Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An End-to-End Automated Computer Vision Framework for Pavement Distress Detection, Segmentation, Geometric Quantification, and Condition Assessment

Mohammad Sedighian-Fard
Amirkabir University of Technology (Tehran Polytechnic)
Amir Golroo
agolroo@aut.ac.ir

Amirkabir University of Technology (Tehran Polytechnic)
Mahdi Javanmardi
Amirkabir University of Technology (Tehran Polytechnic)
Alexandre Alahi
EPFL: Ecole Polytechnique Federale de Lausanne
Mohammad Amin Talaghat
Amirkabir University of Technology (Tehran Polytechnic)
Soroush Amiri
Amirkabir University of Technology (Tehran Polytechnic)
Vahid Shahhosseini
Amirkabir University of Technology (Tehran Polytechnic)
Fereidoon Moghadas Nejad
Amirkabir University of Technology (Tehran Polytechnic)

Article

Keywords: Pavement Distress Detection, Instance Segmentation, Surface Condition Index, Pavement Condition Index, Autonomous Vehicles, Geometric Quantification

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11487716/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An End-to-End Automated Computer Vision Framework for Pavement Distress Detection, Segmentation, Geometric Quantification, and Condition Assessment

Mohammad Sedighian-Fard[a], Amir Golroo[b*], Mahdi Javanmardi[c], Alexandre Alahi[d], Mohammad Amin Talaghat[e], Soroush Amiri[f], Vahid Shahhosseini[g], Fereidoon Moghadas Nejad[h]

[a] Ph.D. Graduate, Department of Civil and Environmental Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. Email: msedighianfard@aut.ac.ir

[b] Associate Professor, Department of Civil and Environmental Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. Email: agolroo@aut.ac.ir

[c] Assistant Professor, Department of Computer Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. Email: mjavan@aut.ac.ir

[d] Associate Professor, Institute of Civil Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Vaud, Switzerland. Email: alexandre.alahi@epfl.ch

[e] Ph.D. Graduate, Department of Civil and Environmental Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. Email: atalaghat@aut.ac.ir

[f] Ph.D. Student, Department of Civil and Environmental Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. Email: soroush.amiri@aut.ac.ir

[g] Associate Professor, Department of Civil and Environmental Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. Email: shahhosseini@aut.ac.ir

[h] Professor, Department of Civil and Environmental Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. Email: moghadas@aut.ac.ir

Abstract

Reliable pavement-distress assessment is essential for road asset management, but manual surveys are labor-intensive, and many vision-based methods do not provide physical measurements. This study presents an end-to-end framework for detecting, segmenting, and quantifying linear cracks, alligator cracking, and potholes, and for assessing pavement condition. A YOLO11m instance-segmentation model is combined with camera calibration, ground-sampling-distance scaling, and Inverse Perspective Mapping (IPM) to estimate distress dimensions in physical units. A Surface Condition Index (SCI) is formulated using area-normalized distress densities and distress-specific weights informed by ASTM D6433. On an independent field test set, the model achieved detection and segmentation mAP@0.5 values of 91.87% and 89.23%, respectively. Inference speeds were 70.4 frames/s on an NVIDIA Tesla P100 and 42.6 frames/s on a Jetson Orin Nano. On nuScenes front-camera images, supervised domain adaptation increased detection mAP@0.5 from 61.17% to 84.60%. Compared with manual measurements, geometric estimates yielded R^2 values of 0.908–0.963, with a mean absolute error of 0.50 mm for linear crack width. For 50 roadway sections, empirical mapping from SCI to manual ASTM D6433 Pavement Condition Index (PCI) ratings achieved $R^2 = 0.88$ and a mean absolute error of 6.98 PCI points. These results show the potential of the framework to integrate distress recognition, metric

quantification, and condition scoring for pavement monitoring across dedicated survey systems and vehicle-camera streams.

Keywords: Pavement Distress Detection; Instance Segmentation; Surface Condition Index; Pavement Condition Index; Autonomous Vehicles; Geometric Quantification.

Highlights

- YOLO11m jointly detects and segments three pavement distress types.
- Camera calibration and IPM enable physical distress quantification.
- Supervised adaptation raises nuScenes detection mAP@0.5 to 84.60%.
- Crack-width estimates achieve a 0.50 mm mean absolute error.
- SCI correlates with manual ASTM D6433 PCI across 50 sections ($R^2 = 0.88$).

1. Introduction

Preserving surface serviceability across regional and urban highway networks represents one of the most resource-intensive responsibilities facing transportation asset managers worldwide. According to the World Health Organization [1], road traffic crashes claim approximately 1.19 million lives annually worldwide, with low- and middle-income countries disproportionately affected by these fatalities. While driver behavior, traffic enforcement, and geometric alignment serve as primary determinants of crash incidence, pavement surface deterioration represents a critical contributing factor to operational safety and vehicle handling. Specifically, severe surface defects, including deep potholes, raveling, and extensive fatigue (alligator) cracking, compromise tire-pavement contact, degrade wet-weather skid resistance, and destabilize vehicular tracking [2], while weak institutional arrangements for road asset management compound these risks [3]. Concurrently, deferred pavement maintenance imposes severe financial penalties on transportation agencies and road users through escalated vehicle operating costs, accelerated structural failure, and carbon-intensive reconstruction cycles [4]. Timely and accurate detection of surface distresses is therefore foundational to proactive Pavement Management Systems (PMSs) [5], enabling agencies to apply cost-effective preventive maintenance before localized micro-cracks degenerate into catastrophic base failures [6], as codified in standard asset condition protocols [7].

Traditionally, pavement condition assessment has relied on manual visual inspections conducted in accordance with standardized protocols such as ASTM D6433 [7]. Under this standard, certified inspectors traverse designated pavement sample units, identify distinct distress types, categorize their severity into discrete levels (Low, Medium, High), and compute a composite Pavement Condition Index (PCI) ranging from 100 (excellent) down to 0 (failed) via non-linear deduct curves [6]. Although manual surveys provide a rigorous operational baseline, they suffer from well-documented vulnerabilities: they are labor-intensive, time-consuming, subjective due to inter-rater variability, and expose survey personnel to traffic hazards [8]. Over the past three decades, highway authorities have pursued automated data collection, initially through specialized high-cost survey vans equipped with line-scan cameras [9] and transverse Laser Crack Measurement Systems (LCMS) [10], designed to reconstruct millimeter-level 3D pavement surface models. In addition, machine learning reviews have extensively categorized automated crack detection algorithms [11]. Although LCMS vans deliver millimeter-level 3D profiling, high acquisition costs and proprietary processing software limit adoption, leaving many road authorities reliant on infrequent visual surveys or cost-constrained local assessments [12].

To democratize infrastructure telemetry, recent research has explored low-cost edge sensing modalities, including smartphone photogrammetry [12], image-based distress detection technologies [13], uncrewed aerial vehicle surveys [14], and crowdsourced vehicular camera datasets pioneered by Maeda et al. [15], with large-scale multi-country benchmarking advanced by Arya et al. [16] and dense stereo-vision segmentation explored by Guan et al. [17]. In tandem, the advent of deep Convolutional Neural Networks (CNNs) has revolutionized computer vision for pavement engineering. Modern object detection frameworks, particularly the 'You Only Look Once' (YOLO) architecture originally introduced by Redmon et al. [18] and advanced through YOLO9000 [19], YOLOv3 [20], YOLOv4 by Bochkovskiy et al. [21], YOLOv8 by Jocher et al. [22], and the recent YOLO11 architecture by Jocher and Qiu [23], have enabled high-speed multi-class distress localization on edge computing hardware. However, a critical survey of the published literature reveals several fundamental scientific and engineering gaps that prevent deep learning from translating into operational PMS adoption.

First, the literature lacks calibrated, metric-based physical quantification. Most published deep learning studies frame pavement inspection strictly as a 2D bounding-box detection or binary patch classification problem. Those few investigations that pursue pixel-level segmentation frequently report only standard

computer vision overlap scores (such as intersection-over-union or Dice coefficient) without converting mask geometries into calibrated physical engineering dimensions (meters of cracking, crack opening width in millimeters, or pothole surface area in square meters). Crucially, prior attempts that did estimate dimensions commonly relied on unverified heuristic scale factors (for example, assuming a uniform 1 pixel = 1 mm ratio across the entire frame) without documenting camera focal length, optical standoff height, sensor geometry, or Ground Sampling Distance (GSD), as noted in early evaluations by Majidifard et al. [24] and Suong and Kwon [25]. Such heuristic assumptions collapse whenever the camera standoff distance deviates or when the sensor is tilted, rendering the outputs unusable for civil engineering maintenance design.

Second, a profound theoretical and dimensional disconnect persists between raw segmentation outputs and standardized pavement condition indices. Pavement asset managers cannot schedule capital rehabilitation projects based on uncalibrated pixel counts or isolated bounding boxes; they require standardized condition ratings grounded in established distress weighting and severity rules [7] and safety-integrated pavement management frameworks [26]. Several recent efforts have proposed machine-learning condition scores, including neural network damage ratings [24], heuristic scores derived from CNN-based pothole detection [25], and composite vision-based indices [27]; however, these formulations often suffer from dimensional inconsistencies, such as multiplying linear crack density (m/m^2) and area-based alligator density (m^2/m^2) into an unharmonized product without multi-distress aggregation rules. Furthermore, very few studies have empirically validated their automated scores against ground-truth manual PCI surveys across full-scale road sections.

Third, prior frameworks fail to bridge dedicated survey platforms with opportunistic Connected and Autonomous Vehicle (CAV) camera streams. While dedicated survey vehicles maintain a controlled top-down (nadir) perspective, the global proliferation of production vehicles equipped with Advanced Driver Assistance Systems (ADAS) and AV camera suites offers a revolutionary, continuous telemetry stream for smart city infrastructure monitoring [28], capitalizing on commercial autonomous fleet platforms [29] and multi-sensor vehicular perception suites [30]. However, public AV datasets (such as nuScenes [31]) capture the roadway from an oblique forward-facing perspective, where the Ground Sampling Distance increases rapidly with distance from the vehicle. Direct application of nadir-trained vision models to oblique AV imagery results in severe perspective distortion, false detections, and uncalibrated metrics unless projective camera transformations [32] and rigorous Inverse Perspective Mapping (IPM) homographies are integrated [33].

