Deep Learning for Automated Crack Recognition: Insights, Challenges, and Future Directions

Ghodsiyeh Rostami[1], Po-Han Chen[1], Mahdi S. Hosseini[2]

[1]Department of Building, Civil, and Environmental Engineering, Concordia University, Canada
[2]Department of Computer Science and Software Engineering, Concordia University, Canada
rose.rostami@concordia.ca, pohan.chen@concordia.ca, mahdi.hosseini@concordia.ca

Abstract – Crack detection in concrete structures is a critical component of infrastructure health monitoring, as cracks pose significant risks to structural integrity. Traditional manual inspection methods, while effective, are resource-intensive, time-consuming, and prone to human error. Automated crack detection using vision-based techniques has become a focal point of research, aiming to enhance accuracy, efficiency, and consistency in inspections. Deep learning (DL) methods have revolutionized crack detection by enabling end-to-end systems that automatically learn from data and perform pixel-level segmentation. This study provides an overview of recent advances in deep learning-based crack recognition, highlighting the mainstream detection methods, and revealing ongoing challenges including the need for large and diverse datasets, the development of lightweight models for real-time performance, and model generalization across diverse conditions.

Keywords –
Crack Detection; Deep Learning; Computer Vision; Infrastructure Health Monitoring

1 Introduction

Civil infrastructure, such as buildings, bridges, and dams, develop defects over time due to material weathering, cyclic loads, environmental factors, and natural hazards. In concrete structures, cracks are the most common defect which can degrade structural integrity. Therefore, early detection of cracks is crucial to prevent catastrophic damages, making regular visual inspections critical for safety and timely maintenance planning. However, traditional manual crack inspections rely on specialized equipment, and skilled professionals, and often disrupt regular operations, making the process resource-intensive, costly, time-consuming, and prone to human error. To address these challenges, automated vision-based crack recognition methods have become a focal point of research, aiming to provide efficient, cost-effective, and consistent inspection methods.

Early efforts in vision-based crack recognition relied on digital image processing techniques, however, these techniques demonstrated limited efficacy in handling the complex and irregular geometries of cracks, as well as the varying illumination and background noise in real-world images. The introduction of machine learning (ML) provided a more adaptive approach, allowing algorithms to recognize patterns from image data and improve detection accuracy. Nevertheless, ML algorithms rely on manual feature engineering, making the process time-consuming and difficult to scale to more complex environments, which limits their broader application.

The advent of deep learning (DL), particularly Convolutional Neural Networks (CNNs), has revolutionized crack detection, enabling models to automatically learn features from data without manual intervention. CNN-based methods have shown superior performance in crack segmentation tasks, to identify cracks at the pixel level, overcoming the limitations of traditional image processing and ML algorithms. In addition, the introduction of Vision Transformers (ViTs) has enabled deep models to capture global and long-range dependencies, further enhancing the ability to detect cracks in complex backgrounds and under varying conditions. This has led to a shift in focus toward deep models capable of segmenting cracks with higher accuracy, robustness, and speed.

This study aims to provide an overview of the recent advancements in automated crack recognition, with a focus on deep learning-based methods that have revolutionized the field through their accuracy, scalability, and ability to handle complex patterns in diverse scenarios. Challenges introducing potential future directions for research and development are also explored and discussed.

2 Background

Over the past two decades, image-based crack detection has been a critical and active research area for achieving efficient and reliable automated inspections

[1,2]. Crack recognition, which involves identifying cracks in digital images or videos and quantifying their dimensions (i.e., length and width), is being extensively studied within the fields of image processing and computer vision. By definition, images are composed of pixels, which can be represented as numerical matrices, where the dimensions of the matrix are determined by the resolution of the image, and the values within the matrix represent the color or intensity of each pixel. This matrix representation enables the application of various mathematical and computational operations for tasks such as object detection and segmentation [3]. Crack detection involves highlighting pixels representing the crack, which can be analyzed at three levels: image, region, and pixel [4]. At the image level, images are classified based on whether they contain cracks (image classification) [5]. At the region level, cracks are identified and localized with bounding boxes (object detection). At the pixel level, segmentation identifies the exact boundary of cracks by highlighting pixels that represent the crack area (Figure 1) [1,6]. The goal of crack recognition is not only to detect cracks but also to measure their dimensions, where researchers explore various approaches to improve accuracy and robustness in crack segmentation. However, crack detection is challenging due to the irregular geometry and non-rigid topology of cracks, which vary in shape and size [7]. Additionally, images captured by inspection devices, such as UAVs, present diverse conditions including variations in illumination, noise, motion blur, and background complexity. For years, digital image processing techniques were explored for crack detection, however, due to their limitations remained impractical.

