Journal of Infrastructure Intelligence and Resilience 4 (2025) 100144

Contents lists available at ScienceDirect

Journal of Infrastructure Intelligence and Resilience

journal homepage: www.journals.elsevier.com/journal-of-infrastructure-intelligence-and-resilience

# Deep learning in crack detection: A comprehensive scientometric review

Yingjie Wu [a,d], Shaoqi Li [a,*], Jingqiu Li [b], Yanping Yu [c], Jianchun Li [d], Yancheng Li [d,**]

[a] College of Civil Engineering, Nanjing Tech University, Nanjing, 211816, China
[b] College of New Energy, China University of Petroleum (East China), Qingdao, 266580, China
[c] School of International Language and Cultures, Yunan University of Finance and Economics, Kunming, 650221, China
[d] School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW, 2007, Australia

ARTICLE INFO

Handling Editor: Ms. Yanwei Li

*Keywords:*
Crack detection
Deep learning
Concrete structure
Scientometric analysis

ABSTRACT

Cracks represent one of the common forms of damage in concrete structures and pavements, leading to safety issues and increased maintenance costs. Therefore, timely crack detection is crucial for preventing further damage and ensuring the safety of these structures. Traditional manual inspection methods are limited by factors such as time consumption, subjectivity, and labor intensity. To address these challenges, deep learning-based crack detection technologies have emerged as promising solutions, demonstrating satisfactory performance and accuracy. However, the field still lacks comprehensive scientometric analyses and critical surveys of existing works, which are vital for identifying research gaps and guiding future studies. This paper conducts a bibliometric and critical analysis of the collected literature, providing novel insights into current research trends and identifying potential areas for future investigation. Analytical tools, including VOSviewer and CiteSpace, were employed for in-depth analysis and visualization. This study identifies key research gaps and proposes future directions, focusing on advancements in model generalization, computational efficiency, dataset standardization, and the practical application of crack detection methods.

## 1. Introduction

Cracks are a widespread problem in concrete structures and pavements, mainly caused by material fatigue, stress concentration, and other external factors. The development of cracks in concrete structures is attributed to factors such as concrete shrinkage, thermal expansion and contraction, external loads, and poor structural design. Furthermore, the causes of cracks in pavement structures include material fatigue, traffic loading, climate change, and foundation instability. Cracks can lead to a reduction in structural strength and corrosion resistance, potentially resulting in structural collapse and pavement damage, thereby affecting both structural integrity and traffic safety. Therefore, timely crack detection in structures and pavements is crucial (König et al., 2022).

Manual inspection has been the dominant crack detection method in recent decades; however, it is constrained by inefficiency, subjectivity, and time-consuming processes. Furthermore, variability in inspectors' skills and experience affects detection accuracy, leading to inconsistent results. To address these limitations, researchers have increasingly turned to image processing and computer vision (CV) techniques, which have seen significant advancements in recent years (Nyathi et al., 2024).

Crack detection methods based on image processing typically consist of several key steps: image acquisition, image preprocessing, and crack detection (Munawar et al., 2021). During the image acquisition phase, a variety of cameras are used to capture the necessary data, including photographs from standard cameras, infrared and thermal images, depth images, and ultrasound images (Alexander et al., 2022; Alipour et al., 2019; Kalfarisi et al., 2020). Subsequently, image preprocessing is performed to enhance image quality and extract features. Techniques such as grayscale conversion, median and Gaussian filtering, and histogram equalization are commonly applied to improve quality (Lslam et al., 2022). For feature extraction, methods include the Canny edge detector, Sobel operator, texture analysis techniques such as the Gray Level Co-Occurrence Matrix (GLCM) (Kabir and Rivard, 2007) and Local Binary Patterns (LBP) (Zoubir et al., 2022), morphological features (Li and Zhao, 2021), and wavelet transforms (Akbari et al., 2020). After feature extraction, researchers use machine learning and deep learning-based algorithms to detect cracks, including Support Vector Machines (SVM) (Hasni et al., 2017), Random Forest (RF) (Peng et al., 2020), K-Nearest Neighbors (Wang, 2016), Convolutional Neural Networks (CNNs) (Ali

* Corresponding author.
** Corresponding author.
*E-mail addresses:* shaoqi@njtech.edu.cn (S. Li), yanchen.li@uts.edu.au (Y. Li).

https://doi.org/10.1016/j.iintel.2025.100144
Received 15 October 2024; Received in revised form 5 February 2025; Accepted 20 February 2025
Available online 23 February 2025

et al., 2021), and Recurrent Neural Networks (RNNs) (Ahmed et al., 2019).

Although traditional image processing techniques have laid a solid foundation for crack detection, the evolution of CV techniques has profoundly transformed the field. As shown in Fig. 1, the upper portion of the timeline highlights key milestones in CV development, while the lower portion illustrates the parallel progression of crack detection methods. Additionally, notable crack datasets, including CFD (Shi et al., 2016), CCIC (Özgenel and Sorguç, 2018), and CrackSC (Guo et al., 2023), are also highlighted in the timeline. These datasets have played a pivotal role in improving model training and evaluation by providing diverse and complex crack samples across various surfaces and conditions. The advancements in CV, as marked by AlexNet's (Krizhevsky et al., 2017) success in the 2012 ImageNet competition, demonstrated CNNs' remarkable potential in achieving high image classification accuracy. This success paved the way for further innovations, such as VGGNet (Simonyan and Zisserman, 2014), ResNet (He et al., 2016), and U-Net (Ronneberger et al., 2015) significantly enhanced feature extraction capabilities and network depth. Building on these advancements, AlexNet was first applied to classification tasks by (Zhao and Li, 2017), marking a pivotal moment in integrating deep learning into the field of crack detection. Meanwhile, the emergence of object detection networks like Faster R-CNN (Ren et al., 2017) and YOLO (Redmon, 2016) offered new possibilities for detecting objects within images by using bounding boxes. Leveraging these developments, subsequent studies applied Faster R-CNN (Cha et al., 2018) and YOLOv2 (Mandal et al., 2018) to crack detection tasks, where bounding boxes were used to identify and localize cracks within complex backgrounds. As object detection methods evolved, semantic segmentation networks gained prominence, particularly with the introduction of DeepLabv3+ (Chen et al., 2018), which enhanced pixel-level predictions by incorporating Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale contextual information. Following this development, DeepLabv3+ was applied bt (Ji et al., 2020) to segment cracks, demonstrating its effectiveness in producing precise and pixel-level segmentation results.

Building on the success of Transformers in NLP (Vaswani, 2017), and their revolutionary ability to model long-range dependencies, researchers adapted them to CV field, leading to the development of Vision Transformers (ViT) (Dosovitskiy, 2020). Building on this advancement, various ViT-based variants have emerged in crack detection tasks, including CrackFormer (Liu et al., 2021). More recently, large-scale pre-trained models such as GPT-4 Vision (2023) and the segmentation-specificSegment Anything Model (SAM) (Ravi et al., 2024) have emerged as versatile frameworks capable of addressing diverse visual tasks. These models are characterized by their large-scale training on diverse datasets, enabling strong generalization capabilities and adaptability to various downstream applications. The study by (Ye et al., 2019) demonstrated the potential of SAM in crack detection by efficiently segmenting crack regions within infrastructure monitoring tasks. This development highlights the increasing role of large-scale pre-trained models in solving specialized problems like crack detection. As these models continue to evolve, crack detection methods will become increasingly diverse and effective.

Despite rapid advancements in crack detection methodologies, existing review articles lack a comprehensive synthesis of the field (Spencer Jr et al., 2019). Many reviews either narrowly focus on specific algorithms or offer overly broad overviews, lacking critical analysis of key aspects such as the evolving methodological innovations, the integration of foundational models, and patterns of scholarly collaboration (Ali et al., 2022; Zhou et al., 2023). Moreover, with advances in CV, a growing array of innovative approaches has been applied to crack detection tasks (Awadallah and Sadhu, 2023; Chen et al., 2024; Hang et al., 2023; Hu et al., 2025; Wu et al., 2024; Yang et al., 2023; Ma et al., 2024). These include hybrid network architectures, methods that integrate multimodal features, and large-scale pre-trained models. Despite their increasing relevance and potential, these methods remain underexplored and lack systematic review in existing scientometric literature. This gap in systematic analysis limits understanding of their broader impact and future potential.

To address these gaps, this review systematically examines advancements in CV-based crack detection methods, focusing on research conducted between 2017 and 2024, a period marked by significant progress in computational power and methodological innovation. The main contributions of this review include:

- Comprehensive bibliometric analysis: This review conducts a bibliometric analysis of the collected deep learning-based crack detection papers, identifying influential papers, journals, authors, countries, collaboration patterns, and research keywords within the field.
- Critical synthesis of methodologies: This review systematically examines the advanced methods used in crack detection, focusing on key areas such as classification, object detection, and segmentation.
- Identification of challenges and future directions: This review summarizes the existing methodologies and identifies potential pathways to drive future advancements in crack detection.

## 2. Previous literature reviews

Crack detection is a critical issue in structural health monitoring, driving extensive research efforts. Numerous review papers have synthesized the literature on crack detection, providing comprehensive insights into methodologies, technologies, and challenges in the field. These reviews systematically analyze the progression of detection techniques, from traditional manual inspections to modern CV applications. This section provides a concise synthesis of previous review findings and highlights their key contributions, as summarized in Table 1.

(Cao et al., 2020; Gopalakrishnan, 2018) focused on categorizing and tracing the evolution of crack detection methods, emphasizing the shift

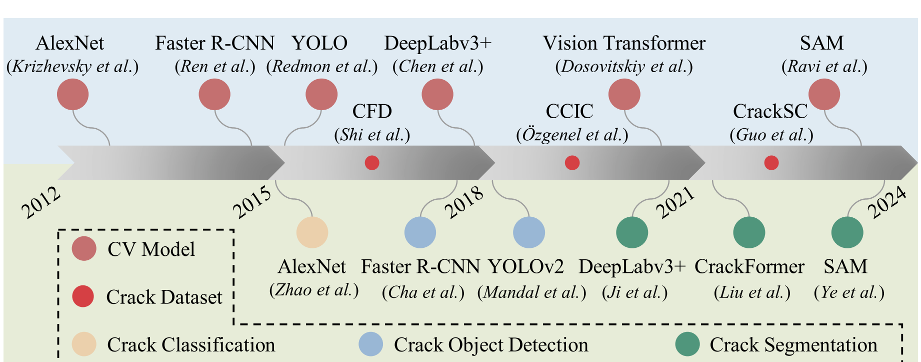


**Fig. 1.** Timeline of key advancements in CV and crack detection methods.

 

**Table 1**
Highlights of previous review in deep learning-based crack detection methods.

| Ref | Major Contributions | Limitations |
|---|---|---|
| Cao et al. (2020) | • Compared traditional crack detection methods with 3D-based crack detection methods.<br>• Classified crack detection methods into three categories. | • Insufficient availability of suitable datasets for comparing the performance of traditional and 3D crack detection methods. |
| Gopalakrishnan (2018) | • Reviewed frameworks for crack detection algorithms.<br>• Emphasized the trend in crack detection shifting from 2D images to 3D analysis. | • Limited the introduction to performance evaluation metrics for crack detection. |
| Ye et al. (2019) | • Conducted a comprehensive review of the development and utilization of deep learning techniques within the area of SHM. | • Failed to include a discussion or review of crack detection algorithms. |
| (Hsieh and Tsai, 2020) | • Provided a systematic review of 68 ML-based crack detection papers.<br>• Highlighted key challenges and trends in crack segmentation research. | • Did not quantitatively compare models across different datasets. |
| Hu et al. (2021) | • Provided a novel summary of the key challenges faced by deep learning-based crack detection.<br>• Discussed new directions for the future development of crack detection technology from three aspects. | • Limited discussion on the available algorithms. |
| Deng et al. (2022) | • Categorized traditional image processing and deep learning approaches.<br>• Suggested future research on lightweight models and attention mechanisms.<br>• Addressed data deficiency through semi-supervised learning and synthetic data. | • Limited discussion on computational costs and real-time feasibility. |
| (Kheradmandi and Mehranfar, 2022) | • Categorized approaches into rule-driven and data-driven techniques.<br>• Focused on the challenges of segmenting complex pavement images. | • Did not propose a standardized evaluation for diverse algorithms.<br>• Insufficient coverage of other pavement distresses beyond cracks, such as potholes or rutting. |
| Ai et al. (2023) | • Discussed various data collection methods, available crack datasets, and performance evaluation metrics.<br>• Provided a new classification based on the development of automatic crack detection algorithms. | • Limited discussion on the balance between computational efficiency and accuracy. |
| Yuan et al. (2024) | • Reviewed 120 research papers and categorized them into three groups.<br>• Emphasized that integrating additional dimensional data can improve crack detection accuracy. | • The articles are not collected in a systematic way. |
| Yang et al. (2024) | • Focused on crack detection of asphalt pavements.<br>• Provided a systematic review summarizing methods for road maintenance. | • Lacked an assessment of algorithm performance. |

from traditional 2D to 3D-based techniques. A key issue highlighted is the lack of suitable datasets to enable performance comparison between these methods (Ye et al., 2019).examined the development of deep learning in structural health monitoring (SHM) but overlooks the inclusion of crack detection algorithms (Hsieh and Tsai, 2020). built on this by reviewing machine learning-based crack detection, identifying the challenges and trends in crack segmentation, but similarly fails to provide quantitative comparisons of models across different datasets (Hu et al., 2021; Deng et al., 2022). further contributed to the discussion by addressing deep learning-based crack detection techniques, introducing lightweight models, attention mechanisms, and methods to address data limitations, such as semi-supervised learning and synthetic data. However, both studies noted the lack of focus on computational costs and real-time feasibility.

