13th International Conference on
Structural Health Monitoring of Intelligent Infrastructure
DOI: 10.3217/978-3-99161-057-1-052

Overview and Challenges of Computer Vision-Based Visual Inspection for the Assessment of Bridge Defects

R.U. Khan, R. Kromanis
Department of Civil Engineering and Management, Faculty of Engineering Technology, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands
Email: r.u.khan@utwente.nl, r.kromanis@utwente.nl

ABSTRACT: Visual inspection remains the most fundamental and widely used method for assessing the condition of bridges. This process involves observation of structural surfaces at a close distance to identify visible signs of deterioration such as cracking, spalling, corrosion, and delamination. Traditionally, human inspectors perform visual inspections manually. This labour-intensive process is associated with many limitations, for example, subjectivity to an inspector's interpretation, difficulty accessing structural components, management of large volumes of unstructured data and the lack of consistent historical records. Recent advancements in computer vision and artificial intelligence have enabled considerable progress toward automating visual inspections. However, the full automation of visual inspections in practical, real-world scenarios remains constrained by several challenges: (i) the continued need for human intervention, (ii) the limited availability of high-quality labelled datasets, (iii) the generalizability of existing models, and (vi) the lack of standardized inspection protocols. In this positioning paper, we present an overview of the current state of automated visual inspection for defects identification in bridges. It reviews key open-source datasets of defects and state-of-the-art deep learning models. We give our forward-looking perspective on fully automated defects identification systems that align with standardized visual inspection guidelines.

KEY WORDS: Visual Inspection; Defects Identification; Condition Assessment; Automated Bridge Inspection; Computer Vision; Structural Health Monitoring (SHM); SHM at Local level.

1 INTRODUCTION

Bridges are critical components of our transportation infrastructure. Rigorous and timely inspections are needed to ensure their long-term performance and operational safety, and to avoid catastrophic failures. In general, bridge inspections are classified as general, principal, and special inspections. Each type of inspection serves a distinct purpose, i.e., from basic visual checks, to more in-depth often involving touching distance examinations, to address specific concerns or unusual events such as accidents (e.g., collisions by heavy loads trucks) or natural disasters (e.g., earthquake, flooding) [1]. Among these, visual inspection is the most employed method, particularly during general and principal inspections. It involves systematic observation of the bridge's surface to detect visible signs of deterioration such as cracking, corrosion, spalling, and delamination. These observable defects serve as initial indicators of potential structural issues and guide/suggest evaluation or maintenance actions.

The traditional visual inspection, while primary for initial defects detection, is inherently labour-intensive and often requires close-proximity access to structural elements, which is frequently unfeasible in hard-to-reach locations [2]. The process is also highly subjective, relying heavily on the individual expertise and judgment of trained inspectors. Research indicates a substantial probability (approximately 50%) of inconsistent classification of concrete defects when different inspectors evaluate the same defects [3]. This poses a significant challenge, as the availability of experienced personnel is steadily declining [4]. Inconsistencies often arise from the disconnect between the on-site surveyor and the off-site expert who interprets and documents findings, leading to a potential mismatch between observed conditions and reported assessments. Also, inspections typically generate a huge amount of unstructured data, including images and reports. Historical inspection reports are rarely utilized in subsequent evaluations due to inadequate data organization and retrieval systems. This lack of continuity poses a challenge for inspectors to accurately visualize and locate previously identified defects, complicating re-localization of defects and trend analysis over time, which involves monitoring defects progression across inspections, identifying deterioration patterns, and anticipating future degradation or necessary interventions. These challenges collectively highlight the need for integrated, data-driven, and automated visual inspection workflows that improve consistency, traceability, and long-term asset management of bridges.

Recent advancements in computer vision (CV) and artificial intelligence (AI) have matured to a level that enables the enhancement and partial automation of visual bridge inspections [5], [6]. These technologies are promising for digitizing inspection workflows, making them significantly fast, reliable, and repeatable. By using AI-driven image analysis, defects identification, and data management systems, bridge visual inspections can be completely automated and independent of subjectivity. Previous literature reviews [7], [8] on CV and deep learning (DL) based bridge visual inspection, have provided foundational overviews of the field up to 2020. While valuable, these reviews have limitations that necessitate an updated perspective. For instance, [7] comprehensively covers structural health monitoring (SHM) with an emphasis on CV-based defects detection (e.g., cracks, spalling, delamination, rust, and bolt loosening). However, its scope is

DOI: 10.3217/978-3-99161-057-1-052

restricted to datasets and algorithms available prior to 2020, thereby omitting crucial post-2020 advancements in both novel defect types and state-of-the-art AI-driven techniques. Similarly, [8] offers a broad exploration of DL-based SHM, encompassing CV, unmanned aerial vehicles (UAVs), vibration-based methods, and physics-informed approaches, effectively linking traditional machine learning with modern DL strategies. Nevertheless, it lacks critical discussion on the practicalities of real-world applications of visual inspection, the availability of experimental/test datasets, and open-source implementations, thus highlighting a significant gap in addressing deployment challenges and reliability of automated visual inspection. This collective gap highlights the urgent need for a contemporary position that re-evaluates automated visual inspections considering recent research advances and emerging trends, with a focus on improvements in the enrichment of defects dataset, DL-based defects identification algorithms, and practical deployment strategies.

In this positioning paper, we critically evaluate the current state-of-the-art in available defects datasets and DL-based defects identification algorithms, systematically identify the principal barriers to real-world implementation, and propose a forward-looking perspective on the future development of automated visual inspection of defects for bridges. The rest of the paper is organized as follows: Section 2 describes the dataset utilized for bridge defects detection, including data sources and characteristics. Section 3 details the baseline algorithms and the proposed enhancements for defects recognition. Section 4 provides a perspective on the future overview of automated visual inspection. Finally, Section 5 summarizes the overall position paper and draws conclusions.

2 RELEVANT DEFECTS DATASETS

Most highway bridges are reinforced concrete (RC) bridges [9]. While a reliable identification of RC defects is essential, existing datasets are often limited in size and class diversity, raising concerns about their real-world applicability and suitability as benchmarks. Over the past decade, significant progress has been made through datasets enabling binary classification [13, 23, 24, 25], multi-class classification [1, 7, 26], multi-label classification [16], binary semantic segmentation [8, 27], and multi-label semantic segmentation [10, 28, 29]. Each dataset contributes to advancing defects assessment methodologies reviewed in this section.