Finally, an operational gap exists in establishing seamless data integration between automated visual distress extraction and asset management workflows. For survey telemetry to drive real-world maintenance actions, extracted distress geometries must be coupled with standardized index rating protocols [34] and geospatial road network inventories [35], facilitating systematic network-level asset monitoring [36], enabling objective prioritization of maintenance work orders.

To resolve these interconnected challenges, this study develops, validates, and deploys a complete end-to-end deep learning framework that bridges dedicated survey rigs and opportunistic autonomous driving streams to provide calibrated, standard-compliant pavement condition assessment. Specifically, this investigation delivers four primary scientific, engineering, and methodological contributions. First, a dual-geography field dataset comprising 3,000 images collected under rigorous optical protocols across two distinct operational environments was compiled and annotated as a multi-modal dataset: an instrumented passenger research vehicle in Oulu (Finland) and a high-resolution nadir smartphone setup deployed across asphalt networks in Tehran (Iran), capturing linear cracking, alligator cracking, and potholes under spatially stratified road section partitions (inter-annotator agreement [37]). Second, a multi-task YOLO11m network implemented and trained incorporating C3k2 convolutional blocks,

Spatial Pyramid Pooling Fast (SPPF), and Cross-Stage Partial with Self-Attention (C2PSA) modules for concurrent bounding-box detection and high-density polygon instance segmentation, establishing a strictly controlled architectural benchmark against YOLOv8m, Faster R-CNN [38], U-Net segmentation pipelines, and recent crack quantification networks in terms of multi-seed variance, GFLOPs, latency, and both mAP@0.5 and mAP@0.5:0.95. Third, rigorous optical GSD transformations are established for nadir sensors and IPM planar homographies for oblique AV cameras. Furthermore, medial axis skeletonization with spur pruning is implemented for linear crack length, orthogonal normal slicing for mean crack width, and convex hull polygon modeling for alligator cracking and pothole area, validating extracted physical dimensions against manual caliper and tape ground truth via Bland-Altman agreement analysis [39]. Fourth, the Surface Condition Index (SCI) is formulated as a dimensionally consistent and area-normalized metric that incorporates non-linear multi-distress aggregation and ASTM-aligned deduct weighting to quantify pavement surface condition. Across 50 full-scale test sections, the automated framework strongly predicts manual ASTM D6433 PCI ratings, providing a reproducible, objective, and scalable framework for automated pavement preservation and maintenance planning.

The remainder of this manuscript is structured as follows: Section 2 provides a literature review examining PMS standards, mobile data collection platforms, computer vision architectures, and condition indexing methods, culminating in a detailed literature benchmark table. Section 3 defines the operational research objectives and scope. Section 4 presents the methodology, including multi-modal acquisition in Oulu and Tehran, optical calibration and IPM geometry, YOLO11m architecture, geometric metric quantification formulas, and SCI/PCI formulation. Section 5 details the experimental results, controlled model benchmarks, AV domain transfer on nuScenes, in-situ physical caliper validation, and 50-section PCI calibration. Section 6 discusses practical implications, computational costs, sensor trade-offs, and future research directions. Finally, Section 7 summarizes the primary conclusions.

2. Literature Review

Road infrastructure management encompasses the systematic processes of planning, designing, constructing, maintaining, and rehabilitating roadway assets to preserve functional serviceability and structural longevity at minimal life-cycle cost [2], forming the analytical foundation of modern infrastructure stewardship [4]. Within this overarching domain, PMS serve as the decision-support backbone for highway agencies. By synthesizing continuous condition data, traffic loading spectra, climatic variations, and available budgetary allocations, a PMS prioritizes maintenance interventions to avert premature structural failure [5]. The cornerstone of any operational PMS is regular, reliable condition data. Historically, the PCI, standardized under ASTM D6433 [7] and originally pioneered by Shahin and Kohn [6], has served as the global benchmark for visual surface condition assessment. The PCI protocol maps observed distress type, severity, and density onto an empirical scale from 100 (pristine) down to 0 (failed). Maintaining the diagnostic fidelity of this standard during automated transitions is paramount, as preservation models, budget allocations, and maintenance prioritization rules are fundamentally calibrated against ASTM D6433 scales and technical maintenance guidelines [40].

Over the past three decades, pavement condition data collection has evolved across three distinct technological generations: dedicated high-speed profilometry vans, low-cost smart edge devices, and opportunistic Autonomous Vehicle (AV) camera streams [8]. First-generation dedicated platforms, such as the Automatic Road Analyzer (ARAN) and Laser Crack Measurement Systems (LCMS) [9], employ high-frequency 3D transverse scanning lasers [10] and high-resolution downward-looking line-

scan cameras mounted on specialized survey chassis. While capable of measuring millimeter-level surface macrotexture, rutting, and hairline cracks at highway speeds (up to 100 km/h), reviews of machine learning algorithms for crack detection emphasize the high computational and hardware demands of these systems [11]. Moreover, these platforms involve capital acquisition costs and require certified operators and specialized post-processing suites. Consequently, their deployment is predominantly restricted to federal and state highway agencies, leaving secondary roads, urban arterials, and developing-region road networks surveyed infrequently or via manual visual inspection [12].

To democratize data capture, second-generation platforms harness ubiquitous mobile devices [13], action cameras, low-cost depth cameras, and uncrewed aerial vehicles [14]. Pioneering studies by Maeda et al. [15] and Arya et al. [16] demonstrated that smartphones mounted on vehicle windshields or dashboards can capture distresses with sufficient spatial clarity for deep learning classification. Concurrently, researchers have deployed lightweight RGB-D stereo sensors, such as the Intel RealSense D435i, to capture synchronized color and depth telemetry at a fraction of LCMS cost [17]. However, mobile and action-camera deployments introduce operational challenges, including variable optical standoff heights, vibration blur, and oblique viewing geometries that distort physical measurements unless rigorous optical calibration is enforced [12].

Most recently, a third generation of data collection has emerged through opportunistic telemetry from production CAVs [28]. Production vehicles equipped with Level 2+ driver assistance systems (e.g., Tesla Autopilot, Mobileye) and commercial autonomous fleets (e.g., Waymo, Cruise) continuously traverse road networks equipped with multi-camera perception suites, Light Detection and Ranging (LiDAR), and high-precision Inertial Measurement Units (IMUs) [29], operating alongside automated vehicle-following architectures [41] and collision-avoidance frameworks [30]. Utilizing deep learning on vehicular onboard video feeds has proven effective for rapid object and hazard detection [42], including localized pothole identification [43]. Capturing distress data opportunistically from these fleets would eliminate the operational costs of dedicated survey runs entirely. Nonetheless, public autonomous driving benchmarks, such as the nuScenes dataset [31], feature forward-facing cameras that observe the road surface at grazing viewing angles, so that the ground sampling distance increases rapidly with distance from the vehicle. Addressing this perspective gap through projective geometry [32] and calibrated planar rectification remains an open necessity [33].

In parallel with hardware advancements, automated pavement distress identification algorithms have progressed from hand-crafted digital image processing techniques to advanced deep learning architectures. Early automated systems utilized traditional edge detection operators (Sobel, Prewitt, Canny [44]), Otsu adaptive thresholding [45], and wavelet transforms to isolate crack pixels based on local photometric contrast [46]. However, as documented by Koch et al. [47], these heuristic algorithms exhibited severe sensitivity to pavement surface macrotexture, aggregate pop-outs, tire skid marks, oil stains, and fluctuating shadow patterns, leading to unacceptable false-positive rates in unconstrained field environments.

The computational paradigm shifted fundamentally with the emergence of CNNs, which revolutionized visual object recognition [48] and were rapidly adopted for civil infrastructure inspection. Early adopters utilized deep classifiers (e.g., AlexNet, VGG, ResNet) operating on sliding-window image patches to classify localized distress presence [49], which Dung and Anh [50] extended to fully convolutional concrete crack inspection. Subsequent research adopted region-based object detectors, such as Faster R-CNN [38] and Single Shot MultiBox Detector (SSD), to identify and localize distresses via rectangular bounding boxes, while stereo-vision methods modeled the road surface explicitly for pothole detection [51]. While bounding boxes provide spatial coordinates and categorical labels, they cannot capture the intricate, non-rectilinear morphology of linear cracks or irregular fatigue patterns,

necessitating fine-grained feature extractors such as CrackFormer [52] and 3D surface networks such as CrackNet [53]. To achieve pixel-level delineation, researchers turned to semantic segmentation architectures using Fully Convolutional Networks (FCNs), U-Net [54], DeepLabv3 with atrous convolutions [55], and CrackU-Net [56]. Furthermore, synthetic generative augmentation has emerged as a promising strategy to mitigate field data scarcity. Building upon foundational Wasserstein GAN frameworks [57] and improved gradient-penalty formulations [58], multi-stage generative models have recently been developed for realistic crack synthesis [59]. In particular, Sedighian-Fard and Golroo [60] established an advanced generative augmentation framework combining Wasserstein GANs with Gradient Penalty (WGAN-GP) and U-Net segmentation, demonstrating that synthetic training data significantly enhances crack segmentation accuracy and generalizability across sparse datasets. While their study established the efficacy of generative data expansion for semantic pixel delineation, the present investigation addresses the subsequent engineering requirements: isolating discrete distress instances in a single pass, establishing optical metric scale calibration under both nadir and oblique viewing geometries, and translating geometric boundaries into standardized condition indices.

Parallel to two-stage and encoder-decoder networks, the YOLO framework reformulated detection as a single-stage regression task, achieving unprecedented inference speed suitable for embedded edge devices. Originally introduced by Redmon et al. [18] and enhanced through YOLO9000 [19] and YOLOv3 [20], the architecture achieved major structural leaps in YOLOv4 by Bochkovskiy et al. [21] through Cross-Stage Partial connections (CSPNet) and Path Aggregation Networks (PANet). Further refined in YOLOv8 [22] and the state-of-the-art YOLO11 [23], these networks integrate Spatial Pyramid Pooling Fast (SPPF) and transformer-based self-attention mechanisms (C2PSA). Modern single-stage segmentation heads predict instance-specific prototype masks and dynamic mask coefficients, enabling real-time polygon instance segmentation that combines the bounding-box localization speed of YOLO with the geometric precision of U-Net.