The more recent works by (Kheradmandi and Mehranfar, 2022; Ai et al., 2023) categorized detection approaches into rule-driven and data-driven methods and explore advanced image segmentation techniques for complex pavement images. Both studies pointed out key gaps, such as the absence of standardized evaluation frameworks for crack detection algorithms and the insufficient attention given to other pavement issues like potholes and rutting (Yuan et al., 2024; Yang et al., 2024). explored the potential of integrating 3D imaging and dimensional data to enhance detection accuracy, with Yang specifically focusing on asphalt pavements and summarizing road maintenance strategies, though both studies lacked a systematic evaluation of algorithm performance. While existing review articles offer valuable insights into crack detection methods, they often lack a systematic approach to collecting and analyzing the relevant literature. Many reviews failed to employ bibliometric or statistical methods to quantify field trends, making it difficult for future researchers to gain a clear understanding of the research landscape. Consequently, significant gaps persisted in understanding the progression and impact of research in this area. To address these gaps, a systematic review has been conducted to provide a clearer perspective on the advancements and methodologies in deep learning-based crack detection.

## 3. Previous literature reviews

This study adopts a mixed-methods approach to conduct both a bibliometric analysis and a critical review of the literature on crack detection algorithms. The analysis focuses on crack detection in concrete buildings, roads, and pavements. Fig. 2 presents a visual representation of the research methodology applied in this study.

In the initial phase, relevant literature is retrieved from the Scopus and Web of Science (WOS) databases, as depicted in Fig. 2. Subsequently, a bibliometric analysis is performed using visualization tools such as CiteSpace and VOSviewer. This analysis examines annual publication trends, influential authors, journals, and papers, as well as notable countries, collaboration patterns, and research keywords. In the final phase, a comprehensive review of the referenced papers is conducted, focusing on their abstracts, methodologies, and results.

In the literature retrieval process, relevant papers were systematically retrieved by performing a keyword search in the Scopus and WOS databases. The decision to use Scopus and WOS was based on their comprehensive coverage of high-quality research articles and their established use in bibliometric analysis across various disciplines, ensuring access to a broad yet relevant dataset. The keywords used in the search, “deep learning,” “crack detection,” and “image,” were chosen to

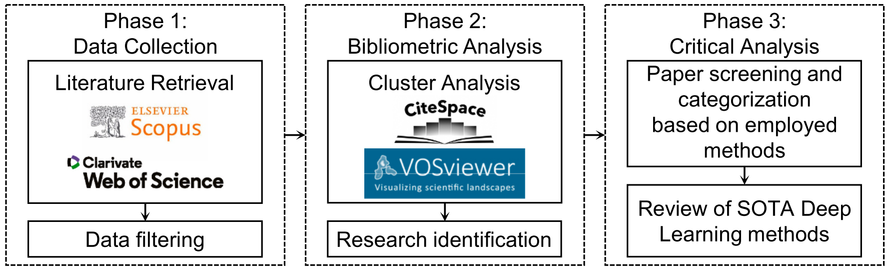


**Fig. 2.** Summary of the research approach.

focus on studies utilizing deep learning techniques specifically for crack detection tasks based on image methods. These keywords were joined using the Boolean operator "AND" to capture papers addressing the intersection of these topics. Additionally, "concrete," "pavement," and "road" were joined using the Boolean operator "OR" to encompass a wide range of application scenarios within the domain of infrastructure, as shown in Table 2.

Following the application of keyword searches, 949 articles were retrieved from Scopus and 672 articles from the WOS. To ensure the relevance of these articles to the research topic and eliminate unrelated works, further filtering criteria were employed. The publication year was restricted to post-2016 for both databases to focus on recent advancements in deep learning, which has experienced significant breakthroughs in the past decade. Subject areas in Scopus were narrowed to Engineering and Computer Science, as these fields are most relevant to crack detection using computational methods. Similarly, in WOS, the subject categories were refined to include those most closely aligned with the research focus, such as Engineering Civil, Construction Building Technology, and Computer Science Artificial Intelligence. Additional restrictions were applied to maintain the quality and relevance of the dataset. In Scopus, only articles in the "Final" publication stage were considered to avoid pre-publication or incomplete studies. In WOS, the index was restricted to "SCI-EXPANDED" to ensure inclusion of articles published in high-impact journals. In both databases, only peer-reviewed articles written in English were included, as English is the dominant language of scientific communication. After applying these restrictions, a total of 994 articles were retrieved, of which 299 were duplicates. Following the removal of duplicates, a final dataset of 695 articles was prepared for bibliometric analysis.

**Table 2**
The search terms, refinements, and outcomes from the Scopus and WOS repositories.

| Search Engine | String and Refinement | Results |
|---|---|---|
| Scopus | (TITLE-ABS-KEY ("crack detection") AND TITLE-ABS-KEY ("deep learning") AND TITLE-ABS-KEY (image) AND TITLE-ABS-KEY ("concrete" OR "pavement" OR "road")) | 949 |
| | AND PUBYEAR >2016 | 943 |
| | AND (LIMITED-TO (SUBJAREA, "ENGINEERING") OR LIMITED-TO (SUBJAREA, "COMPUTER SCIENCE")) | 868 |
| | AND (LIMITED-TO (DOCTYPE, "ARTICLE")) | 542 |
| | AND (LIMITED-TO (LANGUAGE, "ENGLISH")) | 506 |
| | AND (LIMITED-TO (PUBLICATION PHASE, "FINAL") | 471 |
| WOS | TOPIC: ("crack detection") AND TOPIC: ("deep learning") AND TOPIC: ("image") AND TOPIC: ("concrete") OR TOPIC: ("pavement") OR TOPIC: ("road") AND PUBYEAR >2016 WOS Index: SCI-EXPANDED | 672 |
| | Refined by: WOS CATEGORIES: (ENGINEERING CIVIL OR CONSTRUCTION BUILDING TECHNOLOGY OR ENGINEERING MULTIDISCIPLINARY OR TRANSPORTATION SCIENCE TECHNOLOGY OR COMPUTER SCIENCE ARTICIAL INTELLIGENCE OR COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS OR COMPUTER SCIENCE INFORMATION SYSTEMS OR IMAGING SCIENCE PHOTOGRAPHIC TECHNOLOGY) | 548 |
| | Refined by: DOCTYPE: ARTICLE Refined by: LANGUAGE: ENGLISH | 523 |

Sum of papers = 994.
Duplicates = 299.
Remaining = 695.

## 4. Bibliometric analysis

Bibliometric analysis is a quantitative approach to evaluate the scholarly impact of publications, sources, and authors, while simultaneously uncovering research trends in a specific area through various statistical techniques, including publication rate, citation rate, collaboration patterns, and keywords occurrence. In this section, two visualization tools, CiteSpace and VOSviewer, are employed to perform a bibliometric analysis of the collected papers. This analysis focuses on identifying the most productive publications, authors, and journals within crack detection field. Additionally, it will feature a scientific mapping analysis that includes several key subsections. Co-Citation analysis examines the relatedness of sources and authors, quantifying their proximity within the discipline. Co-Authorship analysis explores collaboration patterns among countries and institutions, shedding light on international research networks. Lastly, the analysis of keywords occurrence identifies emerging trends and significant terms related to crack detection.

### *4.1. Overview of the publications*

#### *4.1.1. Overview of the publications*

Fig. 3 illustrates the trends in publication and citation numbers in crack detection research from 2017 to 2024. The data shows a steady rise in publication numbers, increasing from 4 papers in 2017 to a peak of 171 in 2023, highlighting the growing interest in this field. However, this upward trend in publications contrasts sharply with citation trends, which peaked at 5886 in 2020 and then declined significantly to 1180 in 2023 and 206 in 2024. This divergence suggests that while more papers are being published, their overall academic impact, as measured by citations, appears to be diminishing.

The contrasting trends between publication and citation numbers

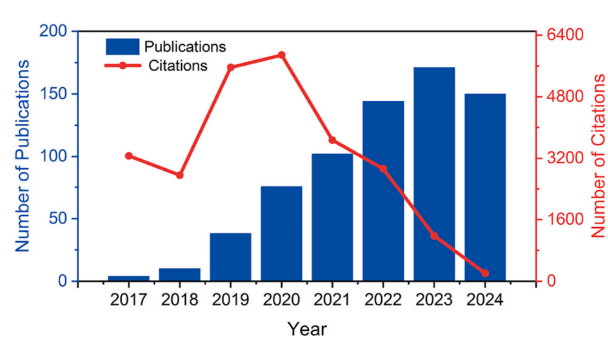


**Fig. 3.** Annual publications and citations trends.

reveal several notable patterns worth examining. In the early years (2017–2020), fewer papers were published, but these had significantly higher citation rates. For instance, the 4 papers published in 2017 generated 3270 citations, averaging 817.5 citations per paper. This indicates that research during this period introduced foundational theories and influential methodologies with lasting academic impact. By 2020, the field appeared to reach its peak, with influential studies contributing to a citation count of 5886 across 76 papers published that year. After 2020, the number of publications continued to increase, with 102 in 2021, 144 in 2022, and 171 in 2023, while citation numbers experienced a significant decline. In 2024, 150 papers were published and collectively received only 206 citations. This sharp decline is primarily due to the limited number of groundbreaking contributions in recent years. Much of the recent work focuses on extending or refining previous research rather than introducing fundamentally new concepts or methods.

Overall, the steady rise in publications demonstrates sustained interest in crack detection research, but the declining citation rates underscore the need for innovative and high-impact contributions. Moving forward, it is essential to prioritize not only the quantity of research output but also its quality to foster significant advancements in the field.

#### 4.1.2. Overview of the publications

This section focuses on identifying and analyzing the most influential papers out of 695 articles, selecting the top ten based on citation numbers for detailed examination. Collectively, these ten papers have received 6947 citations, representing 27.27% of all citations. Table 3 summarizes these top-cited articles, including their publication sources, corresponding authors, publication years, citation counts, and average citations per year.

**Table 3**
Summary of the top-cited articles.

| Ref | Journal | Corresponding Author | Citation | Average Citation Year |
|---|---|---|---|---|
| Cha et al. (2017) | Computer-Aided Civil and Infrastructure Engineering | Young-Jin Cha | 2042 | 291.7 |
| Cha et al. (2018) | Computer-Aided Civil and Infrastructure Engineering | Young-Jin Cha | 980 | 163.3 |
| Dung (2019) | Automation in Construction | Cao Vu Dung | 682 | 136.4 |
| Zhang et al. (2017) | Computer-Aided Civil and Infrastructure Engineering | Allen A, Zhang | 629 | 89.9 |
| Gopalakrishnan et al. (2017) | Construction and Building Materials | Kasthuriangan Gopalakrishnan | 588 | 84 |
| Zou et al. (2018) | IEEE Transactions on Image Processing | Qin Zou | 512 | 102.4 |
| Yang et al. (2019) | IEEE Transactions on Intelligent Transportation systems | Fan Yang | 490 | 122.5 |
| Dorafshan et al. (2018) | Construction and Building Materials | Sattar Dorafshan | 421 | 81.7 |
| Li and Zhao (2019) | Computer-Aided Civil and Infrastructure Engineering | Xuefeng Zhao | 305 | 71 |
| Huang et al. (2018) | Tunnelling and Underground Space Technology | Hongwei Huang | 298 | 49.7 |

Among these, the study published in Computer-Aided Civil and Infrastructure Engineering (Cha et al., 2017) stands out with the highest citation numbers, totaling 2042 and averaging 291.7 citations per year. This study marked a significant milestone in crack detection research as it was the first to apply CNNs to the task of crack classification. By leveraging CNNs, the authors demonstrated their effectiveness in automating crack detection with high accuracy, establishing a foundation for deep learning applications in this field. Another highly influential paper by (Cha et al., 2018), also published in the same journal, received 980 citations. This study was pioneering in applying Faster R-CNN, an advanced object detection framework, to crack detection, enabling the localization and identification of cracks with bounding boxes. A notable contribution was made by (Dung, 2019) with 682 citations, adopting a fully convolutional network (FCN) for semantic segmentation in crack detection. Unlike earlier approaches that focused on classification or object detection, this study addressed pixel-level segmentation, providing a more granular understanding of crack morphology.

Other foundational studies, such as (Zhang et al., 2017) with 629 citations (Zou et al., 2018), with 512 citations, and (Yang et al., 2019) with 490 citations, introduced or refined important methodologies. For example, Zou's work emphasized the importance of data augmentation techniques to improve model robustness, while Yang demonstrated the application of lightweight architectures for efficient crack detection in real-time scenarios.

By comparison, some studies published in the same timeframe received fewer citations. For instance, although (Gopalakrishnan et al., 2017) was published in the same year as Cha's work, its average annual citation count is only 84. Similarly, the studies by (Dorafshan et al., 2018; Li and Zhao, 2019; Huang et al., 2018) exhibited lower average citations per year at 81.7, 71, and 49.7, respectively. These differences highlight the varying impact of research contributions, with the most highly cited studies being those that introduced groundbreaking methods or demonstrated novel applications of deep learning in crack detection.

In summary, these findings highlight a trend where earlier articles, particularly those that introduced CNNs, object detection frameworks, or segmentation methods, tend to have significantly higher citation numbers compared to more recent studies. However, the relatively lower citation counts for more recent studies suggest that further innovation is needed to enhance the scholarly impact and visibility of research in this field. Specifically, future studies may benefit from integrating novel methodologies, such as multimodal approaches and large-scale pre-trained models, to address the growing complexity of crack detection tasks and increase their academic influence.

### 4.2. Influential journals, authors, and countries

#### 4.2.1. The most productive journals

This section analyzes the data collected from 695 articles published across 182 journals, highlighting the top twenty journals by publication count. These journals collectively contributed 400 articles, accounting for 57.55% of the total publications, while the remaining 295 articles were distributed among 172 other journals. Table 4 summarizes these high publication journals, including their names, total publications (TPs), total citations (TCs), average citations (ACs), impact factor (IF), and H-index.

Automation in Construction leads with 52 publications, accumulating 2475 citations and an average of 47.5 citations per article, reflecting its significant influence in the field of crack detection.

