Deep learning is currently the most popular method. Semantic segmentation models are widely used in pavement crack detection because they can classify images pixel by pixel, allowing for quantitative measurement of road crack severity and density while simultaneously expressing the size and location of the targets. Various image segmentation models, including FCN [25], UNet [24], DeepLabv3+ [26], and SegNet [27], have been applied extensively in pavement crack detection research [11][12][13][14][15]. To capture the multi-scale spatial features of cracks, techniques such as spatial pyramids [16][28][29], dilated convolutions [30], and residual connections [31] have been incorporated into network designs. Additionally, attention mechanisms have been widely applied due to their advantages in representing long-range spatial dependencies and inter-channel correlations [17][32][33][34], especially transformer-based approaches [18][35][36][37][38].

Despite the integration of dilated convolutions, attention mechanisms, and various other techniques into CNN-based models, the performance of crack detection based on CNNs is still limited by the local receptive field. This limitation hinders the ability to identify long cracks and capture the entire image background, leading to discontinuous detection results and susceptibility to noise interference. Additionally, due to the difficulty in accurately labeling ground truth and data imbalance, the network may easily converge to treating all pixels as background.

Some studies have already applied the feature learning capability of generative models to crack detection. CrackGAN [39] proposed a crack-patch-only (CPO) supervised GAN network for generating crack-GT images. Deep convolutional GAN [40] is employed to generate a crack image dataset. Conditional GAN [41] is used to first extract the road and then detect cracks. However, due to the simultaneous training of both the generator and discriminator, GANs are challenging to balance, leading to training instability.

2.2. Diffusion Model

The diffusion model is a type of probabilistic generative model that can be divided into two main stages: the forward noise-adding stage and the backward denoising stage. In the forward noise stage, original images are progressively corrupted by adding Gaussian noise to the original data. In the backward denoising stage, the generative model's task is to learn the noise-adding in the forward process and recover the original input data from the noisy data. Currently, diffusion models can be categorized into three main types, namely Denoising Diffusion Probabilistic Models (DDPMs) [20], Score-based Generative Models [42], and Generative Models based on Stochastic Differential Equations [43]. CrackDiff is founded on the architecture of the DDPM.

The inspiration for the DDPM comes from non-equilibrium thermodynamics, and the training process involves two stages: the forward diffusion with noise process and the reverse denoising process. To accelerate the sampling speed, for traditional forward Markov processes, Denoising Diffusion Implicit Models (DDIMs) [44] has demonstrated that using non-Markovian processes can still yield good generative results and improve the sampling speed. The IDDDPM [45] reduced the number of sampling steps by adding cosine noise during the forward noise injection process and introducing learnable variance during the reverse denoising process. In terms of model structure, D2C [46] uses the idea of contrastive representation learning for the diffusion decoder model and improves the quality of representations through contrastive self-supervised learning. Peebles [47] replaced the commonly used UNet network for reverse image generation with transformer networks, achieving better training results. In the realm of conditional diffusion probabilistic models, the IDDDPM [45] incorporates class information as a conditional embedding in the timestamp embedding, generating images of a given class. SRDiff [48] combines diffusion generative models to make predictions for super-high-resolution images (SR) during the reverse process. It uses the low-resolution information encoded by an encoder as conditional noise to progressively denoise high-resolution images and generate super-resolution images. SegDiff [49] adds original image embeddings to the network and generates segmentation results for the original image.

Apart from the applications mentioned above, diffusion models have been successfully applied to various complex visual problems, such as image-to-image translation [50] and image blending [51], demonstrating the excellent capability of diffusion models in capturing latent patterns and relationships between samples.

References

1. Yang, F.; Zhang, L.; Yu, S.; Prokhorov, D.; Mei, X.; Ling, H. Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Trans. Intell. Transp. Syst.* 2019, 21, 1525–1535.
2. Yang, J.; Wang, W.; Lin, G.; Li, Q.; Sun, Y.; Sun, Y. Infrared Thermal Imaging-Based Crack Detection Using Deep Learning. *IEEE Access* 2019, 7, 182060–182077.
3. Li, Q.; Zhang, D.; Zou, Q.; Lin, H. 3D laser imaging and sparse points grouping for pavement crack detection. In *Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, 28 August–2 September 2017; pp. 2036–2040.
4. Kheradmandi, N.; Mehranfar, V. A critical review and comparative study on image segmentation-based techniques for pavement crack detection. *Constr. Build. Mater.* 2022, 321, 126162.
5. Zhao, H.; Qin, G.; Wang, X. Improvement of canny algorithm based on pavement edge detection. In *Proceedings of the 2010 3rd International Congress on Image and Signal Processing*, Yantai, China, 16–18 October 2010; Volume 2, pp. 964–967.
6. Hoang, N.D.; Nguyen, Q.L. Metaheuristic optimized edge detection for recognition of concrete wall cracks: A comparative study on the performances of roberts, prewitt, canny, and sobel algorithms. *Adv. Civ. Eng.* 2018, 2018, 7163580.
7. Zhang, Y.; Lu, Y.; Duan, Y.; Wei, D.; Zhu, X.; Zhang, B.; Pang, B. Robust surface crack detection with structure line guidance. *Int. J. Appl. Earth Obs. Geoinf.* 2023, 124, 103527.

