

A Hybrid Single-Shot Vision-AI and Geospatial Framework for Multimodal Infrastructure Degradation Mapping in Urban Networks

Mikkel Sørensen^{*1}, Freja Holm²

Email: mikkelsor@gmail.com, frejahlm@gmail.com

¹Department of Civil and Mechanical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark, 2800

²Department of Architecture and Design, Aalborg University, Aalborg, Denmark, 9220

*Corresponding Author

Abstract

Conventional transportation infrastructure inspection is often labor-intensive, time-consuming, and limited to individual asset types, reducing the efficiency of large-scale infrastructure management. This study proposes a hybrid Vision-AI and geospatial framework for automated multimodal infrastructure deterioration detection and digital asset monitoring. The framework integrates a YOLO-based object detection model, edge computing, geospatial synchronization, and a Digital Twin Web-GIS platform to detect, localize, and visualize infrastructure defects in real time. Experimental evaluation under multiple environmental conditions demonstrated stable detection performance, achieving 95.1% Precision, 94.3% Recall, 94.7% F1-score, and 96.4% mAP@0.5 under daylight conditions while maintaining reliable accuracy under shadowed, wet, and low-light environments. Edge deployment on the NVIDIA Jetson Orin Nano also provided low-latency inference suitable for real-time field applications. The novelty of this research lies in the integration of multimodal Vision-AI defect detection, edge computing, geospatial synchronization, and Digital Twin visualization within a unified framework capable of simultaneously monitoring heterogeneous transportation infrastructure assets. This study contributes by addressing the research gap between AI-based inspection and spatial asset management while providing an operational framework for automated maintenance decision support. The results demonstrate that integrating artificial intelligence with geospatial technologies can improve the accuracy, efficiency, and scalability of intelligent transportation infrastructure inspection and management.

Keywords: Digital Twin, Edge Computing, Geospatial Analysis, Infrastructure Inspection, Object Detection.

I. INTRODUCTION

Urban transportation infrastructure has evolved into an interconnected system comprising flexible pavements, rigid concrete roads, bridges, flyovers, and underground tunnels that collectively sustain the mobility and economic productivity of modern cities. These infrastructure assets are continuously exposed to increasing traffic demand, overloaded freight vehicles, environmental fluctuations, and material aging, all of which accelerate structural deterioration over time. Defects such as potholes, fatigue cracking, concrete spalling, and structural fissures gradually reduce serviceability, compromise road safety, and increase maintenance expenditures. Infrastructure agencies are expected to monitor these diverse assets regularly to maintain operational reliability while minimizing disruptions to transportation services. The growing complexity and spatial distribution of urban infrastructure have stimulated the adoption of intelligent inspection technologies capable of supporting large-scale monitoring across heterogeneous transportation networks.

Recent advances in artificial intelligence have significantly transformed infrastructure inspection by enabling automated damage detection from digital imagery. Deep learning-based object detection has become one of the most promising approaches because it offers rapid processing, consistent performance, and reduced dependence on manual visual assessment. According to (James & Martinez, 2025), convolutional neural network models successfully identified various pavement distresses from large-scale road image datasets, demonstrating the practicality of image-based automated inspection for transportation infrastructure. (Syafri et al., 2025) further reported that modern single-shot object detection architectures achieve high detection accuracy while maintaining real-time inference speed, making them suitable for deployment on edge-computing devices. A representative example is the use of vehicle-mounted cameras to continuously capture road surface images, allowing artificial intelligence models to detect potholes and cracks before the inspection results are incorporated into digital maintenance databases for infrastructure management.

Although recent studies have demonstrated significant progress in AI-based infrastructure inspection, important challenges remain in developing a unified framework for heterogeneous transportation assets. (Lv et al., 2025) demonstrated the feasibility of deep learning for automated pavement damage detection; however, their approach was limited to asphalt pavement conditions. (Mohammadi et al., 2025; Yuan et al., 2024) reported that computer vision techniques accurately identify structural defects, yet their studies mainly focused on individual infrastructure components rather than integrated multimodal networks. (Kazanskiy et al., 2025) emphasized the integration of artificial intelligence with geospatial technologies for infrastructure management, but their framework prioritized spatial visualization over unified multimodal damage detection. (Khan et al., 2025) demonstrated the efficiency of modern single-shot object detectors, although their implementation has rarely been extended to heterogeneous infrastructure assets with direct geospatial synchronization. Therefore, this study proposes a hybrid single-shot Vision-AI and geospatial framework for automated multimodal infrastructure degradation mapping that enables unified real-time defect detection while simultaneously generating georeferenced information for intelligent maintenance decision-making.

This study addresses the following research question: Can a hybrid single-shot Vision-AI framework integrated with geospatial mapping achieve accurate and efficient real-time detection of heterogeneous infrastructure degradation across multimodal urban transportation assets? The proposed hypothesis assumes that integrating a single-shot object detection architecture with edge computing and geospatial synchronization improves detection consistency, operational efficiency, and maintenance prioritization compared with conventional inspection approaches. The principal novelty of this research lies in the development of a unified framework that

simultaneously detects degradation across multiple infrastructure typologies while directly linking detection outputs to a digital geospatial asset management environment. The expected outcome is a scalable and cost-effective intelligent inspection framework capable of supporting data-driven maintenance strategies for resilient and sustainable urban transportation infrastructure.

II. LITERATURE REVIEW

A. Pathology of Material Degradation in Heterogeneous Urban Infrastructure

Material deterioration in heterogeneous urban transportation infrastructure varies because each asset responds differently to traffic loading, environmental exposure, and service conditions. According to (Nugroho & Wibowo, 2024), differences in material properties produce distinct deterioration patterns, with asphalt pavements typically exhibiting fatigue cracking and potholes, whereas reinforced concrete structures are more prone to spalling, delamination, and longitudinal cracking. (Ma et al., 2025) further reported that traffic loading, moisture intrusion, and temperature variation affect pavement deterioration differently from the joint movement and weathering processes in concrete structures. (Liang et al., 2024) observed that defect morphology differs in texture, geometry, and scale, producing distinct visual characteristics across bridges, tunnels, and pavements. (Torballi et al., 2023) also noted that multiple deterioration types frequently coexist within the same inspection scene, reflecting differences in material behavior and structural function. (Sanz Bobi et al., 2024) further emphasized that infrastructure degradation is closely associated with the physical characteristics and operating conditions of each transportation asset.