The journal is further supported by a robust impact factor of 9.6 and an H-index of 23. Applied Science-Basel ranks as the second most productive journal, with 39 publications garnering 505 citations and an average of 12.95 citations per article, indicating a solid academic contribution. In third place, IEEE Transactions on Intelligent Transportation Systems contributes 31 publications and has received 1408

**Table 4**
Details of the most productive journals.

| Journal Name | TPs | TCs | ACs | IF | H-Index |
|---|---|---|---|---|---|
| Automation in Construction | 52 | 2475 | 47.5 | 9.6 | 23 |
| Applied Science-Basel | 39 | 505 | 12.95 | 2.5 | 12 |
| IEEE Transactions on Intelligent Transportation systems | 31 | 1408 | 45.42 | 7.9 | 14 |
| IEEE Access | 30 | 878 | 29.27 | 3.9 | 16 |
| Computer-Aided Civil and Infrastructure Engineering | 29 | 6165 | 212.59 | 8.5 | 19 |
| Construction and Building Materials | 28 | 1835 | 65.54 | 7.4 | 14 |
| Structural Health Monitoring-an International Journal | 27 | 760 | 28.15 | 5.7 | 12 |
| Sensors | 25 | 349 | 13.96 | 3.4 | 9 |
| International Journal of Pavement Engineering | 18 | 388 | 21.56 | 3.8 | 8 |
| Structural Control and Health Monitoring | 15 | 900 | 60 | 4.6 | 10 |
| Engineering Applications of Artificial Intelligence | 14 | 205 | 14.64 | 7.5 | 7 |
| Measurement | 13 | 293 | 22.54 | 5.2 | 9 |
| Journal of Civil Structural Health Monitoring | 12 | 256 | 21.33 | 3.6 | 7 |
| Remote Sensing | 11 | 130 | 11.82 | 4.2 | 6 |
| Smart Structures and Systems | 11 | 42 | 3.82 | 2.37 | 4 |
| Journal of Computing in Civil Engineering | 10 | 778 | 77.8 | 4.7 | 7 |
| Journal of Transportation Engineering, Part B: Pavements | 10 | 126 | 12.6 | 1.9 | 8 |
| Tunnelling and Underground Space Technology | 9 | 453 | 50.33 | 6.7 | 6 |
| Advanced Engineering Informatics | 8 | 331 | 41.38 | 8 | 4 |
| Engineering Structures | 8 | 169 | 21.13 | 5.6 | 4 |

citations, averaging 45.42 citations per article, alongside an impact factor of 7.9 and an H-index of 14.

Given the consideration of citation counts, Computer-Aided Civil and Infrastructure Engineering stands out with the highest total citations, accumulating 6165 across 29 publications and achieving an exceptional average of 212.59 citations per article. This demonstrates the substantial impact of its published articles on subsequent research within the field of crack detection. Construction and Building Materials also demonstrates strong performance, with 28 publications and 1835 citations, averaging 65.54 citations per article, along with an impact factor of 7.4 and an H-index of 14. Other notable journals include IEEE Access (30 publications, 878 citations), Structural Health Monitoring - An International Journal (27 publications, 760 citations) and Sensors (25 publications, 349 citations). The International Journal of Pavement Engineering published 18 articles with 388 citations, while Structural Control and Health Monitoring contributed 15 publications and 900 citations.

These journals play a vital role in advancing crack detection research by disseminating influential studies and fostering technological innovation. Journals like Computer-Aided Civil and Infrastructure Engineering and Automation in Construction stand out as authoritative sources, reflecting their pivotal role in promoting foundational theories and methodologies. Their ability to attract and disseminate high-impact research underscores their influence in shaping advancements in crack detection. Overall, these journals provide a strong foundation for future interdisciplinary studies and contribute to the ongoing development of structural health monitoring technologies.

#### *4.2.2. The most productive authors*

In this section, the authors with the highest productivity in this field are analyzed. A total of 1124 authors were identified from the 695 articles collected. The top five authors were selected for detailed examination based on publication numbers and citation numbers. Table 5 summarizes these most productive authors, including their total number

**Table 5**
Details of the most productive authors.

| | Author's Name | TPs | TCs | ACs | As 1st Author | H-index |
|---|---|---|---|---|---|---|
| Based on Publications | Allen A, Zhang | 13 | 1352 | 104 | 3 | 6 |
| | K.C.P, Wang | 11 | 1270 | 115.45 | 0 | 5 |
| | Niannian, Wang | 8 | 122 | 15.25 | 1 | 5 |
| | Lei, Wang | 6 | 112 | 18.67 | 3 | 4 |
| | Gye-Chun, Cho | 6 | 151 | 25.17 | 0 | 6 |
| Based on Citations | Young-jin Cha | 5 | 3667 | 733.4 | 2 | 5 |
| | Wooram, Choi | 3 | 3303 | 1101 | 1 | 3 |
| | Oral Buyukozturk | 2 | 3022 | 1511 | 0 | 2 |
| | Allen A, Zhang | 13 | 1352 | 104 | 3 | 6 |
| | K.C.P, Wang | 11 | 1270 | 115.45 | 0 | 5 |

of publications, total citations, average citations, the number of times they acted as first author, and H-index.

In terms of publication numbers, Allen A, Zhang leads with 13 publications, accumulating 1352 citations and an average of 104 citations per article. Additionally, Zhang has served as first author three times and has an H-index of 6, reflecting a high level of research activity and impact within the field of crack detection. K.C.P. Wang follows closely with 11 publications, totaling 1270 citations and an average of 115.45 citations per article. However, Wang has not acted the first author role, and his H-index is 5. Niannian Wang has published 8 articles that have received a total of 122 citations, with an average of 15.25 citations per article. Although both Lei Wang and Gye-Chun Cho have published 6 articles, their impact on the field differs significantly. Specifically, Lei has 112 citations (averaging 18.67 citations per article) and an H-index of 4, having served as the first author three times. In contrast, Gye-Chun has 151 citations (averaging 25.17 citations per article) but has not published as the first author, and with an H-index of 6.

Regarding citation impact, Young-jin Cha stands out with 5 publications and a remarkable total of 3667 citations, averaging 733.4 citations per article, which highlights the significant impact of Cha's work. Wooram Choi follows with only 3 publications but has garnered 3303 citations, averaging 1101 citations per article, along with an H-index of 3. Similarly, Oral Buyukozturk has only published two articles but has accumulated 3022 citations, averaging 1511 citations per article. This demonstrates the substantial impact of his research, reflected in his H-index of 2. Notably, both Allen A. Zhang and K.C.P. Wang appear in both categories, highlighting their versatility and sustained influence in crack detection research.

The analysis of the most productive authors highlights key contributors to the field of crack detection, with several authors demonstrating both high publication counts and significant citation impact. While Allen A. Zhang and K.C.P. Wang exemplify sustained productivity and versatility, authors like Young-jin Cha, Wooram Choi, and Oral Buyukozturk have made substantial contributions through highly impactful individual publications.

#### *4.2.3. The most productive countries*

In this section, the authors have organized data from the 695 collected articles to analyze the contributions of countries in this field. The analysis reveals that the articles originate from 61 different countries or regions. Fig. 4 illustrates the distribution of these publications, with the size of the circles representing the number of articles published by each country or region. Notably, China has the highest publication count, accounting for 53.53% of the total articles, while the United States ranks second, contributing approximately 14.39%. Additionally, the authors have compiled statistics for the top ten countries by

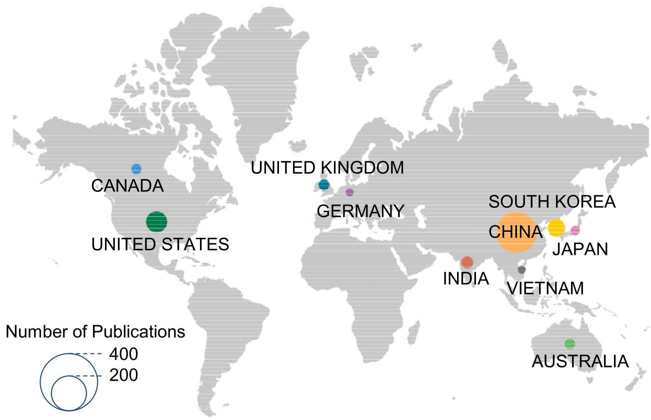


**Fig. 4.** Geographical distribution of publications.

publication volume, as detailed in Table 6. This table includes total publications, total citations, average citations, and the number of articles with citations greater than or equal to 100, 50, and 10, as well as the H-index for each country.

According to Table 6, China leads significantly with 372 total publications and an impressive 10024 citations, resulting in an average of 26.95 citations per article. Additionally, China has the highest H-index of 49, indicating substantial impact within the field. The United States follows with 100 publications and 9571 citations, achieving an average of 95.71 citations per article and an H-index of 36. South Korea ranks third with 68 publications and 2829 citations, averaging 41.6 citations per article, and an H-index of 25. Moreover, China not only leads in total publications but also has 23 articles with over 100 citations and 49 articles with more than 50 citations, reflecting the high impact of its research. The United States has 21 articles cited over 100 times and 31 articles exceeding 50 citations, demonstrating a strong influence in the field. In contrast, countries like India and the United Kingdom show limited high-impact publications, with India having no articles cited more than 100 times. While Canada has a notable average citation score, it has only 6 articles with over 100 citations.

Overall, China and the United States remain the most influential contributors, with China leading in total publications and citations, while the United States excels in average citations per article. South Korea ranks third, making significant research contributions, whereas countries like India and the United Kingdom demonstrate fewer high-impact publications. Despite Canada's limited output, its high average citation performance reflects the quality and impact of its research.

### *4.3. Science mapping analysis*

#### *4.3.1. Co-citation analysis*

Co-Citation Analysis is recognized as an important science mapping technique for evaluating the relationships between cited sources and authors. For instance, when Paper 3 concurrently cites Papers 1 and 2, it establishes a Co-Citation that signifies a scholarly linkage between the two works. Therefore, this section involves a Co-Citation analysis carried out by the authors to investigate the connections between journals and authors. Citing two sources or authors together suggests a shared research field and interest, thereby illuminating the intellectual landscape surrounding crack detection methods in various structures. To ensure the relevance and interpretability of the analysis, a citation threshold of 20 was set. This threshold was chosen to focus on journals with substantial scholarly recognition, as sources with fewer citations often provide limited insights into the field's intellectual structure. Consequently, 18 journals meeting this criterion were included in the analysis.

**Table 6**
Details of the most productive countries.

| Country | TPs | TCs | ACs | ≥100 | ≥50 | ≥10 | H-Index |
|---|---|---|---|---|---|---|---|
| China | 372 | 10024 | 26.95 | 23 | 49 | 162 | 49 |
| United States | 100 | 9571 | 95.71 | 21 | 31 | 59 | 36 |
| South Korea | 68 | 2829 | 41.6 | 6 | 18 | 45 | 25 |
| India | 32 | 414 | 12.94 | 0 | 2 | 10 | 10 |
| United Kingdom | 26 | 388 | 14.92 | 1 | 2 | 8 | 9 |
| Australia | 25 | 910 | 36.4 | 3 | 8 | 14 | 12 |
| Canada | 23 | 4219 | 183.43 | 6 | 8 | 15 | 13 |
| Japan | 19 | 1184 | 62.32 | 2 | 5 | 11 | 9 |
| Vietnam | 14 | 1260 | 90 | 2 | 3 | 9 | 9 |
| Germany | 11 | 105 | 9.54 | 0 | 0 | 3 | 5 |

Table 7 presents the Co-Citation indices of the sources, including citation counts and Total Link Strength, where Total Link Strength refers to the overall measure of Co-Citation frequency. Furthermore, the authors used VOSviewer software to create a visual map of the journals' Co-Citation networks, providing a more intuitive understanding of the link model, as illustrated in Fig. 5. This figure categorizes the different journals into four groups, represented by red, green, blue, and yellow. Each node symbolizes a particular source, with larger nodes denoting higher Co-Citations, and thicker links between nodes indicating a stronger connection between the associated sources.

The red cluster comprises five sources, in which Automation in Construction and Computer-Aided Civil and Infrastructure Engineering contribute the most significantly. These two journals exhibit 17 connections with other literature, although their contributions differ. Specifically, Automation in Construction has 308 citations and a Total Link Strength of 4165, whereas Computer-Aided Civil and Infrastructure Engineering has 249 citations and a Total Link Strength of 3494. Their high Co-Citation link strength indicates their role as key nodes for foundational and interdisciplinary research, bridging civil engineering, computer vision, and advanced data analytics.

Similarly, the green cluster also contains five sources, in which Sensors is the most influential, generating 195 citations and a Total Link Strength of 2466. Construction and Building Materials follows closely, with 152 citations and a Total Link Strength of 2068. The blue and yellow clusters each include four sources, including Journal of Computing in Civil Engineering and IEEE Access, which have 124 citations and a Total Link Strength of 1836, and 145 citations and a Total Link Strength of 1863, respectively.

**Table 7**
Co-citation indices of the sources.

| Source | Citations | Total Link Strength |
|---|---|---|
| Automation in Construction (Autom Constr) | 308 | 4165 |
| Computer-Aided Civil and Infrastructure Engineering (Comput-aided Civ Inf) | 249 | 3494 |
| Sensors | 195 | 2466 |
| Construction and Building Materials (Constr Build Mater) | 152 | 2068 |
| IEEE Access (IEEE Access) | 145 | 1863 |
| Journal of Computing in Civil Engineering (J Comput Civil Eng) | 124 | 1836 |
| IEEE Transactions on Intelligent Transportation Systems (IEEE T Intell Transp) | 143 | 1570 |
| arXiv (Arxiv) | 107 | 1420 |
| Proceedings CVPR IEEE (Proc CVPR IEEE) | 100 | 1258 |
| Structural Health Monitoring (Struct Health Monit) | 86 | 1137 |
| Applied Science (Appl Sci) | 86 | 1111 |
| IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE T Pattern Anal) | 93 | 1059 |
| Neurocomputing (Neurocomputing) | 43 | 639 |
| Measurement | 40 | 622 |
| Structural Control & Health Monitoring (Struct Control Monit) | 34 | 569 |
| Sustainability | 35 | 519 |
| Proceedings ICCV IEEE (Proc ICCV IEEE) | 28 | 407 |
| Communications of the ACM (Commun Acm) | 22 | 335 |

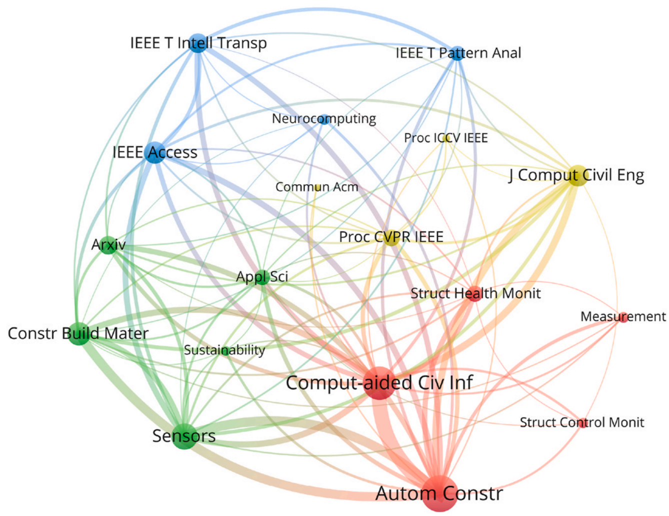


**Fig. 5.** Network visualization of Co-Citation analysis of the sources.

These findings highlight the critical role of journals like Automation in Construction and Computer-Aided Civil and Infrastructure Engineering in advancing foundational research in crack detection field. This analysis can help researchers identify high-impact journals for publishing their findings, thereby driving innovation in the field of crack detection.

