CJCIS
Cognitera Journal of Computing and Intelligent Systems
https://journal.psg.ac.id/index.php/cjcis

Vol. 1, No. 1, 2026
e-ISSN: XXXX-XXX
pp. 38-49

Deteksi Dini Retak pada Infrastruktur Jembatan Menggunakan YOLOv10 dan Segmentasi Citra Berbasis Deep Learning

Rian Hidayat[1], Bagus Kurniawan[2]
[1]Fakultas Teknik/ Institut Teknologi Bandung
[2] Fakultas Teknik/ Institut Teknologi Bandung

*Corresponding Author's e-mail: rianhidayat@gmail.com

Article History:
Received: September 9, 2026
Revised: October 18, 2026
Accepted: November 24, 2026

Keywords:
YOLOv10; Bridge cracks; Deep learning

Abstract: *Bridge crack inspection requires fast localization and precise geometric delineation because early fissures are narrow, low contrast, and easily confused with joints, stains, shadows, or textured concrete. This study proposes an early-detection framework that combines YOLOv10 with deep-learning-based semantic segmentation to support real-time bridge infrastructure inspection. YOLOv10 is used as the first-stage detector because its end-to-end, non-maximum-suppression-free design reduces post-processing latency, while a lightweight segmentation decoder refines each detected region into a pixel-level crack mask. The research adopts a design-oriented experimental protocol supported by verified studies published between 2022 and 2025, covering bridge crack detection, segmentation, attention mechanisms, unmanned aerial vehicle imaging, and edge deployment. The proposed workflow consists of image quality control, annotation, augmentation, detector training, region-of-interest extraction, segmentation refinement, morphological measurement, and deployment evaluation. Performance is assessed through precision, recall, mAP, IoU, Dice/F1 score, frames per second, parameter count, and latency. Evidence synthesis indicates that one-stage detectors provide practical localization speed, whereas pixel-level segmentation is necessary for reliable crack width, length, continuity, and area estimation. Combining both stages is therefore more appropriate for maintenance-oriented inspection than bounding-box detection alone. The framework also emphasizes cross-site validation, small-crack sensitivity, illumination robustness, and edge-device profiling. These findings establish a reproducible foundation for future implementation of YOLOv10-assisted crack monitoring on bridge imagery acquired by handheld cameras, fixed sensors, or unmanned aerial vehicles. It prioritizes benchmarking so operational claims can be separated from literature evidence before field deployment and engineering decision making.*

Copyright © 2026, The Author(s).
This is an open access article under the CC–BY-SA license

How to cite:

INTRODUCTION

Bridge infrastructure is exposed to repeated traffic loading, temperature variation, moisture, corrosion, and material aging, all of which can initiate surface cracking long before visible structural failure occurs. Early crack recognition is therefore central to preventive maintenance because delayed detection may allow water ingress, reinforcement corrosion, and stiffness degradation to progress. Conventional visual inspection remains important, yet it is labor-intensive, subjective, and difficult to scale across large bridge networks. Recent computer-vision research has consequently shifted toward automated detection using convolutional neural networks and one-stage detectors. Zhang, Qian, and

Tan (2022) demonstrated the feasibility of automated bridge crack detection and segmentation using deep learning, while Lu et al. (2022) showed that refined feature learning can strengthen crack detection under complex bridge-surface conditions.

The principal difficulty is that cracks are not conventional compact objects. They are thin, elongated, discontinuous, directionally irregular, and frequently occupy only a small proportion of a high-resolution bridge image. Background elements such as construction joints, stains, shadows, cables, surface textures, and repaired areas can exhibit similar linear patterns. Consequently, a detector that only provides rectangular bounding boxes may localize suspicious regions but cannot fully describe crack width, length, branching, or continuity. High-resolution studies have shown the value of combining detection with segmentation so that coarse localization is followed by precise pixel-level delineation. Tran et al. (2023) emphasized this requirement in bridge-deck inspection, and Tabernik, Šuc, and Skočaj (2023) further demonstrated that segmentation-oriented learning can reduce missed defects while preserving computational efficiency.