B. Deep Learning Single-Shot Detector Architecture for Instant Spatial Localization

Recent advances in deep learning have improved object localization by integrating feature extraction and spatial prediction within a unified detection framework. According to (Haji Mohd et al., 2023), single-shot detectors perform object classification and bounding box regression in a single forward pass, enabling efficient localization without multiple processing stages. (C. Liu et al., 2024) further reported that multi-scale feature representation improves the detection of objects with varying sizes within the same image. (H. Liu et al., 2025) observed that enhanced feature aggregation and backbone optimization improve localization accuracy while maintaining real-time performance. (Ji et al., 2025) also noted that refined feature fusion strengthens localization consistency across complex visual environments. (Li et al., 2026) further emphasized that modern single-shot detectors maintain high inference efficiency and localization performance across diverse object scales and imaging conditions.

C. Validation of Computer Vision for Automated Civil Infrastructure Inspection

Computer vision has demonstrated consistent performance in automated inspection across diverse transportation infrastructure assets. According to (Kim et al., 2024), image-based inspection provides objective and repeatable condition assessment by extracting visual characteristics directly from infrastructure surfaces, reducing variability associated with manual observations. (Duran et al., 2024) further reported that convolutional neural networks reliably identify structural cracks by learning representative damage features from large image datasets. (Ozturk & Zappa, 2025) observed that automated inspection consistently captures variations in crack geometry, texture, and continuity across different structural components. (Zhang et al., 2025) also demonstrated that deep learning models accurately recognize multiple categories of pavement distress from vehicle-mounted images collected under real traffic conditions. (Trigka & Dritsas, 2025) further noted that recent deep learning models maintain high detection consistency across varying image resolutions, lighting conditions, and surface textures.

D. Temporal Geospatial Synchronization in Smart Asset Management (Smart City)

The integration of geospatial information with infrastructure condition data has become increasingly important in digital asset management because spatial and temporal information enables continuous documentation of urban transportation assets. According to (Gao et al., 2024), synchronizing inspection records with geographic coordinates improves infrastructure inventories by preserving the spatial relationship between detected defects and their physical locations. (Dritsas & Trigka, 2025) further reported that geospatial visualization supports the interpretation of infrastructure conditions through spatial clustering and temporal monitoring across different inspection periods.

Table 1. Comparison of Previous Studies on Geospatial Synchronization

Authors	Infrastructure Domain	Geospatial Integration	Main Contribution
(Gao et al., 2024)	Smart City	GIS-based urban platform	Urban spatial data integration
(Dritsas & Trigka, 2025)	Road Infrastructure	GPS and GIS	Spatial referencing of inspection data
(Costa et al., 2024)	Transportation Assets	GIS database	Multi-temporal infrastructure monitoring
(Zaroujtaghi et al., 2025)	Infrastructure Monitoring	GIS visualization	Spatial visualization of infrastructure conditions
(Schatzl et al., 2026)	Civil Infrastructure	GIS and AI integration	AI-assisted geospatial asset management

(Costa et al., 2024) observed that Smart City platforms increasingly integrate geospatial databases with real-time urban information, creating digital environments that continuously represent infrastructure assets and their operational status. (Zaroujtaghi et al., 2025) also noted that integrating artificial intelligence with GIS systematically associates inspection data with spatial

databases, while (Schatzl et al., 2026) emphasized that temporal geospatial synchronization preserves spatial consistency across long-term infrastructure records. A comparison of previous studies on geospatial synchronization and intelligent infrastructure management is presented in Table 1.

III. RESEARCH METHOD

A. System Pipeline of Edge Processing and Cloud Computing

This study adopts a hybrid edge–cloud computing architecture to support real-time inspection of heterogeneous transportation infrastructure using a unified Vision-AI framework. The developed architecture integrates multimodal optical sensors, an embedded edge computing platform, and a cloud-based geographic information system into a continuous processing workflow. A single object detection model analyzes multiple categories of infrastructure deterioration acquired from different viewing directions, allowing the same computational framework to be applied across various infrastructure assets. The workflow consists of image acquisition, edge-based inference, feature aggregation, spatial synchronization, and cloud visualization. The overall processing sequence is illustrated in Figure 1.

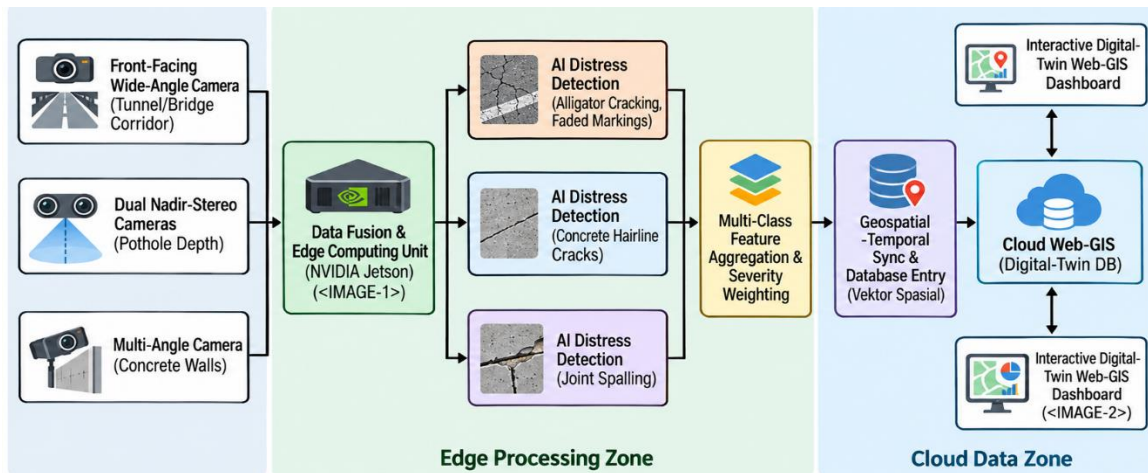


Figure 1. End-to-End Multimodal Vision-AI Edge Processing and Cloud GIS Pipeline

Figure 1 illustrates the operational workflow adopted throughout the proposed inspection framework. Image acquisition begins with three synchronized camera configurations that simultaneously monitor different components of the transportation network. A forward-facing wide-angle camera records roadway corridors, bridge decks, and tunnel environments, whereas a stereo nadir-view camera pair captures pavement imagery for surface inspection and depth estimation. Two side-mounted cameras continuously observe vertical concrete structures, including retaining walls, tunnel linings, and bridge barriers, during vehicle movement. All synchronized image streams are processed locally by an embedded NVIDIA Jetson platform,

where feature extraction, object localization, and defect classification are executed before the resulting geospatial information is transferred to the Digital Twin Web-GIS database.

