Communications Engineering

https://doi.org/10.1038/s44172-026-00749-7

Article in Press

AI-powered bridge defect detection and condition assessment using drone-collected visual data

Received: 31 October 2025

Accepted: 28 July 2026

Published online: 12 August 2026

Cite this article as: Tsegaye, N.T., Cui, Q., Zhang, Y. AI-powered bridge defect detection and condition assessment using drone-collected visual data. *Commun Eng* (2026). https://doi.org/10.1038/s44172-026-00749-7

Natnael Taye Tsegaye, Qingbin Cui & Yunfeng Zhang

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

AI-powered bridge defect detection and condition assessment using drone-collected visual data

Natnael Taye Tsegaye[1] (ORCID: 0009-0008-5119-9677), Qingbin Cui[2], Yunfeng Zhang [3*]

[1]PhD Candidate, Department of Civil and Environmental Engineering, University of Maryland, College Park, MD 20742, USA; email: mypreciousshs@gmail.com, natnael@umd.edu

[2]Professor, Department of Civil and Environmental Engineering, University of Maryland, College Park, MD 20742, USA; email: cui@umd.edu

[3]Professor, Department of Civil and Environmental Engineering, University of Maryland, College Park, MD 20742, USA; email: zyf@umd.edu

*Corresponding author: Dr. Yunfeng Zhang

ARTICLE IN PRESS

Abstract

Maintaining safe bridges requires regular inspections, but traditional methods depend on manual visual surveys that are labor-intensive, subjective, and often limited by access and safety constraints. Recent advances in drones and artificial intelligence (AI) provide opportunities to improve efficiency, consistency, and coverage in bridge monitoring. In this study, we demonstrate a workflow that uses drone-collected data to detect and segment eight types of concrete deterioration. Post-processing methods stabilize detections across video frames and produce machine-readable outputs that align closely with manual tallies. Results from ten bridges in Kansas indicate that the automated defect counts exhibit strong directional consistency with the nationally reported condition ratings. The approach also projects detections into three-dimensional reconstructions for measurement and visualization, and a prototype web service enables near-real-time deployment. Collectively, these findings suggest that drone- and AI-powered inspections can effectively complement traditional inspection practices while offering scalable and data-driven support for infrastructure management.

Introduction

Bridges are critical lifelines within transportation infrastructure networks, enabling mobility, economic activity, and access to essential services. In the United States, more than 623,000 bridges provide the backbone of the national roadway system[1]. Many of these bridges are aging; the American Society of Civil Engineers (ASCE) estimates that 42% are at least 50 years old, and over 42,000 or 6.8% of the total bridges in the United States are rated “Poor” in the National Bridge Inventory (NBI)[2]. Delayed maintenance, limited inspection resources, and increasing traffic demands exacerbate safety risks, particularly in rural areas where funding and technical expertise may be more constrained[3]. Routine inspections are therefore indispensable for maintaining public safety, ensuring serviceability, and guiding resource allocation for rehabilitation and maintenance. Current federal standards mandate regular inspection cycles, with the National Bridge Inspection

ARTICLE IN PRESS

Standards (NBIS) requiring safety inspections at least once every 24 months for highway bridges that exceed 6 m (20 feet)[4].

Conventional bridge inspection methods rely heavily on manual visual inspection or low precision equipment[5]. Although these methods remain the backbone of current practice, they are associated with several limitations. Inspectors must often access difficult or hazardous locations, exposing themselves to safety risks[6]. The process is also labor-intensive, time-consuming, and can be disruptive to traffic, as it may require lane closures or the use of specialized access equipment. Moreover, most existing indices rely on subjective assumptions and are highly dependent on the expertise of inspectors[7], leading to variability in condition assessments across inspectors and agencies. Beyond these technical constraints, agencies face resource limitations[8]: with shrinking budgets and staff shortages, it is increasingly difficult to meet inspection needs, particularly in rural areas that manage extensive inventories of small bridges. Taken together, these challenges underscore the need for scalable, objective, and technology-assisted inspection methods that can augment and enhance conventional practices.

Unmanned Aerial Systems (UAS), or drones, have emerged as promising tools for infrastructure inspection, offering safer and more efficient alternatives to traditional methods. Early applications demonstrated the feasibility of using multicopter Unmanned Aerial Vehicles (UAVs) to capture imagery of bridges in Oregon, enabling inspectors to identify defects such as bolt pattern irregularities, spalling, leaking joints, and cracking without physically accessing hazardous areas[9]. Larger-scale efforts, such as the Minnesota Department of Transportation's program, further validated the technology. More than 60 bridges were inspected using specialized UAVs to access tight and difficult areas, thereby improving safety while reducing inspection costs[10]. UAVs have also been shown to serve as cost-effective supplements to conventional inspections of timber, masonry, and reinforced concrete bridges, demonstrating their potential to lower costs while maintaining diagnostic quality[11,12]. Importantly, safety benefits have been quantified, with UAV use associated with reduced worker compensation risks compared to traditional access methods[13]. These findings underscore that UAVs can extend the reach of inspections, support access to difficult environments, and enhance the efficiency and safety of field operations.

More recently, research has shifted toward automated inspection frameworks that combine UAV imaging with computer vision and machine learning techniques. Studies have proposed UAV-enabled bridge information models that integrate geo-referenced defect detection, localization, and tracking within asset management workflows[14,15]. Such frameworks facilitate element-level documentation while bridging UAV data with digital Bridge Information Modeling (BrIM) environments. Complementary investigations applied image-processing approaches to steel and masonry bridges, demonstrating how UAV-captured imagery coupled with color-based or machine-learning classifiers can quantify defect extents[16]. Flotzinger et al.[17] introduced dacl10k, the largest real-world dataset for multi-label semantic segmentation of reinforced concrete bridge defects, comprising 9,920 images across 12 defect classes and 6 structural components. Their baseline experiments highlight the challenges of overlapping and visually similar defects,

ARTICLE IN PRESS

achieving a best mean IoU of 0.424 and establishing dacl10k as a critical benchmark for advancing automated concrete defect detection. Advances in imaging resolution and sensor payloads have also expanded UAV capabilities. For instance, high-definition (HD) cameras have been demonstrated to detect simulated cracks as small as 0.1 mm when flown within 1–3 m of a structure[18]. Similarly, simple integrations, such as equipping off-the-shelf drones with 360° cameras, have been shown to reveal defects missed by conventional approaches[19]. Beyond surface imagery, UAV platforms have been successfully applied to displacement monitoring of structural components under load, providing dynamic response data not possible with static visual inspections[20–25].

In parallel, UAV imagery has enabled 3D reconstruction of bridge structures, providing geometric information critical for inspection and structural assessment. Frameworks such as ARGO 3D integrate drone data collection with digital twin reconstruction, AI-powered defect detection, and automated reporting, enabling fully remote inspection pipelines[26]. Other efforts demonstrated how UAV-based reconstructions can be used to extract structural dimensions and update analytical models for seismic risk evaluation[27]. Photogrammetric pipelines that combine Structure-from-Motion (SfM) and Multi-View Stereo (MVS) with deep learning (e.g., Mask R-CNN) have achieved high-precision crack segmentation and georeferenced measurements of crack length, width, and area[28]. Recent demonstrations using consumer-grade drones reinforced the practicality of this approach, showing that automated UAV flights coupled with low-cost hardware and open-source tools can generate 3D point clouds and defect documentation[29]. More comprehensive UAS-based methodologies have also been proposed, combining video-based inspections, 3D reconstruction, and systematic data management within a single workflow, demonstrating the potential of multirotor UAVs to support large-scale condition assessment and damage detection programs[30]. Recent studies have further advanced bridge digitalization through the integration of UAV photogrammetry and LiDAR data. Cabral et al.[31] demonstrated a hybrid reconstruction framework that combines UAV imagery with terrestrial and mobile LiDAR point clouds for railway bridge reconstruction. Their results showed improvements in geometric accuracy, texture quality, and model completeness through multi-sensor data fusion, highlighting the growing role of integrated 3D reconstruction workflows for infrastructure monitoring and asset management.

Advances in deep learning, particularly convolutional neural networks (CNNs), have revolutionized computer vision and offer powerful opportunities for automating bridge inspections. Models such as Faster R-CNN[32], Mask R-CNN[33], and the YOLO family[34,35] enable real-time defect detection and segmentation, and when trained on labeled bridge imagery, these models can identify a wide range of deterioration patterns. Instance segmentation, in particular, provides significant advantages over bounding-box detection by delineating the exact geometry of defects[33]. This allows for more precise quantification of defect area and shape, which is essential for structural assessment. More recently, hybrid frameworks combining object detectors with foundation models have been explored. For example, Cabral et al.[36] proposed a YOLO11 and Segment Anything Model (SAM) pipeline for multi-damage segmentation, demonstrating the

ARTICLE IN PRESS

potential of integrating object detection and foundation models for improving the delineation of cracks, efflorescence, and exposed reinforcement in bridge inspection imagery. However, most prior studies remain limited in scope, focusing primarily on still imagery and bounding-box detection while neglecting the more challenging problem of video-based analysis. Few efforts have attempted to benchmark AI-generated defect metrics against established inspection standards such as the NBI. Moreover, to the best of the authors' knowledge, no prior studies have demonstrated the projection of 2D AI-based instance segmentation detections into georeferenced 3D reconstructions of bridge elements, which would enable inspectors to visualize, measure, and track defects over time. Finally, although research prototypes are abundant, very few AI inspection studies have been translated into deployable tools for practitioners. Without accessible platforms such as web services, APIs, or containerized applications, the practical impact of AI remains largely confined to academic demonstrations.

To address these challenges, this project was initiated to explore the integration of drones, AI, and geospatial platforms into operational bridge inspection workflows. Conducted across four counties in Kansas, USA, the project brought together diverse urban and rural contexts, reflecting the varied challenges faced by transportation agencies. Supplementary information Tables 1 and Table 2 provide details on the demographics of the project's service area, as well as bridge deficiency and safety information within the same region. This study introduces a pipeline that spans data acquisition with UAVs, AI-based instance segmentation of eight concrete defect classes, post-processing to generate machine-readable outputs, and projection of some detections into georeferenced 3D reconstructions for measurement and visualization. Results demonstrate that the AI model can reliably detect defects across both images and videos, with temporal post-processing strategies (ByteTrack and IoU filtering) aligning automated defect counts closely with manual tallies. When compared against NBI condition ratings, the AI-derived metrics showed directional consistency, highlighting their potential as supplemental indicators of bridge condition. Further, by deploying the trained model as a prototype web service, the study closes the gap between research and practice, offering inspectors a practical, near-real-time tool for field applications. Collectively, these findings show that drone- and AI-powered inspection can improve safety, efficiency, and consistency in bridge monitoring, contributing to the development of scalable and technology-assisted asset management systems.

Methods

Overview

Our workflow transforms drone imagery, videos, and Ground Control Points (GCPs) into georeferenced information on bridge defects to support asset management. The pipeline has four stages: (i) acquisition, (ii) 2D image instance segmentation, (iii) post-processing and standardized outputs, and (iv) 3D/GIS (Geographic Information Systems) integration for measurement, visualization, and decision support. Acquisition covers flight-path planning, UAS data capture (RGB imagery, videos, LiDAR, GCPs) and basic pre-processing. Instance segmentation uses a

ARTICLE IN PRESS

YOLOv11-based model customized for eight concrete defect classes. Post-processing produces an AI-powered detection API, machine-readable JSON, per-image and per-video counts of defects, and crops for downstream VLM/LLM analysis. Finally, 2D detections are georeferenced and projected to 3D reconstructions (orthomosaics, point clouds, meshes) of bridge elements and surfaced in a GIS environment to support engineering assessments at the element level.

Data acquisition

The USDOT SMART project was implemented in four Kansas counties: Sedgwick, Saline, Cowley, and Cloud in collaboration with NCRPC. The selected bridges are county-owned and some are in fair and poor condition, ensuring that they are representative of typical infrastructure challenges while still being suitable and distinctive enough for initial testing. All bridges are constructed from concrete, a common material in the area, which allows for consistent assessment across similar structures. Additionally, the bridges were ranked based on their span, which helps prioritize those that present unique structural challenges and opportunities for data collection. The drone evaluation was focused on aspects such as drone durability, imaging quality, and operational range to ensure alignment with the project's specific use cases. Fig. 1 shows some of the drones and their use in pilot test flights on selected bridges prior to broader deployment of drone-based bridge inspections across the four counties.



(a)

ARTICLE IN PRESS



(b)

Fig. 1

Before the field tests, a bridge inspection guideline was also developed and provided to county engineers participating in the field test with the aim of leveraging drones to gather a comprehensive dataset of bridge information for developing an AI algorithm[35] to detect structural surface defect issues. This guideline and a training program document outline the process for conducting efficient, safe, and thorough bridge inspections using UAS[37], ensuring structural integrity, identifying potential defects, and adhering to regulatory standards[38]. Based on previous studies [39–41] a flight path strategy was developed to ensure complete visual coverage of all structural components of the bridge while maintaining optimal imaging conditions for defect detection. The planning considered critical parameters such as flight height, lateral offset, camera angle, image overlap, and drone speed based on the desired ground sampling distance (GSD) [42,43]. Distinct paths were defined for the top, sides, and underside of the bridge, ensuring safe drone navigation and consistent data quality. The planning also factored in obstructions like piers and abutments, with flexible path adjustments to capture hard-to-reach areas.