2.1 Binary-class classification

The Cambridge Bridge Inspection Dataset (CDS) [10] combines two primary data sources to enhance defects recognition in RC structures. The first source comprises 21,284 high-resolution images captured from 10 RC highway bridges in Cambridge. The focus is on critical structural elements such as decks, columns, piers, and abutments. Since these bridges are in good condition, the dataset lacks sufficient diversity of defects. To address this limitation, a second set of 22,121 images was incorporated from the U.S. Federal Highway Administration and the Georgia Department of Transportation, enriching the dataset with a variety of examples of defects. All images are categorized into two classes “healthy” and “potentially unhealthy,” providing a foundational binary classification benchmark for structural condition assessment. This hybrid approach ensures broad applicability. SDNET [11] is also a large-scale binary annotated image dataset designed to train, validate, and benchmark AI-driven crack detection models for concrete structures. Comprising over 56,000 images of cracked and non-cracked surfaces of diverse structural elements such as bridge decks, walls, and pavements. The dataset captures a wide range of crack widths, from 0.06 mm to 25 mm, enhancing its applicability to real-world scenarios. SDNET incorporates various challenging conditions such as shadows, surface roughness, scaling, edges, holes, and background debris, simulating shared challenges encountered in visual inspections. While the dataset covers many structural elements, its initial version suffered from labeling inaccuracies, which may require preprocessing or correction for reliable model training. Despite this limitation, SDNET remains a valuable resource for advancing automated structural defects assessment.

The Image-based Concrete Crack Database (ICCD) [12] is developed using 1,455 high-resolution crack images captured via smartphone from suspension bridge towers and anchor chambers in Dalian, China. To ensure diversity, images were taken at varying distances (0.1 – 1.0 m) and under different lighting conditions (daylight, nighttime, direct sunlight, and shaded surfaces), simulating real-world inspection conditions. These images were then cropped into 256 × 256 px patches and manually labeled into two classes – cracked and uncracked concrete. Through data augmentation, the final dataset was expanded to 60,000 images, significantly enhancing its utility for training robust DL models in automated crack detection. This approach improves generalization and addresses variability in real-world visual inspection scenarios. The Bridge Crack Dataset (BCD) [13] is specifically designed for robust crack detection in bridge inspection scenarios. The original dataset consists of 2,068 images of bridge cracks, which were processed and augmented to generate 6,069 image patches, optimizing them for DL applications. To enhance real-world applicability, the dataset intentionally includes challenging conditions such as bridge shading, water stains, and bright light reflections, common obstacles in field inspections. By incorporating these complexities, BCD serves as a valuable benchmark for developing generalizable and noise-resistant crack classification models, ensuring practical utility in automated visual inspection systems. While these datasets provide valuable benchmarks for binary crack detection, they suffer from critical limitations, for example, CDS [10] lacks natural defects diversity, SDNET [11] has labeling errors, ICCD [12] relies on artificial augmentation, and BCD’s [13] small scale and synthetic challenges may not reflect real-world complexity.

2.2 Multi-class classification

The Bearing Condition State Classification dataset [14] comprises 947 annotated images of structural bridge bearings. The annotations adhere to the condition state assessment guidelines outlined by the American Association of State Highway and Transportation Officials [21] and the Bridge Inspector's Reference Manual [22]. The dataset categorizes steel corrosion into four distinct condition states: good, fair, poor, and severe. Detailed annotation guidelines, along with explanatory examples, are provided to ensure consistent and accurate condition assessment. The dataset serves as a valuable resource for visual inspection and deterioration assessment of

DOI: 10.3217/978-3-99161-057-1-052

bridge bearings. The Multi-classifier Dataset (MCDS) [3] consists of 38,408 annotated images capturing various deterioration patterns in concrete structures. Defects are classified into eight distinct categories including spalling, cracks, rust staining, efflorescence, scaling, abrasion/wear, exposed reinforcement, and general defects. This comprehensive taxonomy enables detailed analysis of concrete degradation assessment and supporting visual inspection.

2.3 Multi-label classification

COncrete DEfects BRidge IMage (*CODEBRIM*) dataset [16] is the largest and most realistic multilabel dataset for RC deterioration classification. The dataset categorizes defects into six distinct classes: crack, spalling, exposed reinforcement bar, efflorescence, corrosion, and background. The unbalanced version of *CODEBRIM* consists of 7,729 annotated image patches extracted from 30 bridges, selected to represent diverse deterioration levels, defects sizes, severity, and surface textures. High-resolution images were captured under different weather conditions and using multiple cameras at different scales. To address accessibility challenges, a subset of data was acquired via UAV for defects located at elevated positions. Despite its comprehensiveness, *CODEBRIM* present a key challenge for real-world transferability. The dataset is composed of cropped image patches, where original images are subdivided into rectangular segments based on maximum defects dimensions. This approach may disrupt contextual information, which is critical for holistic defects assessment.

2.4 Binary-class semantic segmentation

The *UAV75* dataset [17] comprises 75 images featuring pixel-wise manual annotations for binary semantic segmentation tasks. This dataset provides precise delineation of target features at the pixel level, enabling detailed analysis of structural characteristics. The fixed image dimensions and binary classification scheme facilitate consistent model training and evaluation, while the limited sample size positions this dataset as a specialized benchmark for targeted applications in structural assessment. Kulkarni et al. [18] introduced *CrackSeg9k*, currently the largest and most diverse binary crack segmentation dataset, comprising 9,255 images aggregated from ten preexisting sub-datasets including *Crack500, Deepcrack, SDNET, CrackTree, GAPs, Volker, Rissbilder, Noncrack, Masonry, and Ceramic*. The dataset addresses key limitations in individual source datasets (e.g., noise, distortion) through standardized preprocessing, while maintaining diversity in surface materials (concrete, masonry) and crack morphologies. Despite its focus on binary crack segmentation as shown in Figure 1, the dataset's practical utility depends on recognizing at least nine distinct defect types. Notably, CrackSeg9k homogenizes acquisition conditions including camera pose, lighting, and hardware across sub-datasets to minimize confounding variables. This curation enhances its reliability for benchmarking semantic segmentation algorithms in visual inspection for RC defects applications.