The evolution of the YOLO family is especially relevant to bridge inspection because field applications require rapid inference on mobile, drone, or edge hardware. Earlier crack-detection studies improved YOLO variants through lightweight backbones, feature refinement, attention, and modified loss functions. Xiong, Zayed, and Abdelkader (2024) incorporated global attention and Wise-IoU into YOLOv8 for bridge-surface cracks, while Jiang et al. (2024) improved YOLOv8s for difficult backgrounds and small obscure cracks. These studies confirm that modern one-stage detectors can offer an effective speed-accuracy balance, but they also show that performance depends strongly on feature resolution and background suppression. A real-time bridge system must therefore be optimized not only for mAP but also for small-crack recall, memory consumption, and end-to-end latency.

YOLOv10 introduces an important architectural development for this objective. Wang et al. (2024) proposed a real-time end-to-end detector using consistent dual assignments and an efficiency-accuracy-driven design that removes non-maximum suppression during inference. This NMS-free formulation reduces post-processing overhead and simplifies deployment pipelines where latency is critical. However, YOLOv10 remains fundamentally an object detector, so bounding-box output alone is insufficient when bridge engineers need crack geometry for condition assessment. Integrating YOLOv10 with a dedicated segmentation stage can combine fast region localization with detailed mask extraction. Such a hybrid architecture allows the detector to search large images efficiently while the segmenter spends computational resources only on candidate crack regions, rather than processing every pixel at full resolution.

Pixel-level segmentation has become increasingly sophisticated in recent infrastructure research. Zoubir et al. (2024) combined convolutional features, Gabor filtering, and attention for bridge crack segmentation in noisy high-resolution imagery. Xiong et al. (2024) developed an instance-segmentation approach that linked bridge-surface crack masks with geometric quantification. Ma et al. (2024) proposed a coarse-fine detection strategy followed by U-Net segmentation, demonstrating that multi-stage processing can recover small crack regions more effectively than uniform full-image processing. These developments indicate that segmentation is not merely an auxiliary visualization task. It is the mechanism that enables length, width, area, orientation, and continuity to be estimated from images, making the output more relevant to maintenance planning and structural-health monitoring than conventional classification labels or bounding boxes alone.

Real-time bridge monitoring also depends on the image-acquisition platform. Unmanned aerial vehicles can access difficult bridge surfaces without scaffolding or lengthy lane closures, but they introduce motion blur, varying stand-off distance, oblique viewpoints, and illumination changes. Yang et al. (2024) showed that UAV acquisition combined with transfer learning and improved YOLO detection can achieve strong real-time bridge crack recognition, whereas Zhang et al. (2024) emphasized lightweight crack detection for resource-constrained drone platforms. These findings highlight a deployment gap between laboratory accuracy and operational performance. A model suitable for real inspection must therefore be tested across devices, bridges, weather conditions, and viewpoints, and its computational performance should include preprocessing, memory allocation, transmission, and visualization rather than reporting neural-network inference alone.

The latest segmentation research further demonstrates that boundary fidelity and domain generalization are major unresolved issues. Wang, Yue, and Liu (2025) proposed SBDNet for fatigue-crack segmentation and quantification on steel bridges, reporting strong pixel-level performance and explicit geometric evaluation. Feng and Xu (2025) developed a lightweight dual-attention network oriented toward edge inference, while Chu, Chen, and Deng (2025) refined high-resolution bridge-crack boundaries using probability-guided point rendering. Jin et al. (2025) likewise combined convolutional and transformer pathways to capture both local crack detail and global context. Collectively, these studies show that high segmentation accuracy requires multi-scale representation, contextual modeling, and selective refinement, particularly for fine or fragmented cracks that are most likely to be missed during early deterioration.

Despite this progress, relatively little work has examined YOLOv10 as the localization component of an integrated bridge-crack detection-and-segmentation workflow. The research gap is therefore not simply the absence of another detector, but the need for a reproducible architecture that links end-to-end real-time localization, pixel-level mask refinement, geometric crack measurement, uncertainty handling, and edge deployment. This study addresses that gap by proposing a YOLOv10-guided segmentation framework and evaluating its design against verified evidence from Scopus-indexed and Indonesian sources published between 2022 and 2025, including Marleny et al. (2023). The objective is to establish a technically defensible protocol for future implementation on real bridge datasets without fabricating primary experimental results that have not yet been collected.