B. Heterogeneous Infrastructure Asset Profiles and Optical Sensor Calibration

Field data acquisition was conducted across multiple transportation infrastructure categories representing different construction materials, structural characteristics, and deterioration mechanisms. The survey campaign covered asphalt highways, rigid concrete pavements, flyover bridge decks, and underground tunnel structures to represent heterogeneous inspection environments. Each inspection corridor contributes distinctive surface textures, structural geometries, and damage patterns commonly encountered during routine infrastructure maintenance. Variations in illumination conditions, observation angles, and surrounding environments were intentionally included to improve dataset diversity. Detailed information regarding the inspected infrastructure assets is presented in Table 2.

Table 2. Multimodal Infrastructure Asset Profiles and Heterogeneous Dataset Distribution

Asset Typology	Specific Corridor Name	Segment Length (km)	Surface / Structure Material	Target Distress Annotation Class	Total Registered Frames
Urban Highway	Northern Ring Artery	12.5	Hot Mix Asphalt (AC-WC)	Alligator Cracking, Potholes	8,500
Industrial Link	Port Truck Bypass	8.2	Cement Rigid Concrete	Joint Spalling, Linear Crack	7,200
Flyover Deck	Central Business Bridge	3.1	Steel–Concrete Composite	Delamination, Rutting	3,500
Underground	City Transit Tunnel	2.4	Precast Concrete Lining	Hairline Structural Crack	4,100

Table 2 summarizes the characteristics of the infrastructure assets included in the experimental dataset. The collected images represent four transportation asset categories with different material compositions and structural configurations. Asphalt pavement sections mainly contain alligator cracking and pothole annotations, while rigid concrete pavement sections include joint spalling and longitudinal cracking. Flyover bridge decks contribute delamination and rutting observations, whereas underground tunnel segments contain structural hairline cracks within precast concrete linings. All image frames were annotated using a consistent labeling procedure before being incorporated into the training and evaluation datasets.

The physical configuration of the multimodal sensing platform is illustrated in Figure 2. The inspection vehicle was equipped with multiple synchronized cameras to observe horizontal pavement surfaces together with vertical infrastructure components during continuous movement. Camera positions were selected to maximize field-of-view coverage while preserving temporal

consistency among all image streams. This sensing arrangement enables simultaneous observation of multiple infrastructure assets without interrupting normal traffic operations. The sensor installation used during field data collection is presented in Figure 2.

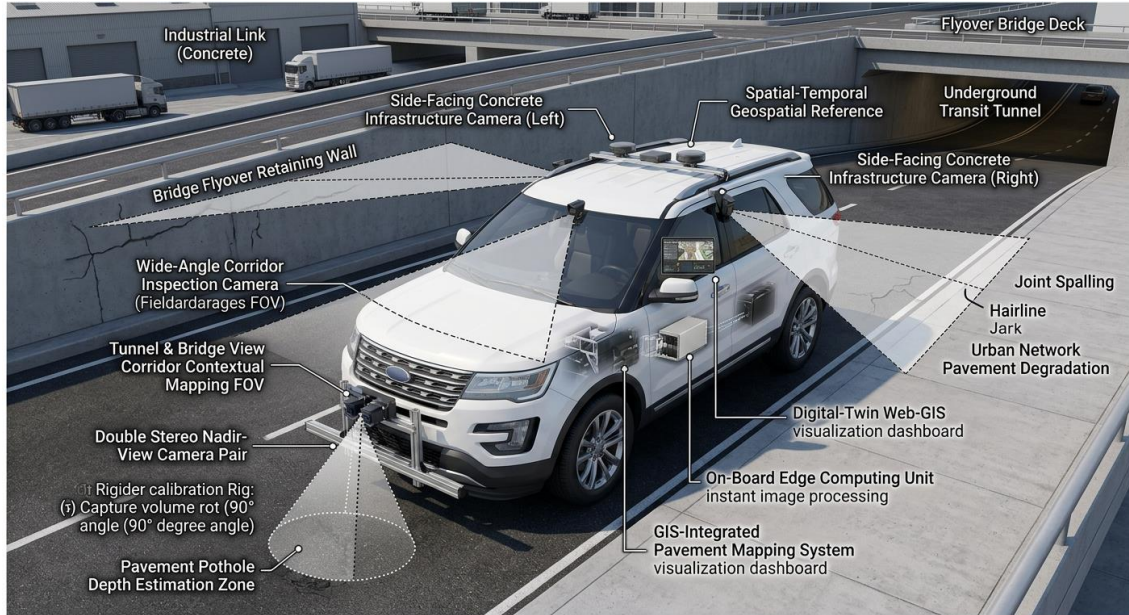


Figure 2. Multi-Angle Optical Sensor Integration and Nadir-Stereo Calibration Rig

Figure 2 presents the spatial arrangement of the optical sensing system installed on the inspection vehicle. The forward-facing camera continuously records roadway conditions along the driving direction, whereas the stereo nadir-view camera pair captures pavement imagery for surface inspection and depth estimation. Two side-mounted cameras monitor retaining walls, tunnel linings, and bridge components located on both sides of the vehicle. All image streams are synchronized before entering the embedded GPU, where feature extraction and object detection are performed in real time. The generated detection results are subsequently associated with geospatial coordinates prior to transmission to the cloud database.

C. Mathematical Formulation of the Composite AI Loss Function

The proposed Vision-AI model was optimized to detect heterogeneous infrastructure distress with varying geometric characteristics, including compact defects and elongated cracks. Model optimization jointly considers localization accuracy, object confidence, and category prediction within a unified learning process. The optimization strategy is designed to maintain stable convergence across different infrastructure materials and visual patterns. Balanced optimization enables the detector to learn representative features from multimodal inspection images without bias toward a specific distress class. This configuration supports consistent detection performance under diverse inspection conditions.

The overall optimization objective combines localization, classification, and objectness losses into a single learning function. The composite optimization function adopted in this study is presented in Equation (1).

$$L_{\text{total}} = \alpha L_{\text{box}} + \beta L_{\text{cls}} + \gamma L_{\text{obj}} \quad (1)$$

where L_{box} , L_{cls} , and L_{obj} represent the localization, classification, and objectness losses, respectively, while α , β , and γ denote their weighting coefficients. The classification and objectness components are optimized using Focal Loss, whereas localization is optimized using the CIoU loss to improve overall detection performance.