Pilot testing was conducted on three bridges in Saline County using two different sizes of drones to evaluate their effectiveness in detailed bridge inspections. These bridges, primarily in fair or poor conditions [44], served demonstration and verification purposes to gather further data and assess drone operational efficiency and reliability. Building on the success of these pilots, the project expanded to a large-scale implementation involving twenty bridges across the four counties in Kansas, representing a wide variety of structural types, ages, span lengths, and condition ratings. Fig. 2 presents the geographic distribution of the inspected bridges, while Supplementary Information Table 3 provides detailed characteristics of each bridge included in the field tests. All drone flights were performed in accordance with established inspection guidelines and planned

flight paths. The algorithm used in this project was developed using a large dataset of reinforced concrete defect images sourced from public repositories[17], plus the images collected from the four counties. This dataset was customized with project-specific damage labeling, allowing for precise detection of structural issues and enabling the drones to identify damage patterns effectively.

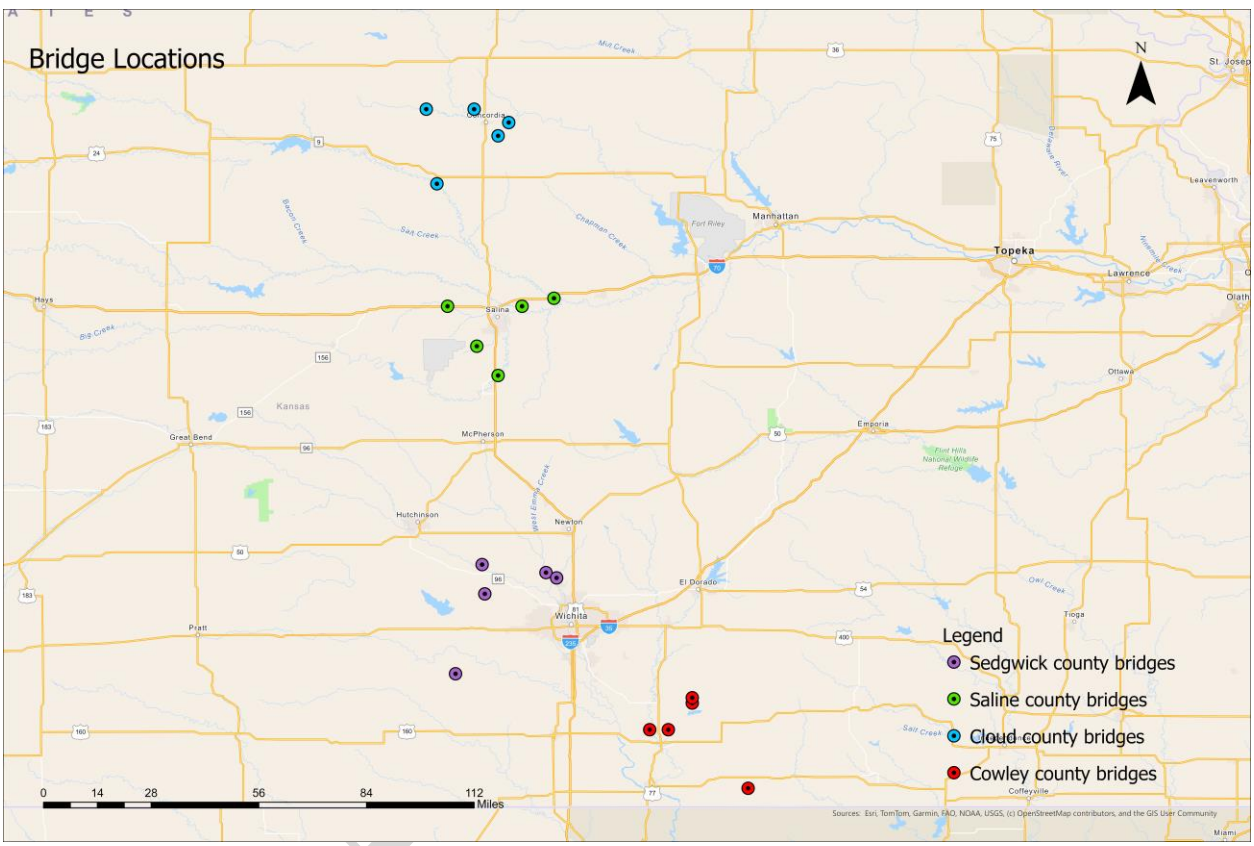


Fig. 2

Segmentation Model & Training

We adopt YOLOv11-seg, an extension of the YOLO family of real-time object detectors[35], for instance segmentation of eight concrete defect classes (Table 1): Crack, ACrack (Alligator Crack), Efflorescence, WConccor (Water-Induced Concrete Washouts), Spalling, Wetspot (Wet spot), Rust, and ExposedRebars (Exposed Reinforcement Bars). Supplementary Information Table 4 illustrates the visual characteristics, causes and presents annotated images for each defect type. The dataset used in this study consisted of 6,625 manually annotated images, including 5,962 selected and curated public imagery[17] adapted and re-annotated to YOLO format, and 663 UAV-collected field images acquired from bridge inspections in Kansas counties. All defect annotations were manually created by human annotators using the LabelMe[45] annotation tool and subsequently converted into YOLO segmentation format. To prepare the dataset for training, images with corresponding annotation files were programmatically divided into training, validation, and test subsets using a

ARTICLE IN PRESS

reproducible random split (random seed = 42). Approximately 73% of the images (4,837 images) were allocated to the training set, 17% (1,126 images) to the validation set, and 10% (662 images) to the test set. Only image-annotation pairs were included in the dataset. The validation set was used for model monitoring during training, while the test set was used for final performance evaluation. Images were normalized and resized; videos were sampled at configurable strides for frame-level supervision prior to training. Metadata were retained to support later geo-projection.

Table 1 | Summary of Defect Classes and Engineering Significance

Class Label	Defect Name	Description	Engineering Significance
Crack	Crack	Linear fracture in concrete, either structural (load-induced) or non-structural (temperature/shrinkage).	May indicate structural stress; allows moisture ingress, accelerating deterioration and corrosion.
ACrack	Alligator Crack	Interconnected, map-like crack patterns caused by shrinkage or temperature effects; typically, non-structural.	Increases permeability, enabling moisture penetration and long-term durability issues.
Efflorescence	Efflorescence	White crystalline deposits on the concrete surface caused by leaching of calcium compounds through moisture movement.	Indicates moisture ingress and permeability issues; may signal conditions that promote corrosion and long-term deterioration.
WConccor	Water-Induced Concrete Washouts, Erosion and Corrosion	Deterioration of concrete due to prolonged exposure to water, often leading to surface degradation.	Suggests sustained moisture exposure and potential chemical attack; can accelerate long-term deterioration.
Spalling	Spalling	Localized breaking, flaking, or chipping of concrete, often exposing aggregates or reinforcement.	Indicates advanced deterioration, commonly due to reinforcement corrosion or overstress; may reduce structural capacity and durability.
Wetspot	Wet spot	Visible darkened or damp areas on the concrete surface due to moisture accumulation.	Indicates active water infiltration or drainage issues; may precede other forms of deterioration.
Rust	Rust	Reddish-brown discoloration on concrete and steel surfaces caused by corrosion of steel.	Indicates reinforcement corrosion; often associated with cracking and potential structural weakening.
ExposedRebars	Exposed Reinforcement Bars	Visible steel reinforcement due to loss of concrete cover.	Represents advanced deterioration; significantly increases risk of structural degradation and requires urgent repair.

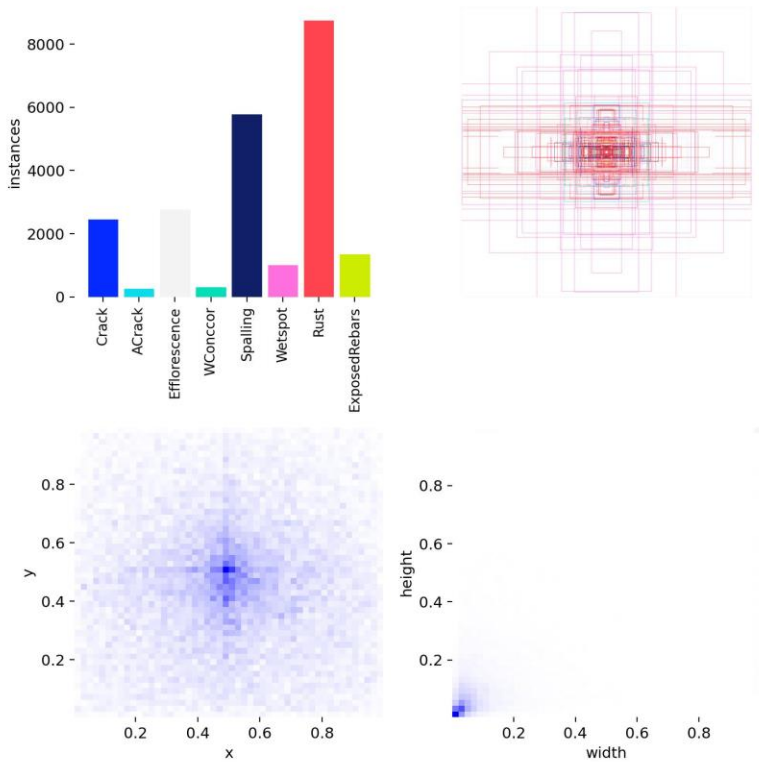
Note: Definitions and engineering significance are based on standard bridge inspection guidelines and prior literature[17,46–48].

The training of the defect detection model was carried out on a dedicated Google Cloud Platform (GCP) virtual machine equipped with an NVIDIA A100 GPU (80 GB). Datasets were initially stored on a solid-state drive (SSD) and subsequently uploaded to Google Cloud Storage for persistent access. Separate folders were created for training, validation, and test images and videos,

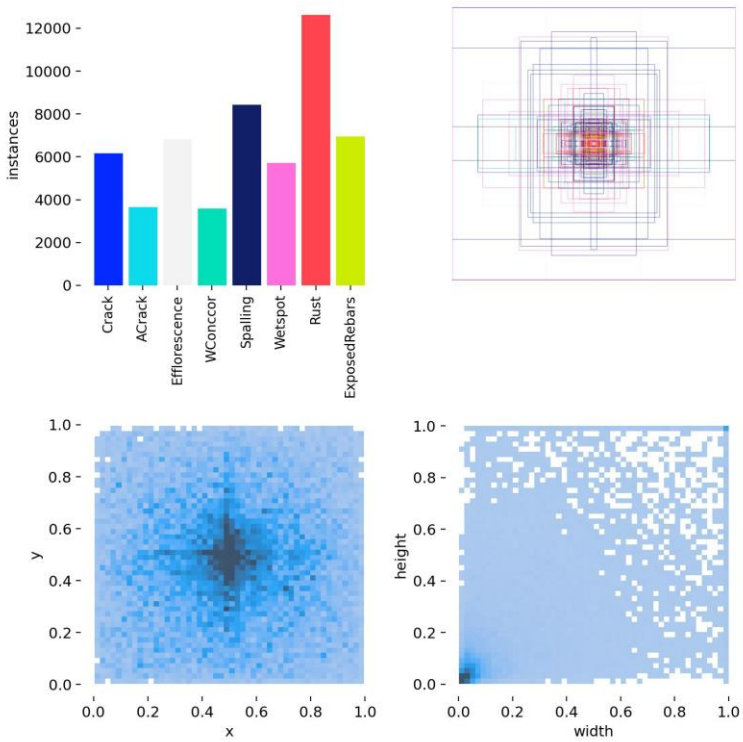
and corresponding annotations in YOLO format were also transferred. These datasets were later downloaded to the compute environment for training.

A custom virtual machine instance was provisioned using the Google Compute Engine, configured with Ubuntu 24.04 LTS (NVIDIA-optimized image), 170 GB persistent disk storage, and full API access to cloud services. A Python 3.12 virtual environment was established to ensure isolated dependency management. Within this environment, the Ultralytics YOLO package[35,49] and other necessary libraries (e.g., libGL for OpenCV) were installed. The YOLOv11m-seg pretrained weights were verified to ensure correct environment configuration.

The dataset structure was defined through a data.yaml[50] configuration file, specifying the relative paths for training and validation sets and enumerating the eight defect categories. To address class imbalance, targeted augmentation strategies were applied to minority classes. Standard augmentations included random rotations, translations, scaling, shear, perspective transformations, horizontal and vertical flipping, and HSV color-space perturbations[51]. These transformations increased variability and reduced overfitting. To address class imbalance in the dataset, we applied class-balanced augmentation, oversampling minority classes with additional transformations to increase their representation during YOLO training. Fig. 3 summarizes the original defect class distributions and the balanced distributions obtained after augmentation.