In addition, the authors also analyzed the Co-Citation network of the cited authors to identify key contributors and collaborative relationships within the field. Based on citation numbers, the top fifteen authors were selected and summarized in Table 8, which includes the cited authors, citation numbers, and Total Link Strength. Similarly, VOSviewer software was employed to generate a visual map of the authors' Co-Citation networks, as presented in Fig. 6. The authors are grouped into three clusters, represented by red, green, and blue.

The red cluster consists of six authors, with Young-Jin Cha being the most prominent, having 421 citations and a Total Link Strength of 1943. In addition, Cha's work has the highest citation number among all the authors, which indicates that Cha's work is not only widely referenced but also significantly interconnected with other research in the field, suggesting his influential role in advancing the study of crack detection. The second-ranked author is Cao Vu Dung, who has been cited 180 times and has a Total Link Strength of 920. Other influential authors include Alex Krizhevsky, Karen Simonyan, Sattar Dorafshan, and Yann LeCun, with Total Link Strengths of 951, 938, 784, and 832, respectively.

The green cluster also comprises six authors, with K.C.P. Wang having the highest citation count at 325 and a Total Link Strength of 1671. Additionally, Olaf Ronneberger and Lei Zhang demonstrate

**Table 8**
Author's Co-Citation indices.

| Cited Author | Citations | Total Link Strength |
|---|---|---|
| Young-Jin, Cha | 421 | 1943 |
| K.C.P, Wang | 325 | 1671 |
| Kaiming, He | 310 | 1716 |
| Qin, Zou | 268 | 1482 |
| Olaf Ronneberger | 219 | 1146 |
| Lei, Zhang | 218 | 1160 |
| Yong, Shi | 187 | 1008 |
| Cao Vu, Dung | 180 | 920 |
| Alex Krizhevsky | 178 | 951 |
| Joseph Redmon | 173 | 833 |
| Karen Simonyan | 171 | 938 |
| Sattar Dorafshan | 167 | 784 |
| Ross Girshick | 157 | 885 |
| Yann LeCun | 148 | 832 |
| Vijay Badrinarayanan | 140 | 809 |

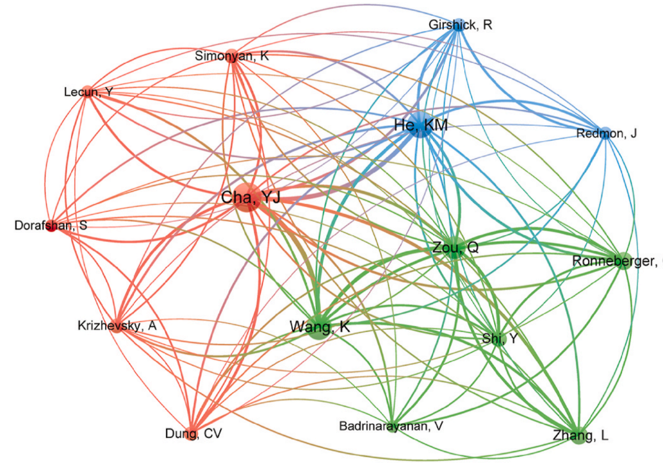


**Fig. 6.** Network visualization of Co-Citation analysis of the authors.

closely aligned impacts within the red cluster, with citation counts of 219 and 218, respectively, and Total Link Strengths of 1146 and 1160. This proximity in their metrics suggests that both Ronneberger and Zhang have made comparable contributions to the field, highlighting their significance alongside Wang in advancing research on crack detection.

The blue cluster contains only three authors: Kaiming He, Joseph Redmon, and Ross Girshick. Kaiming He's influence significantly surpasses that of the others, with citation counts exceeding those of Redmon and Girshick by 137 and 153, respectively. Similarly, Kaiming He's Total Link Strength is higher by 833 and 831 compared to the others. This demonstrates Kaiming He's prominent role in the cluster and emphasizes the impact of his contributions to the field.

The Co-Citation analysis highlights the significant influence of deep learning advancements on crack detection research. Several authors in the clusters, such as Kaiming He, Alex Krizhevsky, and Yann LeCun, are prominent figures in the field of CV, having pioneered foundational architectures like ResNet and CNNs. Their contributions have been instrumental in shaping the methodologies applied in crack detection, particularly in areas like feature extraction, object detection, and semantic segmentation. Similarly, researchers like Young-Jin Cha and K.C. P. Wang have made creative use of deep learning techniques in crack detection, highlighting how innovative applications of these methods have significantly contributed to advancements in this field.

The analysis highlights the integration of CV and civil engineering, demonstrating that advancements in deep learning-based crack detection are closely tied to the broader development of CV technologies. This connection underscores the importance of leveraging progress in CV to drive further innovation in structural health monitoring and crack detection.

### 4.3.2. *Co-authorship analysis*

Co-Authorship analysis is a vital bibliometric method to clarify the collaborative relationships among researchers. Collaboration is essential in research as it promotes the generation of innovative ideas and simplifies the implementation of research tasks. This section presents a Co-Authorship analysis of the gathered particles, with countries serving as the unit of analysis. A threshold of 20 published articles per country was established, and ten countries met this criterion, from which the top five were selected for detailed examination. Table 9 summarizes the number of articles published by these five countries along with their Total Link Strength.

Among all countries, China has published the most articles, totaling 325. Its collaboration with other nations is also closer, with a Total Link Strength of 61, indicating that China occupies an important position in international cooperation in the field of crack detection. Australia contributes 48 articles to the green cluster; however, its Total Link Strength is only 7, indicating limited academic cooperation with other nations. The United States contributes 111 articles, significantly fewer than China; however, its Total Link Strength is only 8 points lower than that of China. This indicates that the United States has closer collaboration with other countries in the field of crack detection. South Korea and the United Kingdom contributed 94 and 54 articles, with Total Link Strengths of 10 and 21, respectively.

### *4.3.3. Co-occurrence analysis of keywords*

The Co-Occurrence analysis of keywords aims to reveal the relationships and trends among research topics. This is accomplished by examining the frequency with which keywords appear together in the literature, thus reflecting the hotspots and current issues in the research field. This section details a Co-Occurrence analysis conducted by the authors on the gathered articles, identifying a total of 2219 keywords from 695 articles. A threshold of 28 was established based on the frequency of keyword occurrence, resulting in 18 keywords that met this criterion. VOSviewer software was utilized to create a scientific landscape of the Co-Occurrence networks, as shown in Fig. 7. The size of each node represents the frequency of keyword occurrence, larger nodes indicate that the keywords are used more frequently and are more significant. The lines connecting the nodes represent the frequency with which pairs of keywords co-occur. A thicker line suggests that the two keywords are mentioned together more often. All keywords were categorized into three clusters: red, green, and blue, summarized in Table 10. Additionally, Table 11 presents the top ten keywords by frequency along with their corresponding Total Link Strengths.

All keywords were categorized into three clusters: red, green, and blue, as summarized in Table 10. Additionally, Table 11 presents the top ten keywords by frequency along with their corresponding Total Link Strengths. The red cluster comprises nine keywords, with “Deep Learning” having the largest node size, indicating that it appears most frequently, totaling 446 occurrences, and has a Total Link Strength of 759. Other frequently occurring keywords include “Image Segmentation” and “Semantic Segmentation,” which appear 75 and 43 times, respectively.

In the green cluster, “Crack Detection” is the leading keyword, appearing 216 times with a Total Link Strength of 529. Other keywords in this cluster include “Computer Vision,” “Image Processing,” “Crack Segmentation,” and “Machine Learning,” with frequencies of 57, 43, 32, and 28, respectively. The blue cluster contains only four keywords, with “CNN” appearing most frequently at 153 times and having a Total Link Strength of 370. This suggests that CNN-based algorithms continue to be a predominant method for crack detection.

The keyword analysis reveals the continuous progression of crack detection research, with deep learning and segmentation techniques remaining central to its development. Additionally, the significant role of machine learning and image processing across various fields highlights potential opportunities for integrating advanced technologies, such as transformer-based models and multimodal data fusion.

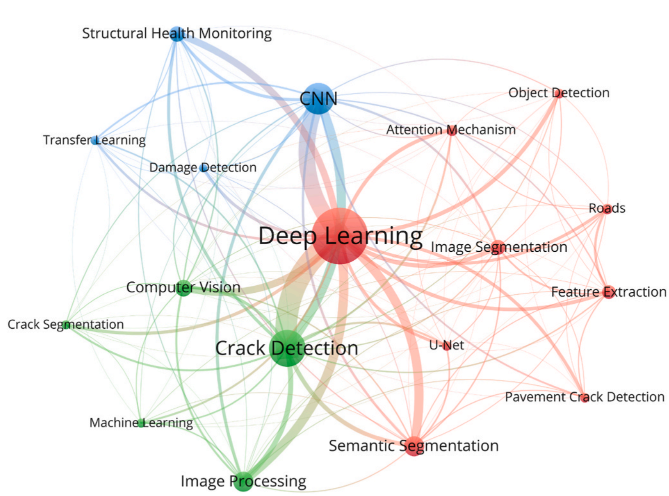


**Fig. 7.** Network visualization of Co-Occurrence analysis of the keywords.

**Table 9**
Co-Authorship indices of the countries.

| Country | Documents | Total Link Strength |
|---|---|---|
| China | 325 | 61 |
| United States | 111 | 53 |
| South Korea | 94 | 10 |
| United Kingdom | 54 | 21 |
| Australia | 48 | 7 |

**Table 10**
Summary of the resulting clusters related to keyword analysis.

| Cluster Color | Keywords | No. of Keywords |
|---|---|---|
| Red | Deep Learning, Image Segmentation, Attention Mechanism, U-Net, Semantic Segmentation, Object Detection, Roads, Feature Extraction, Pavement Crack Detection | 9 |
| Green | Crack Detection, Computer Vision, Image Processing, Crack Segmentation, Machine Learning | 5 |
| Blue | CNN, Damage Detection, Transfer Learning, Structural Health Monitoring | 4 |

**Table 11**
Details of the top 10 keywords.

| Keywords | Frequency | Total Link Strength |
|---|---|---|
| Deep Learning | 446 | 759 |
| Crack Detection | 216 | 529 |
| CNN | 153 | 370 |
| Image Processing | 73 | 192 |
| Convolution | 43 | 172 |
| Image Segmentation | 43 | 159 |
| Concretes | 37 | 155 |
| Semantic Segmentation | 75 | 152 |
| Computer Vision | 57 | 139 |
| Structural Health Monitoring | 51 | 132 |

## 5. Critical analysis

Following the analysis of the collected articles and the bibliometric examination of keywords, the authors concluded that deep learning-based crack detection technology remains a prominent research area, drawing considerable interest from researchers both now and in the future. Therefore, in this section, the authors conducted a comprehensive study of deep learning-based crack detection technology and performed a critical analysis of the relevant articles. They selected 40 articles from the collected data and categorized them according to the type of computer vision employed, specifically classification, object detection, and semantic segmentation. Subsequently, the authors analyzed these articles based on their problem statements, research methods, and results, raising the following questions.

Q1. What deep learning-based method is used in the paper?
Q2. What backbones does the deep learning-based method use?
Q3. What dataset is used in the paper?
Q4. What defects does the paper consider?
Q5. Which loss function is used in the paper?
Q6. What performance levels are reached in the paper?

The answers to these questions are summarized in Tables 12–14 for papers in various categories.