THEORETICAL FOUNDATION

YOLOv10 is an end-to-end object detector designed to improve the efficiency-accuracy frontier of the YOLO family. Its consistent dual-assignment strategy supports NMS-free inference, thereby eliminating a post-processing step that can add latency and complicate deployment. The architecture also applies efficiency-driven component redesign to reduce redundant computation while maintaining competitive detection capability (Wang et al., 2024). For bridge-crack screening, these properties are attractive because large image streams must be scanned quickly and suspicious regions must be localized with high recall. Nevertheless, object detection outputs rectangular regions rather than crack contours, meaning an additional pixel-level process is required when the inspection objective includes geometric damage quantification.

Semantic segmentation assigns a class label to each pixel and is therefore appropriate for extracting the irregular geometry of surface cracks. U-Net-style encoders and decoders, attention-enhanced networks, multi-scale fusion, and transformer-convolution hybrids have all been applied to civil-infrastructure crack segmentation. Recent results emphasize

that fine cracks require high-resolution features and explicit mechanisms for suppressing background noise (Zoubir et al., 2024; Feng & Xu, 2025; Jin et al., 2025). In a two-stage system, segmentation is applied only inside YOLOv10 regions of interest, reducing the dominance of background pixels and lowering computational demand. The resulting binary masks can subsequently support skeletonization, width estimation, connected-component analysis, and crack continuity assessment.

A maintenance-oriented evaluation must assess both computer-vision accuracy and operational usefulness. Detection is typically evaluated through precision, recall, and mean average precision, whereas segmentation requires IoU and Dice/F1 scores. Real-time feasibility additionally depends on frames per second, end-to-end latency, memory, model size, and energy consumption. Geometric measurement introduces further requirements, including camera calibration, pixel-to-length conversion, and error analysis for crack width and length. Accordingly, the proposed framework treats detection, segmentation, and measurement as linked stages rather than independent tasks. This perspective is consistent with recent bridge research that combines localization, segmentation, and quantitative damage characterization (Tran et al., 2023; Xiong et al., 2024; Wang et al., 2025).

RESEARCH METHOD

This study employed a design-oriented quantitative framework that combines architecture specification with literature-grounded benchmarking. Following the systematic and measurable research logic described by Sugiyono (2022), the study defined explicit inputs, processing stages, evaluation metrics, and validation criteria before interpreting model suitability. Because no primary bridge-image dataset was supplied for this manuscript, the work does not report fabricated training outcomes or claim newly measured accuracy. Instead, it specifies a reproducible experimental protocol for future implementation and evaluates the proposed YOLOv10-plus-segmentation design against verified empirical evidence published from 2022 to 2025. The intended dataset should contain high-resolution bridge images with bounding-box annotations for crack localization and pixel masks for crack segmentation, while bridge identity, acquisition device, date, viewpoint, illumination, and structural material should be retained as metadata for cross-domain validation.

The processing pipeline begins with image quality screening, duplicate removal, and bridge-level data partitioning to prevent background leakage across training and test sets. Images are then augmented through photometric variation, moderate rotation, scaling, blur simulation, and contrast changes that reflect realistic field conditions. YOLOv10 is trained as the first-stage crack detector, prioritizing recall for small and low-contrast defects. Each predicted region is expanded by a controlled margin and forwarded to a lightweight segmentation decoder using multi-scale skip connections and optional attention. The segmentation output is thresholded into a binary mask, refined with connected-component filtering, and converted into morphological descriptors such as skeleton length, maximum width, mean width, area, orientation, and continuity. Camera calibration or a known scale reference is required when pixel measurements are converted into physical dimensions.