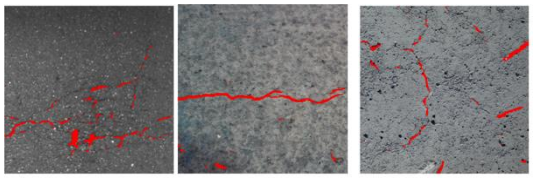

Figure 1. Examples of crack segmentation

The advent of learning-based and data-driven approaches has transformed the field, specifically Deep Learning (DL) become the mainstream approach for handling complex and high-dimensional data. The key innovation is the model's ability to automatically learn features from raw data and complete tasks without the need for manual feature engineering (Figure 2) [8]. In the field of computer vision, DNNs, especially Convolutional Neural Networks (CNNs) and Vision Transformers (ViT), have consistently outperformed traditional image processing techniques. This advancement has led to a paradigm shift in crack recognition research. The focus of many studies has been on deep models that offer faster and more generalized crack detection, where the model learns to extract features from inspection images autonomously and can identify cracks at the pixel level, a process known as semantic segmentation. Given the multi-scale nature of cracks, which can appear in different sizes within an image, and their non-rigid changing forms, several network architectures have been explored to address these challenges.

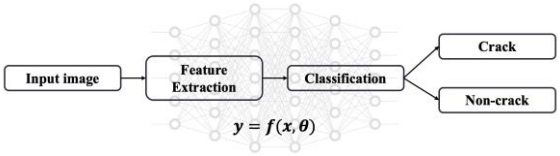

Figure 2. Representation learning by deep neural networks

3 Methodology

This review is based on a systematic content analysis of the existing literature on crack recognition. In the preliminary phase, relevant research papers were collected from databases, including Google Scholar, Scopus, ScienceDirect, and Web of Science, and using backward and forward snowballing techniques.

The keywords "concrete" AND "crack" AND ("segmentation" OR "detection"), "concrete" AND "crack" AND ("segmentation" OR "detection") AND ("measurement" OR "quantification"), and "deep" AND "crack" AND "detection" AND "quantif*" were used to search for the publications. The asterisk (*) wildcard was applied to include variations such as "quantify" and "quantification" in the search results.

During this phase, the collected papers were screened based on their abstracts, focusing on studies employing deep learning methodologies for crack detection. Only the most relevant studies were selected for an in-depth review. The analytical findings from this review were synthesized into a conceptual framework, as illustrated in Figure 3. Overview of the sections and subsections of this study . The remainder of this paper is structured according to the sections outlined in Figure 3. Overview of the sections and subsections of this study .

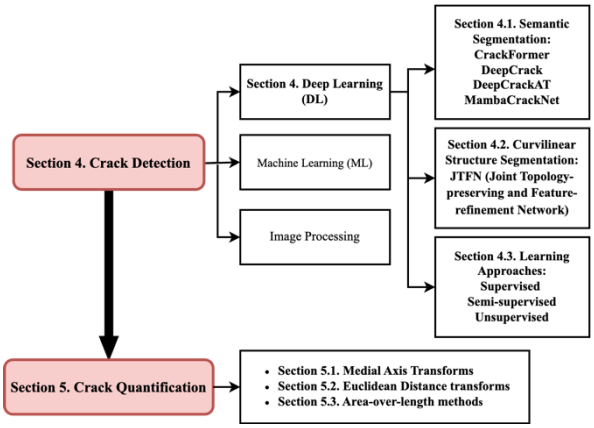

Figure 3. Overview of the sections and subsections of this study

4 Deep Learning-based Crack Detection

This section discusses various crack detection methods based on deep neural networks, with a summary of the selected studies provided in Table 1.

4.1 Semantic Segmentation

Convolutional Neural Networks (CNNs). The work by [9] introduced an early attempt at end-to-end crack segmentation using deep neural networks (DNNs), proposing a Fully Convolutional Network (FCN)-based semantic segmentation model trained on 500 concrete crack images of size 227×227 pixels. While the model outperformed Otsu's method with a 90% F1-score and precision in extracting crack paths, it failed to quantify crack dimensions due to noisy crack-like artifacts in the segmented images, a limitation stemming from the hierarchical processing of FCNs. In FCN, detailed information often from early layers is lost, which is crucial for crack detection as it requires both low-level and high-level features for fine details and semantic context, respectively. Additionally, the inductive bias of CNNs toward local feature extraction limits their ability to capture global contextual information, reducing their effectiveness in distinguishing cracks from complex backgrounds, such as material imprints.