**Table 12**
Summary of Deep Learning techniques for crack classification.

| Ref | Method | Backbone | Dataset | Surface | Loss Function | Performance |
|---|---|---|---|---|---|---|
| Cha et al. (2017) | CNN | – | Own collection | Concrete structure | Softmax | Accuracy = 98% |
| Gopalakrishnan et al. (2017) | DCNN | VGG16 | Own collection | Pavement, Concrete structure | MSE | Accuracy = 97% |
| Dung (2019) | CNN | VGG16 | Own collection | Concrete structure | Binary Cross-Entropy | Accuracy = 97% |
| (Silva and Lucena, 2018) | CNN | VGG16 | Own collection | Concrete structure | – | Accuracy = 92.27% |
| Billah et al. (2019) | CNN | ResNet50 | Own collection | Concrete structure, Road, | – | Accuracy = 94% |
| Paramanandham et al. (2022) | CNN | ResNet50 | Own collection | Concrete structure | Cross-Entropy | Accuracy = 99.92% |
| (Reis, et al., 2021) | ReCRNet | ResNet50 | REB dataset | Concrete structure | Binary Cross-Entropy | Accuracy = 92.3% |
| Li et al. (2023) | CNN | ResNet50 | Own collection | Bridge | SSD | Accuracy = 96.24%, Precision = 97.82% |
| Li et al. (2019) | CNN | AlexNet | Own collection | Bridge | Softmax | Accuracy = 99.09% |
| Palevičius et al. (2022) | CNN | AlexNet | Own collection | Concrete structure | – | Accuracy = 99.41% |
| Islam et al. (2022) | CNN | AlexNet | CCIC | Bridges, Pavement | Cross-Entropy | Accuracy = 99.9%, Precision = 99.92%, Recall = 99.8%, F1-score = 99.89% |

**Table 13**
Summary of Deep Learning techniques for crack object detection.

| Ref | Method | Backbone | Dataset | Surface | Loss Function | Performance |
|---|---|---|---|---|---|---|
| Cha et al. (2018) | Faster R-CNN | – | Own collection | Concrete structure, Steel Corrosion | Regression Loss | mAP = 87.8% |
| (Nie and Wang, 2019) | YOLOv3 | CSPDarknet53 | Own collection | Road | Cross-Entropy | Accuracy = 88% |
| Yao et al. (2021) | YOLOv4 | CSPDarknet53 | Own collection | Concrete structure | Cross-Entropy | mAP = 94.09%, FPS = 44 |
| Ren et al. (2023) | YOLOv5 | Customized CNN | RDD2020 | Road | CIoU | Precision = 79.8% |
| Su et al. (2024) | MOD-YOLO | CSPDarknet53 | Own collection | Concrete structure | – | mAP = 91.1%, FPS = 118.8 |
| Xing et al. (2023) | YOLOv5 | CSPDarknet53 | Own collection | Road | CIoU | mAP = 85% |
| Kao et al. (2023) | YOLOv4 | CSPDarknet53 | SDNET 2018, Own collection | Concrete structure | CIoU | Accuracy = 92% |
| (Qiu and Lau, 2023) | YOLOv2 | ResNet50 | Own collection | Concrete structure | CIoU | Accuracy = 94.54%, FPS = 71.71 |
| Li et al. (2023) | Faster R-CNN | VGG16 | Own collection | Concrete structure | RPNLogLoss, RCNNLogLoss, RPNL1Loss, RCNNL1Loss | Precision = 91.8%, Recall = 93%, F1-score = 92.4% |
| He et al. (2023) | MUENet | YOLOX-S | Own collection | Road | RegLoss, ClsLoss, ObjLoss | mAP = 83.1% |

### 5.1. Classification

Classification is a fundamental task in CV that involves categorizing input images into predefined classes. In crack detection, classification involves identifying and labeling regions in images that contain cracks, and distinguishing them from regions that do not. One of the pioneering efforts in this area was made by (Cha et al., 2017), who proposed a CNN-based method that achieved an impressive accuracy of 98%. This model used a sliding window technique to detect cracks under challenging conditions, such as strong lighting, shadows, and thin cracks. This approach laid the groundwork for subsequent advancements in CNN-based crack detection. Following this (Gopalakrishnan et al., 2017), introduced a pretrained VGG16 network on the ImageNet database, demonstrating transfer learning technique can improve network performance when dealing with diverse pavement surfaces. The VGG16 architecture has stacked convolutional layers that allow it to capture hierarchical features in images, making it well-suited for recognizing crack patterns on varying surfaces. Building on the success of VGG16 (Dung, 2019), compared multiple network architectures, including a shallow CNN, pretrained VGG16 features, and a fine-tuned VGG16 model, to detect cracks. In the fine-tuned VGG16 model, only the final convolutional block and fully connected layer were trained, while the rest of the network was frozen to prevent overfitting and retain pretrained knowledge. This approach was supported by data augmentation and highlighted the effectiveness of transfer learning and fine-tuning for complex crack detection tasks. Similarly, the VGG16 model was also utilized by (Silva and Lucena, 2018) for detecting cracks on concrete surfaces, focusing on transfer learning to address challenges posed by a small dataset. By fine-tuning the learning rate and the number of fully connected layer nodes, the model achieved a promising performance and demonstrated robustness across varying conditions such as lighting, surface texture, and humidity. Building upon the foundation established by VGG16 and other CNN-based methods, the introduction of residual learning with ResNet took crack detection to new heights by enabling deeper architectures without sacrificing accuracy.

The introduction of ResNet (He et al., 2016) was a groundbreaking development that addressed the limitations of traditional CNNs by enabling deeper networks without sacrificing accuracy, significantly advancing the field of crack detection. For instance, ResNet50 was utilized by (Billah et al., 2019) to detect cracks in civil infrastructure, achieving 94% accuracy on the dataset under varying conditions. Similarly, multiple pretrained networks for crack classification, including AlexNet, VGG16, VGG19, and ResNet50, were evaluated by (Paramanandham et al., 2022). The experiments demonstrated that ResNet50 achieved the highest accuracy among these models, highlighting its advantages in terms of performance and efficiency for crack detection. Building on success of ResNet, various improved versions have been proposed to address specific challenges in crack detection. The introduction of ReCRNet by (Reis, et al., 2021) presented a lightweight architecture inspired by ResNet. This network features a streamlined structure, which includes a Stem block, two ResBlocks, a Conv Layers block, and a classifier block. This design enabled fast and accurate classification, outperforming AlexNet and VGG19 in different metrics such as accuracy, precision, recall, and F1-score. The extension

**Table 14**
Summary of Deep Learning techniques for crack segmentation.

| Ref | Method | Backbone | Dataset | Surface | Loss Function | Performance |
|---|---|---|---|---|---|---|
| Cheng et al. (2018) | U-Net | – | CFD, AigleRN | Concrete structure, Pavement | Cross-Entropy | Precision = 92.12%, Recall = 95.7%, F1-score = 93.88% |
| Yu et al. (2022) | RUC-Net | Customized CNN | DeepCrack, Crack500, CFD | Concrete structure, Pavement | Focal Loss | Precision = 88.33%, Recall = 81.2%, F1-score = 84.61%, IoU = 73.33 |
| Al-Huda et al. (2023) | ADDU-Net | Customized CNN | DeepCrack, Crack500, CFD, CrackSC | Concrete structure, Pavement | Dice Loss | Precision = 90.7%, Recall = 92.4%, F1-score = 91.5%, mIoU = 92% |
| Ali et al. (2024) | RS-Net | Customized CNN | DeepCrack, CrackTree, Crack500, CFD | Concrete structure, Pavement | Binary Cross-Entropy | Accuracy = 97.8%, Precision = 72.06%, Recall = 64.41% |
| Fu et al. (2021) | DeepLaBv3+ | – | Own collection | Pavement | – | mIoU = 82.37% |
| Li et al. (2024) | DeepLaBv3+ | – | Crack500, GAPs384 | Pavement | Dice Loss | Accuracy = 98.62%, IoU = 54.91%, Precision = 68.87%, Recall = 72.38% |
| Dung (2019) | FCN | VGG16 | Own collection | Concrete structure | Binary Cross-Entropy | F1-score = 89.3%, AP = 89.3% |
| Yang et al. (2023) | PHCF-Net | U-Net | DeepCrack, CFD | Concrete structure, Pavement | Binary Cross-Entropy | Precision = 96%, Recall = 95.5%, Dice = 90.7%, mIoU = 90.3% |
| (Wang and Su, 2022) | SegCrack | PVT | Own collection | Concrete structure, Pavement | Dice Loss | Precision = 96.66%, Recall = 95.46%, F1-score = 96.05%, mIoU = 92.63% |
| Xiao et al. (2023) | CrackFormer | Transformer | Crack500, CrackTree260, CRKWH100, CrackLS315, Stone331, GAPs384 | Concrete structure, Pavement | Equalized Focal Loss | Precision = 93.76%, Recall = 93.52%, F1-score = 93.64% |
| Yu et al. (2024) | CSTF | Swin Transformer | Crack500, CrackTree260, CRKWH100, CrackLS315, DeepCrack, CFD | Concrete structure, Pavement | Focal Loss | mIoU = 81.3% |
| Wang et al. (2024) | SwinCrack | Swin Transformer | Crack500, Crack260, CrackLS315, Stone331, WHCF218 | Concrete structure, Pavement | – | AP = 80.8%, ODS = 74.8%, OIS = 78.1% |
| Wang et al. (2024) | Crackmer | – | DeepCrack, CrackForest, CrackTree260 | Concrete structure, Pavement | Binary Cross-Entropy | Accuracy = 98.49%, F1-score = 85.68%, mIoU = 87.04% |
| Xiong et al. (2024) | DefNet | – | CrackLS315, CrackTree260, DeepCrack537 | Concrete structure, Pavement | BCE_Loss Dice_Loss | Precision = 91.31%, F1-score = 85.57% |
| Lu et al. (2024) | Crack_PSTU | – | CFD | Concrete structure, Pavement | BCELoss | Precision = 96.95%, F1-score = 95.61%, Recall = 94.83% |
| Zhao et al. (2021) | PANet | – | Own collection | Concrete structure | – | Accuracy = 80.95%, F1-score = 66.12%, IoU = 48.82% |
| Beckman et al. (2019) | Faster R-CNN | – | Own collection | Concrete structure | – | AP = 90.79% |
| Lin et al. (2023) | YOLACT++ | – | Own collection | Pavement | – | mAP = 82%, mIoU = 75% |
| Pan et al. (2023) | 3D crack detection | – | CFD, Crack500 | Concrete structure | – | mIoU = 53.73% |

of ResNet application by (Li et al., 2023) involved using ResNet34 and ResNet50 for binary and multi-label classification of pavement distress. The results showed ResNet50 achieved 96.24% accuracy in binary classification and 90.257% in multi-label classification, meeting the Chinese standard (JTG H202018) for road distress detection.

AlexNet, a foundational network in crack detection classification tasks, has been widely employed and optimized for enhanced performance in various specialized applications. The study by (Li et al., 2019) explored the impact of different parameter settings, such as learning rates, on the performance of AlexNet. Through extensive experimentation and continuous adjustments to the architecture and hyperparameters, the network achieved a validation accuracy of 99.06% on high-resolution images, demonstrating its robust crack detection performance. The model was further integrated into a smartphone application, offering a mobile tool for on-site crack detection. The generalization of AlexNet under complex lighting conditions was improved by (Palevičius et al., 2022) through the introduction of a shadow augmentation technique, which generated a dataset rich in shadowed crack samples. By training AlexNet on this augmented dataset, the network demonstrated superior performance in classifying cracks, outperforming its performance on non-augmented datasets. Additionally, the effectiveness of transfer learning was demonstrated by comparing AlexNet with deeper networks like ResNet18 and DenseNet161 (Islam et al., 2022). By leveraging a pretrained AlexNet on a large-scale dataset, the network achieved superior accuracy of 99.90% on the CCIC dataset.

In conclusion, advancements in crack classification have been significantly influenced by the integration of CNNs, transfer learning, and data augmentation techniques. Transfer learning has proven particularly effective, with pretrained models like VGG16 and AlexNet achieving strong performance across varying pavement surfaces. The introduction of deeper architectures such as ResNet has further advanced the field by enabling deeper models without sacrificing accuracy. Additionally, data augmentation techniques have proven essential in improving network performance. By increasing the dataset's diversity through transformations like shadow enhancement, rotation, and others, the network will exhibit more generalization and robustness across varied scenarios.

### 5.2. Object detection

The application of object detection techniques in crack detection has gained significant attention due to their ability to localize damage through bounding-box predictions. Pioneering works like Faster R-CNN and YOLO revolutionized object detection by introducing bounding-box-based localization, making it possible to simultaneously identify and localize cracks within large-scale images. Faster R-CNN adopts a two-stage process, where a region proposal network first generates potential object regions, followed by a classification step. In contrast, YOLO streamlines the process using a single-stage detection framework, enabling real-time object detection. Building on these foundational innovations, a crack detection network based on Faster R-CNN was developed by (Cha et al., 2018) to detect five types of structural damage, including concrete cracks. By generating bounding boxes around regions of interest (ROIs) and identifying the most likely damage areas through a defined threshold, the proposed network demonstrated superior performance across various defect categories, achieving an average precision of 87.8%. In contrast, a YOLOv3-based method was proposed by (Nie and Wang, 2019) to overcome the limitations of traditional detection techniques, which often suffer from lower accuracy and slower real-time performance. Compared to previous CNN-based methods and conventional approaches, the YOLOv3 model demonstrated faster detection speeds and improved accuracy, highlighting its effectiveness for crack detection in road maintenance applications.

With the continuous development of the YOLO series of networks, more crack detection studies based on YOLO have been widely proposed. A lightweight crack detection network based on YOLOv4 with CSPDarknet53 as its backbone was introducted by (Yao et al., 2021). By incorporating separable convolutions and an optimized spatial pyramid pooling (SPP) module, the model achieved a mAP of 94.09%. Separable convolutions reduce computational complexity by breaking down the standard convolution operation into smaller, more efficient tasks. Meanwhile, the optimized SPP enhances the model's ability to capture multi-scale features, improving overall performance in detecting cracks. YOLOv5 was enhanced for large-scale road inspection by (Ren et al., 2023) through the incorporation a Generalized Feature Pyramid Network (Generalized-FPN), which facilitates effective cross-scale feature fusion. The Generalized-FPN uses two types of connections: skip-layer and cross-scale connections. The skip-layer connection preserves low-level features, while the cross-scale connection fuses information from different layers, enhancing the model's ability to detect road damage at multiple scales. MOD-YOLO was developed by (Su et al., 2024) through the introduction of innovative modules, including the Maintaining the Original Information-Deeply Separable Convolution (MODSConv) and Global Receptive Field-Space Pooling Pyramid-Fast. These modules improved channel information retention and multi-scale feature extraction, leading to a 27.5% accuracy increase compared to YOLOX while reducing computational complexity.