2.5 Multi-label semantic segmentation

The Structural Defects Dataset (S2DS) [20] is the first multi-class semantic segmentation dataset for RC defects analysis, containing 743 annotated images of RC bridges. It classifies five defect types: cracks, spalling, corrosion, efflorescence, and vegetation along with control points for georeferencing as shown in Figure 2. While the dataset is pioneering in enabling multi-class segmentation and features high-quality manual annotations by a trained expert, its limited size and diversity raise concerns about real-world applicability. The dacl10k dataset [5] represents the first large-scale benchmark for multi-label semantic bridge defects segmentation, featuring 9,920 annotated images sourced from bridge inspections in Germany between 2000 and 2020. Developed to support AI-assisted defects recognition and documentation, the dataset aligns with structural inspection guidelines, focusing on defects that can be legally assessed. It includes 19 classes as shown in Figure 3 categorized into concrete defects, general defects, and objects, capturing complex real-world scenarios where multiple defects often overlap. The dacl1k dataset [19] addresses critical limitations in existing RC defects datasets by providing 1,474 uncropped, real-world inspection images with multi-label annotations across five damage classes including Crack, Efflorescence, Spalling, Bars Exposed, Rust, and a No Damage category, derived from diverse sources including authorities and engineering offices. Unlike datasets such as *CODEBRIM* (which uses cropped patches) or *MCDS*, dacl1k preserves raw image heterogeneity varying in camera types, poses, lighting, and resolutions to better reflect real inspection challenges. However, while its diversity enhances practical applicability, the dataset's small size (1,474 images) and moderate label count (2,367 total labels) raise concerns about statistical robustness and class balance. Despite the shortcomings, dacl1k represents a step toward bridging the gap between controlled research datasets and actual field conditions, though larger-scale, more granular annotations and rigorous benchmarking remain unmet needs for reliable *RC defects* assessment.

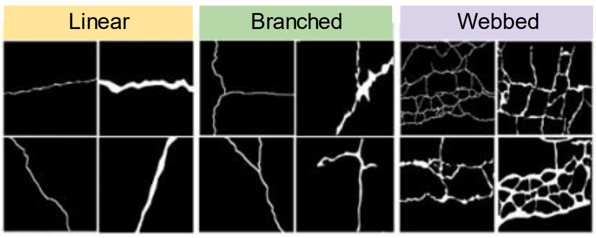

Figure 1. CrackSeg9k's categorization of crack types [18]

3 DFECTS IDENTIFICATION MODELS

Dong et al. study [7] provide a detailed review of computer vision-based structural health monitoring at local level (CV-SHM-LL), outlining various applications and methodological approaches. The work scrutinizes a wide range of models designed for localized analysis, including both patch-based and pixel-based techniques. Additionally, it discusses traditional methods alongside data-driven machine learning approaches. However, it is limited to approximately eight types of structural defects. Cha et al. [8] provides a broad overview of DL-based SHM techniques applied to various infrastructure systems, such as bridges, and different construction materials like concrete and steel. The study extensively reviews the historical development of DL architecture, as illustrated in Figure 4. However, it does not explore deeply into the precise

DOI: 10.3217/978-3-99161-057-1-052

Figure 2. S2DS Classes and Labels: crack (black), spalling (red), corrosion (orange), efflorescence (blue), vegetation (green), and control point (purple) [20]

application of visual inspection methods for bridges defects identification and their real-world implementation challenges.

Hüthwohl et al. [10] proposed an automated method for detecting intact concrete areas using image segmentation and classification. By applying morphological operations to generate boundary masks, their approach filters out defects-free regions, optimizing inspection efficiency by concentrating analysis on potential damage zones. Dorafshan et al. [11] introduced SDNET dataset and benchmarked performance using AlexNet [23], validating the dataset's utility for algorithm development. Li et al. [12] developed a convolutional neural network (CNN) based crack detection method using an enhanced AlexNet [23], overcoming traditional limitations like noise sensitivity. Their model achieved 99.06% validation accuracy and was deployed as a smartphone app for real-world use. Xu et al. [13] developed an end-to-end CNN-based crack detection model. Achieving 96.37% accuracy without pre-training, it outperformed traditional methods and showed potential as a versatile feature extraction module for other networks.

Hüthwohl et al. [3] developed a three-stage multi-classifier system for concrete defects detection in bridges using fine-tuned deep neural networks trained on multi-source inspection data. Their approach first identifies five defect types along with defects-free areas, then detects exposed reinforcement, and finally recognizes rust staining. The approach achieves 85% average classification accuracy. Mundt et al. [16] developed a meta-learning approach for automated CNN design targeting multi-defects concrete classification. Using their CODEBRIM dataset containing images with overlapping defects, they employed MetaQNN [24] and ENAS [25] reinforcement learning methods to generate optimized architectures. The resulting CNNs achieved higher multi-target accuracy than manually designed models while using fewer parameters, with validation accuracy serving as the reinforcement learning (RL) controller's reward signal. Benz et al. [17] presents CRACKNAUSNET, a transfer learning-based CNN for crack detection in unmanned aircraft system (UAS) imagery,

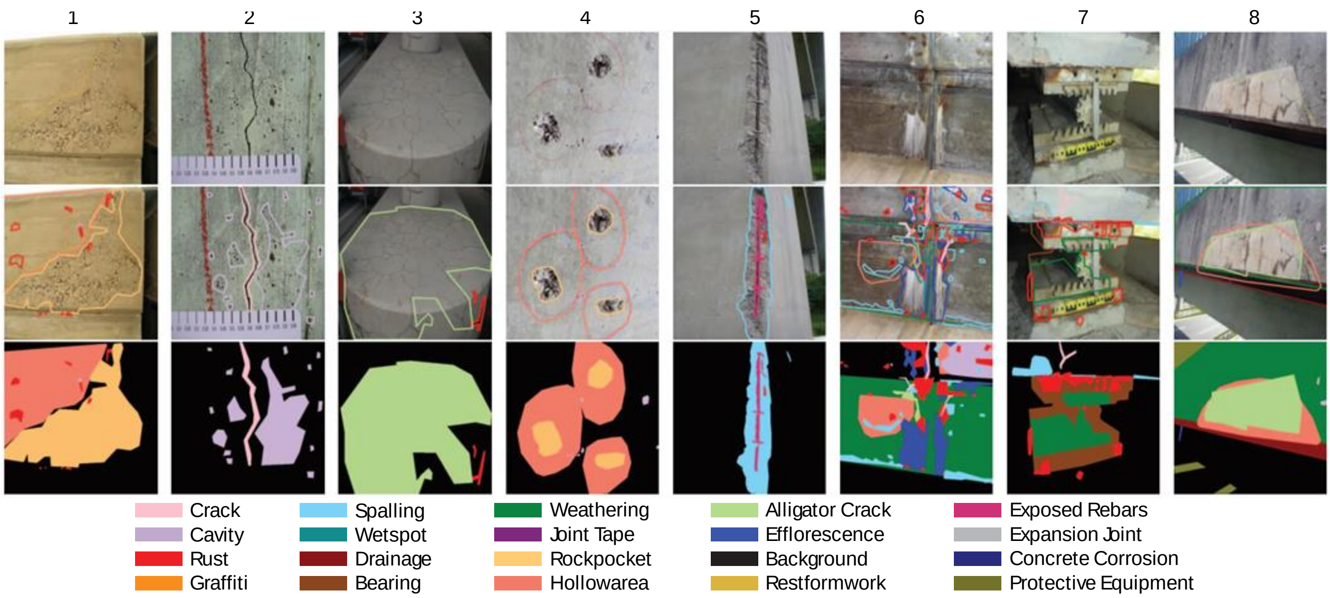

Figure 3. dacl10k dataset and labels [5]. The font size of the caption has been adjusted for clarity.