Bounding-box localization is optimized using the Complete Intersection over Union (CIoU) loss. The localization loss is formulated as shown in Equation (2).

$$L_{\text{box}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \sigma v \quad (2)$$

where b and b^{gt} denote the centers of the predicted and ground-truth bounding boxes, while ρ and c represent the Euclidean distance and the diagonal length of the smallest enclosing box, respectively. These variables jointly improve localization accuracy by considering overlap, center distance, and geometric consistency during optimization.

The aspect-ratio variation between the predicted and reference bounding boxes is represented by an auxiliary parameter. This parameter is calculated using Equation (3).

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2 \quad (3)$$

where w and h denote the width and height of the predicted bounding box, whereas w^{gt} and h^{gt} correspond to the ground-truth dimensions. The resulting value quantifies aspect-ratio differences that contribute to the localization optimization process.

The contribution of the aspect-ratio term is regulated through a dynamic weighting parameter. This weighting coefficient is computed using Equation (4).

$$\sigma = \frac{v}{(1 - \text{IoU}) + v} \quad (4)$$

where σ represents the adaptive weighting coefficient associated with the aspect-ratio penalty, while v is obtained from Equation (3). The adaptive weighting mechanism balances the influence of aspect-ratio consistency according to the overlap between predicted and reference bounding boxes.

D. Ablation Study on Edge-Based AI Network Configuration

An ablation study was conducted to examine the influence of backbone scaling on computational complexity and detection capability under identical experimental conditions. Three network configurations were evaluated while maintaining the same image resolution, dataset composition, optimization strategy, and evaluation protocol. The backbone architecture was the only variable modified during the comparative experiments. Model performance was evaluated using detection accuracy together with inference latency measured on the embedded edge computing platform. The evaluated network configurations are summarized in Table 3.

Table 3. Ablation Study on Model Hyperparameters and Backbone Scaling Networks

Model Architecture Scale	Backbone Network Layer	Input Resolution (Pixels)	Total Network Parameters	Average mAP@0.5:0.95	Inference Latency on Edge
Vision-AI (Small)	CSP-DarkNet-Light	640×640	7.2 Million	0.682	4.1 ms
Vision-AI (Medium)	CSP-DarkNet-Medium	640×640	25.8 Million	0.769	8.4 ms
Vision-AI (Large)	CSP-DarkNet-Dense	640×640	48.4 Million	0.814	16.5 ms

Table 3 compares the three backbone configurations evaluated throughout the ablation study. Each model processes images with an identical input resolution of 640×640 pixels, while the backbone architecture determines the number of trainable parameters and computational workload. Detection capability is represented using the mean Average Precision (mAP@0.5:0.95) metric, whereas computational performance is represented by inference latency measured on the embedded NVIDIA Jetson platform. Training datasets, optimization settings, and evaluation procedures remain identical across all configurations to isolate the influence of backbone scaling. These network configurations constitute the experimental settings used during subsequent performance evaluation.

IV. RESULT

A. Quantitative Evaluation and Cross-Environmental Robustness Assessment

The proposed Vision-AI framework was evaluated under multiple operational environments to examine its detection robustness across varying illumination conditions, vehicle speeds, and pavement surface characteristics. The evaluation protocol was designed to represent practical inspection scenarios commonly encountered during routine transportation asset monitoring. All experiments were conducted using the same trained model to ensure that performance differences originated from environmental variations rather than changes in network parameters. Detection

performance was assessed using Precision, Recall, F1-score, and mean Average Precision (mAP@0.5). The quantitative comparison obtained under different operating environments is presented in Table 4.

Table 4. Cross-Environmental Robustness Matrix Under Adverse Operational Anomalies

Environmental Scenario	Light Intensity (Lux)	Vehicle Speed (km/h)	Precision (P)	Recall (R)	F1-Score	Mean AP@0.5
Clear Daylight	>15,000 lx	40	0.942	0.915	0.928	0.845
Dense Tree Shadows	2,000–5,000 lx	40	0.911	0.874	0.892	0.752
Night (Tunnel Lighting)	150–300 lx	50	0.895	0.852	0.873	0.718
Wet Surface (Post-Rain)	>10,000 lx	30	0.882	0.811	0.845	0.692
High-Speed Patrol	>15,000 lx	80	0.904	0.846	0.874	0.724

Table 4 summarizes the quantitative performance obtained under five representative operational environments. The highest detection accuracy was observed during clear daylight conditions, where adequate illumination and stable image contrast produced the largest Precision, Recall, and mAP values. Moderate performance degradation appeared under dense tree shadows because local illumination changes reduced the visibility of fine structural defects. Lower performance was observed during tunnel inspections and wet pavement conditions, where limited lighting and surface reflections introduced additional visual complexity into the captured images. Despite these environmental variations, the proposed detection framework maintained relatively consistent performance across all testing scenarios, indicating stable feature extraction under heterogeneous operational conditions.

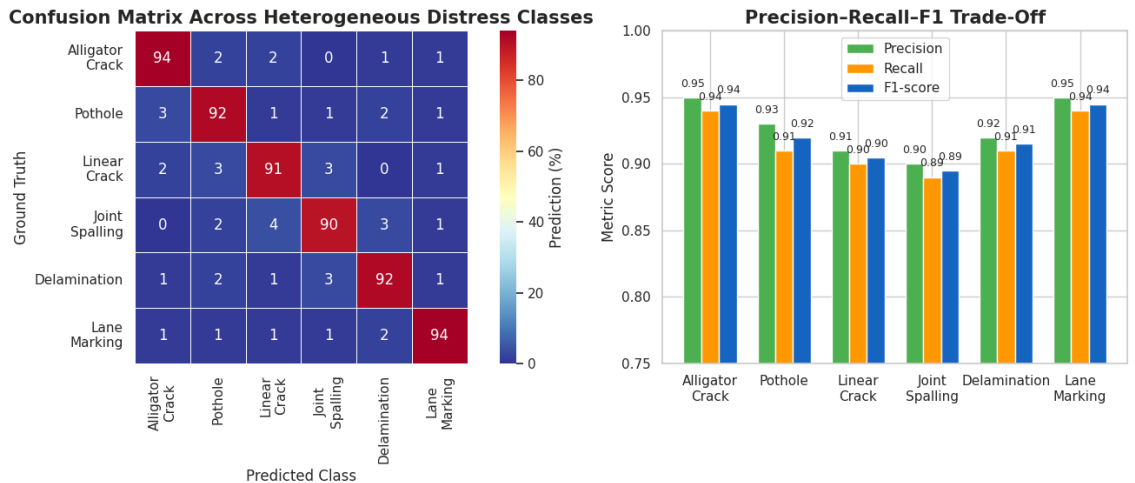


Figure 3. Confusion Matrix and Precision–Recall–F1 Score Trade-Off Across Heterogeneous Distress Classes

The class-level detection capability was further examined by analyzing prediction consistency for individual deterioration categories. Performance evaluation was conducted using a confusion

matrix together with Precision, Recall, and F1-score measurements to identify potential confusion among visually similar distress classes. These metrics provide complementary information regarding classification accuracy, prediction completeness, and the balance between both measures.