As the YOLO series continues to evolve and improve, its integration with unmanned aerial vehicles (UAVs) has opened up new possibilities for large-scale and real-time crack detection in infrastructure. UAVs offer mobility and high-resolution imaging, making them ideal for applications where traditional inspection methods are limited. Recent developments have explored the potential of integrating UAVs with YOLO models to perform crack detection tasks. The YOLOv5 architecture was enhanced by (Xing et al., 2023) through the integration a Swin Transformer and Bidirectional Feature Pyramid Network (BIFPN) for multi-scale feature fusion, enabling pixel-level precision in detecting cracks as narrow as 1.2 mm. The proposed method achieved a crack detection accuracy of 90% and a real-time processing speed of 43.5 FPS during performance evaluation using UAV-captured highway pavement images. Challenges in bridge inspections, such as poor lighting and limited accessibility, were addressed by (Kao et al., 2023) through the use of a UAV-mounted camera and a YOLOv4 model to detect cracks and quantify their dimensions using image processing and scaling methods. Expanding on this concept, a UAV-based crack detection system using YOLO was developed by (Qiu and Lau, 2023), who investigated the impact of different YOLO architectures on detection performance. The authors replaced YOLO's feature extraction network with ResNet50 and, through ablation experiments, demonstrated that YOLOv2 and YOLOv4-Tiny exhibited superior performance. In addition, both models achieved real-time detection speeds of over 80 FPS, making them suitable for real-time monitoring tasks. In bridge inspections, the DJI M210-RTK UAV was utilized by (Li et al., 2023) in combination with the Faster R-CNN model based on VGG16 to detect structural cracks and investigate the optimal flight distance of the UAV for capturing surface details of the structure. Experimental results demonstrated that the proposed method outperformed other networks used for comparison across various performance metrics. Additionally, the study concluded that the optimal flight distance for the UAV is 8.2 m, at which the network is capable of detecting cracks as small as 0.2 mm. Meanwhile, a UAV-based road crack detection algorithm utilizing the advanced MUENet architecture was proposed by (He et al., 2023). The network leverages the proposed Main and Auxiliary Dual-Path Module (MADPM) to effectively extract the complex morphological features of cracks and enhance detection accuracy. Evaluated on UAV-captured near-far scene images, MUENet demonstrated significantly superior performance in both accuracy and detection speed compared to other detection algorithms.

Overall, crack detection through object detection has evolved into a rapidly advancing domain, driven by innovations in network design, loss optimization, and real-world deployment strategies. CSPDarknet53 has been widely adopted in YOLO-based models due to its ability to balance high-resolution feature extraction and computational efficiency. This backbone enhances detection accuracy by capturing both fine-grained and global features while maintaining real-time processing capabilities, making it ideal for complex crack detection tasks. Another critical factor is the strategic use of loss functions like CIoU and Cross-Entropy, which have effectively addressed the challenges of accurate object localization and robust classification by minimizing false positives and improving precision. Moreover, the integration of UAVs has expanded the practical applications of crack detection systems. UAV-based systems equipped with advanced object detectors like YOLOv5 and Faster R-CNN allow for the inspection of hard-to-reach areas, improving operational efficiency and enabling large-scale, real-time monitoring. As research continues with hybrid architectures and feature fusion techniques, future crack detection systems are expected to be more adaptive and seamlessly integrated into broader structural health monitoring frameworks.

### 5.3. Segmentation

Semantic segmentation plays a crucial role in crack detection tasks by providing pixel-level classification, where each pixel in an image is assigned a label corresponding to either crack or non-crack regions. Unlike object detection methods that rely on bounding boxes for localization, segmentation approaches allow for precise delineation of crack boundaries, making them particularly effective for applications in civil infrastructure maintenance, where accurate crack morphology is critical for quantification and structural assessments. Recent research has introduced various deep learning models and their extensions to effectively detect cracks across diverse environmental conditions and structural scenarios.

An automatic crack detection method using the U-Net architecture was introducted by (Cheng et al., 2018), leveraging an encoder-decoder structure to process entire images without the need for mini-patches. To enhance pixel-level segmentation precision, a novel cost function based on distance transform was employed. Evaluations on road crack datasets demonstrated accuracies exceeding 92%, outperforming traditional methods. Despite its success, U-Net exhibited limitations, such as blurry feature maps and errors in skip connections, which hindered performance in more complex scenarios. To address these limitations, RUC-Net was developed by (Yu et al., 2022), combining U-Net with ResNet and introducing spatial-channel squeeze and excitation (scSE) attention modules to enhance relevant feature extraction. Additionally, RUC-Net employed a focal loss function to mitigate class imbalance issues commonly found in crack segmentation tasks, resulting in improved segmentation performance under complex backgrounds. Building on attention-based improvements, the Asymmetric Dual-Decoder U-Net (ADDU-Net) was introducted by (Al-Huda et al., 2023) to tackle the segmentation of both thick and thin cracks on pavement surfaces under diverse environmental conditions. By integrating a dual attention module and an asymmetric dual-decoder design, ADDU-Net effectively captured features of both large and fine cracks. Its comprehensive design demonstrated superior segmentation accuracy and robustness across various datasets, further extending the capabilities of U-Net-based architectures. The Residual Sharp U-Net (RS-Net) was proposed by (Ali et al., 2024) to address remaining challenges related to feature extraction and skip connections. RS-Net incorporated residual blocks to enhance feature extraction and replaced standard skip connections with a sharpening kernel filter to refine feature maps. This design resulted in significant performance improvements, and demonstrated superior results in both crack segmentation and severity assessment compared to previous U-Net variants.

In parallel, the DeepLab family of architectures has also made significant contributions to crack segmentation. The DeepLabv3+ network was enhanced by (Fu et al., 2021) by incorporating a densely connected ASPP module to expand the receptive field and improve pixel-level detail extraction. The model demonstrated superior segmentation accuracy in bridge crack detection. A pavement crack segmentation approach was propsoed by (Li et al., 2024) through the integration of an External Attention Block into the ASPP module of the DeepLabV3+. This modification allowed for improved long-range feature extraction and better context aggregation, which is critical for accurate segmentation in complex environments. Additionally, Fully Convolutional Network (FCN)-based models have also been influential in advancing crack segmentation. A deep FCN model was explored in (Dung, 2019), which used VGG16 as the encoder backbone for semantic segmentation of concrete cracks. To further address issues related to incomplete and discontinuous cracks, PHCF-Net was proposed in (Yang et al., 2023), which integrates progressive context fusion (PCF) and hierarchical context fusion (HCF) blocks. By aggregating local and global context information efficiently and enhancing feature extraction at various scales using a multi-scale context fusion (MCF) block, PHCF-Net consistently outperformed mainstream segmentation networks on publicly available datasets.

Although CNN-based models have demonstrated strong performance in crack segmentation, they still face limitations in capturing long-range dependencies and contextual information due to their relatively small receptive fields. To address these challenges, recent advancements have turned toward transformer-based architectures, which can effectively model global dependencies while maintaining pixel-level detail. A transformer-based model SegCrack was proposed by (Wang and Su, 2022) to focus on diverse field inspection scenarios. The hierarchically structured transformer encoder of SegCrack generates multiscale features that are progressively up-sampled and fused through a top-down pathway with lateral connections. An online hard example mining strategy was applied to prioritize difficult samples during training, improving detection accuracy in challenging environments. SegCrack laid the foundation for integrating multiscale feature extraction and robust training strategies into transformer-based segmentation networks. Building on this concept, CrackFormer, a transformer-based hybrid-window attentive vision framework, was introduced by (Xiao et al., 2023) to address challenges posed by complex pavement crack patterns and environmental conditions. CrackFormer's hybrid-window self-attention mechanism combines dense local windows for fine-grained feature extraction with sparse global windows for capturing contextual information. Further extending the capabilities of hybrid attention mechanisms, CSTF, a segmentation network that leverages the Swin Transformer encoder, was proposed by (Yu et al., 2024). Unlike traditional models that struggle with detecting long and thin cracks, CSTF utilizes hierarchical and shifted window attention mechanisms to capture both local and global semantic information. A feature pyramid pooling module and dual-branch decoder further enhance its ability to detect cracks at multiple scales while preserving crucial details. SwinCrack, an end-to-end detection network proposed by (Wang et al., 2024, simulates long-range interactions while preserving local details through adaptive spatial aggregation. SwinCrack differs from CSTF by integrating advanced convolutional modules, such as the Convolutional Patch Embedding Layer (CPEL) and Convolutional Swin Transformer Block (CSTB), to improve spatial context modeling. The Depth-convolution Forward Network (DFN) enhances local feature extraction, while the Convolutional Attention Gated Skip Connection (CAGSC) suppresses background interference, leading to superior performance in complex real-world crack segmentation tasks.

While CNNs and transformers have individually driven significant advancements in crack segmentation, they face inherent limitations: CNNs struggle with limited receptive fields, and transformers often fail to capture fine local details, which can hinder performance in complex real-world scenarios. CNNs are effective in extracting localized features but have difficulty capturing global context, while transformers excel in modeling long-range dependencies but often underperform in extracting fine-grained details. By combining these strengths, recent studies highlight the potential of hybrid networks for more effective crack segmentation. A dual-path network for pavement crack segmentation was developed by (Wang et al., 2024), demonstrating the benefits of combining CNNs and transformers. The network pairs a lightweight CNN encoder, which efficiently captures localized features, with a transformer encoder using a fully convolutional high-low frequency attention (FCHiLo) mechanism for long-range contextual information. Experimental results revealed that this hybrid architecture outperformed both CNN-based and transformer-based models, achieving superior segmentation accuracy and robustness. The need for effective local and global feature extraction was addressed by (Xiong et al., 2024) through the proposal of DefNet, a multi-scale dual-encoding fusion network. DefNet integrates a dual-branch attention transformer (DBAT) that combines large kernel attention (LKA) and channel attention to capture rich global context. Meanwhile, the CNN encoding branch uses depthwise separable convolutions to extract fine-grained local details. A feature enhancement module further integrates multi-scale features by aggregating spatial and channel information from different receptive fields. Crack_PSTU, proposed by (Lu et al., 2024), integrates the U-Net framework with a pre-trained Swin Transformer encoder to handle the complexity of irregular crack patterns. The Swin Transformer focuses on long-range feature extraction, while the U-Net decoder refines segmentation boundaries. This combination enables the network to capture both global context and local details, enhancing segmentation performance.

Accurate crack segmentation provides detailed contours of damage, but in real-world structural health monitoring, it is equally vital to quantify the geometric characteristics of cracks. Measurements such as crack width, depth, and length play a crucial role in assessing damage severity and informing targeted maintenance and repair strategies. To address these challenges, recent works have explored diverse quantification techniques, ranging from segmentation-based 2D measurements to integrating depth data and 3D crack profiling. One of the foundational approaches in 2D crack quantification was introduced by (Zhao et al., 2021), who extended the PANet model with a semantic branch to enhance crack segmentation and ensure continuous representations. This refinement reduces errors caused by discontinuous crack geometries, addressing a common challenge in traditional segmentation-based methods. A key innovation of this approach is the integration of the A* algorithm, which accurately calculates crack length and width from segmented binary images. Unlike traditional methods that often face difficulties in handling fragmented or complex crack paths, the A* algorithm excels in tracing optimal paths along crack contours, enabling precise and continuous 2D measurements.

Expanding beyond 2D segmentation, several studies have explored 3D quantification techniques to provide more comprehensive geometric measurements. An early approach for detecting and quantifying concrete spalling using a Faster R-CNN model integrated with a depth sensor was proposed by (Beckman et al., 2019). Unlike methods limited to 2D crack measurements, this approach automatically detects multiple instances of spalling and computes their respective volumes. By merging segmentation outputs with depth data and generating 3D point clouds, the method enables precise volume estimation through geometric reconstruction techniques, offering an efficient and scalable solution for monitoring surface damage in concrete structures. Building on the use of depth data, a cost-effective crack detection and quantification method was developed by (Lin et al., 2023), integrating an RGB-D camera with an instance segmentation algorithm. The 2D segmentation provides precise crack contours, while the depth data from the RGB-D camera enables the generation of 3D point clouds. This combination bridges the gap between traditional 2D methods and depth-based quantification, allowing for accurate measurements of crack length, width, and depth and significantly improving the accuracy of pavement distress evaluation. Further advancing the use of 3D spatial data, A one-stage crack detection and quantification method was proposed by (Pan et al., 2023), utilizeing high-resolution 3D pavement profiles generated by a low-cost stereo imaging system. This method directly extracts crack maps and measures length, width, and depth from the 3D profiles. Crack depth is calculated using point cloud fitting and depth estimation techniques, providing precise measurements that outperform traditional 2D-based approaches. The integration of stereo disparity data with crack contours further enhances the accuracy of crack morphology representation, making the method highly effective for practical applications in real-world scenarios.

In conclusion, the evolution of crack segmentation methods reveals a progression from CNN-based methods to advanced transformer-based and hybrid models. Initially, CNN-based approaches focused on localized feature extraction, but their limitations in capturing long-range dependencies and global context were gradually addressed through the development of transformer architectures. Transformers introduced the ability to model global dependencies, but they encountered challenges in preserving local details. To overcome these limitations, hybrid models that combine the strengths of CNNs and transformers have emerged, offering superior performance by balancing local feature extraction with global context understanding. This integration has enhanced segmentation robustness, particularly under challenging conditions involving discontinuous or irregular cracks. Simultaneously, traditional 2D measurements have evolved into comprehensive 3D crack profiling through the integration of depth information, such as that provided by RGB-D cameras, stereo imaging systems, and depth sensors. These technologies enable accurate assessments of crack width, depth, and length, which are essential for evaluating damage severity and developing targeted maintenance strategies. As segmentation and quantification methods converge, advanced systems are now capable of transforming pixel-level classifications into actionable geometric insights, bridging the gap between surface-level analysis and holistic structural health monitoring.

## 6. Future outlook

### *6.1. Enhancing model generalization*

A key challenge in current crack detection research is the limited generalization of models across diverse environments and materials. Most existing models are optimized for specific datasets, leading to performance degradation when applied to new scenarios, such as different infrastructure types or varying environmental conditions. While techniques like data augmentation and transfer learning have been employed to improve generalization, they are insufficient for addressing the full scope of this issue. To address this limitation, future research should focus on the development of generalized architectures capable of handling diverse environments. One promising approach is the application of large-scale pre-trained models, such as SAM and GPT-4 Vision, which are trained on extensive datasets and provide rich, transferable feature representations. By fine-tuning these models for crack detection tasks, their generalization capabilities across various structure types and environmental conditions can be significantly enhanced, reducing the reliance on task-specific data collection and model retraining.