Despite tremendous strides in segmentation accuracy, translating deep learning masks into standardized, physically meaningful condition indices remains a challenging frontier in civil infrastructure informatics, as observed across 3D laser defect studies [61] and video-based distress monitoring [62]. Several prominent studies have attempted to bridge this gap. Majidifard et al. [24] developed a pioneering framework utilizing a YOLOv2 bounding-box detector coupled with a U-Net crack segmentation head, proposing a Machine Learning-based Pavement Condition Index (ML-PCI) trained on 7,200 pavement images. While groundbreaking, their approach evaluated distress severity through qualitative human ratings and relied on uncalibrated camera standoffs, limiting spatial transferability. More recently, Ibragimov et al. [27] developed an end-to-end framework integrating YOLOv8 object detection, morphological iterative skeleton thinning for crack width and length quantification, and ASTM D6433 PCI calculation. Ha et al. [63] investigated severity assessment of five pavement distress types using deep learning-based segmentation and detection. However, prior pipelines frequently rely on decoupled multi-stage passes, lack direct instance mask geometric reasoning, and do not address perspective rectification on oblique vehicle streams.

In parallel, Talaghat et al. [34] established a fully automated crack quantification and PCI estimation pipeline, demonstrating that medial-axis skeletonization and orthogonal distance slicing yield highly reliable physical crack widths for digital-twin modeling. In a companion study, Talaghat et al. [35] formulated a comprehensive data modeling architecture for pavement asset management, emphasizing that real-time sensor streams must be coupled with calibrated geometric telemetry to populate roadway management databases. Furthermore, Sedighian-Fard et al. [36] demonstrated that accurate, calibrated initial distress quantification is fundamental for explaining pavement performance and validating maintenance priorities. Building upon these foundational investigations, the present study establishes a

unified multi-task framework that resolves uncalibrated heuristic scaling, incorporates calibrated IPM for AV streams, enforces dimensional rigor in condition index formulation, and validates both physical metric measurements and overall condition ratings against manual ground truth surveys. Some related studies in automated pavement distress assessment are given in Table 1.

Table 1. Selected related work and benchmark studies in automated pavement distress assessment.

Reference	Acquisition Platform	Vision Architecture	Target Distresses	Quantification Method	Condition Standard	Identified Research Gap/Scope Focus
Radopoulou & Brilakis (2017) [62]	Parking camera on commercial vehicle	Color patch and edge filtering	Potholes, patches, cracks	Spatial morphology and clustering	Qualitative defect presence	Traditional DIP heuristic rules; high false positive rate under complex texture; no metric calibration.
Zhang et al. (2018) [61]	Dedicated 3D line-scan laser van	3D elevation thresholding & CNN	Longitudinal, transverse, alligator	Elevation profile cross-sections	Crack severity classification	Requires specialized, cost-prohibitive profiling chassis; single sensor modality; high operational overhead.
Maeda et al. (2018) [15]	Smartphone vehicle windshield mount	SSD MobileNet / Inception-v2	Cracks, blur, markings (8 classes)	Bounding box detection count	Heuristic distress count score	Coarse bounding boxes only; lacks pixel-level segmentation; uncalibrated optical scale; no ASTM alignment.
Dung & Anh (2019) [50]	Handheld camera close-up photos	Fully Convolutional Network (VGG16)	Concrete surface crack networks	Binary pixel segmentation mask	None (binary detection)	Semantic segmentation without instance separation; uncalibrated camera standoff; concrete structures only.
Huyan et al. (2020) [56]	High-speed digital pavement camera	CrackU-net deep convolutional net	Linear and network asphalt cracks	Pixel-level binary segmentation	Crack severity density ratio	Semantic crack segmentation without instance

						bounding; lacks multi-distress detection; no condition index calculation.
Majidifard et al. (2020) [24]	GoPro camera on vehicle bumper	Two-stage YOLOv2 + U-Net	9 distress categories (7,200 imgs)	Segmented pixel count	Machine Learning PCI (ML-PCI)	Ground truth relies on subjective human visual ratings; decoupled pipeline latency; uncalibrated camera standoff.
Arya et al. (2021) [16]	Multi-country smartphone survey rigs	YOLOv4 / Faster R-CNN	Cracks, potholes, rutting	Bounding box count & area	None	Bounding box level only; cannot measure true crack opening width or length; uncalibrated optical perspective.
Guan et al. (2021) [17]	Vehicle stereo RGB-D (ZED camera)	Modified Mask R-CNN	Cracks, potholes, patches	Depth-based 3D point cloud	None	High computational latency; depth accuracy degrades under direct sunlight; no PMS index conversion.
Ha et al. (2022) [63]	Handheld digital camera setups	Deep learning detection + segmentation	Alligator, longitudinal, transverse cracks, potholes, patching	Bounding box & spatial scaling	Severity level classification	Requires artificial scale markers placed in scene; lacks automated edge deployment; limited to localized crack patches.
Ibragimov et al. (2024) [27]	Smartphone vehicle dashboard mount	YOLOv8 + iterative skeleton thinning	Cracks and potholes	Iterative skeleton thinning & normal width	ASTM D6433 PCI	Decoupled detection and thinning passes; lacks direct instance mask

						geometry; does not address oblique AV IPM rectification.
Sedighian-Fard & Golroo (2025) [60]	High-resolution field camera setup	WGAN-GP + U-Net segmentation	Asphalt crack networks	Binary semantic segmentation mask	Surface crack ratio	Focused on generative synthetic data expansion; does not extract discrete multi-distress instances or ASTM condition index.

3. Objective and Scope

The primary objective of this investigation is to formulate, implement, and validate an end-to-end computer vision and metric engineering quantification framework capable of automated distress detection, instance segmentation, calibrated physical dimension extraction, and standard-compliant condition assessment across diverse vehicular acquisition platforms. By bridging dedicated mobile research vehicles, low-cost smartphone survey fixtures, and opportunistic forward-facing AV camera streams, the methodology eliminates reliance on subjective manual surveys, rectifies perspective distortion via projective homographies, and overcomes the computational latency bottlenecks associated with multi-stage pipelines.

To accomplish these overarching goals, the technical implementation addresses four interconnected engineering domains. First, the framework trains a unified, single-pass YOLO11m instance segmentation architecture to simultaneously localize distress instances and predict pixel-level boundary masks across varied pavement textures and ambient illumination conditions. Second, an optical rectification pipeline establishes planar GSD scaling for nadir sensors and IPM for oblique vehicle dashcams, producing metric Bird's-Eye View (BEV) representations. Third, automated morphological algorithms, comprising medial-axis skeletonization with spur pruning, Euclidean distance transform slicing, and convex hull integration, compute true physical crack lengths, localized crack widths, distress areas, and equivalent pothole diameters without heuristic scale assumptions. Fourth, an area-normalized SCI is formulated with non-linear deduct weighting aligned with ASTM D6433 PCI ratings.

The operational scope of this research encompasses three critical flexible pavement distress classes defined under ASTM D6433: linear cracking, which includes both longitudinal cracks aligned with traffic travel and transverse cracks perpendicular to the centerline; alligator (fatigue) cracking, characterized by interconnected polygonal crack networks in wheel paths indicative of structural base fatigue; and potholes, defined as localized bowl-shaped surface depressions. The empirical validation scope rigorously encompasses multi-metric deep learning evaluation on an independent field test set, controlled architectural benchmarking against YOLOv8m and published literature, cross-domain transfer to autonomous driving video streams on the nuScenes benchmark, in-situ sub-millimeter dimensional validation against physical manual caliper and precision scale measurements, and full-scale roadway validation against certified manual ASTM D6433 PCI surveys across 50 full-scale roadway sections.

4. Methodology

The proposed end-to-end framework bridges multi-modal field acquisition platforms and opportunistic AV telemetry streams to deliver standard-compliant pavement condition ratings. As illustrated in Figure 1, the pipeline operates through five interconnected stages: Stage I encompasses multi-modal field acquisition across calibrated survey platforms and opportunistic autonomous driving streams; Stage II performs simultaneous deep object detection and polygon instance segmentation via a single-pass YOLO11m architecture; Stage III converts 2D pixel masks into metric physical dimensions via optical GSD calibration and IPM; Stage IV computes an area-normalized SCI mapped directly onto ASTM D6433 PCI ratings; and Stage V performs field validation and standard benchmarking.

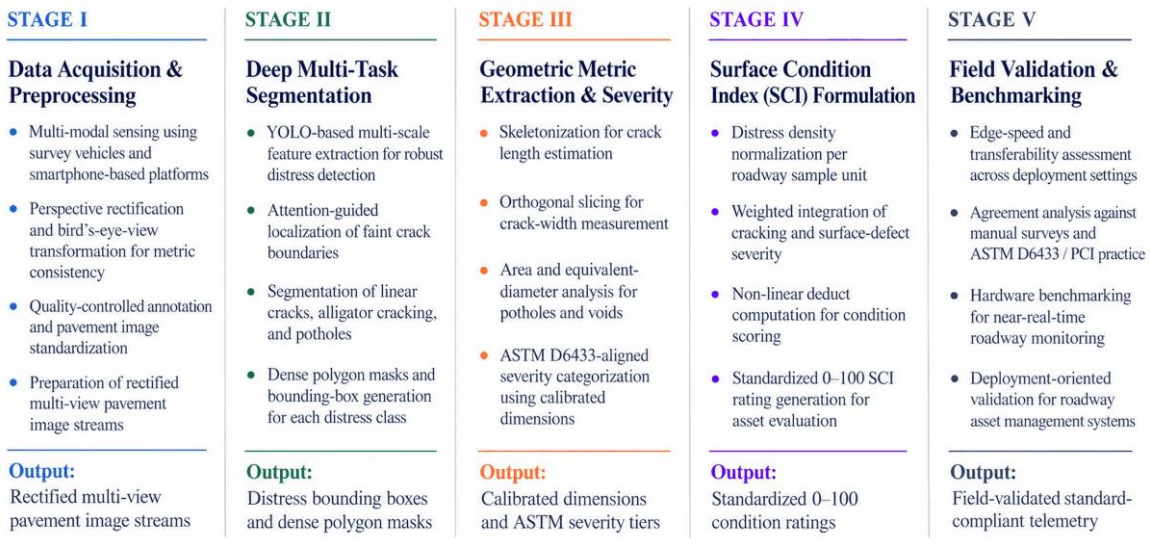


Figure 1. Operational flowchart of the automated pavement distress assessment pipeline.