To overcome these limitations, researchers have explored various mechanisms in designing crack segmentation models. These include dilated convolutions for enlarging the receptive field without losing resolution [10], attention mechanisms for focusing on relevant features [11–13], and Vision Transformers for capturing long-range dependencies in the data [14–19]. For instance, [20] presented a deep multi-scale crack feature learning model, named DeepCrackAT, based on an encoder-decoder architecture employing dilated convolution and attention mechanism. The model was trained on two datasets including 1440 images of masonry cracks and 2736 images of concrete and wall cracks, all resized to 224×224 pixels in size. The final model achieved 76.67% F1-score and 64.44% mIoU, nevertheless, it had limitations in segmenting cracks with significant variations in width.

Vision Transformers (ViT), introduced by [21] in 2020, employ self-attention mechanisms to capture global patterns and long-range dependencies. Demonstrating strong generalization capabilities, researchers have leveraged ViTs to enhance model robustness in new scenes and to improve crack detection in complex backgrounds [14]. For instance, [22] addressed the challenge of detecting fine-grained cracks by proposing a transformer-based segmentation network, CrackFormer. The model adapts SegNet architecture, introducing novel self-attention modules with 1×1 convolutional kernels to capture efficient contextual information across features channels, and efficient positional embeddings. New scaling-attention modules are also proposed to suppress non-semantic features and sharpen semantics when combining encoder-decoder outputs. Their comparative results demonstrated CrackFormer's superiority in multi-scale analysis of pavement crack images with higher performance and fewer floating-point operations (FLOPs).

Vision transformers, however, are dependent on large data and are computationally expensive. Therefore, researchers have proposed more efficient hybrid architectures combining convolutions and transformer operations. For instance, [18] proposed a dual-encoder network fusing Transformers and CNNs to capture local and global context information in crack inspection images with complex backgrounds. The model was trained on the Crack3238 dataset and achieved 76.44% F1-score and 78 FPS, which was higher than that of other networks in the comparative analysis.

Vision Mamba is a state-of-the-art high-performance computational framework that incorporates bidirectional state-space models for data-dependent global visual context modeling and position embeddings for location-aware visual recognition [23]. It is designed for efficient and precise data processing, offering scalability and flexibility for diverse applications. In a recent study, [24] proposed a hybrid architecture, MambaCrackNet, combining Mamba and CNN approaches to improve fine crack segmentation accuracy while reducing misdetections of background stains. Comparative experiments on two crack datasets demonstrated MambaCrackNet's superior performance over conventional segmentation models. Nevertheless, the model has a large number of learnable parameters due to the addition of two feature processing paths, leading to increased model complexity and slower inference speed.

Additionally, the Mamba module imposes restrictions related to input image size and the handling of extra spatial dimensions, which can limit the model's flexibility and adaptability in crack detection applications.

4.2 Curvilinear Structure Segmentation

Another approach in crack detection is curvilinear structure segmentation. These methods focus on identifying elongated, narrow structures such as cracks, blood vessels, roads, or rivers in images [25]. Preserving continuity, connectivity, and shape while minimizing false detections are the emphasis of these methods. Early approaches relied on edge detection and hand-crafted features, while modern techniques leverage deep learning models including CNNs and U-Nets, incorporating multi-scale feature fusion and global context modeling. Despite advances, challenges such as fragmentation, noise sensitivity, and scalability remain [26,27].

4.3 Learning Approaches

The studies on crack detection can also be categorized based on their learning approaches, including supervised, semi-supervised, and unsupervised learning methods. Supervised learning approaches have been widely studied in the literature. However, one of the challenges in crack detection is the lack of large-scale annotated datasets, hindering the full employment of deep learning models to achieve robust and generalized models. The most large-scale crack detection study is [28], which recently introduced OMNICRACK30K, a comprehensive collection of decentralized crack segmentation datasets, aimed at establishing a sustainable benchmark. Therefore, to address the challenge of large-scale labeled datasets, some studies investigated semi-supervised learning methods to leverage both small-scaled annotated data and largely available unlabelled datasets. [29] introduced a semi-supervised framework for crack segmentation on concrete and asphalt surfaces. The framework involves three stages: 1) supervised learning with different fractions of labeled data (2%, 10%, 14%, 20%, and 25%) using transfer learning, 2) consistency regularization for stability between original and perturbed images, and 3) self-training with pseudo-labeling and retraining the model from scratch on the expanded dataset for two times. According to the results, despite using only 2% of the data as labeled data, the model's performance (mIoU) was only 2.6% to 4.7% lower than using 100% of the labeled data, showing the effectiveness of the semi-supervised approach with limited annotations.