### *6.2. Improving computational efficiency*

Many existing crack detection models face computational challenges, including large parameter sizes, high memory consumption, and slow detection speeds. These issues limit the ability of models to perform real-time processing and hinder their deployment in practical engineering applications. Current methods often require crack images to be collected on-site and processed offline using high-powered computers, leading to delays in decision-making. Future research should prioritize the development of lightweight, real-time models optimized for deployment on devices and mobile platforms. By designing models with reduced parameter sizes and computational requirements, on-site crack detection can be achieved efficiently without the need for high-powered external hardware. Techniques such as model pruning, quantization, and knowledge distillation could be explored to create compact models capable of delivering precise results while ensuring fast response times.

### *6.3. Advancing multi-modal data fusion models*

Current crack detection approaches primarily rely on visual data, limiting their ability to detect hidden or subsurface cracks and reducing robustness under challenging environmental conditions. Incorporating multiple data modalities could provide a more comprehensive understanding of crack characteristics and progression. Future work should explore multi-modal data fusion by combining visual data with complementary inputs from depth sensors, thermal cameras, LiDAR, and others. By integrating information from multiple sources, crack detection systems could achieve enhanced accuracy, even in scenarios where visual cues are insufficient.

### 6.4. Developing crack predictive and decision-making models

Current crack detection and quantification models often provide geometric measurements such as crack width, depth, and length. However, these outputs are frequently disconnected from practical repair strategies or risk assessment frameworks, limiting their effectiveness in real-world maintenance planning. Furthermore, most existing models are static, detecting cracks at a specific point in time without accounting for future crack progression or structural deterioration.

Future research should prioritize the development of integrated systems that link crack quantification outputs with risk evaluation models and decision-making algorithms. These systems would automatically translate geometric measurements into actionable maintenance recommendations or prioritize repair actions based on the severity and expected progression of crack.

Additionally, crack evolution prediction models integrating machine learning with CV techniques should be developed to anticipate how cracks will progress over time. By accurately forecasting future damage trajectories, these models can provide maintenance strategies, reducing the risk of structural damage and ensuring long-term infrastructure safety. The combination of predictive models and decision-making frameworks will enable smarter, data-driven interventions, improving the efficiency and reliability of infrastructure maintenance.

### 6.5. Establishing standardized and large-scale datasets

The field of crack detection currently lacks standardized and large-scale datasets that cover diverse surface types, environmental conditions, and crack morphologies. Current datasets are often limited in scope, preventing models from learning robust representations applicable to varied real-world conditions. Additionally, inconsistencies in data collection and annotation limit the reproducibility and hinder fair performance comparisons across various crack detection approaches. Hence, the establishment of comprehensive, standardized datasets is essential for improving model generalization and robustness. These datasets should include a wide range of real-world conditions and infrastructure types, ensuring that models can be effectively trained and evaluated. Standardized data collection protocols and annotation guidelines will further enhance consistency and reproducibility across studies.

## 7. Conclusion

In this paper, the authors present a comprehensive review of deep learning-based crack detection research from 2017 to 2024, combining bibliometric and critical analyses. The bibliometric analysis highlights influential articles, key authors, major journals, and international collaborations that have shaped the field, while keyword analysis confirms deep learning as the dominant approach in crack detection research. The critical review focuses on the significant contributions of deep learning models, including CNNs, YOLO-based object detectors, and transformer networks, in addressing crack classification, object detection, and semantic segmentation tasks. Techniques such as attention mechanisms, feature pyramid networks, and multi-scale context fusion have significantly enhanced detection accuracy and robustness under diverse environmental conditions.

Despite significant advancements, key challenges persist in areas such as model generalization, computational efficiency, multi-modal data integration, and the practical use of crack quantification outputs for decision-making. Addressing these challenges requires the development of generalized models through large-scale pre-trained architectures, lightweight, real-time solutions to improve computational efficiency, and multi-modal data fusion to enhance detection accuracy under diverse conditions. Equally important is the establishment of standardized and large-scale datasets with consistent collection and annotation protocols, which will improve model robustness, reproducibility, and fair performance evaluations. Additionally, integrating crack prediction models with decision-making frameworks and maintenance planning will ensure long-term infrastructure safety and reliability.

Overall, this review offers researchers a comprehensive framework for understanding the evolving landscape of deep learning in crack detection. It summarizes key findings in the field while also identifying critical areas for future investigation, including existing research gaps. The findings of this review are poised to inform both academic pursuits and practical applications, ultimately contributing to the advancement of the crack detection field. By highlighting unresolved challenges and potential directions, this review serves as a valuable resource for future researchers, guiding them in addressing current limitations and inspiring innovative methodologies to enhance the accuracy, robustness, and efficiency of crack detection technologies.

## CRediT authorship contribution statement

**Yingjie Wu:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Shaoqi Li:** Writing – review & editing, Visualization, Project administration, Methodology, Conceptualization. **Jingqiu Li:** Visualization, Writing – review & editing. **Yanping Yu:** Visualization, Writing – review & editing. **Jianchun Li:** Writing – review & editing, Visualization. **Yancheng Li:** Writing – review & editing, Supervision, Project administration, Conceptualization.

## Data availability

Data will be available on request.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

Ahmed, T.U., Hossain, M.S., Alam, M.J., Andersson, K.. An integrated CNN-RNN framework to assess road crack. https://doi.org/10.1109/ICCIT48885.2019.9038607.

Ai, D., Jiang, G., Lam, S.-K., He, P., Li, C., 2023. Computer vision framework for crack detection of civil infrastructure—A review. Eng. Appl. Artif. Intell. 117, 105478. https://doi.org/10.1016/j.engappai.2022.105478.

Akbari, J., Ahmadifarid, M., Amiri, A.K. (Eds.), 2020. Multiple crack detection using wavelet transforms and energy signal techniques. Frat. ed Integrita Strutt 52 (14), 269–280. https://doi.org/10.3221/IGF-ESIS.52.21.

Al-Huda, Z., Peng, B., Algburi, R.N.A., Al-antari, M.A., Rabea, A.-J., Al-maqtari, O., Zhai, D., 2023. Asymmetric dual-decoder-U-Net for pavement crack semantic segmentation. Autom. ConStruct. 156, 105138. https://doi.org/10.1016/j.autcon.2023.105138.

Alexander, Q.G., Hoskere, V., Narazaki, Y., Maxwell, A., Spencer Jr., B.F., 2022. Fusion of thermal and RGB images for automated deep learning based crack detection in civil infrastructure. AI Civ. Eng 1, 3. https://doi.org/10.1007/s43503-022-00002-y.

Ali, L., AlJassmi, H., Swavaf, M., Khan, W., Alnajjar, F., 2024. Rs-net: residual Sharp U-Net architecture for pavement crack segmentation and severity assessment. J. Big Data 11, 116. https://doi.org/10.1186/s40537-024-00981-y.

Ali, L., Alnajjar, F., Jassmi, H.A., Gocho, M., Khan, W., Serhani, M.A., 2021. Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. Sensors 21, 1688. https://doi.org/10.3390/s21051688.

Ali, R., Chuah, J.H., Talip, M.S.A., Mokhtar, N., Shoaib, M.A., 2022. Structural crack detection using deep convolutional neural networks. Autom. ConStruct. 133, 103989. https://doi.org/10.1016/j.autcon.2021.103989.

Alipour, M., Harris, D.K., Miller, G.R., 2019. Robust pixel-level crack detection using deep fully convolutional neural networks. J. Comput. Civ. Eng. 33, 04019040. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000854.

Awadallah, O., Sadhu, A., 2023. Automated multiclass structural damage detection and quantification using augmented reality. J. Infrastruct. Intell. Resil. 2, 100024. https://doi.org/10.1016/j.iintel.2022.100024.

Beckman, G.H., Polyzois, D., Cha, Y.-J., 2019. Deep learning-based automatic volumetric damage quantification using depth camera. Autom. ConStruct. 99, 114–124. https://doi.org/10.1016/j.autcon.2018.12.006.