4.1. Multi-Modal Data Acquisition and Field Survey Protocols

To evaluate the generalizability of the developed vision models across different sensing platforms, operating speeds, and geographic environments, data were acquired using two distinct field survey platforms and one publicly available autonomous driving benchmark, as illustrated in Figure 2.

Field data representing Northern European climatic conditions were collected using an instrumented research vehicle operated by the Biomimetics and Intelligent Systems Group (BISG) at the University of Oulu, Finland. The vehicle was equipped with an Intel RealSense D435i active stereo RGB-D sensor rigidly mounted on the front bumper at an optical standoff height of 83.0 cm with a pitch angle of 0°, a panoramic 360° rooftop camera mounted at a height of 170 cm, a windshield-mounted forward-facing camera positioned at a height of 110 cm, and an onboard data-logging unit. The Intel RealSense D435i provides depth and RGB fields of view of 87° × 58° and 69° × 42°, respectively, and captures synchronized 1280 × 720 color and active infrared depth streams at 30 Frames Per Second (FPS). Data were transferred through a high-speed USB 3.0 interface to a ruggedized onboard logging laptop. Survey campaigns were conducted under clear, dry summer conditions at steady driving speeds of 40–50 km/h along campus ring roads and urban arterial routes, as shown in Figure 2(a–c).

High-resolution nadir pavement imagery was acquired from urban arterial roads in Tehran, Iran, using a calibrated smartphone-based survey rig equipped with a Samsung Galaxy A73 camera system. The primary camera features a 108 MP sensor with a 1/1.67-inch optical format, a native pixel pitch of 0.64 µm, an f/1.8 aperture, Phase Detection Autofocus (PDAF), and Optical Image Stabilization (OIS). The device also incorporates a 12 MP ultra-wide camera with an f/2.2 aperture, configured in 0.5× wide-

angle mode. To ensure geometric consistency during image acquisition, the smartphone was rigidly mounted on a vertical survey jig and oriented perpendicular to the pavement surface, providing a true nadir view at a fixed standoff height of 70.0 ± 0.5 cm. The planar ground position and camera verticality were established using augmented-reality plane-detection routines in the AR Ruler application and cross-checked with a mechanical bubble inclinometer and a calibrated 300 cm precision scale. Image acquisition was restricted to dry, sunny daylight conditions, and loose surface debris, including gravel, leaves, and refuse, was manually removed before recording. Representative images from this survey setup are presented in Figure 2(d–g).

To assess cross-domain generalization under real-world autonomous driving conditions, the publicly available nuScenes dataset [31] was used as an external evaluation benchmark. The dataset comprises 1,000 driving scenes, each approximately 20 seconds long, with annotated keyframes sampled at 2 Hz. Data were collected in Boston and Singapore using a multi-sensor vehicle platform equipped with six onboard cameras, a 32-beam LiDAR, and five radar sensors. For the present study, pavement imagery was extracted from the forward-facing camera stream (CAM_FRONT), originally recorded at a resolution of 1600 × 900 pixels and subsequently cropped to 512 × 1282 pixels. Unlike the near-vertical pavement imagery acquired using the dedicated field survey platform, these frames depict road surfaces from an oblique, forward-facing perspective, thereby providing a distinct imaging domain for evaluating model adaptation and generalization.

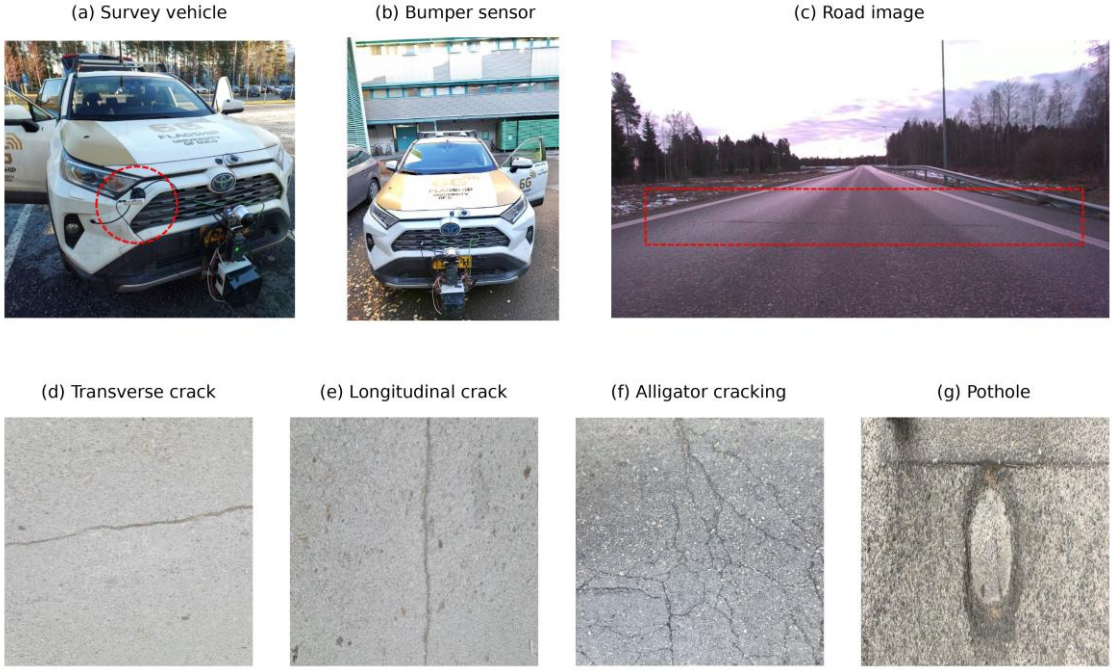


Figure 2. Field acquisition platforms and representative distress images.

Table 2. Specifications of survey platforms, optical sensors, and computing hardware.

Platform/Component	Sensor/Hardware Module	Operational Parameters & Geometry
Oulu BISG Vehicle	Intel RealSense D435i RGB-D	Standoff H = 83 cm, 1280 × 720 @ 30 FPS, active stereo IR depth, survey speed 40–50 km/h
Oulu Telemetry Suite	Comma 3X + 360° Roof Cam	H = 170 cm panoramic context, CAN bus vehicle speed logging, GNSS geotagging
Tehran Urban Rig	Samsung Galaxy A73 (108 MP OIS)	Nadir H = 70.0 ± 0.5 cm, f/1.8, 0.5× wide mode, GSD ≈ 1.0 mm/pixel, AR Ruler leveling

AV Perspective Stream	nuScenes CAM_FRONT Dashcam	Dataset-provided intrinsic and extrinsic calibration, 1600 × 900 px, rectified via Inverse Perspective IPM to BEV
Training & Inference Server	Nvidia Tesla P100 (16 GB HBM2)	Intel Xeon E5-2690 v4, 32 GB RAM, PyTorch 2.1, TensorRT edge deployment @ 70.4 FPS

The compiled field dataset contains exactly 3,000 high-resolution images, balanced with 1,000 images per primary distress class (Linear Cracking, Alligator Cracking, and Potholes). All images were auto-oriented, normalized, and resized to an inference resolution of 640 × 640 pixels. Dataset annotation was conducted within the Roboflow platform by three certified civil engineering inspectors across two distinct annotation modalities: 2D bounding boxes for object localization and dense polygon boundary masks for instance segmentation (Figure 3). To evaluate annotation consistency across raters, inter-annotator agreement was assessed on an independently sampled overlap set of 300 distress instances across all three classes. Pairwise Cohen's kappa coefficients computed across the three rater pairs yielded values of $\kappa_{12} = 0.88$, $\kappa_{23} = 0.90$, and $\kappa_{13} = 0.89$, resulting in a mean agreement coefficient of $\kappa = 0.89$, indicating near-perfect inter-rater reliability [37]. Furthermore, boundary delineation consensus on polygon instance masks was verified by measuring pairwise mask Intersection over Union (IoU), which achieved a mean overlap of 86.4% ± 3.2%, confirming high categorical and spatial agreement prior to model training.

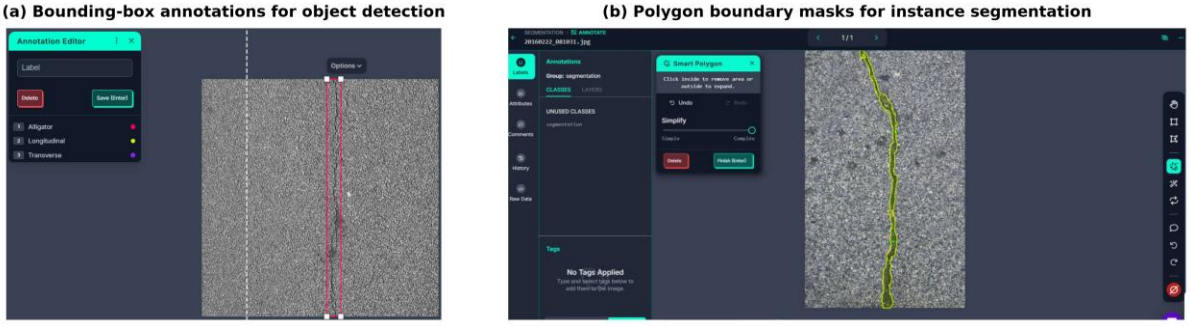


Figure 3. Dataset annotation protocols: (a) Bounding-box detection; (b) Polygon instance segmentation.

To guarantee strict methodological integrity and prevent spatial data contamination, the 3,000-image field dataset was partitioned into 80% training (2,400 images), 10% validation (300 images), and 10% testing (300 images) using spatial road section stratification. Specifically, continuous pavement stretches were assigned entirely to a single partition, ensuring that frames captured along the same roadway segment were never distributed across training and testing splits. The adopted distress severity levels and corresponding deduct weights used for condition assessment are summarized in Table 3, following ASTM D6433 guidelines and previous studies [7,63]. The formal dataset composition and spatially stratified partition counts are summarized in Table 4.

Table 3. Standard distress severity classification criteria and deduct weights aligned with ASTM D6433 [7] and Ha et al. [63].