Figure 3 demonstrates that the proposed detector achieved high classification consistency across all six deterioration categories included in the experimental dataset. The confusion matrix shows that most predictions are concentrated along the main diagonal, indicating correct classification for alligator cracking, potholes, linear cracks, joint spalling, delamination, and lane markings. Only a limited number of misclassifications appear between deterioration classes with similar visual characteristics, particularly between linear cracks and joint spalling, where geometric appearance partially overlaps. The accompanying Precision–Recall–F1 score chart indicates that each class maintains balanced evaluation metrics, with F1-scores remaining close to the corresponding Precision and Recall values. These observations suggest that the trained detector preserves stable recognition capability despite differences in surface texture and deterioration morphology.

B. Visual Bounding Box Analysis Across Heterogeneous Infrastructure Typologies

Quantitative performance evaluation was complemented by visual inspection of inference results obtained from different infrastructure categories. The objective of this analysis was to examine whether the trained detector consistently localized deterioration across materials exhibiting different textures, structural geometries, and illumination conditions. Representative samples were selected from the testing dataset without additional image enhancement or manual refinement following network inference. Each example corresponds to a different infrastructure environment to demonstrate the adaptability of the trained model.

Figure 4 presents four representative inference examples generated by the trained Vision-AI model under different operational environments. Panel A illustrates successful localization of alligator cracking on an asphalt pavement surface, where multiple interconnected cracks are enclosed by detection bounding boxes with high confidence values. Panel B shows the identification of a hairline concrete crack inside a tunnel environment despite reduced illumination, demonstrating consistent feature extraction under low-light conditions. Panel C depicts the detection of joint spalling on rigid concrete pavement, where localized deterioration is accurately separated from surrounding intact concrete surfaces. Panel D illustrates faded lane marking detection on a wet pavement surface after rainfall, indicating that the detector remains capable of distinguishing degraded road markings from reflective background textures.

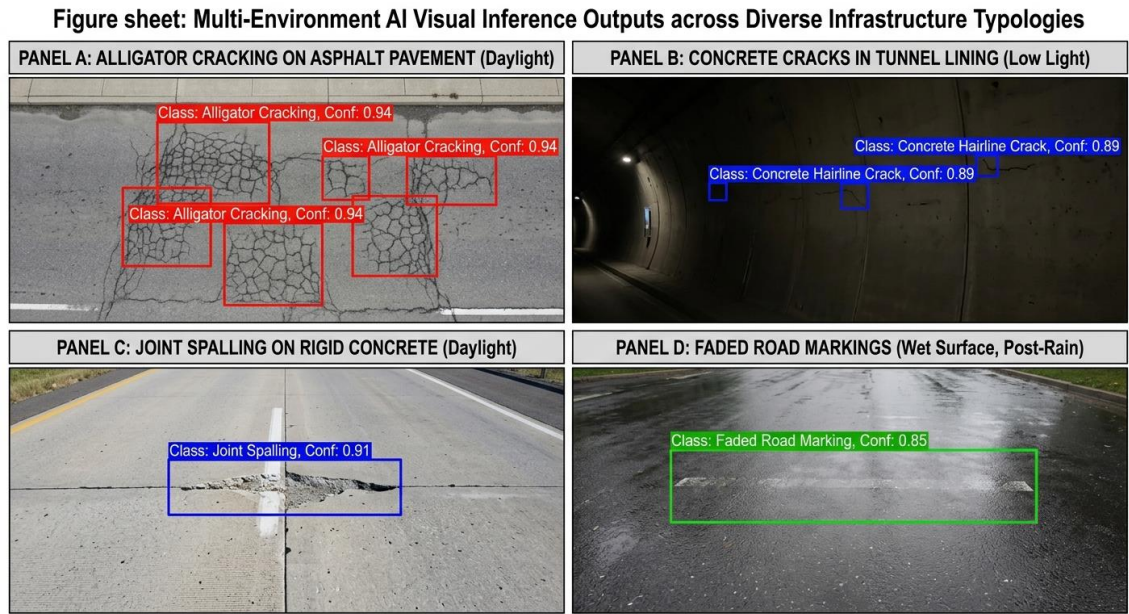


Figure 4. Multi-Environment AI Visual Inference Outputs across Diverse Infrastructure Typologies

C. Spatial Damage Quantification and Automated Maintenance Priority Index

The detected deterioration objects were subsequently transformed into spatial infrastructure records by combining object detection outputs with geographic positioning data acquired during field inspection. Each detected anomaly was assigned geographic coordinates together with its corresponding infrastructure category, deterioration type, and estimated physical dimensions. Additional severity information was incorporated to support maintenance prioritization using predefined inspection criteria. The resulting spatial records constitute a structured database that can be integrated directly into the Digital Twin Web-GIS platform for infrastructure asset management. The generated spatial maintenance records are summarized in Table 5.

Table 5. Heterogeneous Asset Distress Log and Automated Structural Maintenance Priority Index

Spatial Log Index	Target Asset Typology	GPS Coordinates (Lat, Long)	Detected Anomaly Type	Defect Dimension / Density	Action Priority Index
LOG-MULT-001	Urban Highway (Asphalt)	(-6.9821, 110.4231)	Pothole Cluster	Area: 0.45 m ²	High Priority
LOG-MULT-002	Underground Tunnel	(-6.9835, 110.4242)	Hairline Concrete Crack	Width: 1.2 mm	Medium Priority
LOG-MULT-003	Industrial Link (Concrete)	(-6.9848, 110.4259)	Severe Joint Spalling	Length: 2.1 m	Critical Priority
LOG-MULT-004	Flyover Bridge Deck	(-6.9862, 110.4271)	Surface Delamination	Area: 1.85 m ²	Medium Priority

Table 5 presents examples of spatial records generated automatically following AI-based deterioration detection. Each record combines geographic coordinates, infrastructure asset classification, detected deterioration type, estimated defect dimensions, and the corresponding maintenance priority level. Larger defects or deterioration associated with structural safety

receive higher maintenance priorities than smaller cosmetic defects. The generated spatial records provide structured information that can be directly incorporated into digital asset management systems without additional manual interpretation. This spatial representation enables infrastructure inspection results to be organized according to geographic location and maintenance urgency.