Billah, U.H., La, H.M., Tavakkoli, A., Gucunski, N., 2019. Classification of concrete crack using deep residual network. In: Proceedings of the 9th International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII-9), pp. 1–6. Missouri, USA. https://ara.cse.unr.edu/wp-content/uploads/2014/12/SHMII-9_Full-paper-Templatefinalv4.pdf.
Cao, W., Liu, Q., He, Z., 2020. Review of pavement defect detection methods. IEEE Access 8, 14531–14544. https://doi.org/10.1109/ACCESS.2020.2966881.
Cha, Y.J., Choi, W., Büyüköztürk, O., 2017. Deep learning-based crack damage detection using convolutional neural networks. Comput. Aided Civ. Infrastruct. Eng. 32, 361–378. https://doi.org/10.1111/mice.12263.
Cha, Y.J., Choi, W., Suh, G., Mahmoudkhani, S., Büyüköztürk, O., 2018. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. Comput. Aided Civ. Infrastruct. Eng. 33, 731–747. https://doi.org/10.1111/mice.12334.
Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. arXiv: 1802.02611v3. In: . https://doi.org/10.48550/arXiv.1802.02611.
Chen, S.-Y., Wang, Y.-W., Ni, Y.-Q., Zhang, Y., 2024. When transfer learning meets dictionary learning: a new hybrid method for fast and automatic detection of cracks on concrete surfaces. Struct. Control Health Monit. 2024, 3185640. https://doi.org/10.1155/2024/3185640.
Cheng, J., Xiong, W., Chen, W., Gu, Y., Li, Y.. Pixel-level crack detection using U-Net. https://doi.org/10.1109/TENCON.2018.8650059.
Deng, J., Singh, A., Zhou, Y., Lu, Y., Lee, V.C.-S., 2022. Review on computer vision-based crack detection and quantification methodologies for civil structures. Constr. Build. Mater. 356, 129238. https://doi.org/10.1016/j.conbuildmat.2022.129238.
Dorafshan, S., Thomas, R.J., Maguire, M., 2018. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. Constr. Build. Mater. 186, 1031–1045. https://doi.org/10.1016/j.conbuildmat.2018.08.011.
Dosovitskiy, A., 2020. An image is worth 16x16 words: transformers for image recognition at scale. arXiv:2010.11929. . https://doi.org/arXiv:2010.11929.
Dung, C.V., 2019. Autonomous concrete crack detection using deep fully convolutional neural network. Autom. ConStruct. 99, 52–58. https://doi.org/10.1016/j.autcon.2018.11.028.
Fu, H., Meng, D., Li, W., Wang, Y., 2021. Bridge crack semantic segmentation based on improved Deeplabv3+. J. Mar. Sci. 9 (6), 671. https://doi.org/10.3390/jmse9060671.
Gopalakrishnan, K., 2018. Deep learning in data-driven pavement image analysis and automated distress detection: A review. Data 3 (3), 28. https://doi.org/10.3390/data3030028.
Gopalakrishnan, K., Khaitan, S.K., Choudhary, A., Agrawal, A., 2017. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. Constr. Build. Mater. 157, 322–330. https://doi.org/10.1016/j.conbuildmat.2017.09.110.
Guo, F., Qian, Y., Liu, J., Yu, H., 2023. Pavement crack detection based on transformer network. Autom. ConStruct. 145, 104646. https://doi.org/10.1016/j.autcon.2022.104646.
Hang, J., Wu, Y., Li, Y., Lai, T., Zhang, J., Li, Y., 2023. A deep learning semantic segmentation network with attention mechanism for concrete crack detection. Struct. Health Monit. 22, 3006–3026. https://doi.org/10.1177/14759217221126170.
Hasni, H., Alavi, A.H., Jiao, P., Lajnef, N., 2017. Detection of fatigue cracking in steel bridge girders: a support vector machine approach. Archiv. Civ. Mech. Eng. 17, 609–622. https://doi.org/10.1016/j.acme.2016.11.005.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. arXiv:1512.03385. In: . https://doi.org/10.48550/arXiv.1512.03385.
He, X., Tang, Z., Deng, Y., Zhou, G., Wang, Y., Li, L., 2023. UAV-based road crack object-detection algorithm. Autom. ConStruct. 154, 105014. https://doi.org/10.1016/j.autcon.2023.105014.
Hsieh, Y.-A., Tsai, Y.J., 2020. Machine learning for crack detection: review and model performance comparison. J. Comput. Civ. Eng. 34, 04020038. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000918.
Hu, F., Gou, H.-y., Yang, H.-z., Yan, H., Ni, Y.-q., Wang, Y.-w., 2025. Automatic PAUT crack detection and depth identification framework based on inspection robot and deep learning method. J. Infrastruct. Intell. Resil. 4 (1), 100113. https://doi.org/10.1016/j.iintel.2024.100113.
Hu, W., Wang, W., Ai, C., Wang, J., Wang, W., Meng, X., Liu, J., Tao, H., Qiu, S., 2021. Machine vision-based surface crack analysis for transportation infrastructure. Autom. ConStruct. 132, 103973. https://doi.org/10.1016/j.autcon.2021.103973.
Huang, H.-w., Li, Q.-t., Zhang, D.-m., 2018. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. Tunn. Undergr. Sp. Tech. 77, 166–176. https://doi.org/10.1016/j.tust.2018.04.002.
Islam, M.M., Hossain, M.B., Akhtar, M.N., Moni, M.A., Hasan, K.F., 2022. CNN based on transfer learning models using data augmentation and transformation for detection of concrete crack. Algorithms 15 (8), 287. https://doi.org/10.3390/a15080287.
Ji, A., Xue, X., Wang, Y., Luo, X., Xue, W., 2020. An integrated approach to automatic pixel-level crack detection and quantification of asphalt pavement. Autom. ConStruct. 114, 103176. https://doi.org/10.1016/j.autcon.2020.103176.
Kabir, S., Rivard, P., 2007. Damage classification of concrete structures based on grey level co-occurrence matrix using Haar's discrete wavelet transform. Comput. Concr. 4 (3), 243–257. https://doi.org/10.12989/cac.2007.4.3.243.
Kalfarisi, R., Wu, Z.Y., Soh, K., 2020. Crack detection and segmentation using deep learning with 3D reality mesh model for quantitative assessment and integrated visualization. J. Comput. Civ. Eng. 34 (4), 04020010. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000890.
Kao, S.-P., Chang, Y.-C., Wang, F.-L., 2023. Combining the YOLOv4 deep learning model with UAV imagery processing technology in the extraction and quantization of cracks in bridges. Sensors 23 (5), 2572. https://doi.org/10.3390/s23052572.
Kheradmandi, N., Mehranfar, V., 2022. A critical review and comparative study on image segmentation-based techniques for pavement crack detection. Constr. Build. Mater. 321, 126162. https://doi.org/10.1016/j.conbuildmat.2021.126162.
König, J., Jenkins, M., Mannion, M., Barrie, P., Morison, G., 2022. What's cracking? A review and analysis of deep learning methods for structural crack segmentation, detection and quantification. arXiv 2202, 03714. https://doi.org/arXiv.2202.03714.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90. https://doi.org/10.1145/3065386.
Li, D., Duan, Z., Hu, X., Zhang, D., Zhang, Y., 2023. Automated classification and detection of multiple pavement distress images based on deep learning. J. Traffic Transp. Eng. 10 (2), 276–290. https://doi.org/10.1016/j.jtte.2021.04.008.
Li, F., Mou, Y., Zhang, Z., Liu, Q., Jeschke, S., 2024. A novel model for the pavement distress segmentation based on multi-level attention DeepLabV3+. Eng. Appl. Artif. Intell. 137, 109175. https://doi.org/10.1016/j.engappai.2024.109175.
Li, R., Yu, J., Li, F., Yang, R., Wang, Y., Peng, Z., 2023. Automatic bridge crack detection using Unmanned aerial vehicle and Faster R-CNN. Constr. Build. Mater. 362, 129659. https://doi.org/10.1016/j.conbuildmat.2022.129659.
Li, S., Zhao, X., 2019. Image-based concrete crack detection using convolutional neural network and exhaustive search technique. Adv. Civ. Eng. 6520620. https://doi.org/10.1155/2019/6520620.
Li, S., Zhao, X., 2021. Pixel-level detection and measurement of concrete crack using faster region-based convolutional neural network and morphological feature extraction. Meas. Sci.Technol. 32 (6), 065010. https://doi.org/10.1088/1361-6501/abb274.
Li, S., Zhao, X., Zhou, G., 2019. Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. Comput. Aided Civ. Infrastruct. Eng. 34 (7), 616–634. https://doi.org/10.1111/mice.12433.
Lin, W., Li, X., Han, H., Yu, Q., Cho, Y.-H., 2023. A novel approach for pavement distress detection and quantification using RGB-D camera and deep learning algorithm. Constr. Build. Mater. 407, 133593. https://doi.org/10.1016/j.conbuildmat.2023.133593.
Liu, H., Miao, X., Mertz, C., Xu, C., Kong, H.. Crackformer: transformer network for fine-grained crack detection. https://doi.org/10.1109/ICCV48922.2021.00376.
Lslam, M.M., Hossain, M.B., Akhtar, M.N., Moni, M.A., Hasan, K.F., 2022. CNN based on transfer learning models using data augmentation and transformation for detection of concrete crack. Algorithms 15 (8), 287. https://doi.org/10.3390/a15080287.
Lu, W., Qian, M., Xia, Y., Lu, Y., Shen, J., Fu, Q., Lu, Y., 2024. Crack _ PSTU: crack detection based on the U-Net framework combined with Swin Transformer. Structures 62, 106241. https://doi.org/10.1016/j.istruc.2024.106241.
Ma, X., Li, Y., Yang, Z., Li, S., Li, Y., 2024. Lightweight network for millimeter-level concrete crack detection with dense feature connection and dual attention. J. Build. Eng. 94, 109821. https://doi.org/10.1016/j.jobe.2024.109821.
Mandal, V., Uong, L., Adu-Gyamfi, Y., 2018. In: Automated road crack detection using deep convolutional neural networks. Data (Big Data, Seattle, USA, pp. 5212–5215. https://doi.org/10.1109/BigData.2018.8622327.
Munawar, H.S., Hammad, A.W., Haddad, A., Soares, C.A.P., Waller, S.T., 2021. Image-based crack detection methods: a review. Infrastructure 6 (8), 115. https://doi.org/10.3390/infrastructures6080115.
Nie, M., Wang, C.. Pavement Crack Detection based on yolo v3. https://doi.org/10.1109/IICSPI48186.2019.9095956.
Nyathi, M.A., Bai, J., Wilson, I.D., 2024. Deep learning for concrete crack detection and measurement. Metrology 4 (1), 66–81. https://doi.org/10.3390/metrology4010005.
Özgenel, Ç.F., Sorguç, A.G., 2018. Performance comparison of pretrained convolutional neural networks on crack detection in buildings. In: 2018 Proceedings of the 35th ISARC, pp. 1–8. https://doi.org/10.22260/ISARC2018/0094. Berlin, Germany.
Palevičius, P., Pal, M., Landauskas, M., Orinaitė, U., Timofejeva, I., Ragulskis, M., 2022. Automatic detection of cracks on concrete surfaces in the presence of shadows. Sensors 22 (10), 3662. https://doi.org/10.3390/s22103662.
Pan, Z., Guan, J., Yang, X., Fan, K., Ong, J.C., Guo, N., Wang, X., 2023. One-stage 3D profile-based pavement crack detection and quantification. Autom. ConStruct. 153, 104946. https://doi.org/10.1016/j.autcon.2023.104946.
Paramanandham, N., Koppad, D., Anbalagan, S., 2022. Vision based crack detection in concrete structures using cutting-edge deep learning techniques. Trait. du Signal 39 (2), 485–492. https://doi.org/10.18280/ts.390210.
Peng, C., Yang, M., Zheng, Q., Zhang, J., Wang, D., Yan, R., Wang, J., Li, B., 2020. A triple-thresholds pavement crack detection method leveraging random structured forest. Constr. Build. Mater. 263, 120080. https://doi.org/10.1016/j.conbuildmat.2020.120080.
Qiu, Q., Lau, D., 2023. Real-time detection of cracks in tiled sidewalks using YOLO-based method applied to unmanned aerial vehicle (UAV) images. Autom. ConStruct. 147, 104745. https://doi.org/10.1016/j.autcon.2023.104745.
Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., 2024. Sam 2: segment anything in images and videos. arXiv: 2408.00714. . https://doi.org/arXiv:2408.00714.
Redmon, J., 2016. You only look once: unified, real-time object detection. arXiv: 1506.02640. In: . https://doi.org/10.48550/arXiv.1506.02640.
Reis, H.C., Khoshelham, K., 2021. ReCRNet: a deep residual network for crack detection in historical buildings. Arabian J. Geosci. 14, 2112. https://doi.org/10.1007/s12517-021-08491-4.
Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Ren, M., Zhang, X., Chen, X., Zhou, B., Feng, Z., 2023. YOLOv5s-M: a deep learning network model for road pavement damage detection from urban street-view imagery. Int. J. Appl. Earth Obs. Geoinf. 120, 103335. https://doi.org/10.1016/j.jag.2023.103335.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. arXiv:1505.04597. In: . https://doi.org/10.48550/arXiv.1505.04597.

Shi, Y., Cui, L., Qi, Z., Meng, F., Chen, Z., 2016. Automatic road crack detection using random structured forests. IEEE Trans. Intell. Transp. Syst. 17 (12), 3434–3445. https://doi.org/10.1109/TITS.2016.2552248.

Silva, W.R.L.d., Lucena, D.S.d., 2018. Concrete cracks detection based on deep learning image classification. Proceedings 2 (8), 489. https://doi.org/10.3390/ICEM18-05387.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. https://doi.org/arXiv:1409.1556.

Spencer Jr., B.F., Hoskere, V., Narazaki, Y., 2019. Advances in computer vision-based civil infrastructure inspection and monitoring. Engineering 5 (2), 199–222. https://doi.org/10.1016/j.eng.2018.11.030.

Su, P., Han, H., Liu, M., Yang, T., Liu, S., 2024. MOD-YOLO: rethinking the YOLO architecture at the level of feature information and applying it to crack detection. Eng. Appl. Artif. Intell. 237, 121346. https://doi.org/10.1016/j.eswa.2023.121346.

Vaswani, A., 2017. Attention is all you need. arXiv.1706.03762. https://doi.org/10.48550/arXiv.1706.03762.

Wang, C., Liu, H., An, X., Gong, Z., Deng, F., 2024. SwinCrack: pavement crack detection using convolutional swin-transformer network. Digit. Signal Process. 145, 104297. https://doi.org/10.1016/j.dsp.2023.104297.

Wang, D., 2016. K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: revisited. Mech. Syst. Signal Pr. 70, 201–208. https://doi.org/10.1016/j.ymssp.2015.10.007.

Wang, J., Zeng, Z., Sharma, P.K., Alfarraj, O., Tolba, A., Zhang, J., Wang, L., 2024. Dual-path network combining CNN and transformer for pavement crack segmentation. Autom. ConStruct. 158, 105217. https://doi.org/10.1016/j.autcon.2023.105217.

Wang, W., Su, C., 2022. Automatic concrete crack segmentation model based on transformer. Autom. ConStruct. 139, 104275. https://doi.org/10.1016/j.autcon.2022.104275.

Wu, Y., Li, S., Zhang, J., Li, Y., Li, Y., Zhang, Y., 2024. Dual attention transformer network for pixel-level concrete crack segmentation considering camera placement. Autom. ConStruct. 157, 105166. https://doi.org/10.1016/j.autcon.2023.105166.

Xiao, S., Shang, K., Lin, K., Wu, Q., Gu, H., Zhang, Z., 2023. Pavement crack detection with hybrid-window attentive vision transformers. Int. J. Appl. Earth Obs. Geoinf. 116, 103172. https://doi.org/10.1016/j.jag.2022.103172.

Xing, J., Liu, Y., Zhang, G.-Z., 2023. Improved yolov5-based uav pavement crack detection. IEEE Sens. J. 23 (14), 15901–15909. https://doi.org/10.1109/JSEN.2023.3281585.

Xiong, B., Hong, R., Wang, J., Li, W., Zhang, J., Lv, S., Ge, D., 2024. DefNet: a multi-scale dual-encoding fusion network aggregating Transformer and CNN for crack segmentation. Constr. Build. Mater. 448, 138206. https://doi.org/10.1016/j.conbuildmat.2024.138206.

Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X., Ling, H., 2019. Feature pyramid and hierarchical boosting network for pavement crack detection. IEEE Trans. Intell. Transport. Syst. 21, 1525–1535. https://doi.org/10.1109/TITS.2019.2910595.

Yang, L., Huang, H., Kong, S., Liu, Y., 2023. A deep segmentation network for crack detection with progressive and hierarchical context fusion. J. Build. Eng. 75, 106886. https://doi.org/10.1016/j.jobe.2023.106886.

Yang, X., Guo, R., Li, H., 2023. Comparison of multimodal RGB-thermal fusion techniques for exterior wall multi-defect detection. J. Infrastruct. Intell. Resil. 2 (2), 100029. https://doi.org/10.1016/j.iintel.2023.100029.

Yang, X., Zhang, J., Liu, W., Jing, J., Zheng, H., Xu, W., 2024. Automation in road distress detection, diagnosis and treatment. J. Road Eng. https://doi.org/10.1016/j.jreng.2024.01.005.

Yao, G., Sun, Y., Wong, M., Lv, X., 2021. A real-time detection method for concrete surface cracks based on improved YOLOv4. Symmetry 13 (9), 1716. https://doi.org/10.3390/sym13091716.

Ye, X., Jin, T., Yun, C., 2019. A review on deep learning-based structural health monitoring of civil infrastructures. Smart Struct. Syst. 24 (5), 567–585. https://doi.org/10.12989/sss.2019.24.5.567.

Yu, G., Dong, J., Wang, Y., Zhou, X., 2022. RUC-Net: a residual-Unet-based convolutional neural network for pixel-level pavement crack segmentation. Sensors 23 (1), 53. https://doi.org/10.3390/s23010053.

Yu, Z., Chen, Q., Shen, Y., Zhang, Y., 2024. Robust pavement crack segmentation network based on transformer and dual-branch decoder. Constr. Build. Mater. 453, 139026. https://doi.org/10.1016/j.conbuildmat.2024.139026.

Yuan, Q., Shi, Y., Li, M., 2024. A review of computer vision-based crack detection methods in civil infrastructure: progress and challenges. Remote Sens. 16 (16), 2910. https://doi.org/10.3390/rs16162910.

Zhang, A., Wang, K.C., Li, B., Yang, E., Dai, X., Peng, Y., Fei, Y., Liu, Y., Li, J.Q., Chen, C., 2017. Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network. Comput. Aided Civ. Infrastruct. Eng. 32 (10), 805–819. https://doi.org/10.1111/mice.12297.

Zhao, S., Zhang, D., Xue, Y., Zhou, M., Huang, H., 2021. A deep learning-based approach for refined crack evaluation from shield tunnel lining images. Autom. ConStruct. 132, 103934. https://doi.org/10.1016/j.autcon.2021.103934.

Zhao, X., Li, S., 2017. A method of crack detection based on convolutional neural networks. Struct. Health Monit.. https://doi.org/10.12783/shm2017/13959.

Zhou, S., Canchila, C., Song, W., 2023. Deep learning-based crack segmentation for civil infrastructure: data types, architectures, and benchmarked performance. Autom. ConStruct. 146, 104678. https://doi.org/10.1016/j.autcon.2022.104678.

Zou, Q., Zhang, Z., Li, Q., Qi, X., Wang, Q., Wang, S., 2018. Deepcrack: learning hierarchical convolutional features for crack detection. IEEE Trans. Image Process. 28 (3), 1498–1512. https://doi.org/10.1109/TIP.2018.2878966.

Zoubir, H., Rguig, M., El Aroussi, M., Chehri, A., Saadane, R., 2022. Concrete bridge crack image classification using histograms of oriented gradients, uniform local binary patterns, and kernel principal component analysis. Electronics 11 (20), 3357. https://doi.org/10.3390/electronics11203357.