 DOI: 10.3217/978-3-99161-057-1-052

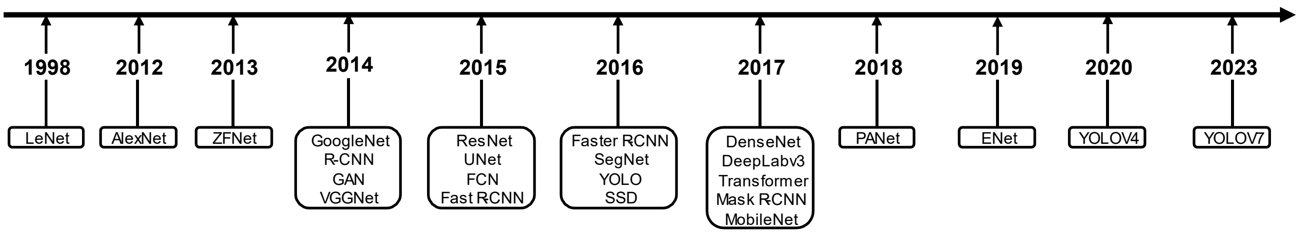

Figure 4. Popular architectures of DL over the years [8]

adapting TernausNet [26] with VGG16's [27] first 16 layers as a pre-trained encoder. The model addresses UAS-specific challenges (low resolution, faint cracks, planking artifacts) using the novel UAV75 dataset, which includes a dedicated class to mitigate planking-induced false positives, where inspection planks were mistakenly identified as defects. While achieving 75% average accuracy on UAV75 outperforming existing approaches, the model's performance declined on external datasets, indicating dataset-dependent effectiveness.

Kulkarni et al. [18] proposed an advanced crack segmentation framework combining YOLOv5 [28] for detection, DINO for unsupervised feature extraction, and FPN-based models [29] for precise segmentation. Their approach utilizes self-supervised transformer attention to enhance CNN performance, overcoming challenges like crack variability and background noise. The method was validated on a newly compiled, meticulously annotated dataset demonstrating improved generalization across diverse crack types and surfaces.

The study [20] proposed a hierarchical multi-scale attention (HMA) model for multi-label semantic segmentation using an HRNet-OCR backbone to oversee objects of varying sizes. The model employs contrastively learned attention maps to dynamically fuse multi-scale features, enhancing pixel-level and contextual representations through transfer learning from Cityscapes. A significant contribution is the line-based tolerant IoU metric designed specifically for crack detection, addressing the shortcomings of conventional area-based metrics (IoU/F1). The evaluation compared CNN-based architectures (DeepLabV3+, FPN with MobileNetV3 / EfficientNet encoders, some with auxiliary losses) and Transformer-based SegFormer (trained with Dice loss). All models used Adam optimization with cosine learning rate scheduling, ImageNet initialization, and 512×512 input resolution over 30 epochs, demonstrating the approach's effectiveness particularly for challenging crack segmentation tasks.

The dacl-challenge [6] aimed to advance automated defects identification in bridges using its large, real-world dataset. It benchmarked CV models as shown in Figure 5 for accurate, detailed detection and classification of bridge defects and components.

a) *Baseline Model:* The dacl-challenge baseline employed SegFormer [30] MiT-b1, pre-trained on ImageNet-1k [23]. This model features a multi-label segmentation head and a compact encoder with 13.1M parameters. For the challenge, SegFormer was trained for 10 epochs on the development set and 30 epochs on the final test set.
b) *First Place Approach (Sheoran):* The top-performing solution by Sheoran utilized an ensemble of predictions from several models. Initially trained with MMSegmentation, the models were adapted to the segmentation-models-pytorch library for multi-label handling. The training process incorporated diverse augmentations and the RangersLars optimizer. Predictions from six distinct models were aggregated for specific classes, leading to enhanced overall performance.
c) *Second Place Approach (Bridge Protector):* Bridge Protector's approach involved training the Mask2Former model [31] with an InternImage-H [32] backbone using the MMSegmentation framework. Pre-trained weights from the ADE20K dataset [33] were utilized. To address the multi-label nature of the data, the problem was divided into 19 individual binary segmentation models, one for each class. The outputs of these 19 models were then combined.
d) *Third Place Approach (Winning Wieners):* Winning Wieners combined a feature pyramid network (FPN) [29] with a multi-axis vision transformer (MaxViT) [34] as the backbone. MaxViT integrates convolutional blocks with the attention mechanism of vision transformers. The xlarge version of MaxViT, pre-trained on ImageNet, was used. The model was trained using a five-fold cross-validation strategy, resulting in an ensemble of five models. Prediction-level threshold optimization was performed for the final ensemble.

Top-performing approaches heavily used transfer learning and adapted existing architectures, potentially limiting exploration of novel structural defects segmentation techniques. Their performance is more sensitive to training configurations than architectural innovation. While ensemble learning improved the results (see Figure 5) their computational cost warrants investigate efficient high-performance strategies. The bar chart presented here shows only the top performance values for defects (objects excluded), compared to the baseline performance. While many deep learning-based approaches for structural defects identification in bridge infrastructure achieve high accuracy using CNNs, transformers, and ensemble methods, their generalizability is often constrained by dataset dependency, high computational demands, and reliance on pre-trained architectures. Future research should prioritize the development of specialized models and training strategies tailored to visual structural assessment, moving beyond generic transfer learning paradigms.