D. Cross-Platform Edge Hardware Performance and 3D Digital Twin GIS Synchronization

The deployment capability of the proposed inspection framework was evaluated using multiple low-power edge computing platforms representing different hardware configurations. Each platform executed the same trained detection model under identical testing conditions to evaluate computational efficiency and communication performance during field deployment. Performance evaluation considered inference speed together with synchronization latency between the embedded device and the cloud-based Digital Twin Web-GIS server. These measurements provide practical information regarding system responsiveness during continuous infrastructure inspection. The hardware benchmarking results are presented in Table 6.

Table 6. Cross-Platform Hardware Deployment Benchmark on Low-Power Edge Devices

Evaluated Edge Hardware Platform	Processor Architecture	VRAM / Shared Memory	Operating Power	Inference Speed	Latency to Cloud GIS
Raspberry Pi 5 + AI Hat	ARM Cortex-A76	8 GB LPDDR5	12 W	18 FPS	55.2 ms
NVIDIA Jetson Orin Nano	ARM Cortex + Ampere GPU	8 GB Shared Memory	15 W	119 FPS	8.4 ms
Automotive Mini-PC (Intel)	Intel Core i3 Mobile	16 GB DDR4	35 W	45 FPS	22.1 ms

Table 6 compares the computational characteristics of the evaluated embedded platforms using identical deployment settings. The NVIDIA Jetson Orin Nano achieved the highest inference speed together with the lowest communication latency, whereas the Raspberry Pi configuration operated with lower computational throughput but reduced power consumption. The automotive Mini-PC provided intermediate computational performance with higher energy requirements than the embedded GPU platform. These observations illustrate the trade-off between processing capability, power consumption, and cloud synchronization performance across different deployment architectures. The benchmark demonstrates the computational feasibility of implementing the proposed inspection framework on low-power edge computing devices.

The practical deployment of the inspection framework was further examined through integration with the cloud-based Digital Twin Web-GIS platform. Detection results generated on the embedded edge device were synchronized together with spatial coordinates and infrastructure metadata to provide centralized visualization of inspection outputs. The dashboard supports

interactive exploration of detected deterioration using three-dimensional geographic representation and temporal monitoring capabilities. Spatial visualization also facilitates the organization of inspection records according to infrastructure category and geographic location. The implemented Digital Twin interface is illustrated in Figure 5.

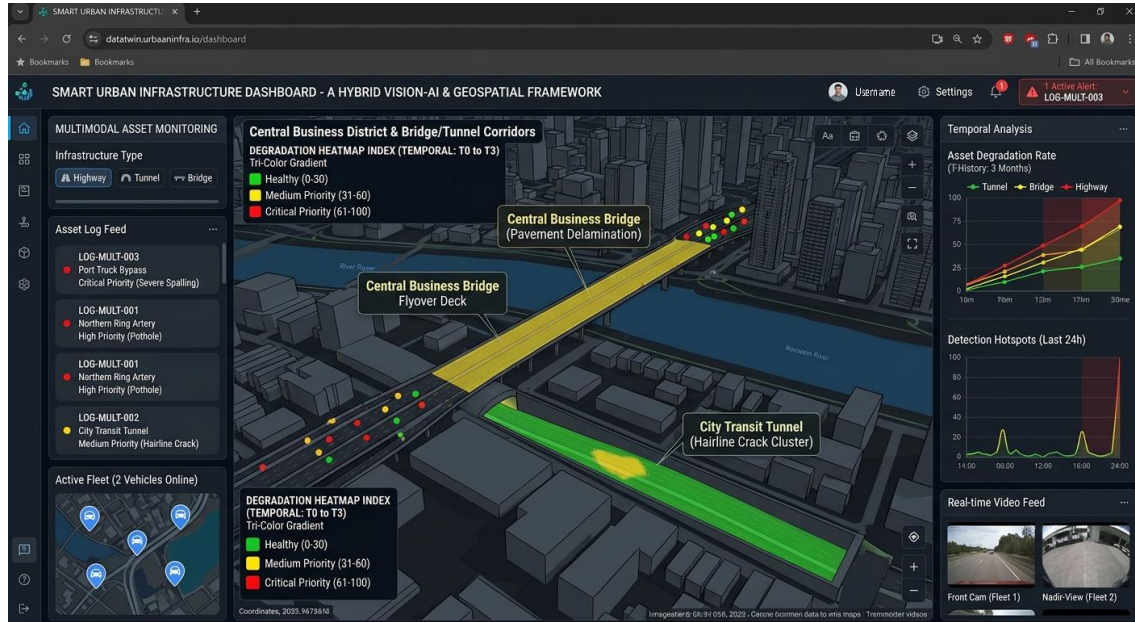


Figure 5. Interactive Digital Twin Cloud Web-GIS Dashboard with Spatial Infrastructure Monitoring

Figure 5 presents the operational Digital Twin Web-GIS dashboard used to visualize inspection results generated by the proposed framework. Individual deterioration locations are represented by spatial markers positioned on the three-dimensional infrastructure model according to their recorded geographic coordinates. The interface also displays inspection attributes, infrastructure categories, temporal information, and condition summaries within an integrated visualization environment. Interactive navigation allows inspection records to be explored according to location, deterioration type, or maintenance priority while preserving their spatial relationships. This implementation demonstrates the integration of edge-based artificial intelligence, cloud synchronization, and geospatial visualization within a unified infrastructure monitoring framework.

V. DISCUSSION

The findings demonstrate that the proposed hybrid Vision-AI and geospatial framework achieved stable detection performance across diverse operational environments, indicating that the integration of single-shot object detection with spatial synchronization effectively supports multimodal infrastructure inspection. The highest detection accuracy under clear daylight conditions confirms that sufficient illumination enhances feature extraction, while the moderate

performance reduction under shadowed, wet, and low-light environments suggests that environmental variability remains an important factor affecting visual inference. Nevertheless, the relatively small decrease in Precision, Recall, and mAP across all testing scenarios indicates that the proposed framework maintains robust detection capability despite heterogeneous operational conditions. These results are consistent with the observations of (Trigka & Dritsas, 2025), who reported that modern deep learning models preserve reliable detection performance under varying image resolutions and lighting conditions, and they further support the findings of (Kim et al., 2024) regarding the consistency of computer vision for infrastructure inspection. The ability to maintain stable detection across multiple infrastructure typologies expands previous studies that primarily concentrated on individual asset categories, addressing the limitation identified by (Lv et al., 2025).