Furthermore, a novel representation learning framework was introduced by [30]. Addressing the ambiguity of marginal non-crack regions in crack segmentation, this study proposed a clustering-inspired representation learning framework (CIRL) consisting of two phases, where in phase 1, marginal non-crack regions are localized by introducing an Ambiguity-aware segmentation loss (Aseg Loss) to learn prediction variances. Next, to prevent the disturbance of marginal non-crack during model training, in the second phase, a Clustering-inspired loss (CI Loss) was proposed for ambiguous pixels based on the idea that features that are located close/farther in feature space should have consistent/inconsistent prediction. Their comparative analysis of their CIRL with other state-of-the-art crack segmentation models and loss functions showed the superiority of CIRL on four different datasets.

Table 1 Summary of the selected studies on crack detection

Category	Ref.	Model	Approach	Dataset	Performance	Advantages	Limitations
CNN-Based Approaches	[9]	FCN-based semantic segmentation	Supervised	500 concrete crack images (227×227)	90% F1-score	Early work on end-to-end crack segmentation, Outperforms Otsu's method	Fails to quantify crack dimensions, loss of crack details
	[20]	DeepCrackAT (Encoder-Decoder + Dilated Conv + Attention)	Supervised	1440 masonry, 2736 concrete/wall cracks (224×224)	76.67% F1-score, 64.44% mIoU	Captures multi-scale features	Struggles with variable crack widths

Vision Transformer-Based	[22]	CrackFormer (Transformer + SegNet)	Supervised	Two pavement crack datasets (CrackTree260, CrackLS315), one stone crack (Stone331)	Higher accuracy with lower FLOPs (88% average precision)	Multi-scale analysis, effective contextual learning	Computationally expensive
Hybrid Approaches	[18]	DTrC-Net with a Dual-encoder (Transformer + CNN)	Supervised	Crack3238	76.44% F1-score, 78 FPS	Balances local and global feature extraction	High computational cost
	[24]	MambaCrack Net (Mamba + CNN)	Supervised	CCSD (458 high-resolution concrete crack), CD (776 pavements/building walls cracks)	Higher performance than conventional models	Improved fine crack detection	High parameter count, slower inference
Curvilinear Structure Segmentation	[26]	JTFN (Joint Topology-preserving and Feature-refinement Network)	Supervised	concrete/pavement cracks (Crack500, CrackTree260,), 40 retina images (DRIVE), 1160 Road cracks	84% F1-score	Preserves connectivity and shape	Fragmentation, noise sensitivity
Learning Approaches	[29]	U-Net	Semi-supervised framework (transfer learning + pseudo-labeling)	Concrete & asphalt datasets (2%-25% labeled)	mIoU only 2.6%-4.7% lower than fully supervised	Effective with limited labels	Dependence on pseudo-labeling quality
	[30]	JTFN + CIRL	CIRL (Clustering-inspired representation learning)	Concrete/pavement cracks (CrackTree260, Crack500, CrackSeg5K), retina images (DRIVE)	Outperforms state-of-the-art methods on ambiguous crack pixels	Reduces marginal non-crack disturbances	Computationally intensive due to the clustering algorithm

5 Crack Analysis

As the literature review revealed, most studies have focused on enhancing crack detection models, while only a few have addressed methods for crack analysis (i.e., width measurement). This aspect is challenging, as high-quality crack segmentation is crucial for reliable and accurate measurement, especially considering variations in crack inspection images, such as scene complexity, background noise, and texture. Various methods have been proposed for crack quantification, including skeleton extraction techniques such as thinning algorithms, medial axis transforms, and distance transforms [31,32], as well as area-over-length methods [33] and edge detection approaches [34]. An overview of these methods is provided in Table 2. Relevant studies with respect to each approach are further discussed in the following sections.