8. Lin, J.; Liu, Y. Potholes detection based on SVM in the pavement distress image. In Proceedings of the International Symposium DCABES, Hong Kong, China, 10–12 August 2010; pp. 544–547.
9. O'Byrne, M.; Schoefs, F.; Ghosh, B.; Pakrashi, V. Texture analysis based damage detection of ageing infrastructural elements. *Comput.-Aided Civ. Infrastruct. Eng.* 2013, 28, 162–177.
10. 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.
11. David Jenkins, M.; Carr, T.A.; Iglesias, M.I.; Buggy, T.; Morison, G. A Deep Convolutional Neural Network for Semantic Pixel-Wise Segmentation of Road and Pavement Surface Cracks. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Roma, Italy, 3–7 September 2018; pp. 2120–2124.
12. 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.
13. Alipour, M.; Harris, D.K.; Miller, G.R. Robust pixel-level crack detection using deep fully convolutional neural networks. *J. Comput. Civ. Eng.* 2019, 33, 04019040.
14. Chen, T.; Cai, Z.; Zhao, X.; Chen, C.; Liang, X.; Zou, T.; Wang, P. Pavement crack detection and recognition using the architecture of segNet. *J. Ind. Inf. Integr.* 2020, 18, 100144.
15. Sun, X.; Xie, Y.; Jiang, L.; Cao, Y.; Liu, B. DMA-Net: DeepLab with multi-scale attention for pavement crack segmentation. *IEEE Trans. Intell. Transp. Syst.* 2022, 23, 18392–18403.
16. Ji, A.; Xue, X.; Wang, Y.; Luo, X.; Xue, W. An integrated approach to automatic pixel-level crack detection and quantification of asphalt pavement. *Autom. Constr.* 2020, 114, 103176.
17. Jiang, X.; Mao, S.; Li, M.; Liu, H.; Zhang, H.; Fang, S.; Yuan, M.; Zhang, C. MFPA-Net: An efficient deep learning network for automatic ground fissures extraction in UAV images of the coal mining area. *Int. J. Appl. Earth Obs. Geoinf.* 2022, 114, 103039.
18. Xiao, S.; Shang, K.; Lin, K.; Wu, Q.; Gu, H.; Zhang, Z. Pavement crack detection with hybrid-window attentive vision transformers. *Int. J. Appl. Earth Obs. Geoinf.* 2023, 116, 103172.
19. Ali, R.; Chuah, J.H.; Talip, M.S.A.; Mokhtar, N.; Shoaib, M.A. Structural crack detection using deep convolutional neural networks. *Autom. Constr.* 2022, 133, 103989.
20. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. In Proceedings of the 34th International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 6–12 December 2020.
21. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* 2018, 35, 53–65.
22. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* 2013, arXiv:1312.6114.
23. Han, L.; Zhao, Y.; Lv, H.; Zhang, Y.; Liu, H.; Bi, G.; Han, Q. Enhancing Remote Sensing Image Super-Resolution with Efficient Hybrid Conditional Diffusion Model. *Remote Sens.* 2023, 15, 3452.
24. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; pp. 234–241.
25. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39, 640–651.
26. 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: 15th European Conference, Berlin/Heidelberg, Germany, 8–14 September 2018; pp. 833–851.
27. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39, 2481–2495.
28. 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 (CVPR), Los Alamitos, CA, USA, 21–26 July 2017; pp. 6230–6239.
29. Ren, M.; Zhang, X.; Chen, X.; Zhou, B.; Feng, Z. YOLOv5s-M: A deep learning network model for road pavement damage detection from urban street-view imagery. *Int. J. Appl. Earth Obs. Geoinf.* 2023, 120, 103335.
30. Song, W.; Jia, G.; Zhu, H.; Jia, D.; Gao, L. Automated pavement crack damage detection using deep multiscale convolutional features. *J. Adv. Transp.* 2020, 2020, 6412562.

31. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2020, 43, 3349–3364.
32. Pan, Y.; Zhang, G.; Zhang, L. A spatial-channel hierarchical deep learning network for pixel-level automated crack detection. *Autom. Constr.* 2020, 119, 103357.
33. Cui, X.; Wang, Q.; Dai, J.; Xue, Y.; Duan, Y. Intelligent crack detection based on attention mechanism in convolution neural network. *Adv. Struct. Eng* 2021, 24, 1859–1868.
34. Zhu, W.; Zhang, H.; Eastwood, J.; Qi, X.; Jia, J.; Cao, Y. Concrete crack detection using lightweight attention feature fusion single shot multibox detector. *Knowl.-Based Syst.* 2023, 261, 110216.
35. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Proc. Adv. Neural Inf. Process. Syst.* 2021, 34, 12077–12090.
36. Liu, H.; Miao, X.; Mertz, C.; Xu, C.; Kong, H. Crackformer: Transformer network for fine-grained crack detection. In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 11–17 October 2021; pp. 3763–3772.
37. Xu, Z.; Guan, H.; Kang, J.; Lei, X.; Ma, L.; Yu, Y.; Chen, Y.; Li, J. Pavement crack detection from CCD images with a locally enhanced transformer network. *Int. J. Appl. Earth Obs. Geoinf.* 2022, 110, 102825.
38. Tao, H.; Liu, B.; Cui, J.; Zhang, H. A Convolutional-Transformer Network for Crack Segmentation with Boundary Awareness. In *Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP)*, Kuala Lumpur, Malaysia, 9–12 October 2023; pp. 86–90.
39. 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.* 2021, 22, 1306–1319.
40. Liu, Y.; Gao, W.; Zhao, T.; Wang, Z.; Wang, Z. A Rapid Bridge Crack Detection Method Based on Deep Learning. *Appl. Sci.* 2023, 13, 9878.
41. Kyslytsyna, A.; Xia, K.; Kislitsyn, A.; Abd El Kader, I.; Wu, Y. Road Surface Crack Detection Method Based on Conditional Generative Adversarial Networks. *Sensors* 2021, 21, 7405.
42. Song, Y.; Durkan, C.; Murray, I.; Ermon, S. Maximum likelihood training of score-based diffusion models. *Adv. Neural Inf. Process. Syst.* 2021, 34, 1415–1428.
43. Song, Y.; Sohl-Dickstein, J.; Kingma, D.P.; Kumar, A.; Ermon, S.; Poole, B. Score-based generative modeling through stochastic differential equations. *arXiv* 2020, arXiv:2011.13456.
44. Song, J.; Meng, C.; Ermon, S. Denoising diffusion implicit models. *arXiv* 2020, arXiv:2010.02502.
45. Nichol, A.Q.; Dhariwal, P. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning, Virtual*, 18–24 July 2021; Volume 139, pp. 8162–8171.
46. Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; Vaughan, J.W. D2C: Diffusion-decoding models for few-shot conditional generation. *Adv. Neural Inf. Process. Syst.* 2021, 34, 12533–12548.
47. Peebles, W.; Xie, S. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, 2–3 October 2023; pp. 4195–4205.
48. Li, H.; Yang, Y.; Chang, M.; Chen, S.; Feng, H.; Xu, Z.; Li, Q.; Chen, Y. Srdiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing* 2022, 479, 47–59.
49. Amit, T.; Shaharbany, T.; Nachmani, E.; Wolf, L. Segdiff: Image segmentation with diffusion probabilistic models. *arXiv* 2021, arXiv:2112.00390.
50. Saharia, C.; Chan, W.; Chang, H.; Lee, C.; Ho, J.; Salimans, T.; Fleet, D.; Norouzi, M. Palette: Image-to-image diffusion models. In *Proceedings of the ACM SIGGRAPH 2022 Conference Proceedings*, Vancouver, BC, Canada, 7–11 August 2022; pp. 1–10.
51. Meng, Q.; Shi, W.; Li, S.; Zhang, L. PanDiff: A Novel Pansharpening Method Based on Denoising Diffusion Probabilistic Model. *IEEE Trans. Geosci. Remote Sens.* 2023, 61, 5611317.