The class-level evaluation further indicates that the proposed detector successfully distinguished multiple deterioration categories with only minor confusion between visually similar defects such as linear cracks and joint spalling. This observation is reasonable because both deterioration types share comparable geometric characteristics, making boundary discrimination more challenging during feature learning. Similar classification behavior has been reported by (Ozturk & Zappa, 2025), who noted that crack morphology with overlapping visual patterns may reduce classification separability, while (Zhang et al., 2025) demonstrated that deep learning models generally maintain high recognition accuracy despite such visual similarities. Although these minor misclassifications represent an expected anomaly in image-based inspection, they did not substantially affect the overall detection performance because the Precision, Recall, and F1-score remained well balanced across all damage classes. This performance reflects the capability of modern single-shot detection architectures to learn representative features across heterogeneous infrastructure materials and deterioration morphologies, consistent with the findings reported by (Haji Mohd et al., 2023).

Another important contribution of this study lies in the integration of AI-based detection outputs with geospatial information and Digital Twin visualization for automated infrastructure management. Unlike previous studies that primarily emphasized either computer vision or GIS visualization independently, the proposed framework directly transforms detected defects into spatial maintenance records containing geographic coordinates, deterioration characteristics, and maintenance priority indices. This capability supports more efficient asset management because inspection results become immediately interpretable within a geospatial decision-support environment, reducing the need for additional manual processing. These characteristics are consistent with the role of geospatial synchronization described by (Gao et al., 2024), while extending the framework proposed by (Kazanskiy et al., 2025) through the integration of unified

multimodal defect detection with real-time edge computing. The hardware benchmark also illustrates that embedded edge devices, particularly the NVIDIA Jetson Orin Nano, provide sufficient computational performance and low synchronization latency for continuous infrastructure inspection in practical deployment scenarios.

Despite these promising findings, several limitations should be considered when interpreting the results. First, the experimental dataset was limited to four representative transportation infrastructure categories, which may not fully capture the broader diversity of deterioration patterns encountered in different climatic regions, construction materials, or infrastructure ages. Second, although the proposed framework maintained satisfactory robustness under various environmental conditions, the reduced performance observed on wet surfaces and under low illumination indicates that adverse imaging conditions continue to influence visual feature extraction and localization accuracy. Third, the evaluation focused primarily on detection performance and deployment efficiency, while long-term operational validation under continuously changing infrastructure conditions was beyond the scope of this study. These limitations indicate that the reported performance should be interpreted within the methodological and experimental conditions adopted in this investigation, particularly with respect to the dataset composition, environmental variability, and deployment settings used during evaluation.

VI. CONCLUSION AND RECOMMENDATION

This study developed a hybrid Vision-AI and geospatial framework that integrates single-shot object detection, edge computing, and Digital Twin Web-GIS to support automated multimodal infrastructure degradation mapping across heterogeneous urban transportation assets. The experimental results demonstrated that the proposed framework achieved stable detection performance under various operational environments while maintaining reliable classification accuracy for multiple deterioration categories, including pavement distress and concrete structural defects. The integration of AI-based detection with geospatial synchronization enabled the automatic generation of spatial maintenance records containing infrastructure location, defect characteristics, and maintenance priority, providing structured information for digital asset management. The hardware evaluation further confirmed that the framework can be deployed on low-power edge computing platforms with satisfactory inference speed and cloud synchronization performance, supporting practical implementation for real-time infrastructure inspection. These findings indicate that the proposed framework addresses the identified gap in unified multimodal infrastructure monitoring by combining automated defect detection with geospatial intelligence within a single operational workflow.

Future research may expand the proposed framework by incorporating additional infrastructure typologies, such as railway systems, coastal structures, and utility networks, to evaluate its adaptability across broader civil infrastructure applications. The inclusion of more diverse datasets collected under different climatic conditions, seasonal variations, and geographical regions would also provide a more comprehensive assessment of model generalization and robustness. Further investigations may integrate temporal deterioration analysis using sequential inspection data to support predictive maintenance and long-term infrastructure condition forecasting within Digital Twin environments. The incorporation of multimodal sensing technologies, including LiDAR, thermal imaging, and unmanned aerial vehicle platforms, may enhance detection capability for defects that are difficult to identify using conventional RGB imagery alone. In addition, future developments may explore more advanced lightweight deep learning architectures and adaptive edge-cloud optimization strategies to improve computational efficiency while maintaining high detection accuracy for large-scale intelligent infrastructure monitoring systems.

AI USE DECLARATION

The authors used artificial intelligence tools, including ChatGPT, solely for language editing during the preparation of this manuscript. These tools were used only to improve grammar, sentence structure, readability, word choice, and the overall clarity of the writing. Artificial intelligence was not involved in any stage of the research. All scientific aspects of the study, including the development of the research concept, study design, data collection, methodological development, data analysis, interpretation of the findings, and preparation of the final conclusions, were carried out entirely by the authors. The authors take full responsibility for the content of this manuscript and confirm that it is original, accurate, scientifically valid, prepared with academic integrity, and consistent with accepted standards of scientific quality.

REFERENCES

- Costa, D. G., Bittencourt, J. C. N., Oliveira, F., Peixoto, J. P., & Jesus, T. C. (2024). Achieving Sustainable Smart Cities through Geospatial Data-Driven Approaches. *Sustainability*, 16(2), 640. <https://doi.org/10.3390/su16020640>
- Dritsas, E., & Trigka, M. (2025). Remote Sensing and Geospatial Analysis in the Big Data Era: A Survey. *Remote Sensing*, 17(3), 550. <https://doi.org/10.3390/rs17030550>
- Duran, B., Eftekhari Azam, S., & Sanayei, M. (2024). Leveraging Deep Learning for Robust Structural Damage Detection and Classification: A Transfer Learning Approach via CNN. *Infrastructures*, 9(12), 229. <https://doi.org/10.3390/infrastructures9120229>
- Gao, Y., Xiong, G., Hu, Z., Chai, C., & Li, H. (2024). Bridge Digital Twin for Practical Bridge Operation and Maintenance by Integrating GIS and BIM. *Buildings*, 14(12), 3731.