4 FUTURE OVERVIEW OF VISUAL INSPECTION

Over the past decade, different AI and CV-based methodologies have been proposed for the identification of

DOI: 10.3217/978-3-99161-057-1-052

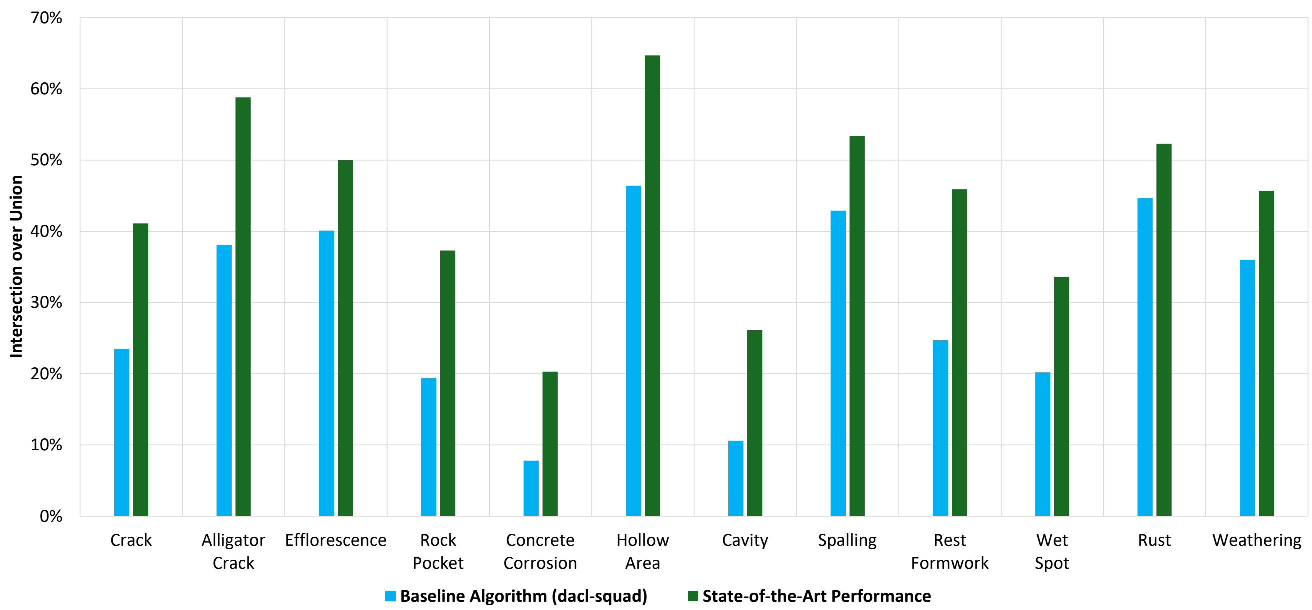

Figure 5. dacl-challenge: state-of-the-art performance achieved vs. baseline algorithm accuracy (% IoU) [6]

local-level defects [7], [8], [36], [37], [38], [39], [40], [41]. However, despite promising results in controlled research settings, many of these solutions have not yet achieved the robustness or scalability approaches for deployment in real-world environments. Most current approaches fall short of industrial standards and remain insufficient for fully automating visual inspections under diverse operational conditions. Thus, while foundational technologies are in place, further developments, validations, and standardizations are necessary to enable their widespread adoption in bridge inspection practices.

Despite all current advancements, the complete automation of visual bridge inspection remains unrealistic at present. This is primarily due to the complexity of standardized inspection protocols such as NEN 2767 in the Netherlands [42], AASHTO in the United States [21], CS 450 in the UK [43], CSA-S6 in Canada [44], AS5100.7 in Australia [45]. These standards define a multi-layered assessment framework, spanning from general structural classification to detailed defects evaluation. For example, NEN 2767 organizes infrastructure assessment across five hierarchical levels. Figure 6 represents inspection levels with an example decomposition for a RC bridge.

Level-1. Element Group (i.e., type of bridge such as reinforced concrete) bridge,
Level-2. Elements (i.e., main parts of that bridge such as handrail),
Level-3. Building Components (i.e., sub-parts of element such as the structural frame of a handrail and protective coating),
Level-4. Materials (i.e., material type of each building component (e.g., steel for the structural frame and paint for the protective coating),
Level-5. Defects (i.e., defect types of building component, like corrosion and rust in case of steel handrail, color peel off in case of paint coating).

During visual inspection, inspectors primarily interact with bridge at material and defects levels. For each observed defect, the inspector manually assigns three critical parameters including severity, extent, and intensity. These judgments, informed by domain expertise, are used to calculate the condition score for each building component (Level-3), which are then aggregated to yield scores for Elements (Level-2) and the overall condition index for the entire Element Group (Level-1). In the NEN 2767 standard the condition score ranges from 1 (very good) to 6 (very poor). It quantifies the current condition of inspected components of the inspected bridge based on observed defects.

While CV techniques have made substantial progress in defects identification (i.e., detection, localization, and classification of defects such as cracks, corrosion, or spalling), the current state of approaches are not yet capable of reliably performing the significant evaluation required by inspection standards. Specifically, automatic quantification of defects parameters (severity, extent, and intensity), contextual cause analysis, and risk assessment still require manual measurements and expert's (e.g., inspector's) interpretation. As a result, human oversight remains essential, and existing AI-based inspection systems are best viewed as decision-support tools rather than fully autonomous solutions.

To fully understand the current limitations in automating bridge inspections, it is important to examine the fundamental challenges associated with defects identification across multiple levels. In SHM, first Rytter in 1993 [46] and then Worden et al. in 2004 [47] outlined hierarchical levels of damage identification including damage detection, localization, classification, assessment, and prediction. While damage refers to changes in structural properties that adversely affect performance, defects in the visual inspection refer to observable surface-level anomalies. An equivalent framework of object detection as previously highlighted in [48] and now in further extended form, as shown in Figure 7,can be applied to CV-based defects identification. The following levels for

DOI: 10.3217/978-3-99161-057-1-052

object (or defects, in this case) identification are defined: detection, localization, classification (or segmentation), quantification, and propagation. Currently, most CV-based approaches [6] in bridge inspection have achieved automation up to the classification (or segmentation) level, identifying and labelling regions of interest such as cracks or corrosion patches [7], [8]. However, progressing to quantification is essential to comply with inspection standards, which require precise evaluation of defects parameters (severity, extent and intensity) to calculate the condition score of a component (refer to Figure 6).