Table 2 Overview of the crack quantification methods

Measurement Method	Relevant Studies	Advantages	Limitations
Medial Axis Transform (MAT)	[35]	- Provides a skeletonized representation of cracks, preserving	- Sensitive to noise, leading to fragmented or broken skeletons.

		topology. - Applicable to thin, continuous cracks.	- Over-simplifies complex crack structures, losing fine details. - Computationally intensive, especially when combined with deep learning.
Euclidean Distance	[36,37]	- Directly measures crack width by computing the shortest distance between two edges. - Works with both 2D images and 3D point clouds.	- Highly dependent on segmentation quality. Errors in crack edges lead to inaccurate width estimation. - Fails for irregular or branching cracks, as it assumes relatively parallel edges. - May overestimate or underestimate width if cracks have varying thicknesses.
Area-over-Length	[33,38]	- Robust against noise since it does not rely on edge detection. - Provides a simple, efficient way to estimate crack width. - Applicable to thin and elongated cracks.	- Inaccurate for irregular crack shapes, as it assumes uniform width. - Affected by small gaps or interruptions in cracks, leading to overestimation. - Struggles with very thick or non-linear cracks, where area may not properly reflect actual width.

5.1 Medial Axis Transform

To enhance microcrack detection in blurry and low-resolution images taken by UAVs, [35] proposed a three-stage crack recognition method by Super-Resolution Reconstruction (SRR), semantic segmentation, and an improved Medial Axis Transform (MAT) for crack width and length quantification. Their study showed the effectiveness of super-resolution reconstruction where reconstructed images from degraded images lead to similar segmentation output as with original high-resolution images. Nevertheless, the method is a two-stage process which cannot fulfill the real-time demand in practice.

5.2 Euclidean Distance

[36] proposed detection and quantification of global concrete cracks in panorama images using a three-stage process, including image classification of cropped panorama images, crack segmentation of positive outputs from the classification stage and crack quantization. For crack measurement, they used the coordinates of pixels to measure the Euclidean distance between two edges of cracks at different heights (lengths). To overcome the challenge of crack-like artifacts in the segmented images, they quantified cracks with a height of 50 pixels and above only. They calculated the measurements in real-world sizes by multiplying actual pixel sizes by the measurements in pixel value. The relative error of crack width measurements was 3.87%. Furthermore, [37] proposed a point cloud-based approach for analyzing cracks and spalls using semantic segmentation of raw point cloud data. The study developed a Normal Vector Enhanced Dynamic Graph CNN (NVE-DGCNN) to enable the direct analysis of the raw point cloud dataset without further conversions. For each identified defect, a minimum bounding box was generated, allowing for precise calculation of defect dimensions, including length, width, and depth. These measurements were determined through Euclidean distance calculations between the bounding box corners. However, achieving real-time crack recognition using point cloud data remains an open research challenge.

5.3 Area-over-Length

[38] proposed a two-stage crack detection method using a CNN classifier (pre-trained SqueezeNet), Deeplabv3+ to first classify crack and non-crack images, then segment cracks in positive images, and crack quantification using the regionprops function in MATLAB. The function calculates the width of each ellipse in the image by dividing its area over the length. Their results showed a lower average error of regionprops function in measuring crack width and length (7 and 2%, respectively) compared to a skeleton-based method such as medial-axis transform (11 and 17%). To leverage a trade-off between Transformers accuracy and computational cost, [33] developed a lightweight crack segmentation model using CNNs and Transformers named CrackTrNet to enable intelligent concrete dam crack patron inspection. Crack's length was

obtained using a thinning algorithm and width was measured by dividing crack's area over the length. According to the results, the model achieved similar performance compared to mainstream segmentation models with lower FLOPs and fewer parameters. 89.10% mIoU was achieved with 6.11 million parameters, which shows the feasibility of high segmentation performance with lightweight models. Yet, the study has limitations due to the non-diverse dataset, which consists of close-up images with relatively simple backgrounds and noise. Addressing these issues could enhance the model's generalization and robustness.

6 Discussion

The findings indicate that for crack detection, while ViT-based models achieve superior accuracy in controlled settings, they require large-scale datasets and substantial computational resources. For real-time applications, hybrid architectures (CNN-Transformers) offer a more effective balance between accuracy and efficiency, making them promising candidates for deployment. Additionally, optimization techniques such as pruning can enhance memory efficiency and reduce computational costs, further improving model feasibility in practical scenarios.

Moreover, accurate crack width measurement has not been achieved. This is due to limitations of the existing measurement algorithms in this task, as they are not specifically designed for crack measurement. For instance, techniques such as the medial axis transform, as discussed by [39], often generate redundant branches due to the irregular and non-rigid shape of cracks, which in turn distorts measurement accuracy. Furthermore, current crack analysis methods rely on perfectly accurate binary segmentation masks, which have not yet been achieved with the current crack detectors proposed in the literature. Imperfect masks, which may include noise, discontinuities in crack structures, or impurities that resemble cracks, affect the precision of the measurement process, making this an area in need of significant improvement.