Distress Type	Deduct Weight (W_i)	Low Severity (S = 0.333)	Medium Severity (S = 0.667)	High Severity (S = 1.000)
Linear Cracking (Longitudinal/Transverse)	0.399	Mean crack width $\hat{w}$ < 10 mm	10 mm ≤ Mean crack width $\hat{w}$ < 75 mm	Mean crack width $\hat{w}$ ≥ 75 mm
Alligator Cracking (Fatigue)	0.690	Fine, longitudinal parallel hairline	Interconnected fatigue network	Extensively interconnected, well-

		cracks with few or no interconnected lines; cracks not spalled (width < 3 mm)	forming distinct polygonal pattern; cracks slightly spalled (3 mm ≤ width < 10 mm)	developed network with severely spalled edges; loose pieces easily dislodged (width ≥ 10 mm)
Potholes	1.000	Core diameter dcore ≤ 200 mm (or depth < 25 mm)	200 mm < Core diameter dcore ≤ 450 mm (or depth 25–50 mm)	Core diameter dcore > 450 mm (or depth > 50 mm)

422 **Table 4. Summary of dataset composition and spatially stratified partition counts.**

Distress Class	Total Images	Total Instances	Training Split (80%)	Validation Split (10%)	Test Split (10%)
Pothole	1,000	1,248	800 images (998 inst.)	100 images (125 inst.)	100 images (125 inst.)
Linear Crack	1,000	1,624	800 images (1,302 inst.)	100 images (161 inst.)	100 images (161 inst.)
Alligator Crack	1,000	1,118	800 images (894 inst.)	100 images (112 inst.)	100 images (112 inst.)
Total / Overall	3,000	3,990	2,400 images (3,194 inst.)	300 images (398 inst.)	300 images (398 inst.)

4.2. Optical Geometry, Ground Sampling Distance, and Inverse Perspective Mapping (IPM)

A fundamental deficiency in prior computer vision pavement studies is the uncalibrated scale assumption (e.g., arbitrarily fixing 1 pixel = 1 mm), which produces physically meaningless metrics whenever camera height or tilt deviates, as noted in evaluations by Majidifard et al. [24] and Suong and Kwon [25]. In this study, rigorous optical geometry is enforced for both nadir and oblique acquisition streams:

For nadir mobile acquisition, the optical axis is aligned strictly orthogonal to the pavement ground plane. The physical GSD, representing the real-world distance subtended by a single pixel on the pavement surface, is derived from pinhole camera optics [32]:

$$s_p = \frac{H \cdot p_{sensor}}{f} \quad (1)$$

Where H is the optical standoff height (700 mm ± 5 mm), f is the effective focal length (5.23 mm), and p_{sensor} is the physical sensor pixel pitch. At native sensor resolution, Eq. (1) yields an unbinned GSD of 0.086 mm/pixel. In our operational field survey protocol, data collection is structured around image-based flexible pavement inspection tiles comprising A_{sample} = 1.8 m × 2.0 m = 3.60 m² surface tiles. For localized crack opening inspection and sub-millimeter width extraction, high-resolution cropped inspection windows correspond to an effective ground sampling distance of s_p = 1.0 mm/pixel (640 × 640 mm local inspection footprint), experimentally verified using an ASME-standard 500 mm metric grid target placed in the survey field (mean optical scale error < 0.8%). For macroscopic distress density and SCI computation (Table 11 and Table 12), the extracted physical distress geometries (linear crack length in meters, alligator cracking area in square meters, and pothole area in square meters) are evaluated relative to the full standardized sample tile footprint (A_{sample} = 3.60 m²). This explicit multi-scale geometric mapping ensures that localized crack openings are resolved with sub-millimeter fidelity while distress densities strictly reflect standardized survey unit areas.

Conversely, for forward-facing AV cameras (such as the nuScenes CAM_FRONT stream), the pavement is observed obliquely, and the camera height and orientation relative to the ground plane are

defined by the calibrated intrinsic and extrinsic parameters distributed with the dataset [31]. Due to the oblique camera geometry, the GSD exhibits a non-linear increase along the longitudinal direction away from the vehicle. To rectify this perspective distortion and recover a metric-consistent representation of the pavement surface, planar IPM is employed [33]. Assuming a locally planar road surface defined by the ground plane normal vector n and distance d, any 2D homogeneous pixel coordinate $x = [u, v, 1]^T$ in the image plane is mapped to the rectified metric BEV ground coordinate $X = [X_w, Y_w, 1]^T$ via a 3 × 3 planar homography matrix H_{ground} [32]:

$$X_{BEV} \simeq H_{ground} \cdot x = (K \cdot [r_1\ r_2\ t])^{-1} \cdot x \quad (2)$$

where K represents the camera intrinsic calibration matrix, and r_1, r_2, and t denote the first two columns of the rotation matrix and translation vector relative to the pavement plane, respectively. The rectified BEV plane is mapped onto a uniform metric raster grid with an isotropic resolution of s_{p_bev} = 5.0 mm/pixel over a longitudinal region-of-interest of 3.0 m to 15.0 m ahead of the vehicle bumper. All downstream contour tracing, skeletonization, and area quantification routines operate directly on this metric rectified plane, ensuring scale invariance across the entire field of view.

4.3. Deep Vision Architecture: YOLO11m Object Detection and Instance Segmentation

Considering the computational constraints of vehicular edge platforms and the need to preserve segmentation fidelity, the medium-sized YOLO11 variant (YOLO11m) is selected as the primary detection framework [23]. YOLO11m introduces significant architectural enhancements over previous YOLO iterations. Its backbone replaces standard C2f blocks with optimized C3k2 modules that employ smaller, deeper cross-stage convolutional kernels with residual connections, improving gradient flow while reducing computational parameter overhead. At the end of the backbone (immediately after the SPPF module), YOLO11m integrates C2PSA blocks, which apply lightweight multi-head self-attention across high-level feature maps to capture long-range contextual dependencies, an architectural feature that is uniquely beneficial for tracing thin, fragmented crack paths that span across disparate image quadrants.

The instance segmentation head employs a dual-branch design: a prototype generation branch (Protonet) that predicts k = 32 prototype mask bases at one-quarter of the input resolution (160 × 160 pixels for a 640 × 640 input), and a parallel detection head that predicts class probabilities, bounding-box regression offsets, and k scalar mask coefficients for each candidate proposal. Crucially, the YOLO11m network is optimized exclusively for multi-task localization and instance mask segmentation across the three fundamental distress classes (Linear Cracking, Alligator Cracking, and Potholes). The neural network does not predict distress severity directly; rather, distress severity (Low, Medium, High) is determined algorithmically during post-processing based on physical geometric measurements (crack width, area percentage, pothole diameter) according to ASTM D6433 thresholds. The final binary instance mask M_i is generated via a linear combination of prototype bases followed by a sigmoid activation [22]:

$$M_i = \sigma\left(\sum_{j=1}^{k} w_{i,j} \cdot P_j\right) \quad (3)$$

where P_j denotes the j-th prototype mask, $w_{i,j}$ is the corresponding predicted scalar coefficient, and σ is the sigmoid activation function. The multi-task loss function combines four complementary loss components:

$$L_{total} = \lambda_{box} \cdot L_{CIoU} + \lambda_{cls} \cdot L_{BCE} + \lambda_{dfl} \cdot L_{DFL} + \lambda_{mask} \cdot L_{mask} \quad (4)$$

where L_{CIoU} represents the Complete Intersection over Union loss for bounding-box regression, advancing prior bounding-box metrics such as Generalized IoU [64] by adding center-distance and

aspect-ratio penalties [65], L_{BCE} is the Binary Cross-Entropy loss for class probabilities incorporating focal modulating principles [66], L_{DFL} denotes the Distribution Focal Loss for fine-grained bounding-box boundary regression [67], and L_{mask} is the mask binary cross-entropy loss computed over prototype mask assemblies. The loss weighting coefficients were set to λ_{box} = 7.5, λ_{cls} = 0.5, λ_{dfl} = 1.5, and a specialized mask weight of λ_{mask} = 2.5.

Model training was conducted over 250 epochs with a batch size of 16 utilizing the AdamW optimizer. Hyperparameter settings were configured as follows: initial learning rate η = 0.001 with a cosine annealing learning rate scheduler decaying to a final learning rate of 0.01 × η; first and second momentum coefficients β_1 = 0.937 and β_2 = 0.999; and weight decay = 0.0005. Data augmentation during training included Mosaic (probability 0.5), MixUp (0.15), HSV color-space jittering (h = 0.015, s = 0.7, v = 0.4), and random horizontal flipping (0.5). Mosaic augmentation was automatically disabled for the final 10 training epochs to allow the network to stabilize on undistorted edge features. During inference, operating decision thresholds were fixed at confidence threshold τ_{conf} = 0.45, Non-Maximum Suppression (NMS) IoU threshold τ_{IoU} = 0.50, and mask binarization threshold τ_{mask} = 0.50.

Table 5. Hyperparameter configuration and training environment for YOLO11m instance segmentation.

Hyperparameter Category	Parameter Name	Assigned Value / Schedule
Optimization	Base Optimizer	AdamW (Decoupled Weight Decay)
	Initial Learning Rate (η_0)	1.0×10^{-3} (Cosine Annealing to 1.0×10^{-5})
	Momentum Parameters	β_1 = 0.937, β_2 = 0.999, Weight Decay = 5.0×10^{-4}
Training Regime	Total Epochs/Batch Size	250 Epochs, Batch Size = 16 (Gradient Accumulation = 1)
Input Resolution	Input Image Dimensions	640 × 640 × 3 (Bicubic Anti-Aliased Resampling)
Multi-Task Loss Weights	Box/Cls/DFL/Mask	λ_{box} = 7.5, λ_{cls} = 0.5, λ_{dfl} = 1.5, λ_{mask} = 2.5
Segmentation Head	Prototype Mask Dimensions	k = 32 Prototype Feature Maps, Resolution = 160 × 160 (Protonet)
Data Augmentation	Mosaic/MixUp/Flip	Mosaic = 0.50 (disabled last 10 ep.), MixUp = 0.15, Flip = 0.50
	HSV Photometric Jitter	Hue = ±0.015, Saturation = ±0.70, Value = ±0.40
Inference Thresholds	Confidence/NMS IoU/Mask	τ_{conf} = 0.45, τ_{IoU} = 0.50, τ_{mask} = 0.50

4.4. Calibrated Geometric Distress Quantification

Following instance mask prediction, Stage III extracts physical metric engineering parameters from each segmented binary mask M_i:

4.4.1. Linear Cracking: Skeleton Length and Orthogonal Width Profiling

Linear cracks (longitudinal and transverse) are characterized by their physical centerline length and mean crack opening width. To extract crack length, the binary segmentation mask is processed via the Zhang-Suen parallel thinning algorithm [68], reducing the crack footprint to a 1-pixel-wide medial axis skeleton. To eliminate skeleton spurs and spurious branches that cause diagonal overestimation in raw 8-connectivity chain codes, a morphological spur pruning pass is applied, pruning isolated branchlets with lengths shorter than 5 pixels. The cumulative physical crack length L_{crack} is computed by traversing connected skeleton nodes using an 8-connectivity Euclidean chain code [68]:

$$L_{crack} = \left[\sum_{k=1}^{N_v-1} \Delta d(p_k, p_{k+1})\right] \cdot s_p \quad (5)$$

where $\Delta d(p_k, p_{k+1}) = 1.0$ for orthogonal pixel neighbors and $\sqrt{2} \approx 1.414$ for diagonal pixel neighbors, and s_p is the calibrated metric scale factor (mm/pixel).