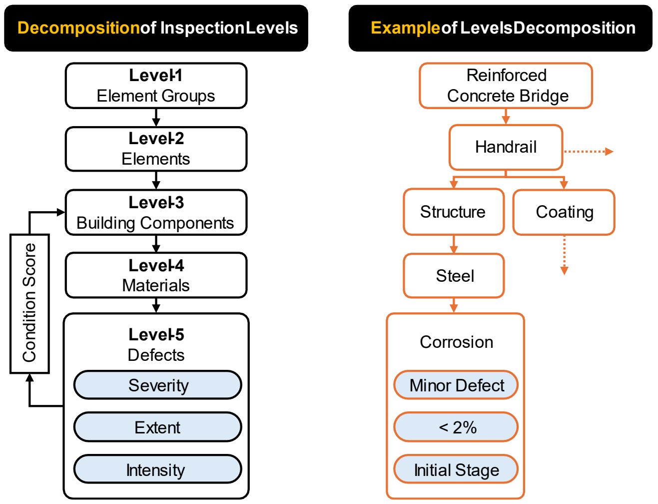

Figure 6. Decomposition and example of levels of inspection process in NEN 2767 standard

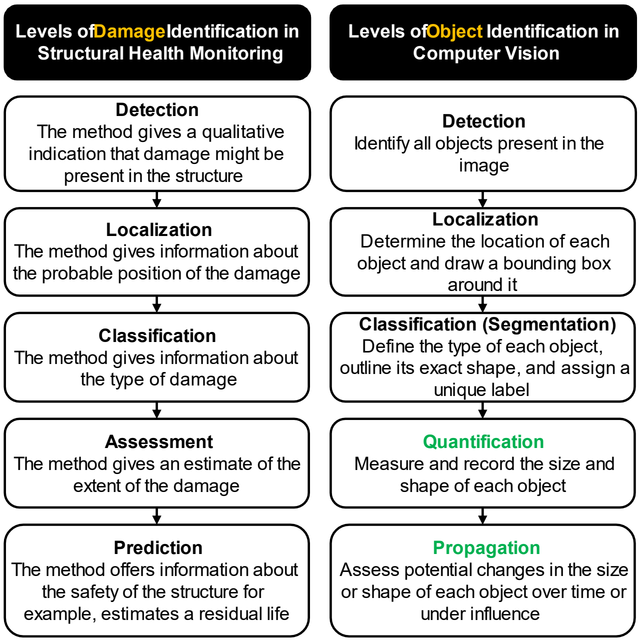

Figure 7. An identification analogy between damage identification [24], [25] and object (or defects in this case) identification

While limited success has been achieved in quantifying specific types of defects (e.g., crack width and length) [49], [50], [51], [52], [53], the methods are not generalized across the wide spectrum of defects encountered in practice. For instance, NEN standard alone defines approximately 128 distinct defect types relevant to bridge visual inspection, each with distinct characteristics and assessment criteria [42]. This diversity presents a significant challenge for developing universal quantification algorithms. Once automated systems can reliably assign defects-level (Level-5) condition scores as shown in Figure 6, this would enable near-complete automation of the visual inspection process and facilitate systematic and periodic data collection at the local (defects) level.

To address these limitations and pave the way for practical CV-SHM implementation, this paper proposes an automated visual inspection system with a future direction, as conceptually illustrated in Figure 8. The proposed system is envisioned as a continuously evolving ecosystem, driven by progressive enhancements in defects identification capabilities and adherence to established inspection standards, such as the Dutch NEN2767 standard. It is initially trained on a comprehensive dataset synthesized by merging existing state-of-the-art datasets, specifically curated to encompass a diverse range of defect types. A user-friendly application is developed in close collaboration with visual inspection experts to ensure seamless integration with current inspection workflows. Once it achieves a satisfactory level of performance, the system is deployed across various on-site inspection platforms, including mobile devices, tablets, drones, and augmented reality (AR) headsets. In its operational phase, the system functions as a semi-autonomous tool for human inspectors. It provides real-time defects predictions, allowing inspectors to contribute their expertise through manual annotations and comments. All data, including system predictions and human input, are securely stored on a cloud platform, facilitating continuous expert evaluation and ongoing system refinement. Furthermore, the system integrates AR tools to analyze historical defects data, including location and characteristics, to predict future defects propagation over time. This predictive capability empowers initiative-taking maintenance strategies and enhances the long-term sustainability of bridges and other civil infrastructures.

5 CONCLUSION

This paper reviews RC defects datasets for bridges and state-of-the-art algorithms and proposes an automated visual inspection system for computer vision-based structural health monitoring at local level (CV-SHM-LL) that integrates deep learning methodologies with standardized inspection protocols and human expertise to advance bridge inspections. Addressing current limitations in data availability, holistic component-level defects evaluation, and deployment feasibility, this study aims to translate theoretical advancements into practical solutions for enhanced SHM. The following conclusions can be drawn from this positioning paper:

- To improve the reliability, robustness, and resilience of the inspection systems, it is necessary to utilize and integrate comprehensive and diverse datasets with unique types of defects, and state-of-the-art prediction models.
- For the inspection system to be useful in real-world scenarios, it must be able to quantify defects by determining their extent, severity, and intensity. This will allow for a step-by-step approach to reach level of quantification as shown in Figure 7.

 DOI: 10.3217/978-3-99161-057-1-052

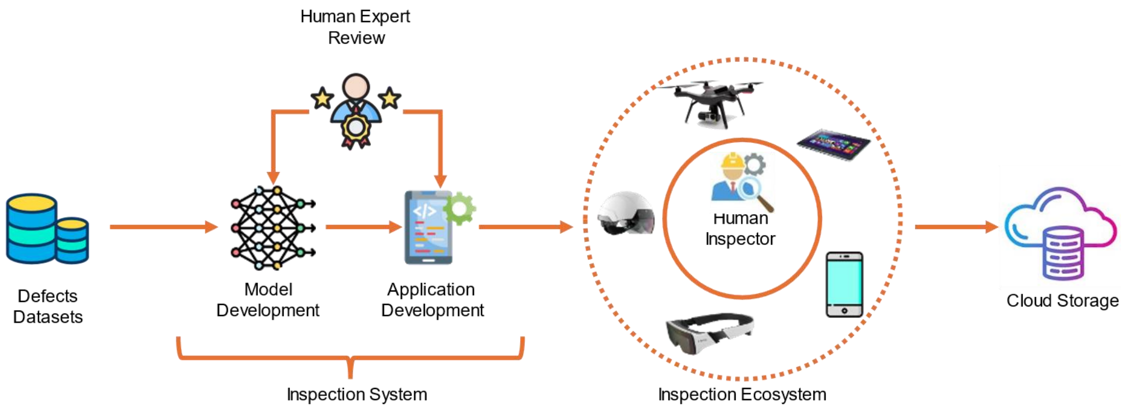

Figure 8. Proposed visual inspection system and future overview of CV-SHM-LL

- Inspection standards must be applied to connect surface-level observations with a deeper understanding of overall structural behaviour. This will bridge the gap to achieving insights into global structural performance.

Future research should focus on the deployment of lightweight neural networks and integration of augmented reality (AR) features on handheld edge devices. This interface should enable the visual overlay of previously identified structural anomalies onto the physical infrastructure. This capability would facilitate targeted inspection efforts and the identification of how defects propagate over time. By leveraging computational models, such an AR-enhanced system should aim to provide inspection personnel with intuitive, real-time data to enable efficient and comprehensive structural evaluations.