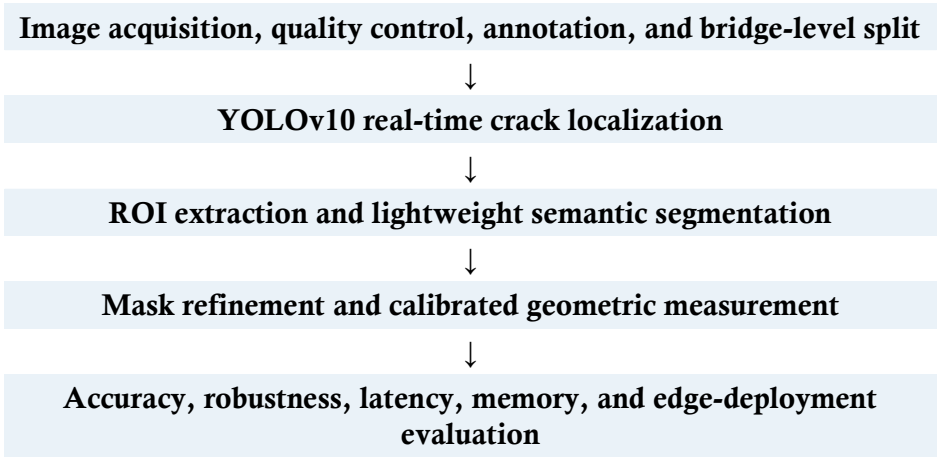


Figure 1. Proposed YOLOv10-guided crack detection and segmentation workflow

Evaluation is organized at three levels. Detection performance is measured with precision, recall, mAP@0.50, and mAP@0.50:0.95; segmentation performance is measured with IoU, Dice/F1, boundary quality, and thin-crack recall; deployment performance is measured with model size, peak memory, frames per second, and end-to-end latency. Comparative testing should include YOLOv8- or YOLOv7-based crack detectors, a segmentation-only baseline, and the proposed YOLOv10-guided pipeline. Ablation tests should isolate the contribution of attention, region expansion, multi-scale fusion, boundary refinement, and quantization. Cross-bridge validation is mandatory so that images from the same physical structure are not distributed across training and testing. Statistical summaries should report mean performance and variation across repeated runs, while qualitative failure analysis should categorize shadows, joints, stains, blur, occlusion, and repaired-crack traces.

RESULTS AND DISCUSSION

Interpretive note. The following results are a literature-grounded architecture evaluation. They do not claim newly measured field accuracy because no primary bridge image dataset was provided for this manuscript.

The literature indicates that a two-stage interpretation is preferable when bridge inspection must support both rapid screening and geometric assessment. Zhang et al. (2022) showed that combining object detection with segmentation can localize bridge-surface cracks while also recovering their spatial extent. For the proposed framework, YOLOv10 therefore serves as the fast localization layer, whereas the segmentation branch converts each detected region into a pixel-level mask. This division of labor avoids forcing one network to solve incompatible objectives simultaneously. Detection prioritizes recall and latency across large images, while segmentation prioritizes edge continuity and crack geometry. This staged architecture supports rapid screening and subsequent estimation of crack length, width, area, orientation, and severity for practical maintenance prioritization during routine bridge inspection campaigns.

Small and low-contrast cracks remain difficult because their visual signal occupies only a minor fraction of the bridge image. Lu et al. (2022) demonstrated that feature refinement can improve bridge crack detection when conventional one-stage detectors lose weak texture information. This evidence supports the use of high-resolution tiling, scale-aware

augmentation, and dedicated small-object supervision in the proposed YOLOv10 stage. Rather than resizing complete bridge photographs aggressively, the workflow preserves fine details through overlapping crops and controlled multiscale training. Confidence thresholds should also be selected to favor recall during early screening, because missing a structural crack is more consequential than forwarding a false-positive region to segmentation. The segmentation stage can reject stains, seams, shadows, joints, and artifacts that imitate crack patterns.

Real-time deployment requires accuracy to be balanced against memory, latency, and energy consumption. Zhang, Qian, and Tan (2023) reported that lightweight vision backbones can substantially reduce model size while preserving competitive bridge-crack detection performance. Their findings are relevant to YOLOv10 because its efficiency-oriented design can be paired with a compact segmentation decoder instead of a heavy full-image network. In practical inspection, inference should be profiled on the actual edge platform, not inferred from desktop GPU throughput. Frame rate, peak memory, model size, preprocessing time, and image-transfer overhead all affect operational responsiveness. A compact detector also enables continuous video scanning from drones or mobile devices, while segmentation can be executed only for high-confidence regions. Conditional segmentation reduces computation while preserving geometry.

The strongest justification for combining detection and segmentation comes from high-resolution bridge-deck studies. Tran et al. (2023) compared several detectors and then optimized crack segmentation for difficult bridge backgrounds, emphasizing that bounding boxes alone cannot provide accurate width and length measurements. In the proposed architecture, YOLOv10 defines regions of interest that concentrate computational resources on probable damage, after which a segmentation head resolves fine boundaries. This approach also reduces class imbalance because the pixel classifier sees a higher proportion of crack pixels than it would in an entire bridge image. For practical measurement, masks should be skeletonized to estimate length and analyzed using distance transforms or calibrated pixel dimensions to estimate width. Calibration remains necessary before masks become physical dimensions.