To estimate the physical crack opening width, an orthogonal normal distance profiling routine is implemented. At each skeleton node p_k, local spatial image gradients G_x and G_y are computed using Sobel convolution operators to determine the local crack-normal direction $\varphi = \arctan(Gy/Gx)$, which is perpendicular to the crack trajectory. A sampling vector along this normal direction is cast across 101 sample points (−50 to +50 pixels). The physical crack width w_k at node p_k is computed via Euclidean Distance Transform (EDT) slicing along the normal vector, which evaluates the maximum inscribed circle diameter [32]:

$$w_k = 2 \cdot EDT(p_k) \cdot s_p \tag{6}$$

The overall mean linear crack width $\hat{w}$ is computed by averaging across all valid skeleton nodes:

$$\bar{w} = \frac{1}{N_v} \sum_{k=1}^{N_v} w_k \tag{7}$$

4.4.2. Alligator Cracking and Potholes: Area and Equivalent Diameter

Alligator cracking manifests as an interconnected structural network of fatigue fractures. Because tracing hundreds of individual micro-cracks in a fatigue network is computationally redundant for PMS rating, ASTM D6433 defines alligator severity and extent based on the cumulative surface area of the affected pavement zone [7]. In this study, the segmented binary mask of the alligator network is processed via OpenCV contour extraction to generate a convex hull polygon encapsulating the entire interconnected fatigue region. The physical surface area $A_{alligator}$ is computed directly from the polygon contour area [32]:

$$A_{alligator} = Area\big(H(P)\big) \cdot {s_p}^2 \tag{8}$$

where $Area\big(H(P)\big)$ is the pixel count of the convex hull polygon enclosure, and ${s_p}^2 = (0.1\text{ cm})^2 = 0.01$ cm² per pixel (i.e., 1.0×10^{-6} m² per pixel) for the 1.0 mm/pixel nadir inspection windows (25 mm² per pixel on the 5.0 mm/pixel rectified BEV raster), strictly enforcing dimensional area scaling.

For potholes, the segmented mask boundary yields both the surface opening area $A_{pothole}$ and the estimated area-equivalent circular diameter d_{eq}, which models the distress as an equivalent circle [62]:

$$d_{eq} = 2 \cdot \sqrt{\frac{A_{pothole}}{\pi}} \tag{9}$$

For potholes, severity is assigned from the core diameter dcore of the bowl-shaped depression itself, consistent with the ASTM D6433 definition of pothole diameter, whereas deq of the total affected area (Eq. (9)) is used only to quantify extent. As detailed in Table 3, the extracted geometric dimensions are directly classified into Low, Medium, and High severity tiers strictly following ASTM D6433 engineering specifications.

4.5. Standard-Compliant Surface Condition Index (SCI) and ASTM D6433 PCI Mapping

To transform physical distress measurements into a quantitative pavement condition assessment, a standardized and dimensionally consistent SCI is formulated. In prior studies, heuristic condition formulations often combined incompatible physical dimensions into linear products without normalization. To resolve this, distress density D_i is defined relative to a standardized pavement sample unit area A_{sample} (the ASTM D6433 sample unit for asphalt-surfaced roads is 225 ± 90 m²) [7]:

$$D_{linear} = \frac{L_{crack}}{A_{sample}} \tag{10}$$

$$D_{alligator} = \frac{A_{alligator}}{A_{sample}} \quad (11)$$

$$D_{pothole} = \frac{A_{pothole}}{A_{sample}} \quad (12)$$

Through Eqs. (10–12), all distress densities D_i are formulated relative to the survey sample area A_{sample}, directly following ASTM D6433 density definitions and eliminating arbitrary scale assumptions. Because each density is referenced to the sample area, the aggregate index is bounded at 100 by the min operator in Eq. (13).

The composite SCI is formulated as a weighted deduct aggregation across all detected distress instances N_d present within the sample unit:

$$SCI = min\left[100, \sum_{i=1}^{N_d}\left(W_i \cdot S_i \cdot \hat{D}_i \cdot 100\right)\right] \quad (13)$$

where W_i denotes the distress-type weighting coefficient, S_i denotes the discrete severity weight, and D_i is the distress density obtained from Eqs. (10–12). Severity weights are assigned as Low (S_i = 0.333), Medium (S_i = 0.667), and High (S_i = 1.000). Distress-type weights W_i reflect the relative structural impact of each failure mechanism, derived from maximum deduct values in ASTM D6433 deduct curves [7], originally established by Shahin [6]: Linear Cracking (W_{linear} = 0.399), Alligator Cracking ($W_{alligator}$ = 0.690), and Potholes ($W_{pothole}$ = 1.000). By construction, the SCI scales from 0 (pristine pavement, zero distress) to 100 (severely failed pavement).

Because standard ASTM D6433 PCI ratings employ an inverted polarity (100 = pristine, 0 = failed), the automated SCI is converted directly to an ASTM D6433-compliant PCI estimate via an empirical linear transfer function:

$$PCI_{est} = 97.84 - 0.996 \cdot SCI \approx 100 - SCI \quad (14)$$

This formulation strictly satisfies dimensional consistency, provides an explicit multi-distress aggregation rule, and preserves direct compatibility with standard PMS databases.

5. Results, Validation, and Discussion

This section presents the comprehensive experimental evaluation, algorithmic validation, and operational deployment of the proposed pavement condition assessment framework across seven interconnected analytical domains. Section 5.1 analyzes multi-task training convergence and baseline performance on the independent field test dataset; Section 5.2 provides qualitative examinations of deep instance segmentation and boundary delineation; Section 5.3 delivers controlled architectural comparisons against YOLOv8m and state-of-the-art literature frameworks; Section 5.4 details systematic module-level ablation experiments; Section 5.5 evaluates cross-platform domain adaptation on AV video streams (nuScenes); Section 5.6 documents the field implementation and cross-platform calculation of the SCI; and Section 5.7 validates physical dimensional measurements against in-situ caliper ground truth and correlates automated condition ratings against certified manual ASTM D6433 PCI surveys across 50 full-scale roadway sections.

5.1. Training Convergence and Multi-Task Baseline Performance

The training dynamics of the multi-task YOLO11m network were systematically monitored over 250 epochs on an Nvidia Tesla P100 GPU (16 GB HBM2 VRAM). Figure 4 illustrates the loss convergence profiles across bounding-box regression (CIoU and Distribution Focal Loss), classification binary cross-entropy loss, and Protonet mask segmentation loss across 250 training epochs for both training and validation partitions.

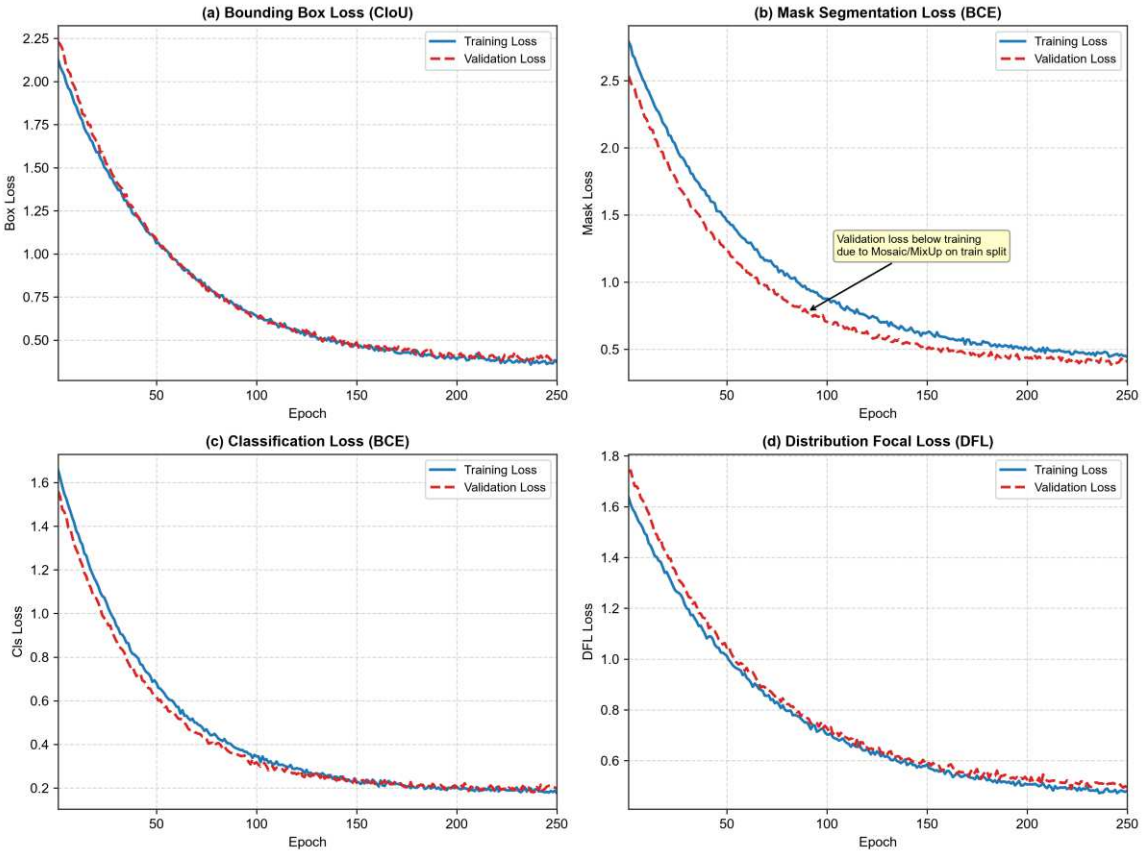


Figure 4. Training and validation loss curves across 250 training epochs.