(a)



(b)

Fig. 3

The YOLOv11 segmentation model was trained using the yolo11l-seg.pt initialization weights[34,35]. Training hyperparameters were selected based on preliminary tuning: input size of 640 × 640 pixels, an automatically selected batch size, initial learning rate of 0.001 with cosine learning-rate scheduling and a final learning-rate factor of 0.01, momentum of 0.937, and weight decay of 0.0005. The training process was set for 250 epochs with an early stopping patience of 30 epochs to prevent overfitting. Warm-up epochs (n = 3) stabilized initial convergence[50]. All training experiments were performed using mixed precision to maximize GPU efficiency. To support instance segmentation rather than bounding-box detection, the original YOLOv11 codebase was customized to generate polygonal masks without color-shaded overlays, ensuring precise boundary visualization of irregularly shaped defects[50]. This modification was critical for engineering interpretation and for downstream processing, where defect areas were later quantified and projected onto 3D reconstructions.

Model performance was evaluated on both still images and videos. For static imagery, the trained model was applied to a held-out validation dataset that was not used during training in order to assess detection performance on unseen images. Model-generated prediction outputs were saved

ARTICLE IN PRESS

both as visualization images and as text files containing bounding box and polygon coordinates of the defects, polygon area in square pixels and confidence threshold for each defect identified during testing. For dynamic evaluation, defect detection was extended to drone videos using the trained YOLOv11 weights. ByteTrack[52] was incorporated during inference to improve temporal consistency of detections and reduce duplicate counts across successive frames.

GIS and 3D Integration

These geospatial outputs were integrated into a GIS platform to support visualization, analysis, and communication of bridge condition data. The GIS environment served as a centralized repository combining model-predicted imagery and videos, bridge plans, inspection records, and NBI condition ratings. This integration enables engineers and stakeholders to monitor bridge conditions, analyze defect trends, and communicate results within a spatially contextualized framework, supporting both technical decision-making and broader community engagement[53,54].

The drone data acquisition strategy was designed primarily to support bridge inspection while also considering the requirements for 3D reconstruction. For the generated 3D reconstructions, images were extracted from high-resolution UAV video streams and complemented with GCPs prior to reconstruction to improve georeferencing accuracy and spatial consistency. Flight parameters were selected based on standard photogrammetry practices while maintaining sufficient proximity for defect-level inspection. When extracting images from UAV videos for 3D reconstruction, a high image overlap of approximately 75% was maintained in both frontal and side directions to ensure reliable feature matching and stable reconstruction. The UAV was operated at a flight height of approximately 9–12 m (30–40 ft) above the bridge deck for global coverage, while closer distances of 2–4 m (7–13 ft) were used for detailed inspection of structural elements. The drone speed and image acquisition rate were adjusted to maintain the required overlap, with images extracted at regular temporal intervals from video sequences. The flight paths were designed to capture multiple viewing angles, including deck-level, side, and underside perspectives, ensuring sufficient redundancy and coverage for SfM and MVS processing[55,56]. This acquisition strategy was designed to balance both inspection quality and reconstruction reliability.

Three-dimensional reconstructions were generated using Pix4Dmapper[57], which produced orthomosaics, dense point clouds, and textured meshes through SfM and MVS photogrammetry[58,59]. To facilitate algorithmic reproducibility independent of the Pix4Dmapper implementation, the underlying geometric transformation can be described using the standard photogrammetric pinhole camera model. During SfM processing, corresponding image features are matched across overlapping images to estimate camera poses and scene geometry, which are subsequently refined through bundle adjustment together with camera intrinsic parameters. The relationship between a 3D world point $X_w = [X, Y, Z, 1]^T$ in homogeneous coordinates and its image projection $x = [u, v, 1]^T$ is expressed as:

$$s\boldsymbol{x} = \mathbf{K}[R|t]\boldsymbol{X_w} \quad (1)$$

ARTICLE IN PRESS

where K denotes the camera intrinsic matrix, R and t represent the camera rotation matrix and translation vector, respectively, and s is a scale factor. The estimated camera parameters obtained through SfM and GCP-constrained georeferencing establish the correspondence between image space and the reconstructed 3D model.

Where available, LiDAR scans provided higher-fidelity reconstructions, particularly for undersides and shadowed regions[60,61]. All products were georeferenced to state-plane coordinate systems and vertical datums. Defect polygons produced by YOLO inference were mapped from 2D image space into georeferenced 3D reconstructions to enable quantitative assessment of deterioration. This projection was achieved through image-to-model alignment inherent in the photogrammetric workflow. UAV images used for YOLO inference were the same images used in the Pix4Dmapper reconstruction process, allowing each image to be associated with estimated camera poses and georeferenced coordinates derived from SfM and GCPs. Defect polygons identified in image space were therefore indirectly mapped into the 3D model through their correspondence with the georeferenced images. Because UAV surveys contain substantial image overlaps, the same physical defect may be detected in multiple images from different viewing angles. These observations are reconciled through the SfM and bundle-adjustment stages of the photogrammetric reconstruction, which align all images within a common georeferenced coordinate system. Consequently, defect projections from overlapping images are generally mapped to the same reconstructed surface region rather than appearing as separate spatially displaced defects. While minor boundary thickening and localized projection errors may occur due to variations in segmentation predictions across multiple views, the resulting projection generally represents a single defect region.

In cases where image metadata were lost during inference, geotags were restored by synchronizing image timestamps with UAV flight logs to ensure accurate spatial alignment[62]. This approach leverages the internal camera calibration and bundle adjustment processes within Pix4Dmapper, rather than explicit ray-casting or pixel-wise back-projection, to enable spatial visualization and approximate measurement of defects in real-world units (e.g., cm² or m²). The results demonstrate strong potential for quantifying the spatial extent of defects across bridge decks[63]. All outputs were subsequently integrated into a GIS layer stack for interactive visualization and reporting.

Post-Processing and Outputs

To support downstream analysis and bridge asset management integration, the YOLOv11 segmentation model was customized extensively. Inference outputs were modified to overlay defect labels and counts directly on inference images, thereby providing inspectors with immediate visual tallies. Each detection was stored in a structured JSON file containing the defect class label, bounding-box coordinates, polygon vertices, confidence score, and polygon area in pixels. In addition, per-image and per-video class counts were exported to additional text files to support quick checks and high-level statistical summaries. Cropped regions of individual defects were also

ARTICLE IN PRESS

generated to enable downstream analysis with Vision-Language Models (VLMs) and other learning frameworks.

For video analysis, duplicate detection across successive frames can artificially inflate defect counts. To mitigate this, we implemented a hybrid strategy: (i) the ByteTrack algorithm was employed to stabilize temporal associations of defects across frames, and (ii) an Intersection-over-Union (IoU)-based deduplication[64] was applied to suppress over-counting of overlapping detections. Frame-stride sampling was further applied to trade computational efficiency against temporal precision for long drone video sequences[65]. To correlate automated detection with established inspection benchmarks, AI-derived defect counts and severity patterns were compared with the NBI[46,66] condition ratings of the field-tested bridges. This provided a basis for validating whether AI outputs reflect the deterioration levels captured in conventional inspection protocols.

For real-world deployment, trained YOLOv11 weights were exported to the Open Neural Network Exchange (ONNX) format[67]. A lightweight prototype web service was developed using FastAPI[68], exposing /predict endpoints for both images and videos. The service returned standardized JSON outputs following the schema described above. A Swagger/OpenAPI interface was implemented to allow inspectors to upload inspection imagery and videos through a browser and immediately visualize inference results in the field. The application was containerized using Docker[69] to ensure portability across environments. The final service was deployed to the cloud platform Amazon Web Services (AWS), demonstrating near-real-time performance on both still imagery and drone video streams. This deployment represents a proof-of-concept pathway for integrating AI-based bridge inspection into existing infrastructure monitoring workflows at scale.

Results and Discussion

Model Performance Evaluation

The performance of the proposed YOLO-based instance segmentation model was evaluated using standard computer vision metrics, including precision, recall, F1-score, mean Average Precision at IoU = 0.5 (mAP@0.5), and mean Average Precision across IoU thresholds from 0.5 to 0.95 (mAP@0.5-0.95). Both training and independent test datasets were analyzed to assess model learning and generalization capability. The training performance is illustrated in Fig. 4a–d. As shown in Fig. 4a, the overall F1 score reached approximately 0.54 at a confidence threshold of 0.304, indicating a balanced trade-off between precision and recall. Class-wise analysis reveals that Wetspot, WConccor, and Rust achieved the highest F1 scores, while ACrack and Exposed Rebars exhibited comparatively lower performance. The precision-recall curves in Fig. 4b show an overall mAP@0.5 of approximately 0.556, confirming effective learning across most defect classes. Furthermore, Fig. 4c–d demonstrate the expected relationship between confidence threshold, precision, and recall, highlighting the inherent trade-off between detection accuracy and coverage.

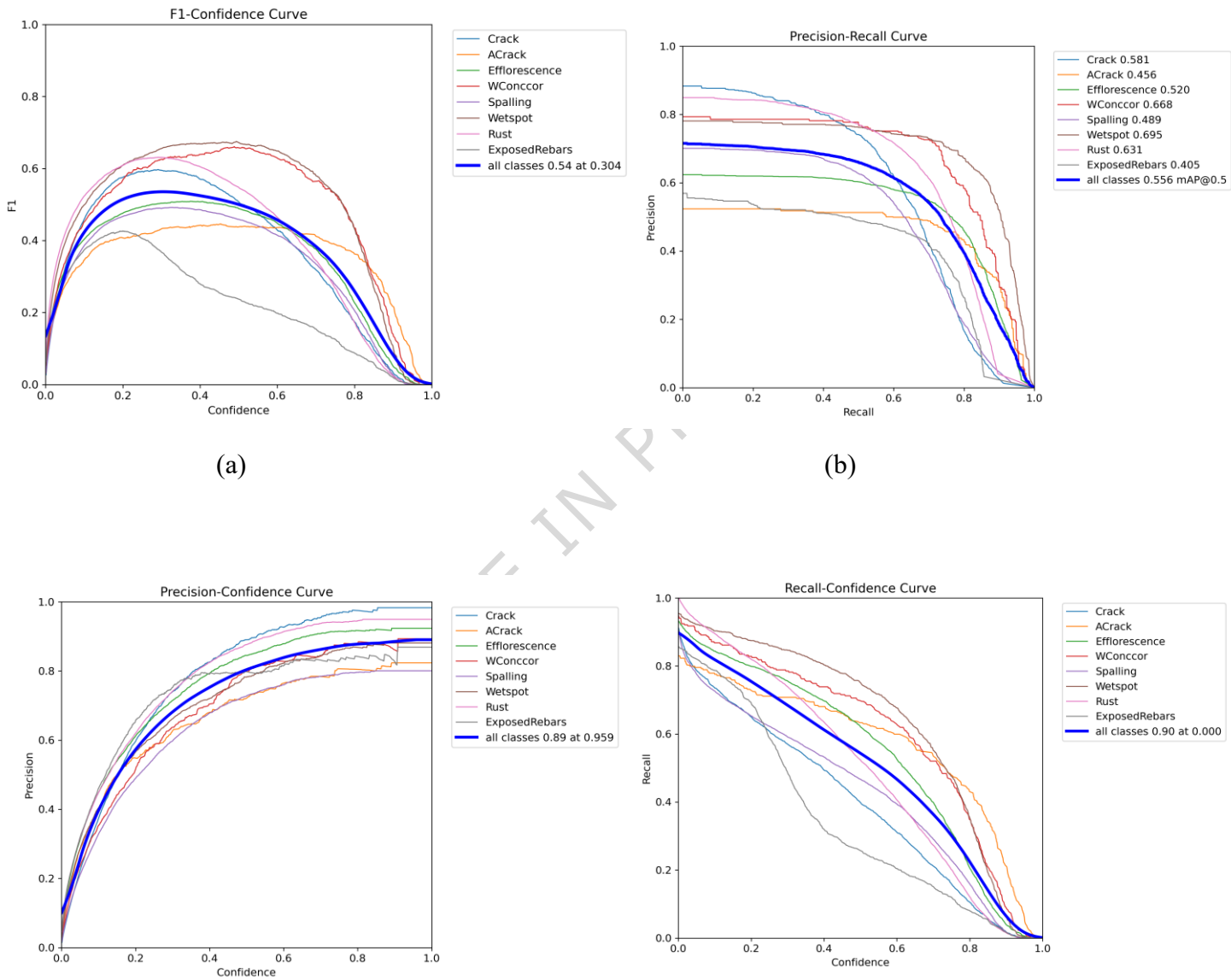


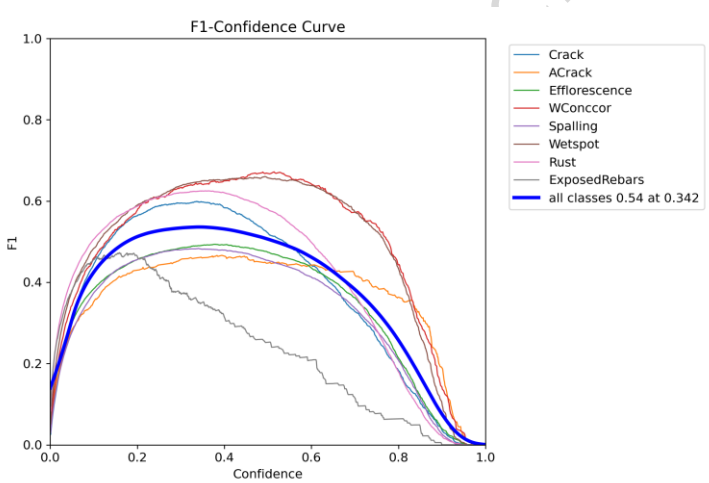
Fig. 4

To evaluate generalization performance and validate manual annotation quality, the trained model was tested on an independent dataset using a confidence threshold of 0.3 and an IoU threshold of 0.5. The results are presented in Fig. 5a–d. As shown in Fig. 5a, the overall F1 score on the test set reached approximately 0.54 at a confidence threshold near 0.342, closely matching the training