7 Challenges, Opportunities, and Future Directions

Fully autonomous crack recognition remains an unsolved issue, as accurate width measurement is dependent on high-quality segmentation, specifically in complex and noisy environments. Therefore, the development of robust crack detection models is essential. One of the main challenges is the lack of large-scale and diverse datasets that can represent the wide range of real-world conditions in which cracks appear [40]. Additionally, the variability in crack structures makes it challenging to design models that can effectively capture the multi-scale and non-rigid structure of cracks. Future research can focus on developing advanced segmentation models and effective operations to enhance detection accuracy and improve generalization across diverse conditions.

Furthermore, in real-world applications, processing speed and computational efficiency are of paramount importance for real-time crack detection during infrastructure inspections or on-site assessments. One challenge is optimizing the trade-off between inference speed, accuracy, and computational cost of the detection models. Moreover, despite advancements in semantic segmentation, multi-stage measurement processes often impact real-time applicability [41]. Developing more robust, noise-resistant quantification methods while improving real-time efficiency can support large-scale, automated infrastructure assessment.

On the other hand, the literature has mainly focused on single-defect detection through binary segmentation whereas, in real-world scenarios, multiple defects (e.g., rust, cracks, and demolition) often coexist, requiring detailed assessments (i.e., area of the damaged surface structure). While some methods have been proposed [42,43], mainly for rough multi-defect localization using object detection and bounding boxes [44,45], future research can shift towards multi-defect segmentation to enable more precise and comprehensive assessment. A major challenge in this direction is the lack of multi-class annotated datasets for multi-defect segmentation. Addressing this limitation through the development of diverse, high-quality datasets and the adoption of multi-task learning frameworks could enhance the versatility of automated inspection systems.

Finally, from a technical standpoint, researchers utilize programming languages such as Python and MATLAB to address these challenges, along with deep learning frameworks including PyTorch or TensorFlow for model development [1]. Additionally, computer vision libraries such as OpenCV and scikit-image facilitate image processing, and automated annotation tools such as LabelMe or CVAT assist in preparing labeled datasets. For practical deployment, optimized frameworks such as TensorFlow Lite and ONNX Runtime enable efficient inference on edge devices. The advancement of these tools allows for the development of more accurate and efficient crack detection solutions.

8 Conclusion

The advancement of deep learning and computer vision technologies has significantly transformed automated crack detection, enabling more precise, and efficient visual crack assessment. This review provided an analysis of the state-of-the-art methodologies in crack

detection and quantification. Key developments, existing challenges, and potential future research directions were provided.

Through a systematic literature review, studies have been analyzed based on their computer vision task, learning approach, and dataset scale, and advantages and limitations were discussed. Furthermore, various crack measurement methods proposed in the literature were reviewed highlighting their effectiveness and shortcomings.

The review revealed that despite the accuracy of ViT-based models in controlled environments, their reliance on large-scale datasets and high computational power limits their practical deployment. Hybrid architectures, such as CNN-Transformers, offer an enhanced balance between accuracy and efficiency, especially when combined with optimization techniques such as pruning to reduce computational overhead. However, accurate crack width measurement remains an unresolved issue due to the limitations of existing algorithms and the reliance on imperfect segmentation masks. Addressing these challenges is crucial for advancing real-world applications.

References

[1] Ai D, Jiang G, Lam S-K, He P, Li C. Computer vision framework for crack detection of civil infrastructure—A review. Engineering Applications of Artificial Intelligence 2023;117:105478. https://doi.org/10.1016/j.engappai.2022.105478.

[2] Halder S, Afsari K. Robots in Inspection and Monitoring of Buildings and Infrastructure: A Systematic Review. Applied Sciences 2023;13:2304. https://doi.org/10.3390/app13042304.

[3] Gonzalez R, Woods R. Digital Image Processing 2017. https://elibrary.pearson.de/book/99.150005/9781292223070 (accessed October 10, 2024).

[4] Szeliski R. Computer Vision: Algorithms and Applications. Springer Nature; 2022.

[5] Prince SJD. Computer Vision: Models, Learning, and Inference. Cambridge University Press; 2012.

[6] Forsyth DA, Ponce J. Computer Vision: A Modern Approach. Prentice Hall Professional Technical Reference; 2002.

[7] Munawar HS, Hammad AWA, Haddad A, Soares CAP, Waller ST. Image-Based Crack Detection Methods: A Review. Infrastructures 2021;6:115. https://doi.org/10.3390/infrastructures6080115.

[8] Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning. MIT Press; 2016.