ACKNOWLEDGEMENT

This project has received funding from the European Union's Horizon 2022 research and innovation programme under grant agreement No. 101119554 "BRIDGITISE" (call HORIZON-MSCA-2022-DN-01).

REFERENCES

[1] "Bridge Inspector Certification Scheme Scheme Manual National Highway Sector Scheme 31 for the Bridge Inspector Certification Scheme Bridge Inspector Certification Scheme", Accessed: May 06, 2025. [Online]. Available: www.lantra-awards.co.uk

[2] J. M. W. Brownjohn, "Structural health monitoring of civil infrastructure," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 589–622, Feb. 2007, doi: 10.1098/RSTA.2006.1925/ASSET/215B12C1-A04E-4972-A61C-5D4B6253F522/ASSETS/GRAPHIC/589FIG13.JPEG.

[3] P. Hüthwohl, R. Lu, and I. Brilakis, "Multi-classifier for reinforced concrete bridge defects," *Autom Constr*, vol. 105, p. 102824, Sep. 2019, doi: 10.1016/J.AUTCON.2019.04.019.

[4] R. van Gils, "The power of monitoring for life extension of civil engineering structures." Accessed: Jun. 04, 2025. [Online]. Available: https://gww-bouw.nl/en/bridge-construction/the-power-of-monitoring-for-life-extension-of-civil-engineering-structures/?utm_source=chatgpt.com

[5] J. Flotzinger, P. J. Rösch, and T. Braml, "dacl10k: Benchmark for Semantic Bridge Damage Segmentation," in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024, pp. 8611–8620. doi: 10.1109/WACV57701.2024.00843.

[6] J. Flotzinger *et al.*, "dacl-challenge: Semantic Segmentation during Visual Bridge Inspections," in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 2024, pp. 716–725. doi: 10.1109/WACVW60836.2024.00084.

[7] C. Z. Dong and F. N. Catbas, "A review of computer vision–based structural health monitoring at local and global levels," 2021. doi: 10.1177/1475921720935585.

[8] Y.-J. Cha, R. Ali, J. Lewis, and O. Büyüköztürk, "Deep learning-based structural health monitoring," *Autom Constr*, vol. 161, p. 105328, 2024, doi: https://doi.org/10.1016/j.autcon.2024.105328.

[9] "Renewal of civil infrastructure. Dutch national forecast for replacement and renovation," 2021. [Online]. Available: www.tno.nl

[10] P. Huethwohl, "Cambridge bridge inspection dataset," 2017.

[11] S. Dorafshan, R. J. Thomas, and M. Maguire, "SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks," *Data Brief*, vol. 21, pp. 1664–1668, 2018, doi: https://doi.org/10.1016/j.dib.2018.11.015.

[12] S. Li and X. Zhao, "Image-Based Concrete Crack Detection Using Convolutional Neural Network and Exhaustive Search Technique," *Advances in Civil Engineering*, vol. 2019, no. 1, p. 6520620, Jan. 2019, doi: 10.1155/2019/6520620.

[13] H. Xu, X. Su, Y. Wang, H. Cai, K. Cui, and X. Chen, "Automatic Bridge Crack Detection Using a Convolutional Neural Network," *Applied Sciences 2019, Vol. 9, Page 2867*, vol. 9, no. 14, p. 2867, Jul. 2019, doi: 10.3390/APP9142867.

[14] E. Bianchi and M. Hebdon, "Trained Model for the Classification of Bearing Condition States," Apr. 2021, doi: 10.7294/16628698.v1.

[15] E. Bianchi and M. Hebdon, "Visual structural inspection datasets," *Autom Constr*, vol. 139, p. 104299, Jul. 2022, doi: 10.1016/J.AUTCON.2022.104299.

[16] M. Mundt, S. Majumder, S. Murali, P. Panetsos, and V. Ramesh, "Meta-learning Convolutional Neural Architectures for Multi-target Concrete Defects Classification with the COncrete Defects BRidge IMage Dataset," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 11188–11197, Apr. 2019, doi: 10.1109/CVPR.2019.01145.

[17] C. Benz, P. Debus, H. K. Ha, and V. Rodehorst, "Crack Segmentation on UAS-based Imagery using Transfer Learning," *International Conference Image and Vision Computing New Zealand*, vol. 2019-December, Dec. 2019, doi: 10.1109/IVCNZ48456.2019.8960998.

[18] S. Kulkarni, S. Singh, D. Balakrishnan, S. Sharma, S. Devunuri, and S. C. R. Korlapati, "CrackSeg9k: A Collection and Benchmark for Crack Segmentation Datasets and Frameworks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13807 LNCS, pp. 179–195, Aug. 2022, doi: 10.1007/978-3-031-25082-8_12.

[19] J. Flotzinger, P. J. Rösch, N. Oswald, and T. Braml, "dacl1k: Real-World Bridge Damage Dataset Putting Open-Source Data to the Test," Sep. 2023, Accessed: Apr. 05, 2025. [Online]. Available: https://arxiv.org/abs/2309.03763v1

[20] C. Benz and V. Rodehorst, "Image-Based Detection of Structural Defects Using Hierarchical Multi-scale Attention," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13485 LNCS, pp. 337–353, 2022, doi: 10.1007/978-3-031-16788-1_21/TABLES/3.

[21] "AASHTO Bridge Element Inspection Guide Manual," 2010.

[22] T. W. Ryan, C. E. Lloyd, M. S. Pichura, D. M. Tarasovich, S. Fitzgerald, and M. B. International, "Bridge Inspector's Reference Manual (BIRM) (2022 NBIS)," Mar. 2023, doi: 10.21949/1503647.

DOI: 10.3217/978-3-99161-057-1-052

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Accessed: Apr. 07, 2025. [Online]. Available: http://code.google.com/p/cuda-convnet/

[24] B. Baker, O. Gupta, N. Naik, and R. Raskar, "Designing Neural Network Architectures using Reinforcement Learning," *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, Nov. 2016, Accessed: Apr. 07, 2025. [Online]. Available: https://arxiv.org/abs/1611.02167v3

[25] H. Pham, M. Y. Guan, B. Zoph, Q. V. Le, and J. Dean, "Efficient Neural Architecture Search via Parameter Sharing," *35th International Conference on Machine Learning, ICML 2018*, vol. 9, pp. 6522–6531, Feb. 2018, Accessed: Apr. 07, 2025. [Online]. Available: https://arxiv.org/abs/1802.03268v2

[26] V. Iglovikov and A. Shvets, "TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation," Jan. 2018, Accessed: Apr. 07, 2025. [Online]. Available: https://arxiv.org/abs/1801.05746v1

[27] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2014, Accessed: Apr. 07, 2025. [Online]. Available: https://arxiv.org/abs/1409.1556v6

[28] G. Jocher *et al.*, "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation", doi: 10.5281/ZENODO.7347926.