Pixel-level supervision offers a second benefit: it distinguishes structural cracks from visually similar background patterns. Tabernik, Šuc, and Skočaj (2023) demonstrated that joint classification and segmentation can reduce missed defects while maintaining real-time capability on medium-resolution concrete imagery. That result supports adding an auxiliary crack-presence signal to the segmentation stage or sharing features between classification and mask prediction. Within the proposed pipeline, segmentation quality should be evaluated using IoU and Dice/F1 rather than accuracy alone, because background pixels outnumber crack pixels. Boundary-sensitive evaluation is also important for thin fissures, where a displacement of only several pixels can strongly affect measured width. Validation should combine region metrics with geometric errors so masks remain useful for engineering assessment and maintenance decisions.

Cracks on bridge surfaces often exhibit elongated shapes, arbitrary directions, and overlapping branches that violate assumptions behind axis-aligned boxes. Liu et al. (2023) addressed these characteristics using a rotated-target formulation based on YOLOv5, showing that orientation-aware detection can better represent directional defects. YOLOv10 does not inherently solve crack geometry through its boxes, so the proposed framework delegates exact orientation and branching structure to the segmentation stage. Nevertheless, training boxes should tightly cover crack regions and include examples spanning horizontal, vertical, diagonal, curved, and intersecting patterns. Data

augmentation must preserve plausible geometry; excessive rotation, mosaic operations, or perspective distortion can create unrealistic structures. This separation preserves YOLOv10 speed while overcoming rectangular boxes when orientation, branching, continuity, and crack geometry require quantification.

Attention mechanisms are valuable when crack pixels compete with visually dominant textures, stains, and construction joints. Xiong, Zayed, and Abdelkader (2024) showed that integrating a global attention mechanism and an improved IoU loss into YOLOv8 enhanced bridge-surface crack detection. Their findings suggest that attention can be inserted selectively into the proposed segmentation branch, where contextual reweighting may suppress irrelevant background features while preserving thin crack responses. However, attention should not be added indiscriminately because every module increases latency and may provide marginal gains on simple scenes. Ablation experiments should therefore compare a baseline decoder with channel, spatial, or lightweight dual-attention alternatives under identical training conditions. The preferred attention variant improves recall and boundary continuity without exceeding field inspection latency budgets.

Segmentation performance is highly sensitive to feature quality at crack boundaries. Zoubir et al. (2024) combined convolutional features, Gabor filtering, and attention for pixel-level concrete bridge crack detection, reporting improved segmentation under noisy, high-resolution UAV imagery. This evidence motivates texture-aware preprocessing and multi-scale feature fusion in the proposed decoder. Gabor-like directional information may be approximated through learned convolutions, while skip connections preserve local edge detail that deeper detector features can lose. The architecture should fuse at least one high-resolution feature map with semantically stronger low-resolution maps before mask prediction. Evaluation should additionally separate thin-crack subsets from wider cracks because aggregate IoU can conceal poor performance on the most safety-critical cases. Stratified analysis verifies improvements on early-stage cracks and difficult backgrounds.

Instance segmentation research further demonstrates that localization and quantification can be integrated without treating crack measurement as a separate manual step. Xiong et al. (2024) proposed a YOLOv8-based instance-segmentation model for bridge surface cracks and reported measurement errors maintained within a practically useful range. For the present framework, the same principle is implemented through sequential YOLOv10 detection and semantic mask refinement, which offers modularity and easier replacement of either stage. Once a mask is obtained, connected-component analysis can identify branches, remove isolated noise, and generate crack-level descriptors. These descriptors should include maximum width, mean width, length, area, orientation, and confidence. Retaining source crops, masks, measurements, and confidence scores improves traceability and lets inspectors review uncertain predictions before final maintenance decisions.