[9] Dung CV, Anh LD. Autonomous concrete crack detection using deep fully convolutional neural network. Automation in Construction 2019;99:52–8. https://doi.org/10.1016/j.autcon.2018.11.028.

[10] Ren Y, Huang J, Hong Z, Lu W, Yin J, Zou L, et al. Image-based concrete crack detection in tunnels using deep fully convolutional networks. Construction and Building Materials 2020;234:117367. https://doi.org/10.1016/j.conbuildmat.2019.117367.

[11] Deng L, Yuan H, Long L, Chun P, Chen W, Chu H. Cascade refinement extraction network with active boundary loss for segmentation of concrete cracks from high-resolution images. Automation in Construction 2024;162:105410. https://doi.org/10.1016/j.autcon.2024.105410.

[12] Liang J, Gu X, Jiang D, Zhang Q. CNN-based network with multi-scale context feature and attention mechanism for automatic pavement crack segmentation. Automation in Construction 2024;164:105482. https://doi.org/10.1016/j.autcon.2024.105482.

[13] Pan Y, Zhang G, Zhang L. A spatial-channel hierarchical deep learning network for pixel-level automated crack detection. Automation in Construction 2020;119:103357. https://doi.org/10.1016/j.autcon.2020.103357.

[14] Asadi Shamsabadi E, Xu C, Rao AS, Nguyen T, Ngo T, Dias-da-Costa D. Vision transformer-based autonomous crack detection on asphalt and concrete surfaces. Automation in Construction 2022;140:104316. https://doi.org/10.1016/j.autcon.2022.104316.

[15] Chen Z, Asadi Shamsabadi E, Jiang S, Shen L, Dias-da-Costa D. An average pooling designed Transformer for robust crack segmentation. Automation in Construction 2024;162:105367. https://doi.org/10.1016/j.autcon.2024.105367.

[16] Wang W, Su C. Automatic concrete crack segmentation model based on transformer. Automation in Construction 2022;139:104275. https://doi.org/10.1016/j.autcon.2022.104275.

[17] Wu Y, Li S, Zhang J, Li Y, Li Y, Zhang Y. Dual attention transformer network for pixel-level concrete crack segmentation considering camera placement. Automation in Construction 2024;157:105166. https://doi.org/10.1016/j.autcon.2023.105166.

[18] Xiang C, Guo J, Cao R, Deng L. A crack-segmentation algorithm fusing transformers and convolutional neural networks for complex detection scenarios. Automation in Construction 2023;152:104894. https://doi.org/10.1016/j.autcon.2023.104894.

[19] Zhang E, Shao L, Wang Y. Unifying transformer and convolution for dam crack detection. Automation in Construction 2023;147:104712. https://doi.org/10.1016/j.autcon.2022.104712.

[20] Lin Q, Li W, Zheng X, Fan H, Li Z. DeepCrackAT: An effective crack segmentation framework based on learning multi-scale crack features. Engineering Applications of Artificial Intelligence 2023;126:106876. https://doi.org/10.1016/j.engappai.2023.106876.

[21] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale 2021. https://doi.org/10.48550/arXiv.2010.11929.

[22] Liu H, Miao X, Mertz C, Xu C, Kong H. CrackFormer: Transformer Network for Fine-Grained Crack Detection. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, p. 3763–72. https://doi.org/10.1109/ICCV48922.2021.00376.

[23] Zhu L, Liao B, Zhang Q, Wang X, Liu W, Wang X. Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model 2024. https://doi.org/10.48550/arXiv.2401.09417.

[24] Han C, Yang H, Yang Y. Enhancing pixel-level crack segmentation with visual mamba and convolutional networks. Automation in Construction 2024;168:105770. https://doi.org/10.1016/j.autcon.2024.105770.

[25] Hu X, Li F, Samaras D, Chen C. Topology-Preserving Deep Image Segmentation. Advances in Neural Information Processing Systems, vol. 32, Curran Associates, Inc.; 2019.

[26] Cheng M, Zhao K, Guo X, Xu Y, Guo J. Joint Topology-preserving and Feature-refinement Network for Curvilinear Structure Segmentation. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada: IEEE; 2021, p. 7127–36. https://doi.org/10.1109/ICCV48922.2021.00706.

[27] Mosinska A, Marquez-Neila P, Kozinski M, Fua P. Beyond the Pixel-Wise Loss for Topology-Aware Delineation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, p. 3136–45. https://doi.org/10.1109/CVPR.2018.00331.

[28] Benz C, Rodehorst V. Omni-Crack30k: A Benchmark for Crack Segmentation and the Reasonable Effectiveness of Transfer Learning. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA: IEEE; 2024, p. 3876–86. https://doi.org/10.1109/CVPRW63382.2024.00392.