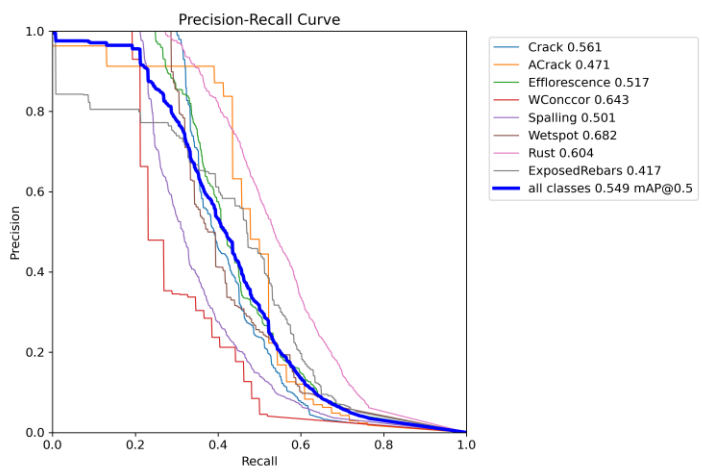
performance. The precision-recall curves in Fig. 5b show an overall mAP@0.5 of approximately 0.549, further confirming consistent performance across datasets. The precision-confidence and recall-confidence curves in Fig. 5c–d demonstrate similar trends to the training results, indicating stable model behavior and minimal overfitting.

A detailed quantitative comparison is provided in Table 2. The model achieved an mAP@0.5 of 0.556 on the training set and 0.549 on the test set, along with an mAP@0.5-0.95 of 0.317 on the test set. Class-wise results indicate that Wetspot and WConccor consistently achieve the highest detection performance, followed by Rust and Crack, while ACrack and Exposed Rebars remain more challenging to detect. The close agreement between training and test metrics demonstrates strong generalization capability and supports the reliability of the manually annotated dataset used in this study.

Although public images constituted the majority of the dataset, 663 UAV images collected from Kansas bridges were also incorporated and independently distributed across the training, validation, and test subsets. Consequently, model development and evaluation included local bridge imagery throughout the learning pipeline. Furthermore, all inference demonstrations, defect localization results, 3D defect projections, and comparisons with standard NBI condition ratings presented in this study were conducted using Kansas UAV imagery. These results provide evidence that the framework successfully captures characteristics present in local bridge inspection data and is not evaluated solely on public datasets.

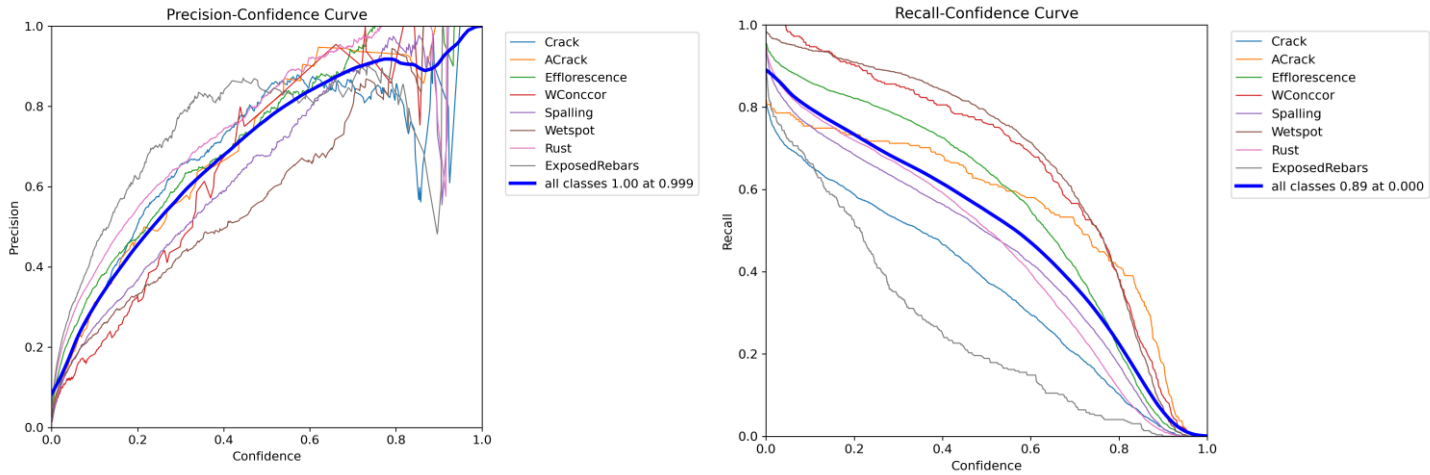


(a)



(b)

ARTICLE IN PRESS



(c) (d)

Fig. 5

Table 2 | Per-class performance comparison between training and test datasets.

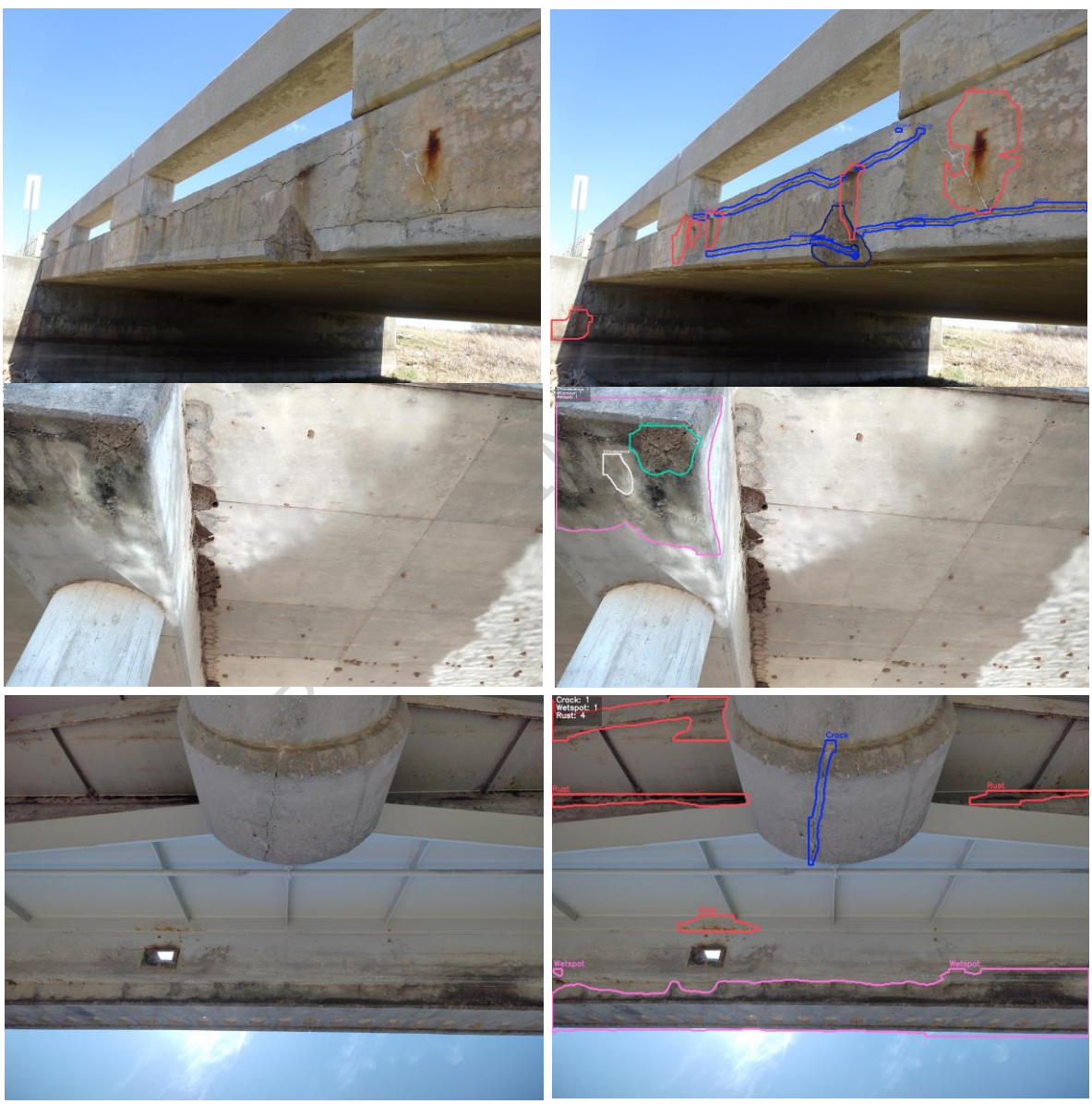
Class	Training (mAP50)	Test Set (mAP50)
Crack	0.581	0.561
ACrack	0.456	0.471
Efflorescence	0.520	0.517
WConccor	0.668	0.643
Spalling	0.489	0.501
Wetspot	0.695	0.682
Rust	0.631	0.604
ExposedRebars	0.405	0.417
All classes	0.556	0.549

Summary of class-wise and overall detection performance for the training and test datasets, reported in terms of Average Precision at IoU = 0.5 (mAP@0.5). The table highlights consistent performance across datasets and provides a comparative evaluation of model accuracy for each defect class.

YOLOv11 Segmentation on Images and Videos

The YOLOv11-seg model achieved reliable detection of the eight concrete defect categories across the drone field-collected images and videos. The customized post-processing allowed model inference outputs to include polygonal outlines, defect class labels, and per-image counts directly overlaid on imagery, giving inspectors immediate visual feedback and tallies during analysis. Model predicted images produced during inference clearly showed the added value of polygonal segmentation over bounding-box detection, as irregular defect shapes such as cracks, spalling

edges, and rust stains were accurately outlined rather than approximated by rectangular boxes. This level of detail was particularly important for estimating defect areas and for subsequent 3D projection into reconstructed bridge surfaces. Fig. 6 presents drone field-collected image ground truth and proof-of-concept results showing how the model can support bridge condition assessment by predicting different categories of concrete damage on the bridges.



(a) (b)

Fig. 6

ARTICLE IN PRESS

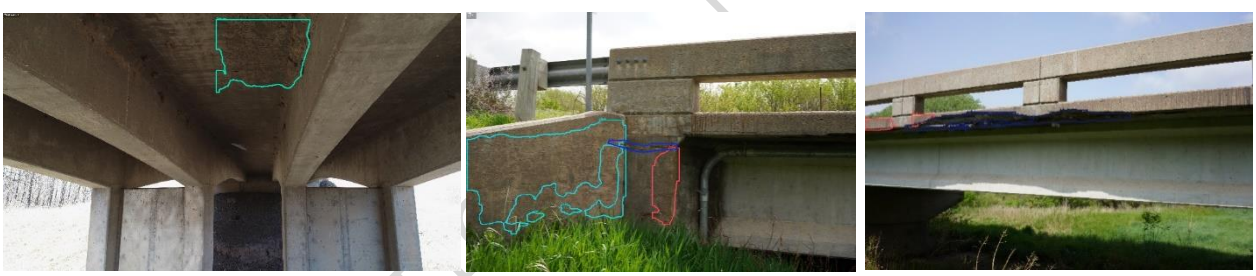
Using drone field-collected imagery, bridges across the study counties were inspected with the YOLOv11-seg model. Fig. 7 presents representative model predictions, collectively illustrating the eight target defect classes (Crack, ACrack, Efflorescence, WConccor, Spalling, Wetspot, Rust, Exposed Rebars). Panel (a) shows bridges where polygons delineate WConccor, Rust, Spalling, and Efflorescence on deck underside and girder surfaces. Panel (b) highlights bridges with detections of Wetspot and Exposed Rebars also on deck undersides and girders; the model also successfully identified multiple coexisting defects, such as Spalling, Efflorescence, and Exposed Rebars, within the same region. Panel (c) illustrates other examples where ACrack and Crack were localized on superstructure components, demonstrating sensitivity to thin, elongated geometries. Finally, Panel (d) presents challenging scenarios, showing that the model maintained robust performance when detecting defects that were either very small and difficult to discern (two examples on the left) or located in shadowed and low-light conditions (two examples on the right). The left examples highlight small spalls and cracks beneath a bridge deck, while the right examples capture spalling, rust, and exposed rebars in dark underside imagery. Across these examples, a consistent observation is that the quality of human-annotated ground truth directly influenced detection performance. High-resolution, carefully delineated annotations used during training enabled the model to learn fine-grained boundaries of irregular defects, resulting in accurate segmentation even in noisy or low-light conditions. In contrast, defects associated with coarse or inconsistent annotations were more prone to misclassification or incomplete segmentation in the resulting model predictions. This underscores the importance of consistent, high-quality labeling protocols as much as model architecture or training strategy in achieving reliable performance. While the results overall are promising, opportunities remain to further improve robustness. Expanding the training dataset with more high-quality annotations, broader defect types, and varied severity levels will help make the model more reliable across a wider range of conditions. Moreover, close-up drone imagery generally yielded more precise detections than wider contextual views, underscoring the role of resolution and detail in supporting effective instance segmentation of local damage.