[29] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," Dec. 2016, Accessed: Apr. 07, 2025. [Online]. Available: https://arxiv.org/abs/1612.03144v2

[30] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," *Adv Neural Inf Process Syst*, vol. 15, pp. 12077–12090, May 2021, Accessed: Apr. 07, 2025. [Online]. Available: https://arxiv.org/abs/2105.15203v3

[31] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention Mask Transformer for Universal Image Segmentation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2022-June, pp. 1280–1289, Dec. 2021, doi: 10.1109/CVPR52688.2022.00135.

[32] W. Wang *et al.*, "InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2023, pp. 14408–14419. doi: 10.1109/CVPR52729.2023.01385.

[33] W. Xia, Z. Cheng, Y. Yang, and J.-H. Xue, "Cooperative Semantic Segmentation and Image Restoration in Adverse Environmental Conditions," Nov. 2019, Accessed: Apr. 07, 2025. [Online]. Available: https://arxiv.org/abs/1911.00679v3

[34] Z. Tu *et al.*, "MaxViT: Multi-Axis Vision Transformer," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13684 LNCS, pp. 459–479, Apr. 2022, doi: 10.1007/978-3-031-20053-3_27.

[35] A. Howard *et al.*, "Searching for MobileNetV3," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-October, pp. 1314–1324, May 2019, doi: 10.1109/ICCV.2019.00140.

[36] X. W. Ye, C. Z. Dong, and T. Liu, "A Review of Machine Vision-Based Structural Health Monitoring: Methodologies and Applications," *J Sens*, vol. 2016, no. 1, p. 7103039, Jan. 2016, doi: 10.1155/2016/7103039.

[37] X. W. Ye, T. Jin, and C. B. Yun, "A review on deep learning-based structural health monitoring of civil infrastructures," *Smart Struct Syst*, vol. 24, no. 5, SI, pp. 567–585, Nov. 2019, doi: 10.12989/sss.2019.24.5.567.

[38] Q.; Yuan *et al.*, "A Review of Computer Vision-Based Crack Detection Methods in Civil Infrastructure: Progress and Challenges," *Remote Sensing 2024, Vol. 16, Page 2910*, vol. 16, no. 16, p. 2910, Aug. 2024, doi: 10.3390/RS16162910.

[39] A. Gomez-Cabrera and P. J. Escamilla-Ambrosio, "Review of Machine-Learning Techniques Applied to Structural Health Monitoring Systems for Building and Bridge Structures," *APPLIED SCIENCES-BASEL*, vol. 12, no. 21, Nov. 2022, doi: 10.3390/app122110754.

[40] M. Azimi, A. D. Eslamlou, and G. Pekcan, "Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review," *SENSORS*, vol. 20, no. 10, May 2020, doi: 10.3390/s20102778.

[41] L. Sun, Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah, "Review of Bridge Structural Health Monitoring Aided by Big Data and Artificial Intelligence: From Condition Assessment to Damage Detection," *JOURNAL OF STRUCTURAL ENGINEERING*, vol. 146, no. 5, May 2020, doi: 10.1061/(ASCE)ST.1943-541X.0002535.

[42] "NEN 2767 Part-2 Defects List - 297633".

[43] "CS 450 - Inspection of highway structures." Accessed: Jun. 01, 2025. [Online]. Available: https://www.standardsforhighways.co.uk/search/c5c2c3e5-f7f3-4c94-8254-184e41ccd1a0

[44] "CSA S6:19, Canadian Highway Bridge Design Code - CSA Group." Accessed: Jun. 01, 2025. [Online]. Available: https://www.csagroup.org/canadian-highway-bridge-design-code/

[45] "AS 5100.7:2017 - Standards Australia." Accessed: Jun. 01, 2025. [Online]. Available: https://www.standards.org.au/standards-catalogue/standard-details?designation=as-5100-7-2017

[46] A. Rytter, "Vibrational Based Inspection of Civil Engineering Structures," 1993. Accessed: Apr. 07, 2025. [Online]. Available: https://vbn.aau.dk/en/publications/vibrational-based-inspection-of-civil-engineering-structures

[47] K. Worden and J. M. Dulieu-Barton, "An Overview of Intelligent Fault Detection in Systems and Structures," *Struct Health Monit*, vol. 3, no. 1, pp. 85–98, 2004, doi: 10.1177/1475921704041866.

[48] S. Wang, S. A. Zargar, and F. G. Yuan, "Augmented reality for enhanced visual inspection through knowledge-based deep learning," *Struct Health Monit*, vol. 20, no. 1, pp. 426–442, Jan. 2021, doi: 10.1177/1475921720976986;PAGE:STRING:ARTICLE/CHAPTER.

[49] J. Zhang, S. Qian, and C. Tan, "Automated bridge surface crack detection and segmentation using computer vision-based deep learning model," *Eng Appl Artif Intell*, vol. 115, Oct. 2022, doi: 10.1016/j.engappai.2022.105225.

[50] G. Xu, Q. Yue, and X. Liu, "Deep learning algorithm for real-time automatic crack detection, segmentation, qualification," *Eng Appl Artif Intell*, vol. 126, Nov. 2023, doi: 10.1016/j.engappai.2023.107085.

[51] X. Wang, Q. Yue, and X. Liu, "SBDNet: A deep learning-based method for the segmentation and quantification of fatigue cracks in steel bridges," *Advanced Engineering Informatics*, vol. 65, May 2025, doi: 10.1016/j.aei.2025.103186.

[52] E. Mohammed Abdelkader, O. Moselhi, M. Marzouk, and T. Zayed, "Entropy-Based Automated Method for Detection and Assessment of Spalling Severities in Reinforced Concrete Bridges," *Journal of Performance of Constructed Facilities*, vol. 35, no. 1, Feb. 2021, doi: 10.1061/(asce)cf.1943-5509.0001544.

[53] S. Wang, J. Wan, S. Zhang, and Y. Du, "Automatic Detection Method for Concrete Spalling and Exposed Steel Bars in Reinforced Concrete Structures Based on Machine Vision," *Buildings*, vol. 14, no. 6, Jun. 2024, doi: 10.3390/buildings14061580.