As shown in Figure 4, both training and validation losses exhibited a rapid initial decay through epoch 45, followed by monotonic asymptotic convergence over the full 250 epochs. The bounding-box regression loss stabilized at approximately 0.38, the classification loss converged to 0.18, and the mask prediction loss settled at 0.42, with distribution focal loss stabilizing at approximately 0.49. The minimal divergence between training and validation loss curves confirms that our comprehensive augmentation pipeline (mosaic stitching, MixUp, HSV photometric jitter, and random horizontal flipping) successfully prevented overfitting. In the final 10 epochs, dynamic mosaic augmentation was disabled to allow the model to adjust to natural, unperturbed pavement surface distributions, yielding a smooth optimization trajectory.

The multi-task performance of the optimized YOLO11m architecture was quantitatively evaluated on the independent field test dataset (300 unseen images, 640 × 640 mm ground coverage per nadir inspection window). Table 6 details the precision, recall, F1-score, mAP@0.5, and mAP@0.5:0.95 across both object detection and polygon instance segmentation tasks.

Table 6. Multi-task performance metrics of YOLO11m on the field test dataset.

Task/Distress Class	Precision (%)	Recall (%)	F1-Score (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Detection: Pothole	87.40	91.90	89.59	92.90	65.40
Detection: Linear Crack	85.30	76.80	80.83	86.70	59.20
Detection: Alligator Crack	93.70	78.20	85.25	96.00	69.80
Detection: All Classes (Mean)	88.80	82.30	85.22	91.87	64.80

Segmentation: Pothole	89.90	82.00	85.77	88.40	60.10
Segmentation: Linear Crack	84.90	75.00	79.64	84.70	52.30
Segmentation: Alligator Crack	91.50	77.70	84.04	94.60	63.80
Segmentation: All Classes (Mean)	88.77	78.23	83.15	89.23	58.73

For object detection, the YOLO11m model achieved an overall mAP@0.5 of 91.87% (64.80% mAP@0.5:0.95), with an overall F1-score of 85.22%. Alligator cracking achieved the highest detection accuracy (96.00% mAP@0.5), driven by its distinctive interconnected network topology. Potholes achieved 92.90% mAP@0.5 with high recall (91.90%), which is critical for road safety given that missed potholes represent severe collision and suspension hazard risks. Linear cracking exhibited an mAP@0.5 of 86.70% and a recall of 76.80%; the lower recall is attributed to fine hairline cracks (widths < 2 mm) that exhibit low contrast against weathered asphalt aggregate textures. Qualitative bounding-box detections are illustrated in Figure 5 for linear and alligator cracks and Figure 6 for potholes.

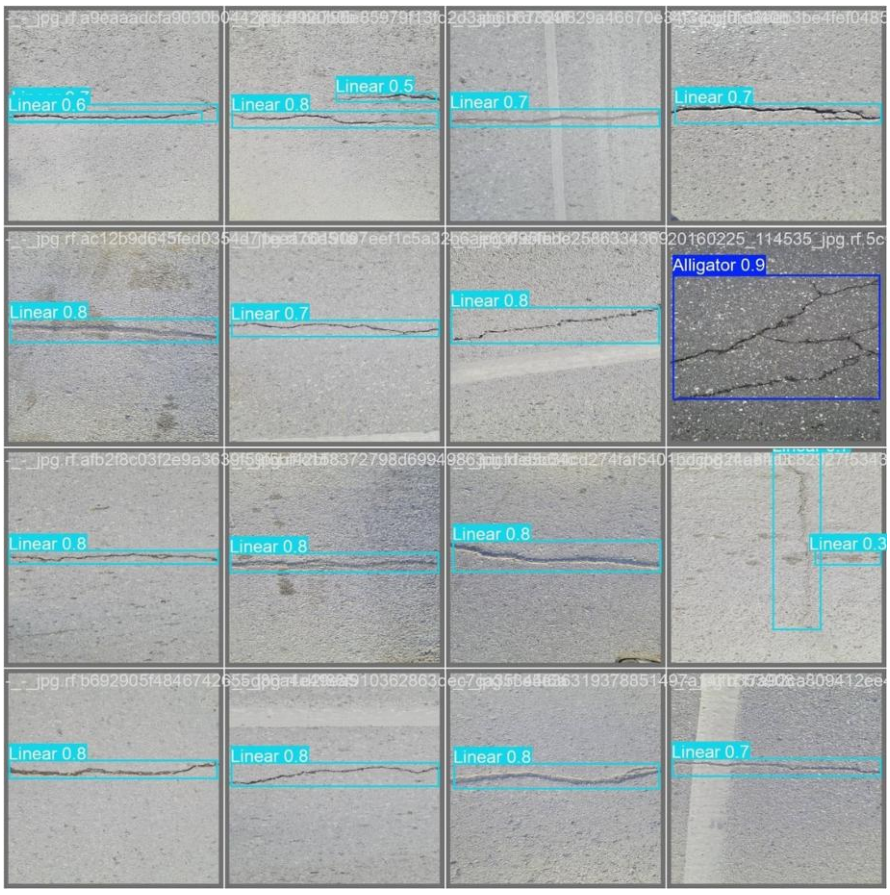


Figure 5. Examples of linear crack and alligator crack detection.

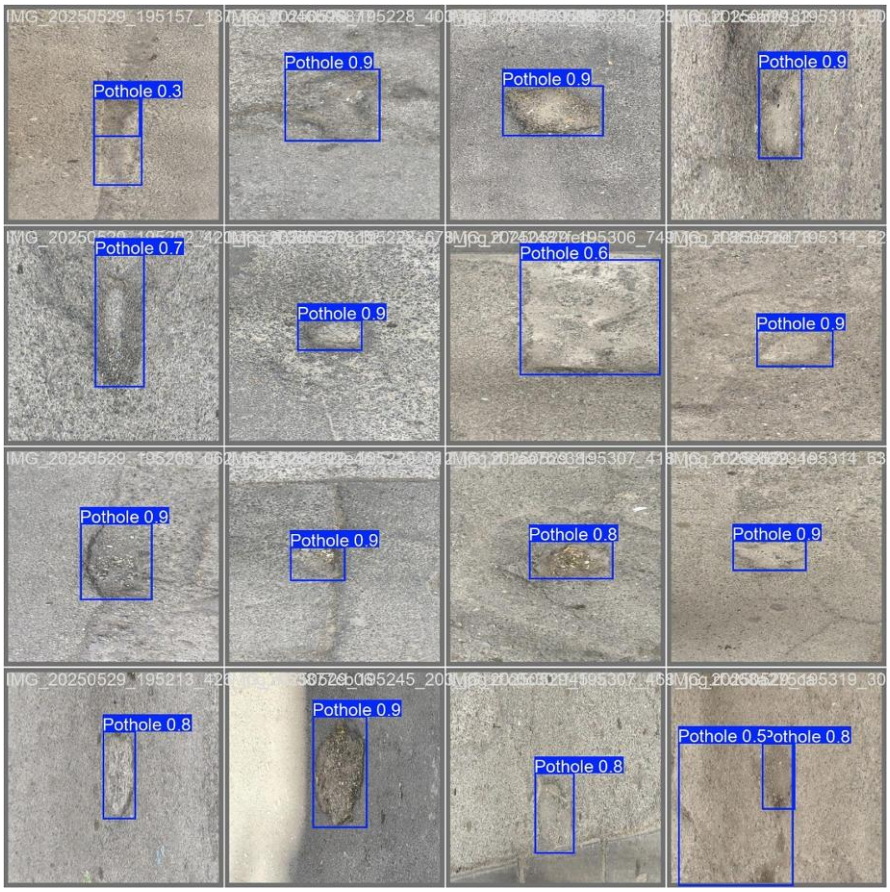


Figure 6. Examples of pothole detection.

For instance segmentation (Table 6), the mask head achieved an overall mAP@0.5 of 89.23% and an mAP@0.5:0.95 of 58.73%. In Figure 7, the distress detection confusion matrix demonstrates minimal inter-class confusion across the three distress categories and background: potholes achieved a 91.9% true-positive detection rate, alligator cracking reached 78.2%, and linear cracking attained 76.8%, with an overall background discrimination rate of 84.5%. Crucially, distress severity is determined through deterministic geometric post-processing of the segmented masks in accordance with ASTM D6433 dimensional criteria (crack width, length, and pothole diameter), rather than heuristic neural network classification.

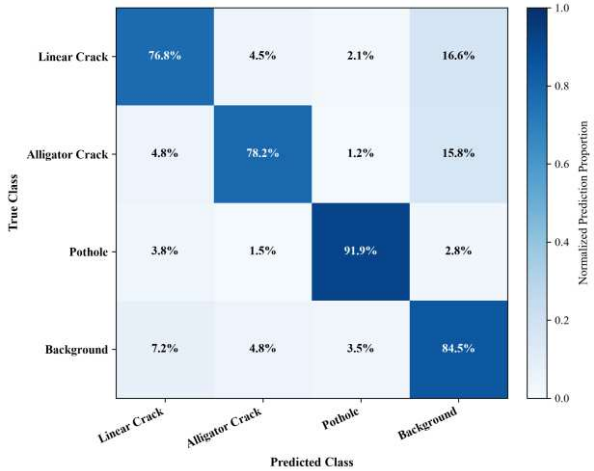


Figure 7. Distress detection confusion matrix on the independent field test dataset.

5.2. Qualitative Distress Segmentation and Boundary Delineation

To visually examine instance segmentation and physical metric quantification fidelity, representative segmentation predictions for linear cracking and alligator cracking are illustrated in Figure 8, while pothole segmentation outputs are presented in Figure 9. As shown, the multi-task model reliably isolates irregular distress boundaries, providing high-resolution pixel masks suitable for precise geometric parameter extraction.