(a)



(b)



(c)



(d)

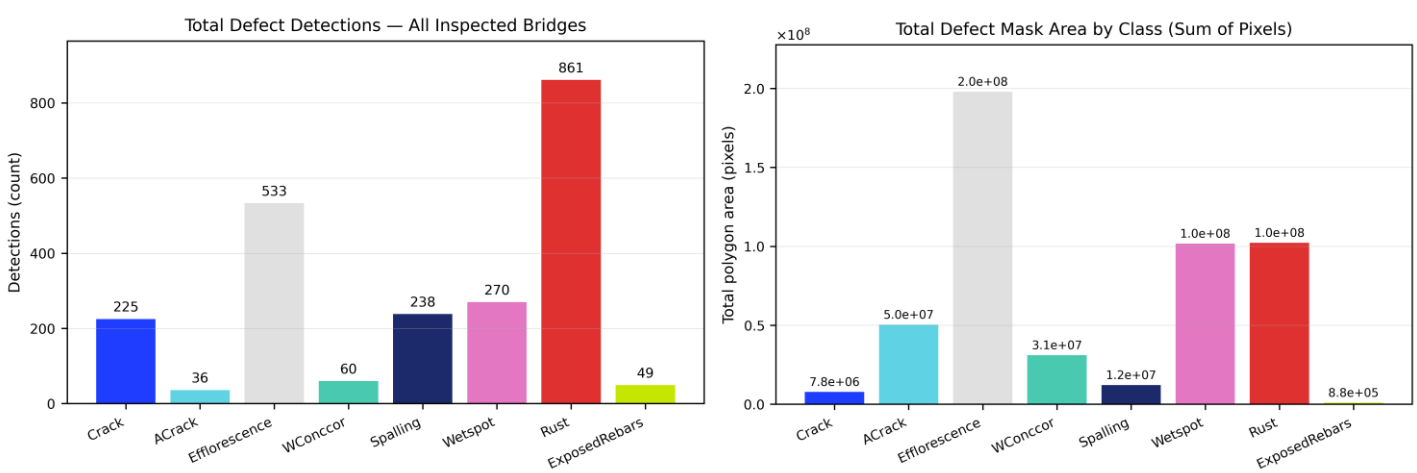
Fig. 7

Machine-Readable Outputs for Data Integration

To enable seamless integration with asset management and inspection platforms, model outputs were stored in structured formats. Each detection was saved as a JSON file containing class labels, bounding-box coordinates, polygon vertices, confidence scores, and polygon areas measured in

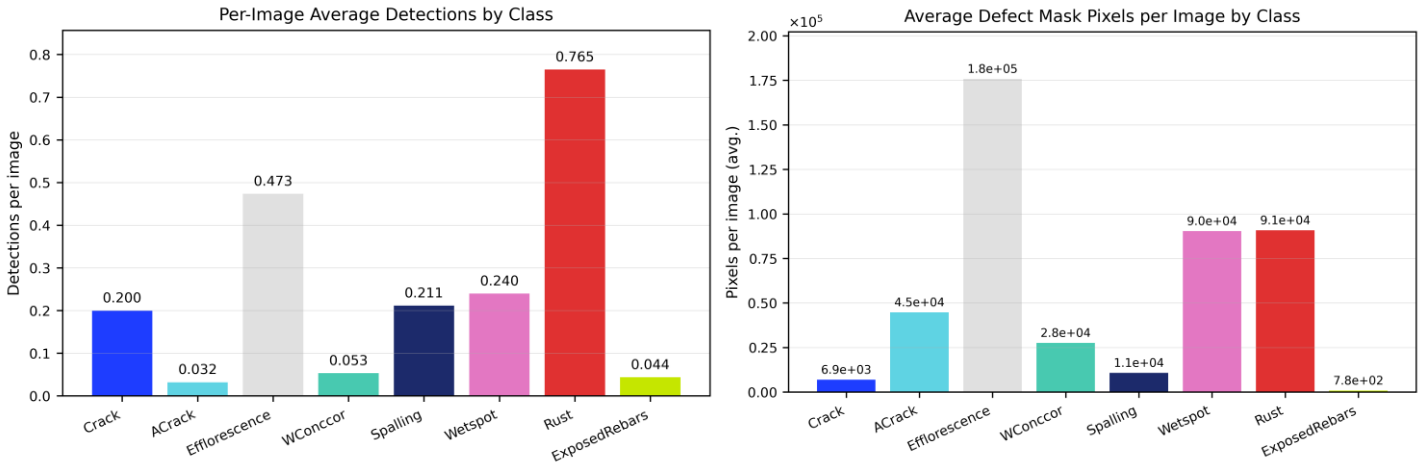
pixel units. These files serve as a foundation for subsequent integration into GIS platforms and automated condition assessment dashboards.

In addition to these structured JSON files, per-image and per-video defect counts were stored in sidecar text files, enabling quick summaries of inspection results. Aggregating these files across all bridges in the Kansas study area revealed clear trends in defect distribution (Fig. 8a). Rust, which includes both steel and concrete rust, was by far the most frequently detected class, accounting for 861 detections, followed by WConccor (533 detections) and Wetspot (270 detections). Crack and Spalling were detected at moderate levels (225 and 238 detections, respectively), while ACrack, Efflorescence, and Exposed Rebars were comparatively rare. When normalized by the number of images, Rust again showed the highest prevalence, nearly double that of WConccor and far higher than Crack or Spalling. These patterns suggest that corrosion and water-related deterioration are the dominant forms of visible concrete distress in the inspected bridges. Beyond frequency, polygon mask areas provided additional insights into defect severity and surface extent (Fig. 8b). Efflorescence and ACrack, although not always the most frequent, contributed disproportionately large polygon areas (approximately 2.0 × 10^8 and 5.0 × 10^7 pixels, respectively), reflecting their tendency to cover broad surface regions. By contrast, cracks and Exposed Rebars, while visually critical, occupied relatively small polygonal extents due to their thin, elongated geometries. On a per-image basis, Efflorescence again dominated with an average of 1.8 × 10^5 pixels per image, compared to about 9.0 × 10^4 for Rust and Wetspot. These results emphasize that different defect types should be interpreted not only in terms of frequency but also in terms of area, as widespread surface degradation may impose different maintenance demands than localized but structurally significant defects like cracks. Finally, cropped regions of each detected defect were automatically saved during inference, enabling future integration with VLMs for automatic captioning of deterioration, defect description, and potentially cross-training auxiliary models for automated inspection reports.



(a)

ARTICLE IN PRESS



(b)

Fig. 8

ARTICLE IN PRESS

YOLOv11 Applied to Drone Video Sequences

The study counties in Kansas collected UAV-based inspection videos for ten bridges, enabling an evaluation of the YOLOv11 model under dynamic inspection conditions. Representative proof-of-concept outputs are shown in Fig. 9, where bounding polygons and class labels are overlaid on bridge components. These results highlight the ability of the model to segment multiple concrete defect types directly from drone-collected videos, including instances where several types of deterioration coexisted in the same structural region. For example, Fig. 9a shows detections from a Kansas bridge (Structure No. 000000000870090), where cracks, rust patches, spalling, efflorescence, and ACrack are all identified along the barrier and substructure elements. Fig. 9b and Fig. 9c present two bridges (Structure Nos. 000000000180190 and 000000000180820), where combinations of spalling, efflorescence, rust, WConccor and exposed rebars are delineated on deck undersides and girders. These examples demonstrate the model's potential to capture deterioration signatures across multiple bridge components, even when defects overlap spatially.

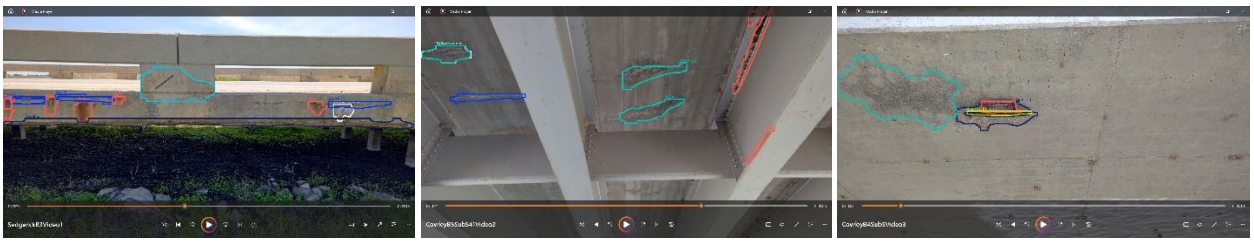
To evaluate detection performance quantitatively, per-video defect counts were exported to sidecar text files with class labels and tallies, and subsequently normalized by video length (seconds) to allow comparison across bridges of varying inspection durations. While still images yielded precise localizations, the transition to video analysis introduced additional complexity. Frame-by-frame inference with YOLO initially inflated defect counts because the same damage was repeatedly detected in consecutive frames. To resolve this, we incorporated a hybrid temporal post-processing strategy. First, the ByteTrack algorithm stabilized defect identities over time, ensuring consistent tracking across successive frames. Second, an Intersection-over-Union (IoU) based

ARTICLE IN PRESS

deduplication filter suppressed overlapping detections to avoid double counting. Together, these approaches reduced inflated counts and brought automated tallies closer to manual ground-truth defect counts.

The performance of these correction methods is shown in Fig. 10, which compares raw YOLO outputs, post-processed counts (ByteTrack + IoU), and ground truth across selected seven bridges in four Kansas counties. The results confirm that uncorrected YOLO inference produced artificially high defect rates, in some cases overestimating counts by nearly a factor of two. After applying the hybrid strategy, defect rates normalized and tracked much more closely with ground truth. For example, in Kansas bridge SN 000870785006028, raw YOLO detected close to 0.7 defects per second of video, while the corrected output (approximately 0.53) was nearly identical to the manually tallied ground truth (approximately 0.52). Similarly, in another bridge with SN 000000000180190, post-processed outputs (0.09) matched the ground truth (0.06) with a minimal deviation.

Some discrepancies remained in certain cases, largely driven by differences in annotation protocols. Specifically, manual ground truth counts of rust were restricted to reinforcement bars or major steel members, as these are critical indicators in National Bridge Inventory (NBI) ratings. In contrast, the automated counts included all rust patches, both in concrete surfaces and on steel, because model predictions did not differentiate between material types. This discrepancy explains why post-processed YOLO counts slightly exceeded ground truth in structures with several rust defects both in concrete and steel members. Future work will refine the human manual annotation protocol by explicitly distinguishing between corrosion on structural steel members and cosmetic rust staining on concrete surfaces. This can be achieved by introducing additional class labels or hierarchical attributes (e.g., material type: steel vs. concrete), enabling the model to differentiate between structurally significant corrosion and surface-level discoloration. Such refinements would improve alignment between automated detections and inspection criteria used in NBI-based assessments. In addition, incorporating multi-class or multi-attribute learning frameworks could allow the model to jointly predict defect type and associated material context, further improving interpretability and engineering relevance. Overall, these results demonstrate that the proposed temporal post-processing framework substantially improves the accuracy and interpretability of AI-driven defect counts in UAV-based video inspections. The ability to reduce inflated detections while retaining sensitivity across multiple defect types provides a promising pathway for operational deployment in county-level bridge asset management.