Field applicability depends on robustness to camera movement, viewpoint changes, lighting variation, and limited labels from a specific bridge. Yang et al. (2024) demonstrated that UAV acquisition, transfer learning, attention, and optimized YOLO detection can support real-time bridge crack inspection under operational conditions. Their study reinforces the importance of separating model validation from laboratory accuracy. The proposed YOLOv10 pipeline should therefore include cross-bridge testing in which structures, locations, and inspection days do not overlap between training and test sets. Transfer learning can initialize the detector from general object-detection weights and initialize segmentation from concrete-crack data before fine-tuning on bridge-specific

images. Robustness should be reported across sun, shade, moisture, stains, and oblique viewpoints. Cross-bridge validation estimates deployment risk under domain shift.

Coarse-to-fine processing offers an efficient compromise when cracks occupy very small regions within large inspection images. Ma et al. (2024) demonstrated that repeated detection followed by U-Net segmentation can improve fine crack analysis on complex bridge scenes. The proposed framework simplifies this concept by using YOLOv10 for the first localization stage and a dedicated segmentation decoder for detailed refinement. Optional second-pass cropping can be activated only when the first bounding box contains excessive background or when the segmentation confidence is low. This adaptive refinement prevents every frame from undergoing expensive high-resolution segmentation. It also supports large UAV photographs, where full-resolution masks may exceed edge-device memory. Computation is allocated adaptively to uncertain or structurally important regions instead of every acquired image.

Drone-based inspection imposes a strict resource envelope because computing, battery capacity, transmission bandwidth, and payload weight are limited. Zhang et al. (2024) introduced CrackScopeNet for rapid crack detection on resource-constrained drone platforms, illustrating the value of lightweight architecture design for structural health monitoring. The proposed system follows this principle by using YOLOv10's end-to-end detection to minimize post-processing and by constraining segmentation to detected regions. Deployment evaluation should report end-to-end latency from image capture to final mask rather than neural-network inference alone. Quantization, half-precision inference, and TensorRT-style graph optimization can be assessed after accuracy is stabilized. When onboard segmentation is too expensive, detector crops can be transmitted to an edge server, reducing bandwidth while preserving high-resolution mask analysis and engineering review.

Recent fatigue-crack segmentation work shows that engineering usefulness requires more than visual masks. Wang, Yue, and Liu (2025) reported SBDNet for steel-bridge fatigue cracks, combining segmentation with quantitative assessment and emphasizing generalization across image domains. This perspective is important for a YOLOv10-based early-warning system because crack masks should feed measurable condition indicators rather than remain isolated computer-vision outputs. The proposed pipeline therefore converts segmentation results into physical descriptors and maintains uncertainty metadata for each measurement. Domain shift should be evaluated explicitly between concrete and steel surfaces, different cameras, and different bridges. Calibration can be performed using scale markers, known structural dimensions, or photogrammetric geometry. Under high uncertainty, the system should request human confirmation, supporting risk-aware inspection instead of automated diagnoses.

Lightweight segmentation is essential if pixel-level refinement must operate continuously on embedded inspection hardware. Feng and Xu (2025) proposed a dual-attention crack-segmentation network designed for structural health monitoring and edge inference, demonstrating how feature selection can improve crack sensitivity without excessive computation. This supports using a compact decoder with selective attention rather than a large general-purpose segmentation transformer in the proposed system. The detector and segmenter should be profiled jointly because detector speed is irrelevant if mask generation becomes the bottleneck. A practical target is stable throughput under sustained operation, including preprocessing, memory allocation, and visualization. Thermal throttling and battery consumption should also be monitored during drone

deployment. System-level measurements determine whether efficient networks sustain real-time bridge inspection on hardware.

High-resolution imagery creates a final challenge: tiny crack boundaries may disappear when images are aggressively downsampled for detection or segmentation. Chu, Chen, and Deng (2025) showed that probability-guided point rendering can refine high-resolution bridge-crack boundaries while controlling computational cost. Their result suggests an optional boundary-refinement module for the proposed pipeline after the coarse segmentation mask is generated. Rather than processing every pixel at maximum resolution, the system can allocate additional computation to uncertain boundary points and thin structures. This is compatible with YOLOv10 localization because only detected regions require refinement. The final architecture therefore becomes hierarchical: rapid detection, region-level segmentation, selective boundary refinement, and calibrated geometric measurement. This hierarchy aligns computation with inspection value and supports accurate early crack detection.