[29] Asadi Shamsabadi E, Erfani SMH, Xu C, Dias-da-Costa D. Efficient semi-supervised surface crack segmentation with small datasets based on consistency regularisation and pseudo-labelling. Automation in Construction 2024;158:105181. https://doi.org/10.1016/j.autcon.2023.105181.

[30] Chen Z, Lai Z, Chen J, Li J. Mind marginal non-crack regions: Clustering-inspired representation learning for crack segmentation. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA: IEEE; 2024, p. 12698–708. https://doi.org/10.1109/CVPR52733.2024.01207.

[31] Bae H, An Y-K. Computer vision-based statistical crack quantification for concrete structures. Measurement 2023;211:112632. https://doi.org/10.1016/j.measurement.2023.112632.

[32] Lee JS, Hwang SH, Choi IY, Choi Y. Estimation of crack width based on shape-sensitive kernels and semantic segmentation. Structural Control and Health Monitoring 2020;27:e2504. https://doi.org/10.1002/stc.2504.

[33] Li M, Yuan J, Ren Q, Luo Q, Fu J, Li Z. CNN-Transformer hybrid network for concrete dam crack patrol inspection. Automation in Construction 2024;163:105440. https://doi.org/10.1016/j.autcon.2024.105440.

[34] Ding W, Yang H, Yu K, Shu J. Crack detection and quantification for concrete structures using UAV and transformer. Automation in Construction 2023;152:104929. https://doi.org/10.1016/j.autcon.2023.104929.

[35] Xiang C, Wang W, Deng L, Shi P, Kong X. Crack detection algorithm for concrete structures based on super-resolution reconstruction and segmentation network. Automation in Construction 2022;140:104346. https://doi.org/10.1016/j.autcon.2022.104346.

[36] Song L, Sun H, Liu J, Yu Z, Cui C. Automatic segmentation and quantification of global cracks in concrete structures based on deep learning. Measurement 2022;199:111550. https://doi.org/10.1016/j.measurement.2022.111550.

[37] Bahreini F, Hammad A. Dynamic graph CNN based semantic segmentation of concrete defects and as-inspected modeling. Automation in Construction 2024;159:105282. https://doi.org/10.1016/j.autcon.2024.105282.

[38] Teng S, Chen G. Deep Convolution Neural Network-Based Crack Feature Extraction, Detection and Quantification. J Fail Anal and Preven 2022;22:1308–21. https://doi.org/10.1007/s11668-022-01430-9.

[39] Yu Z, Shen Y, Sun Z, Chen J, Gang W. Cracklab: A high-precision and efficient concrete crack segmentation and quantification network. Developments in the Built Environment 2022;12:100088. https://doi.org/10.1016/j.dibe.2022.100088.

[40] Deng W, Mou Y, Kashiwa T, Escalera S, Nagai K, Nakayama K, et al. Vision based pixel-level bridge structural damage detection using a link ASPP network. Automation in Construction 2020;110:102973. https://doi.org/10.1016/j.autcon.2019.102973.

[41] Miao P, Srimahachota T. Cost-effective system for detection and quantification of concrete surface cracks by combination of convolutional neural network and image processing techniques. Construction and Building Materials 2021;293:123549. https://doi.org/10.1016/j.conbuildmat.2021.123549.

[42] Al Deen Taher SS, Dang J. Autonomous Multiple Damage Detection and Segmentation in Structures Using Mask R-CNN. In: Wu Z, Nagayama T, Dang J, Astroza R, editors. Experimental Vibration Analysis for Civil Engineering Structures, Cham: Springer International Publishing; 2023, p. 545–56. https://doi.org/10.1007/978-3-030-93236-7_45.

[43] Hoskere V, Narazaki Y, Hoang TA, Spencer Jr. BF. MaDnet: multi-task semantic segmentation of multiple types of structural materials and damage in images of civil infrastructure. J Civil Struct Health Monit 2020;10:757–73. https://doi.org/10.1007/s13349-020-00409-0.

[44] Cha Y-J, Choi W, Suh G, Mahmoudkhani S, Büyüköztürk O. Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types. Computer-Aided Civil and Infrastructure Engineering 2018;33:731–47. https://doi.org/10.1111/mice.12334.

[45] Jiang Y, Pang D, Li C. A deep learning approach for fast detection and classification of concrete damage. Automation in Construction 2021;128:103785. https://doi.org/10.1016/j.autcon.2021.103785.