(a) (b) (c)

Fig. 9

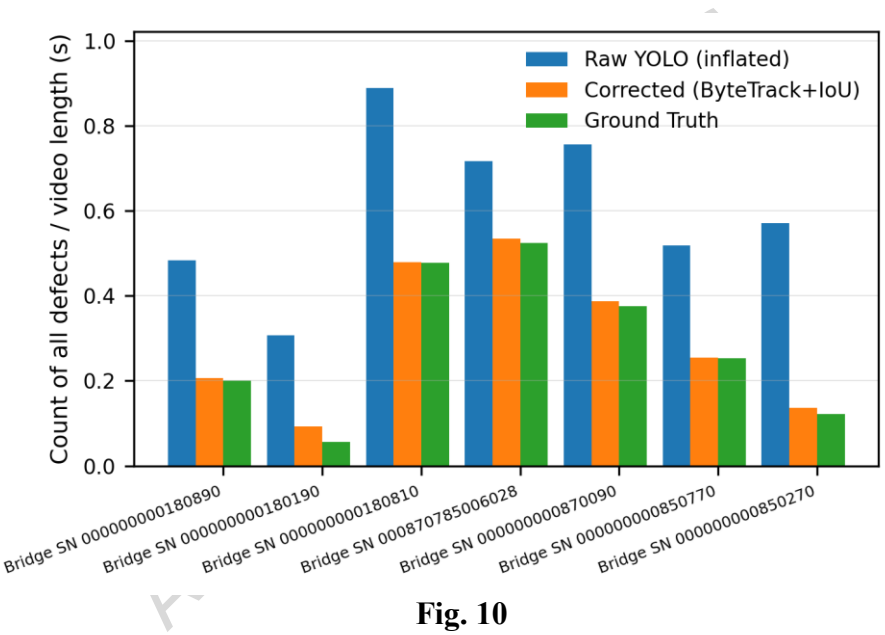


Fig. 10

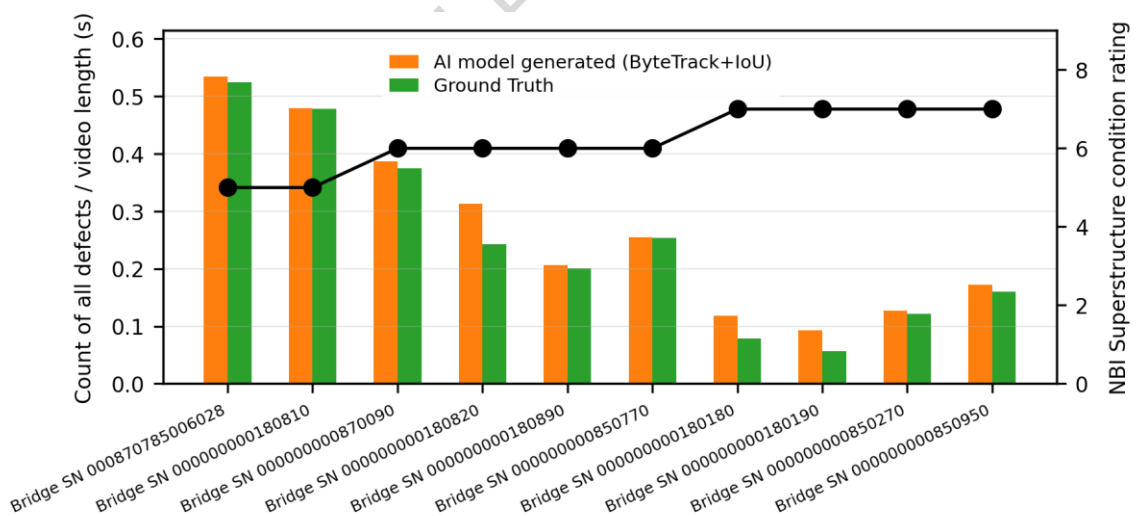
Comparison with NBI Condition Ratings

To evaluate the consistency of AI-generated outputs with existing inspection practice, we compared model-derived defect metrics with the National Bridge Inventory (NBI) condition ratings for ten bridges in the Kansas study counties where UAV-based video inspections were collected. For each bridge, sidecar text files recorded per-video defect counts by class, which were normalized by video length (s) to allow comparability across sites. These AI-based metrics were then validated against manual ground-truth metrics and compared to standard NBI condition ratings for the same bridges.

Fig. 11a presents a dual-axis comparison between AI-generated defect rates (orange bars), manual ground-truth counts (green bars), and NBI superstructure condition ratings (black line). Across the

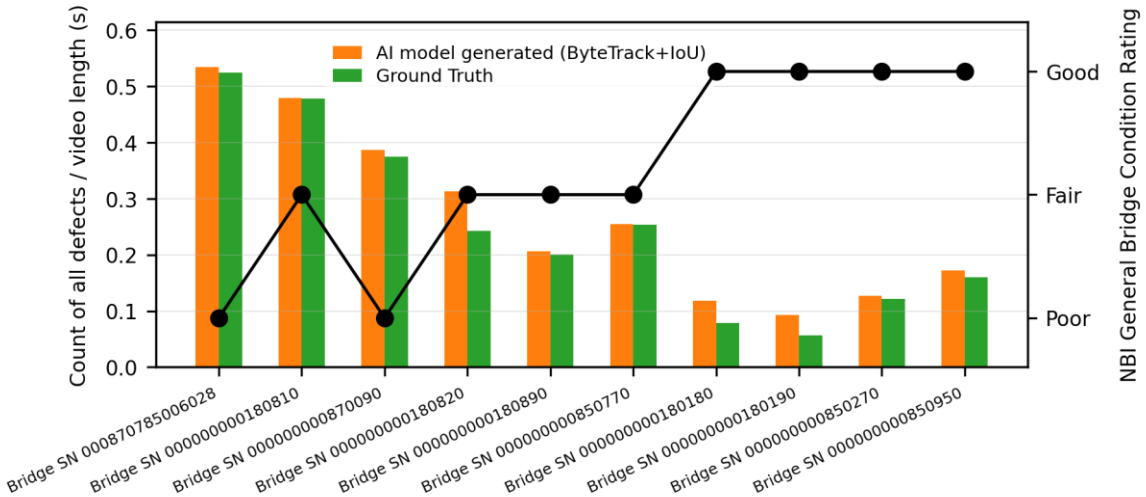
ten bridges, the defect rates produced by the hybrid post-processing strategy (ByteTrack + IoU) consistently aligned with the ground-truth benchmarks, with only minor deviations observed. When compared to NBI ratings, the results revealed an encouraging trend: bridges with higher defect rates generally corresponded to lower NBI ratings. As shown in Fig. 11a, two bridges with normalized defect rates above 0.45 detections/s had NBI superstructure scores clustered around 5 ("Poor to Fair"), while bridges with defect rates below 0.2 detections/s were associated with NBI ratings of 7 or greater ("Good"). This relationship suggests that AI-derived detection methods have great potential to capture deterioration severity in a way that is directionally consistent with inspector judgment.

Fig. 11b further reinforces this alignment by categorizing NBI ratings into Poor, Fair, and Good. Here, the bridges with "Poor" ratings exhibit the highest normalized defect rates, while those with "Good" ratings show minimal AI and ground-truth detections. Importantly, the spread between AI-generated and ground-truth counts remained relatively narrow across all categories, supporting the robustness of the automated pipeline. However, the analysis also highlights a critical nuance; NBI scores reflect holistic structural condition, which for instance, incorporates inspector expertise, load-carrying capacity, section loss, corrosion and context-specific knowledge beyond surface-visible deterioration[44]. For example, a bridge with localized spalling but sound substructure capacity might receive a higher NBI rating than raw defect counts alone would suggest.



(a)

ARTICLE IN PRESS



(b)

Fig. 11

In summary, these results indicate that AI-derived video analytics can serve as a complementary tool for condition assessment, providing objective, repeatable metrics that align with established inspection practice. While the current study was limited to a pilot set of bridges in the central Kansas counties, expanding the dataset to a broader population could strengthen statistical correlations and pave the way for integrating AI-based metrics as supplementary indicators in future NBI rating protocols. However, it must be emphasized that NBI ratings capture holistic structural conditions, including inspector judgment and factors beyond surface-visible deterioration. Thus, while AI outputs align with observed deterioration patterns, they should be considered as supplemental tools rather than replacements for human judgment.

3D Reconstruction and Geospatial Integration

A key contribution of this study was the projection of 2D defect detections from YOLOv11-seg into georeferenced 3D reconstructions of bridge decks. This integration demonstrated that AI-generated inferences can be extended beyond static image analysis into spatial 3D models, with the potential to enable defect visualization, measurement, and long-term deterioration tracking. Unlike conventional model predictions, which provide only qualitative evidence of deterioration, projecting polygons into a 3D environment allows inspectors to interact with defects in measurable units, including length, surface area, and volumetric estimates. These capabilities offer the potential for county engineers to establish a quantitative baseline for deterioration and to track changes over multiple inspection cycles.

ARTICLE IN PRESS

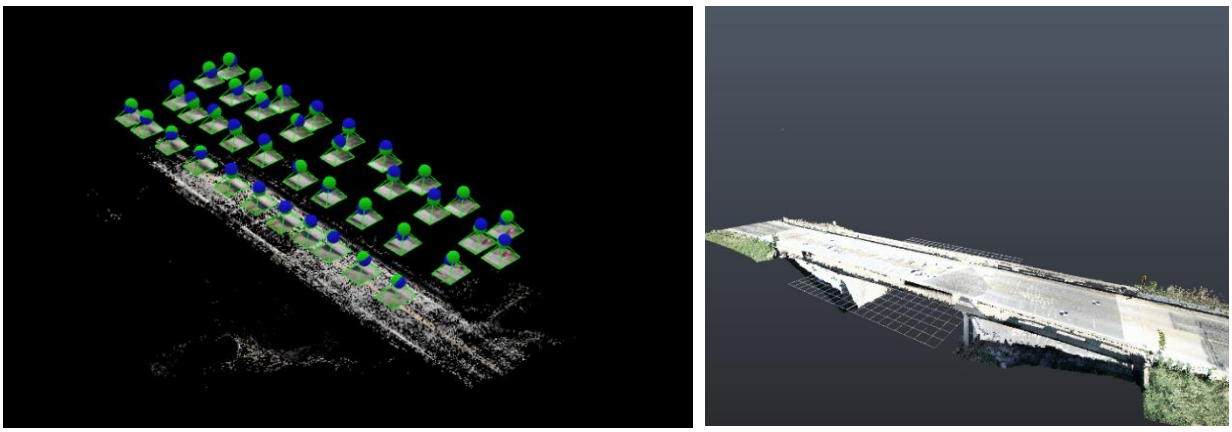
Using Pix4Dmapper[57], drone imagery of selected bridges was processed into orthomosaics, dense point clouds, and textured meshes. These reconstructions were spatially aligned with GCPs to maintain positional accuracy and minimize errors. Where metadata were lost during inference, a synchronization algorithm restored geotags by matching image timestamps with UAV flight logs, ensuring proper placement of defect polygons. Fig. 12a and 12b illustrate the integration of YOLO-inferred bridge deck images (bridge SN: 000000000180190), showing spalling defects with restored geotags alongside the corresponding 3D reconstructions generated in Pix4Dmapper. By projecting YOLO-generated polygons onto these reconstructions, defect areas initially measured in pixels were translated into real-world units (cm² or m²). For example, as shown in Fig. 12c, a spalling/pothole region was measured to have a maximum length of 1.24 m, while another pothole in the same deck exhibited a total length of 5.29 m and surface area of 1.24 m².

The results highlight the advantages of this integration: defects such as spalling, cracking, and surface delamination can be visualized in three dimensions, allowing for clearer communication between engineers, decision-makers, and the public. LiDAR data, when available, further enhanced reconstruction fidelity, particularly in shadowed or occluded regions such as deck undersides. These combined methods demonstrated that 3D-based measurements can approximate deterioration extents at the element level. The reconstructed models, generated using photogrammetry with GCPs and GPS geotagged imagery, provide geometrically consistent representations of bridge structures at the bridge scale. However, the derived defect measurements should be considered approximate, as they were not formally validated against ground-truth field measurements or inspection records in this study. While the reconstructed dimensions were consistent with expected deterioration patterns, future work will involve systematic comparison with field-based measurements to quantify accuracy, calibrate the methodology, and establish confidence levels for practical deployment.

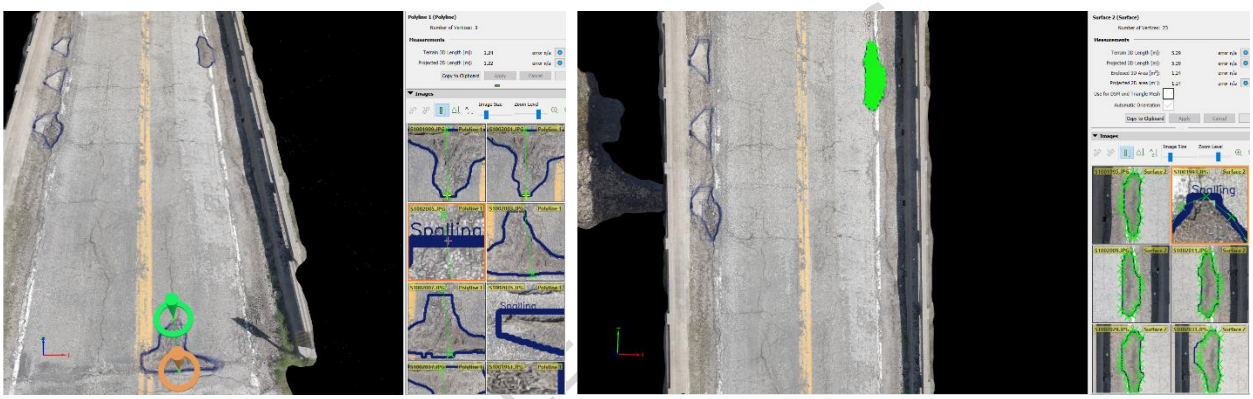
ARTICLE IN PRESS



(a)



(b)



(c)

Fig. 12

ARTICLE IN PRESS

Finally, the geospatial outputs were consolidated into a GIS platform, which served as a central repository for all inspection data. Model inferenced imagery, defect polygons, NBI ratings, and inspection metadata were layered together to provide an interactive dashboard for county engineers. This not only streamlined technical decision-making but also enabled transparent communication with the public through web-based story maps and interactive visualizations.