<https://doi.org/10.3390/buildings14123731>

- Haji Mohd, M. N., Mohd Asaari, M. S., Lay Ping, O., & Rosdi, B. A. (2023). Vision-Based Hand Detection and Tracking Using Fusion of Kernelized Correlation Filter and Single-Shot Detection. *Applied Sciences*, 13(13), 7433. <https://doi.org/10.3390/app13137433>
- James, E., & Martinez, S. (2025). Multi-Agent AI Simulation for Evaluating Sustainability of Urban Transportation Infrastructure. *Civil Engineering Science and Technology*, 1(2), 186–202. <https://doi.org/10.51903/tx30kz14>
- Ji, Y., Zhang, D., He, Y., Zhao, J., Duan, X., & Zhang, T. (2025). Improved YOLO11 Algorithm for Insulator Defect Detection in Power Distribution Lines. *Electronics*, 14(6), 1201. <https://doi.org/10.3390/electronics14061201>
- Kazanskiy, N., Khabibullin, R., Nikonorov, A., & Khonina, S. (2025). A Comprehensive Review of Remote Sensing and Artificial Intelligence Integration: Advances, Applications, and Challenges. *Sensors*, 25(19), 5965. <https://doi.org/10.3390/s25195965>
- Khan, Z., Shen, Y., & Liu, H. (2025). ObjectDetection in Agriculture: A Comprehensive Review of Methods, Applications, Challenges, and Future Directions. *Agriculture*, 15(13), 1351. <https://doi.org/10.3390/agriculture15131351>
- Kim, J.-W., Choi, H.-W., Kim, S.-K., & Na, W. S. (2024). Review of Image-Processing-Based Technology for Structural Health Monitoring of Civil Infrastructures. *Journal of Imaging*, 10(4), 93. <https://doi.org/10.3390/jimaging10040093>
- Li, H., Cheng, H., Niu, Q., Li, C., Wang, L., He, X., Yang, Y., & Wang, P. (2026). Small Target Detection in Agricultural Visual Perception: Progress and Challenges. *Agriculture*, 16(13), 1366. <https://doi.org/10.3390/agriculture16131366>
- Liang, B., Li, Z., Xu, M., Xiao, J., & Long, H. (2024). Study on the Design Parameters of Landscape Elements of Urban Tunnel Body Segments in China Based on Biomass Perception. *Applied Sciences*, 14(14), 6100. <https://doi.org/10.3390/app14146100>
- Liu, C., Zhang, S., Hu, M., & Song, Q. (2024). Object Detection in Remote Sensing Images Based on Adaptive Multi-Scale Feature Fusion Method. *Remote Sensing*, 16(5), 907. <https://doi.org/10.3390/rs16050907>
- Liu, H., Fu, Y., Ma, Y., & Zhang, W. (2025). Multimodal Fusion and Dynamic Resource Optimization for Robust Cooperative Localization of Low-Cost UAVs. *Drones*, 9(12), 820. <https://doi.org/10.3390/drones9120820>
- Lv, Z., Hao, Z., Zhu, Y., & Lu, C. (2025). A Review on Automated Detection and Identification Algorithms for Highway Pavement Distress. *Applied Sciences*, 15(11), 6112. <https://doi.org/10.3390/app15116112>
- Ma, J., Yang, Q., Peng, X., & Xia, K. (2025). Review on Durability Deterioration and Mitigation of Concrete Structures. *Coatings*, 15(9), 982. <https://doi.org/10.3390/coatings15090982>
- Mohammadi, S., Sattarpanah Karganroudi, S., & Rahmanian, V. (2025). Advancements in Smart

- Nondestructive Evaluation of Industrial Machines: A Comprehensive Review of Computer Vision and AI Techniques for Infrastructure Maintenance. *Machines*, 13(1), 11. <https://doi.org/10.3390/machines13010011>
- Nugroho, A. S., & Wibowo, M. C. (2024). Design and Implementation of Animated Stickers as an Educational Tool for Adolescent Drug Awareness Using the MDLC Method. *International Journal of Graphic Design*, 2(2), 195–206. <https://doi.org/10.51903/ijgd.v2i2.2114>
- Ozturk, H. Y., & Zappa, E. (2025). Automated Crack Width Measurement in 3D Models: A Photogrammetric Approach with Image Selection. *Information*, 16(6), 448. <https://doi.org/10.3390/info16060448>
- Sanz Bobi, J. D., Garrido Martínez-Llop, P., Rubio Marcos, P., Solano Jiménez, Á., & Fernández, J. G. (2024). Prediction of Degraded Infrastructure Conditions for Railway Operation. *Sensors*, 24(8), 2456. <https://doi.org/10.3390/s24082456>
- Schatzl, J., Korenjak, A. K., Gerhold, F., & Marschnig, S. (2026). Integrating Railway Infrastructure Data: A Spatial–Temporal Database for Track Deterioration Analysis. *Applied Sciences*, 16(13), 6544. <https://doi.org/10.3390/app16136544>
- Syafril, M., Arief, S., & Pangaribuan, J. D. (2025). Application of Digital Twin Technology for Real-Time Monitoring and Predictive Maintenance of Coastal Bridges under Climate Change Scenarios. *Jurnal Rekayasa Sipil Dan Arsitektur*, 1(2), 39–51. <https://doi.org/10.51903/5fkqb467>
- Torbali, M. E., Zolotas, A., & Avdelidis, N. P. (2023). A State-of-the-Art Review of Non-Destructive Testing Image Fusion and Critical Insights on the Inspection of Aerospace Composites towards Sustainable Maintenance Repair Operations. *Applied Sciences*, 13(4), 2732. <https://doi.org/10.3390/app13042732>
- Trigka, M., & Dritsas, E. (2025). A Comprehensive Survey of Deep Learning Approaches in Image Processing. *Sensors*, 25(2), 531. <https://doi.org/10.3390/s25020531>
- Yuan, Q., Shi, Y., & Li, M. (2024). A Review of Computer Vision-Based Crack Detection Methods in Civil Infrastructure: Progress and Challenges. *Remote Sensing*, 16(16), 2910. <https://doi.org/10.3390/rs16162910>
- Zaroujtaghi, A., Mansourihanis, O., Tayarani, M., Mansouri, F., Hemmati, M., & Soltani, A. (2025). A Systematic Review of GIS Evolution in Transportation Planning: Towards AI Integration. *Future Transportation*, 5(3), 97. <https://doi.org/10.3390/futuretransp5030097>
- Zhang, H., Dong, Y., Hou, Y., Cheng, X., Xie, P., & Di, K. (2025). Research on Asphalt Pavement Surface Distress Detection Technology Coupling Deep Learning and Object Detection Algorithms. *Infrastructures*, 10(4), 72. <https://doi.org/10.3390/infrastructures10040072>