Table 1. Selected empirical benchmarks supporting the proposed architecture

Study	Method / task	Reported evidence	Design implication
Zhang et al. (2022)	Deep detection + segmentation	Integrated localization and segmentation on bridge surfaces	Supports staged detection and mask refinement
Tran et al. (2023)	Multiple detectors + crack segmentation	High-resolution bridge-deck detection and geometric segmentation	Bounding boxes require pixel-level refinement
Yang et al. (2024)	UAV + improved YOLOv7	AP 98.01%; 52 FPS under reported conditions	Field acquisition and transfer learning are important
Zoubir et al. (2024)	CNN + Gabor + attention segmentation	IoU 60.62%; F1 74.49%	Boundary-sensitive evaluation is necessary
Wang et al. (2025)	SBDNet steel-bridge segmentation	Average IoU 76.8%; geometric error below 3%	Masks should support quantitative measurements

CONCLUSION

This study formulated a real-time early bridge-crack inspection framework that combines YOLOv10 localization with deep-learning-based image segmentation. The architecture addresses two complementary requirements: rapid scanning of large bridge images and precise pixel-level representation of crack geometry. Evidence from recent bridge-inspection studies shows that one-stage detectors provide strong operational speed, whereas segmentation is indispensable when maintenance decisions require crack width,

length, area, continuity, or branching information. YOLOv10 is particularly attractive as the localization stage because its NMS-free end-to-end design reduces post-processing complexity and supports efficient deployment.

The analysis also demonstrates that model accuracy cannot be separated from deployment conditions. Small cracks, noisy concrete textures, shadows, repaired surfaces, viewpoint variation, and image compression can substantially alter detection and segmentation quality. Therefore, a credible validation protocol must include bridge-level data separation, thin-crack subsets, cross-device testing, end-to-end latency, and calibrated geometric error. Attention, multi-scale fusion, and selective boundary refinement should be introduced only when ablation results justify their additional computational cost. This principle protects the real-time objective while maintaining engineering usefulness.

Future work should implement the proposed architecture on a publicly documented and locally acquired bridge dataset, train YOLOv10 and the segmentation decoder under identical experimental controls, and validate physical measurements against manual crack gauges or calibrated photogrammetry. Additional research should investigate temporal crack tracking, uncertainty calibration, active learning for rare defects, and optimized deployment on UAV or edge accelerators. A field-tested version could then connect image-based detections with bridge-management systems so that repeated observations support condition trends and maintenance prioritization rather than isolated visual predictions.

REFERENCES

1. Sugiyono. (2022). Metode penelitian kuantitatif, kualitatif, dan R&D (2nd ed., 29th printing). Alfabeta.
2. Zhang, J., Qian, S., & Tan, C. (2022). Automated bridge surface crack detection and segmentation using computer vision-based deep learning model. Engineering Applications of Artificial Intelligence, 115, 105225. https://doi.org/10.1016/j.engappai.2022.105225
3. Yu, L. (2022). Intelligent crack detection and quantification in the concrete bridge: A deep learning-assisted image processing approach. Advances in Civil Engineering, 2022, 1813821. https://doi.org/10.1155/2022/1813821
4. Lu, G., He, X., Wang, Q., Shao, F., Wang, J., & Jiang, Q. (2022). Bridge crack detection based on improved single shot multi-box detector. PLOS ONE, 17(10), e0275538. https://doi.org/10.1371/journal.pone.0275538
5. Zhang, J., Qian, S., & Tan, C. (2023). Automated bridge crack detection method based on lightweight vision models. Complex & Intelligent Systems, 9, 1639–1652. https://doi.org/10.1007/s40747-022-00876-6
6. Tran, T. S., Nguyen, S. D., Lee, H. J., & Tran, V. P. (2023). Advanced crack detection and segmentation on bridge decks using deep learning. Construction and Building Materials, 400, 132839. https://doi.org/10.1016/j.conbuildmat.2023.132839
7. Tabernik, D., Šuc, M., & Skočaj, D. (2023). Automated detection and segmentation of cracks in concrete surfaces using joined segmentation and classification deep neural network. Construction and Building Materials, 408, 133582. https://doi.org/10.1016/j.conbuildmat.2023.133582
8. Liu, Y., Zhou, T., Xu, J., Hong, Y., Pu, Q., & Wen, X. (2023). Rotating target detection method of concrete bridge crack based on YOLO v5. Applied Sciences, 13(20), 11118. https://doi.org/10.3390/app132011118
9. Marleny, F., Fitriansyah, M., Sa'adah, Saputri, W. A. N., Ansari, R., & Mambang. (2023). Segmentasi citra keretakan dinding beton menggunakan teknik

perbandingan evaluasi metrik. TEMATIK, 10(1), 28–33. https://doi.org/10.38204/tematik.v10i1.1261

10. Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). YOLOv10: Real-time end-to-end object detection. Advances in Neural Information Processing Systems, 37, 107984–108011. https://doi.org/10.52202/079017-3429
11. Xiong, C., Zayed, T., & Abdelkader, E. M. (2024). A novel YOLOv8-GAM-Wise-IoU model for automated detection of bridge surface cracks. Construction and Building Materials, 414, 135025. https://doi.org/10.1016/j.conbuildmat.2024.135025
12. Zoubir, H., Rguig, M., El Aroussi, M., Saadane, R., & Chehri, A. (2024). Pixel-level concrete bridge crack detection using convolutional neural networks, Gabor filters, and attention mechanisms. Engineering Structures, 314, 118343. https://doi.org/10.1016/j.engstruct.2024.118343
13. Xiong, C., Zayed, T., Jiang, X., Alfalah, G., & Abelkader, E. M. (2024). A novel model for instance segmentation and quantification of bridge surface cracks—The YOLOv8-AFPN-MPD-IoU. Sensors, 24(13), 4288. https://doi.org/10.3390/s24134288
14. Yang, Y., Li, L., Yao, G., Du, H., Chen, Y., & Wu, L. (2024). An modified intelligent real-time crack detection method for bridge based on improved target detection algorithm and transfer learning. Frontiers in Materials, 11, 1351938. https://doi.org/10.3389/fmats.2024.1351938
15. Ma, K., Hao, M., Meng, X., Liu, J., Meng, J., & Xuan, Y. (2024). Coarse–fine combined bridge crack detection based on deep learning. Applied Sciences, 14(12), 5004. https://doi.org/10.3390/app14125004
16. Zhang, T., Qin, L., Zou, Q., Zhang, L., Wang, R., & Zhang, H. (2024). CrackScopeNet: A lightweight neural network for rapid crack detection on resource-constrained drone platforms. Drones, 8(9), 417. https://doi.org/10.3390/drones8090417
17. Jiang, S., Zou, X., Yang, J., Li, H., Huang, X., Li, R., Zhang, T., Liu, X., & Wang, D. (2024). Concrete bridge crack detection method based on improved YOLO v8s in complex backgrounds. Journal of Traffic and Transportation Engineering, 24(6), 135–147. https://doi.org/10.19818/j.cnki.1671-1637.2024.06.009
18. Wu, Y., Shi, J., Ma, W., & Liu, B. (2024). Bridge crack recognition method based on YOLOv5 neural network fused with attention mechanism. International Journal of Intelligent Information Technologies, 20(1), 1–25. https://doi.org/10.4018/IJIIT.361575
19. Wang, X., Yue, Q., & Liu, X. (2025). SBDNet: A deep learning-based method for the segmentation and quantification of fatigue cracks in steel bridges. Advanced Engineering Informatics, 65, 103186. https://doi.org/10.1016/j.aei.2025.103186
20. Feng, M., & Xu, J. (2025). Lightweight dual-attention network for concrete crack segmentation. Sensors, 25(14), 4436. https://doi.org/10.3390/s25144436
21. Yao, Y., Zhang, Y., Liu, Z., & Yuan, H. (2025). A bridge crack segmentation algorithm based on fuzzy C-means clustering and feature fusion. Sensors, 25(14), 4399. https://doi.org/10.3390/s25144399
22. Jin, Y., Liu, H., Cheng, X., & Du, C. (2025). Crack segmentation in civil infrastructure using conditional modulation and hierarchical dual-path networks. Engineering Structures, 343, 121162. https://doi.org/10.1016/j.engstruct.2025.121162

23. Chu, H., Chen, W., & Deng, L. (2025). Refined segmentation of high-resolution bridge crack images via probability map-guided point rendering technique. Computer-Aided Civil and Infrastructure Engineering, 40(27), 4946–4969. https://doi.org/10.1111/mice.70088