Prototype Web Service Deployment

The trained YOLOv11 segmentation model was exported to the ONNX format and deployed as a containerized FastAPI web service. The deployment was hosted using Elastic Container Service (ECS) with a load-balanced endpoint. The API exposed endpoints for both still images and video inputs, returning prediction outputs in near-real-time. An interactive Swagger/OpenAPI interface was integrated to facilitate browser-based interaction, enabling inspectors to upload imagery or videos and receive inference results, along with sidecar JSON and text files. Performance testing demonstrated that the service achieved image inference times of less than five seconds per

ARTICLE IN PRESS

uploaded image, including file upload, preprocessing, inference, visualization, and output generation. For video, it reached a throughput of approximately 15–20 fps. Fig. 13 illustrates the Swagger/OpenAPI interface and model predictions.

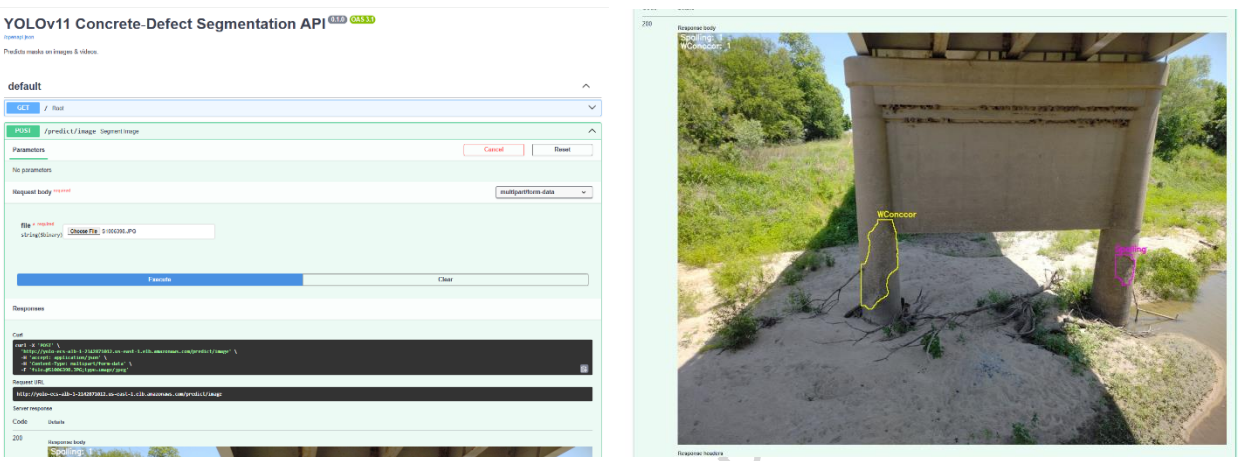


Fig. 13

By demonstrating a scalable deployment pathway, this prototype closed the loop from AI model development to operational use. The containerized service can be replicated across counties and agencies, providing a practical foundation for adoption in infrastructure monitoring workflows. To ensure reproducibility and long-term accessibility, the source code and deployment instructions have been archived on a public repository[70].

Broader Implications and Challenges

This work demonstrates the feasibility of using drones and AI to transform bridge inspection practice from a manual, resource-intensive process into an automated, scalable, and data-rich workflow. By aligning AI-generated metrics with established NBI ratings, the study provides evidence that AI can complement inspector judgment and enhance decision-making. The integration of 2D detections into 3D reconstructions of bridge decks and GIS dashboards illustrates how raw imagery can evolve into actionable information for engineers, policymakers, and the public. Furthermore, deploying the trained model as a containerized web service closes the gap between academic research and practical field applications, paving the way for county- and state-level adoption.

However, several challenges remain. Model performance is sensitive to annotation quality, highlighting the need for consistent and rigorous labeling protocols. The dataset is geographically limited to concrete bridges in four Kansas counties, which restricts generalizability. While the

ARTICLE IN PRESS

model was also applied to bridges incorporating steel girders and demonstrated the ability to detect corrosion-related defects (Fig. 6a, Fig. 6b, and Fig. 7a), and showed robustness under challenging visual conditions such as shadowed and low-light environments (Fig. 7d), a systematic evaluation across a broader range of bridge materials (e.g., timber or steel truss systems) and environmental conditions (e.g., snow, heavy rain) was not conducted.

Addressing these limitations through expanded datasets covering diverse structural types and environmental conditions, improved manual human annotation standards, and multimodal sensing will be essential to rigorously assess model robustness and generalizability. Future work will also refine annotation protocols to distinguish between structurally significant corrosion and surface-level staining, improving alignment with engineering assessment criteria.

Conclusion

This study introduced and tested a drone- and AI-enabled bridge inspection pipeline under the USDOT SMART project in Kansas. Results demonstrated that the AI segmentation model reliably detected eight types of concrete defects across images and video, with post-processing methods correcting inflated video counts and aligning them closely with manual tallies. Comparisons with NBI condition ratings showed that AI-derived metrics were directionally consistent with inspector assessments, supporting their use as supplementary indicators of condition. The projection of AI-based instance segmentation detections into 3D reconstructions of bridge decks further demonstrated the feasibility of measurement, enhanced visualization of deterioration and long-term deterioration tracking, while GIS integration provided a pathway for asset management. Finally, deployment of the model as a web service confirmed the technical feasibility of near-real-time inference in operational contexts. Together, these outcomes highlight the potential of drone-AI integration to improve safety, efficiency, and consistency in bridge inspections.

Despite these advances, challenges remain. Model performance was sensitive to annotation quality, underscoring the need for rigorous labeling protocols. The dataset was geographically limited to four Kansas counties, focusing solely on concrete bridges, which restrict generalizability across different materials and environments. Addressing these issues through larger datasets, multimodal sensing, and improved manual annotation standards will be crucial for scaling adoption. Overall, this work establishes drones and AI not as replacements for traditional inspection but as complementary tools that can enhance safety, efficiency, and consistency in infrastructure monitoring. By bridging the gap between research prototypes and deployable field systems, this study supports the progression toward future technology-assisted inspection programs that can improve asset management.

Data availability

A representative subset of the dataset used in this study, including manually annotated images, selected videos, and corresponding YOLO-format annotations, is publicly available on Zenodo at:

ARTICLE IN PRESS

- Tsegaye, N. T. Curated Dataset and Inference Results for AI-Based Bridge Defect Detection Using Images and Videos [Data set]. Zenodo (2026). https://doi.org/10.5281/zenodo.19268973

The full dataset is expected to be made available in accordance with the USDOT SMART program's Data Management Plan. The corresponding author will respond to reasonable requests for additional data not publicly available.

Code availability

The training and deployment workflows are made publicly available via Zenodo[50,70]:

- Tsegaye, N. T. natnaeltaye/concrete-defect-yolov11: First public release (training workflow). Zenodo (2025). https://doi.org/10.5281/zenodo.17488232
- Tsegaye, N. T. natnaeltaye/yolo11-fastapi-app: First public release (FastAPI + ONNX deployment for YOLOv11 concrete defect detection). Zenodo (2025). https://doi.org/10.5281/zenodo.17488363

Funding

This work was supported by the United States Department of Transportation (USDOT) Strengthening Mobility and Revolutionizing Transportation (SMART) Grant Program.

Acknowledgements

The authors gratefully acknowledge the assistance provided by participating Kansas county engineers and NCRPC staff. The opinions and conclusions expressed in this paper are solely those of the authors and do not necessarily reflect the views of the sponsor.

Declaration of Interest Statement

The authors declare no conflict of interest.

Author contributions

Natnael Taye Tsegaye: Methodology; investigation; software; validation; data curation; formal analysis; visualization; writing -original draft.

Yunfeng Zhang: Conceptualization; funding acquisition; methodology; investigation; validation; supervision; writing - review and editing.

Qingbin Cui: Funding acquisition; project administration; supervision; data curation; resource; writing - review and editing.

ARTICLE IN PRESS

References

1. Federal Highway Administration. *Bridge Condition by Highway System 2024*. U.S. Department of Transportation (2024). https://www.fhwa.dot.gov/bridge/nbi/no10/condition24.cfm.

2. American Society of Civil Engineers. *2025 Report Card for America's Infrastructure* (American Society of Civil Engineers, 2025). Available at: https://infrastructurereportcard.org.

3. Gralund, M. S. & Puckett, J. A. System for bridge management in a rural environment. *Journal of Computing in Civil Engineering* **10**, 97–105 (1996).

4. Federal Highway Administration. *National Bridge Inspection Standards.* U.S. Department of Transportation. https://www.fhwa.dot.gov/bridge/nbis.cfm (2022).

5. Tzortzinis, G., Filippatos, A., Wittig, J., Gude, M., Provost, A., Ai, C. & Gerasimidis, S. Structural integrity of aging steel bridges by 3D laser scanning and convolutional neural networks. *Communications Engineering* **3**, 106 (2024).

6. Panigati, T., Zini, M., Striccoli, D., Giordano, P. F., Tonelli, D., Limongelli, M. P. & Zonta, D. Drone-based bridge inspections: Current practices and future directions. *Autom. Constr.* **173**, 106101 (2025).

7. Phares, B. M., Washer, G. A., Rolander, D. D., Graybeal, B. A. & Moore, M. Routine highway bridge inspection condition documentation accuracy and reliability. *Journal of Bridge Engineering* **9**, 403–413 (2004).

8. Abdallah, A. M., Atadero, R. A. & Ozbek, M. E. A state-of-the-art review of bridge inspection planning: Current situation and future Needs. *Journal of Bridge Engineering* **27**, 03121001 (2022).

9. Gillins, M. N., Gillins, D. T. & Parrish, C. Cost-effective bridge safety inspections using unmanned aircraft systems (UAS). *In Geotechnical and Structural Engineering Congress 2016* 386–395 (American Society of Civil Engineers, 2016). https://doi.org/10.1061/9780784479742.165.

10. Wells, J. L. & Lovelace, B. *Improving the Quality of Bridge Inspections Using Unmanned Aircraft Systems (UAS)*. Research Report MN/RC 2018-26 (Minnesota Department of Transportation, 2018).

11. Seo, J., Duque, L. & Wacker, J. Drone-enabled bridge inspection methodology and application. *Autom. Constr.* **94**, 112–126 (2018).

ARTICLE IN PRESS

ARTICLE IN PRESS

12. Seo, J., Wacker, J. P. & Duque, L. *Evaluating the Use of Drones for Timber Bridge Inspection.* General Technical Report FPL–GTR–258 (U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, 2018).

13. Hubbard, B. & Hubbard, S. Unmanned aircraft systems (UAS) for bridge inspection safety. *Drones* **4**, 40 (2020).

14. Perry, B. J., Guo, Y., Atadero, R. & van de Lindt, J. W. Streamlined bridge inspection system utilizing unmanned aerial vehicles (UAVs) and machine learning. *Measurement* **164**, 108048 (2020).

15. Xu, Y. & Turkan, Y. BrIM and UAS for bridge inspections and management. *Engineering, Construction and Architectural Management* **27**, 785–807 (2020).

16. Potenza, F., Rinaldi, C., Ottaviano, E. & Gattulli, V. A robotics and computer-aided procedure for defect evaluation in bridge inspection. *J. Civ. Struct. Health Monit.* **10**, 471–484 (2020).

17. Flotzinger, J., Rosch, P. J. & Braml, T. dacl10k: Benchmark for semantic bridge damage segmentation. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 8611–8620 (IEEE, 2024).

18. Hiasa, S., Karaaslan, E., Shattenkirk, W., Mildner, C. & Catbas, F. N. Bridge inspection and condition assessment using image-based technologies with UAVs. In *Structures Congress 2018*, 217–228 (ASCE, 2018).

19. Humpe, A. Bridge Inspection with an off-the-shelf 360° Camera Drone. *Drones 2020, Vol. 4, Page 67* **4**, 67 (2020).

20. Hoskere, V., Park, J.-W., Yoon, H. & Spencer, B. F. Jr. Vision-based modal survey of civil infrastructure using unmanned aerial vehicles. *J. Struct. Eng.* **145**, 04019062 (2019).

21. Garg, P., Moreu, F., Ozdagli, A., Taha, M. R. & Mascareñas, D. Noncontact dynamic displacement measurement of structures using a moving laser Doppler vibrometer. *J. Bridge Eng.* **24**, 04019089 (2019).

22. Garg, P., Nasimi, R., Ozdagli, A., Zhang, S., Mascarenas, D. D. L., Taha, M. R. & Moreu, F. Measuring transverse displacements using unmanned aerial systems laser Doppler vibrometer (UAS-LDV): Development and field validation. *Sensors* **20**, 6051 (2020).

23. Catt, S., Fick, B., Hoskins, M., Praski, J. & Baqersad, J. Development of a semi-autonomous drone for structural health monitoring of structures using digital image correlation (DIC). *Conference Proceedings of the Society for Experimental Mechanics Series* **6**, 49–57 (2019).

ARTICLE IN PRESS

ARTICLE IN PRESS

24. Yoon, H., Shin, J. & Spencer, B. F. Jr. Structural displacement measurement using an unmanned aerial system. *Computer-Aided Civil and Infrastructure Engineering* **33**, 183–192 (2018).

25. Perry, B. J. & Guo, Y. A portable three-component displacement measurement technique using an unmanned aerial vehicle (UAV) and computer vision: A proof of concept. *Measurement* **176**, 109222 (2021).

26. Buggiani, D., Malavisi, M. & Baldacci, A. Digital technologies for bridge damage detection: Drones and AI for remote inspections. In *Bridge Maintenance, Safety, Management, Digitalization and Sustainability* (eds Jensen, Frangopol & Schmidt) 3161–3168 (CRC Press, 2024).

27. Wang, X., Demartino, C., Narazaki, Y., Monti, G. & Spencer, B. F. Rapid seismic risk assessment of bridges using UAV aerial photogrammetry. *Eng. Struct.* **279**, 115589 (2023).

28. Ayele, Y. Z., Aliyari, M., Griffiths, D. & Lopez Droguett, E. Automatic crack segmentation for UAV-assisted bridge inspection. *Energies* 13, 6250 (2020).

29. Mohammed, M. *Bridge Inspection Using Drone & AI* (Master's thesis, Purdue University, 2025).

30. Mandirola, M., Casarotti, C., Peloso, S., Lanese, I., Brunesi, E. & Senaldi, I. Use of UAS for damage inspection and assessment of bridge infrastructures. *Int. J. Disaster Risk Reduct.* **72**, 102824 (2022).

31. Cabral, R., Santos, R., Correia, J. & Ribeiro, D. Optimal reconstruction of railway bridges using a machine learning framework based on UAV photogrammetry and LiDAR. *Struct. Infrastruct. Eng.* (2025). https://doi.org/10.1080/15732479.2025.2531562.

32. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2015).

33. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2980–2988 (2017).

34. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 779–788 (2016).

35. Jocher, G. et al. Ultralytics YOLO. GitHub repository. https://github.com/ultralytics/ultralytics (accessed 11 November 2025).

ARTICLE IN PRESS

36. Cabral, R., Santos, R., Correia, J. A. F. O. & Ribeiro, D. A Hybrid YOLO and Segment Anything Model pipeline for multi-damage segmentation in UAV inspection imagery. *Sensors 2025,* **25**, 6568 (2025).

37. Neubauer, K., Bullard, E. & Blunt, R. *Collection of Data with Unmanned Aerial Systems (UAS) for Bridge Inspection and Construction Inspection.* FHWA-HRT-21-086 (Federal Highway Administration, 2021). https://doi.org/10.21949/1521674.

38. Dorafshan, S., Thomas, R. J., Coopmans, C. & Maguire, M. A practitioner's guide to small unmanned aerial systems for bridge inspection. *Infrastructures* **4**, 72 (2019).

39. Li, X., Chen, Y., Chen, Z. & Huang, Z. Coverage path planning of bridge inspection with unmanned aerial vehicle. *Eng. Appl. Artif. Intell.* **156**, 111253 (2025).

40. Debus, P. & Rodehorst, V. Multi-scale flight path planning for UAS building inspection. In *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering (ICCCBE 2020)* (eds Toledo Santos, E. & Scheer, S.) *Lecture Notes in Civil Engineering* **98** (Springer, Cham, 2021). https://doi.org/10.1007/978-3-030-51295-8_74.

41. Bono, A., D'alfonso, L., Fedele, G., Filice, A. & Natalizio, E. Path planning and control of a UAV fleet in bridge management systems. *Remote Sens.* **14**, 1858 (2022).

42. Farah, M. & Alruwaili, A. Optimizing ground sampling distance for drone-based GIS mapping: A case study in Riyadh, Saudi Arabia. In *Proc. 9th Int. Conf. Robotics and Automation Engineering (ICRAE)* (2024).

43. Felipe-García, B., Hernández-López, D. & Lerma, J. L. Analysis of the ground sample distance on large photogrammetric surveys. *Applied Geomatics* **4**, 231–244 (2012).

44. Federal Highway Administration. *Specifications for the National Bridge Inventory* (FHWA-HIF-22-017, U.S. Department of Transportation, 2022).

45. Russell, B. C., Torralba, A., Murphy, K. P. & Freeman, W. T. LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* **77**, 157–173 (2008).

46. Ryan, T. W., Lloyd, C. E., Pichura, M. S., Tarasovich, D. M. & Fitzgerald, S. *Bridge Inspector's Reference Manual (BIRM)* (FHWA-NHI-23-024, Federal Highway Administration, 2023).

47. Washington State Department of Transportation. *Washington State Bridge Inspection Manual* (M 36-64.01, 2009).

48. Hüthwohl, P., Lu, R. & Brilakis, I. Multi-classifier for reinforced concrete bridge defects. *Autom. Constr.* **105**, 102824 (2019).

49. Ultralytics. *Ultralytics Documentation*. Available at: https://docs.ultralytics.com/ (accessed 15 October 2025).

ARTICLE IN PRESS

50. Tsegaye, N. T. natnaeltaye/Concrete-defect-yolov11: First public release (training workflow). Zenodo https://doi.org/10.5281/zenodo.17488232 (2025).

51. Shorten, C. & Khoshgoftaar, T. M. Survey on image data augmentation for deep learning. *J. Big Data* **6**, 1–48 (2019).

52. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W. & Wang, X. ByteTrack: Multi-object tracking by associating every detection box. In *Proc. Eur. Conf. Comput. Vis. (ECCV),* 1–21 (2022).

53. Javadnejad, F., Gillins, D. T., Higgins, C. C. & Gillins, M. N. BridgeDex: Proposed web GIS platform for managing and interrogating multi-year and multi-scale bridge inspection images. *J. Comput. Civ. Eng.* **32**, 04018042 (2018).

54. Ibrahim, N. & Yunus, M. Z. M. Bridge monitoring and management system using GIS. *IOP Conf. Ser.: Mater. Sci. Eng.* **527**, 012076 (2019).

55. Özyeşil, O., Voroninski, V., Basri, R. & Singer, A. A survey of structure from motion. *Acta Numerica* **26**, 305–364 (2017).

56. Schonberger, J. L. & Frahm, J. M. Structure-from-Motion Revisited. *In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* 4104–4113 (2016).

57. Pix4D SA. Pix4Dmapper [software]. Lausanne, Switzerland: Pix4D SA; 2020. Available at: https://www.pix4d.com.

58. Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J. & Reynolds, J. M. 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology* **179**, 300–314 (2012).

59. Snavely, N., Seitz, S. M. & Szeliski, R. Photo tourism: Exploring photo collections in 3D. *ACM Trans. Graph.* **25**, 835–846 (2006).

60. Chen, S.-E., Liu, W., Bian, H. & Smith, B. 3D LiDAR scans for bridge damage evaluation. In *Forensic Engineering 2012* (ASCE, 2013), 487–495.

61. Liu, W., Chen, S. & Hauser, E. LiDAR-based bridge structure defect detection. *Exp. Tech.* **35**, 27–34 (2011).

62. Turner, D., Lucieer, A. & Watson, C. An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SfM) point clouds. *Remote Sens.* **4**, 1392–1410 (2012).

63. Nex, F. & Remondino, F. UAV for 3D mapping applications: a review. *Applied Geomatics* **6**, 1–15 (2014).

ARTICLE IN PRESS

ARTICLE IN PRESS

64. Neubeck, A. & Van Gool, L. Efficient non-maximum suppression. In *Proc. 18th International Conference on Pattern Recognition (ICPR)*, 850–855 (IEEE, 2006).

65. Brkic, M., Filali Razzouki, A., Tevissen, Y., Guetari, K. & El-Yacoubi, M. A. Frame sampling strategies matter: a benchmark for small vision language models. *arXiv* 2509.14769 (2025).

66. Federal Highway Administration. *National Bridge Inventory*. Available at: https://www.fhwa.dot.gov/bridge/nbi.cfm (accessed August 2025).

67. ONNX Contributors. *ONNX: Open neural network exchange*. Available at: https://onnx.ai/ (accessed June 2025).

68. Ramírez, S. *FastAPI.* Available at: https://fastapi.tiangolo.com/ (accessed June 2026).

69. Merkel, D. Docker: lightweight Linux containers for consistent development and deployment. *Linux J.* **2014**, 2 (2014).

70. Tsegaye, N. T. *natnaeltaye/yolo11-fastapi-app: First public release (FastAPI + ONNX deployment for YOLOv11 concrete defect detection)*. Zenodo https://doi.org/10.5281/zenodo.17488363 (2025).

ARTICLE IN PRESS

Figure Captions

Fig. 1 | Pilot testing on selected Kansas bridges to evaluate drone durability, imaging quality, and operational range prior to large-scale deployment. a, b Drones conducting field test flights over county-owned bridges during pilot inspections.

Fig. 2 | Map showing bridge locations inspected using drones across four counties in Kansas. Basemap © OpenStreetMap contributors, licensed under the Open Database License (ODbL). See https://www.openstreetmap.org/copyright for licensing details.

Fig. 3 | Distribution of defect classes before and after class-balanced augmentation. a Original distribution of defect class instances and corresponding anchor box statistics, showing imbalance across categories. **b** Augmented distribution after class-balanced oversampling, reflecting increased representation of minority classes alongside updated anchor and spatial statistics.

Fig. 4 | Training performance curves of the YOLOv11 segmentation model. (a) F1-confidence curve showing the trade-off between precision and recall across confidence thresholds, (b) precision-recall (PR) curves illustrating class-wise detection performance and overall mAP@0.5, (c) precision-confidence curve indicating increasing prediction reliability at higher confidence levels, and (d) recall-confidence curve demonstrating decreasing detection completeness with increasing confidence. The results show stable learning behavior with consistent performance across defect classes.

Fig. 5 | Validation performance curves on the test dataset. (a) F1-confidence curve, (b) precision-recall (PR) curves, (c) precision-confidence curve, and (d) recall-confidence curve evaluated on the test dataset using a

ARTICLE IN PRESS

confidence threshold of 0.3 and IoU threshold of 0.5. The curves exhibit trends consistent with training, confirming stable generalization and reliable detection performance across unseen data.

Fig. 6 | Model predictions on bridge inspection images. a Original field-collected images of bridges with structure numbers 000000000870090, 000000000180890 and 000000000870670 respectively. **b** Corresponding model predictions generated by the trained YOLOv11 model, showing detected categories of concrete defects.

Fig. 7 | Model predictions on representative bridges from three Kansas counties. a Bridges with structure numbers 000000000180180, 000000000180810, and 000000000180820 (left to right) showing polygons outlining WConccor, Rust, Spalling, and Efflorescence. **b** Bridges 000000000870400, 000870785006028, and 000870785006028 (left to right) with detections of Wetspot and Exposed Rebars on deck undersides and girder webs under varying conditions. **c** Bridges 000850771004546, 000000000850950, and 000000000850950 (left to right) illustrating ACrack and Crack localized on superstructure components. **d** Challenging scenarios: two left panels (bridges 000000000180890 and 000870785006028) show detections of very small spalls and cracks on underside surfaces that are difficult to discern with the naked eye; two right panels (bridges 000000000850770 and 000000000850950) show spalling, rust, and exposed rebars in shadowed and low-light conditions. In all panels, defect labels and per-image class counts are displayed directly on the model predictions, illustrating the range of field conditions captured during drone inspections.

Fig. 8 | Defect frequency and polygonal area distributions across bridges in the Kansas study area. a Distribution of detected defect classes by total frequency (left) and distribution of total polygon mask areas (right). **b** Normalized prevalence per image (left) and average polygon mask area per image (right).

Fig. 9 | Representative video inference results from UAV inspections in Kansas counties. a Bridge 000000000870090 showing multiple defect types, including cracks, spalling, rust, efflorescence, and ACrack. **b** Bridge 000000000180190 illustrating rust on girders and additional deterioration features. **c** Bridge 000000000180820 showing co-occurring rust, exposed rebars, spalling, and WConccor defects on underside girders.

Fig. 10 | Comparison of raw YOLO, post-processed (ByteTrack + IoU), and ground truth defect rates for selected bridges. Each bar represents the count of all detected defects normalized by video length (s). Raw YOLO outputs consistently overestimated defect counts, while the hybrid strategy corrected inflated detections and brought results in closer alignment with manual ground truth tallies.

Fig. 11 | Comparison of AI-derived defect counts with NBI condition ratings. a Normalized defect counts per second of video for ten bridges, comparing AI model outputs corrected using ByteTrack + IoU (orange bars) against ground truth (green bars), with NBI superstructure ratings shown as a secondary axis (black line). **b** Normalized defect counts by bridge grouped by categorical NBI ratings (Poor, Fair, Good). Results demonstrate that bridges with higher defect rates generally correspond to lower NBI ratings, highlighting consistency between AI metrics and inspector-based assessments.

Fig. 12 | Projection of YOLOv11 detections onto georeferenced 3D bridge reconstructions. a UAV field-collected images showing YOLO-detected spalling and cracking polygons overlaid on bridge deck surfaces. **b** Reconstruction of the same bridge deck in Pix4Dmapper, producing dense point clouds and textured mesh models aligned to ground control points (GCPs). **c** Examples of quantitative defect measurements performed within the 3D environment, including defect length (1.24 m) and surface area (1.24 m^2), demonstrating the ability to translate pixel-level detections into real-world dimensions.

Fig. 13 | Screenshot of the prototype FastAPI deployment, showing Swagger/OpenAPI interface with image upload and model predictions.