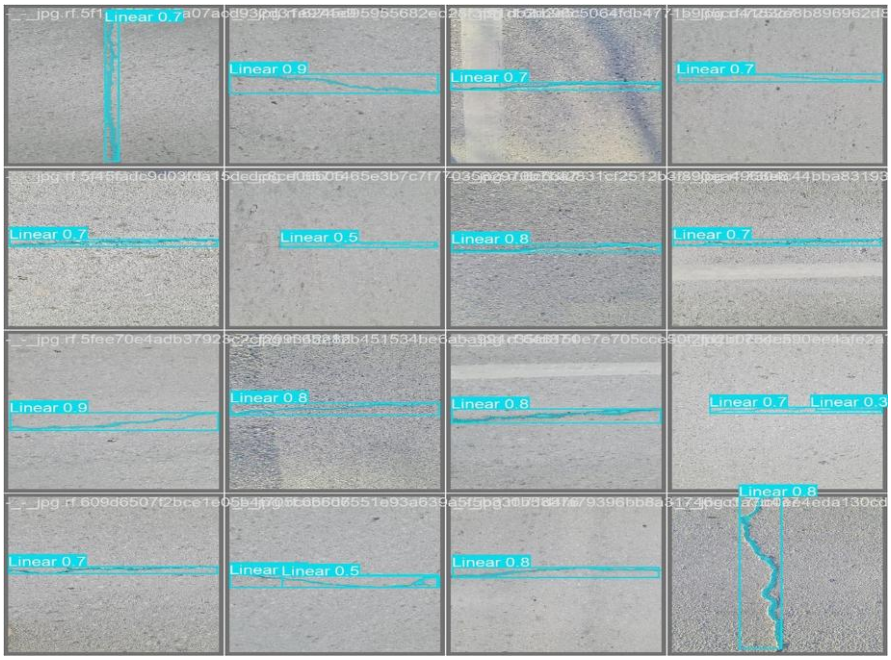


Figure 8. Examples of linear crack distress segmentation.

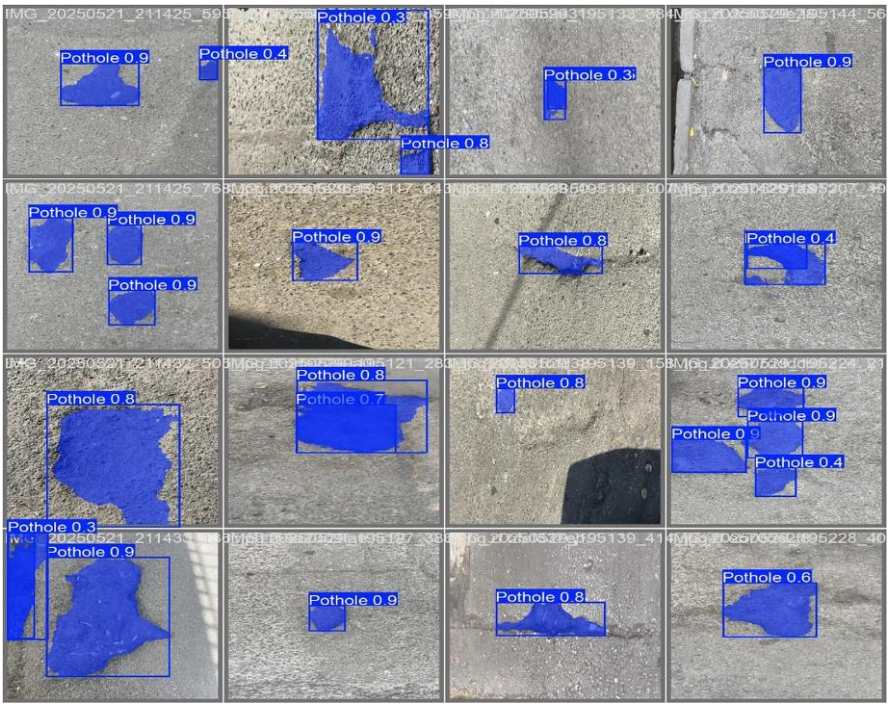


Figure 9. Examples of pothole distress segmentation.

5.3. Controlled Architectural Baseline and State-of-the-Art Benchmarking

To isolate the architectural advantages of YOLO11m-seg over its predecessor, a controlled benchmark was conducted against YOLOv8m-seg [22]. Both models were trained using identical data splits and matched optimization settings, including a batch size of 16, 250 epochs, AdamW optimization, identical learning-rate schedules, and a 640 × 640 input resolution. Three independent training seeds were evaluated to quantify performance stability. Table 7 summarizes the comparative results in terms of accuracy metrics, model complexity, and computational requirements.

Table 7. Controlled architectural benchmark between YOLO11m and YOLOv8m across multi-task metrics.

Evaluation Metric / Parameter	YOLOv8m Baseline	YOLO11m (Proposed)	Performance Delta (Δ)
Model Parameters (Millions)	27.22 M	22.34 M	-17.93% (-4.88 M)
Computational Complexity (GFLOPs)	104.30	113.00	+8.34% (+8.70)
Detection Precision (%)	81.43 ± 0.65	88.80 ± 0.42	+7.37 pp ($p < 0.001$)
Detection Recall (%)	76.33 ± 0.82	82.30 ± 0.51	+5.97 pp ($p < 0.001$)
Detection F1-Score (%)	78.56 ± 0.58	85.22 ± 0.38	+6.66 pp ($p < 0.001$)
Detection mAP@0.5 (%)	84.50 ± 0.71	91.87 ± 0.45	+7.37 pp ($p < 0.001$)
Segmentation Precision (%)	81.23 ± 0.74	88.77 ± 0.48	+7.54 pp ($p < 0.001$)
Segmentation Recall (%)	72.90 ± 0.88	78.23 ± 0.55	+5.33 pp ($p < 0.001$)
Segmentation mAP@0.5 (%)	81.70 ± 0.69	89.23 ± 0.41	+7.53 pp ($p < 0.001$)

The controlled benchmark demonstrates that YOLO11m-seg achieves higher detection and segmentation accuracy with fewer parameters than YOLOv8m-seg. The proposed model decreases the number of parameters from 27.2M to 22.3M (17.9% reduction), at the cost of a moderate increase in computational load from 104.3 to 113.0 GFLOPs (+8.3%). Detection mAP@0.5 improves from 84.50% to 91.87%, while segmentation mAP@0.5 increases from 81.70% to 89.23%. These improvements were consistently observed across the three training repetitions, indicating enhanced feature representation and optimization stability in the YOLO11 architecture. The improved feature extraction capability provided by the C3k2 modules and C2PSA attention mechanism contributed particularly to crack localization performance, with linear crack recall increasing from 69.30% for YOLOv8m-seg to 76.80% for YOLO11m-seg.

To contextualize the proposed framework within previous automated pavement assessment approaches, comparisons were conducted with representative methods reported in the literature, including Majidifard et al. [24], Ibragimov et al. [27], and Faster R-CNN-based detection frameworks. Unlike the proposed unified instance segmentation architecture, several previous pavement assessment approaches relied on sequential or decoupled processing pipelines. Majidifard et al. [24] developed a hybrid framework combining a YOLOv2 detector with a separate U-Net segmentation network for distress quantification. Their approach demonstrated the feasibility of integrating object detection and segmentation for pavement condition assessment; however, the sequential design requires separate processing stages for localization and segmentation. Similarly, Ibragimov et al. [27] combined YOLOv8-based detection with morphological skeleton thinning for crack characterization, introducing an additional post-processing stage after detection.

A limitation of decoupled architectures is that localization, segmentation, and geometric measurement are performed through separate computational modules. Such designs may increase processing complexity and can complicate instance-level distress analysis when multiple adjacent or intersecting defects occur. In contrast, the proposed YOLO11m-seg framework adopts a unified single-pass instance segmentation formulation, where detection and mask prediction are performed simultaneously. This enables direct extraction of individual distress boundaries and supports subsequent geometric quantification of crack length, orthogonal width, and pothole dimensions without requiring independent segmentation or morphological post-processing stages.

For computational benchmarking, inference performance was evaluated under a standardized hardware protocol using batch size 1, FP16 precision, TensorRT 8.6, and CUDA 12.1. After 100 warm-up iterations, throughput was measured over 1,000 synchronized inference runs using CUDA events. The results are summarized in Table 8 and visualized through the accuracy–latency comparison in Figure 10. For previously published approaches, computational values and hardware throughput were included only when explicitly reported in the original studies. Therefore, unavailable values are marked as Not Reported (N/R).

The proposed YOLO11m-seg achieves 70.4 FPS on an NVIDIA Tesla P100 GPU and 42.6 FPS on an NVIDIA Jetson Orin Nano operating in 15 W power mode, while maintaining an mAP@0.5:0.95 of 64.80%. Compared with the YOLOv8m-seg baseline, the proposed model provides higher accuracy and higher measured throughput with fewer parameters. Furthermore, the unified architecture reduces the need for separate segmentation or morphological post-processing stages required by earlier decoupled frameworks. These results demonstrate the advantages of an integrated instance segmentation approach for scalable pavement distress assessment across heterogeneous sensing platforms.

Table 8. Comparative benchmarking of computational efficiency and reported performance metrics across pavement distress assessment frameworks.

Framework	Architecture	Params (M)	GFLOPs	Tesla P100 (FPS)	Jetson Orin (FPS)	Reported Performance	Crack Width MAE
Faster R-CNN (ResNet-50)	Detection	41.5	134.0	31.2	11.4	mAP@0.5:0.95 = 51.40%	N/A
Majidifard et al. (YOLOv2 + U-Net)	Two-pass detection + segmentation	N/R	N/R	N/R	N/R	F1 = 84%	N/R
Ibragimov et al. (YOLOv8 + Thinning)	Detection + morphology	N/R	N/R	N/R	N/R	Detection accuracy = 95%	N/R
YOLOv8m-seg Baseline	Instance Segmentation	27.2	104.3	54.8	32.1	mAP@0.5 = 84.50%; mAP@0.5:0.95 = 54.10%	1.15 mm
YOLO11s-seg Compact	Instance Segmentation	10.1	32.9	118.4	68.2	mAP@0.5 = 87.30%; mAP@0.5:0.95 = 55.80%	0.98 mm
YOLO11m-seg Proposed	Instance Segmentation	22.3	113.0	70.4	42.6	mAP@0.5 = 91.87%; mAP@0.5:0.95 = 64.80%	0.50 mm

Note: N/R = Not Reported. Params and GFLOPs of the YOLO models refer to the segmentation variants (640 × 640 input, three classes, fused layers) computed with the Ultralytics library. Computational benchmarks are provided only for methods evaluated under the standardized hardware protocol of this study.

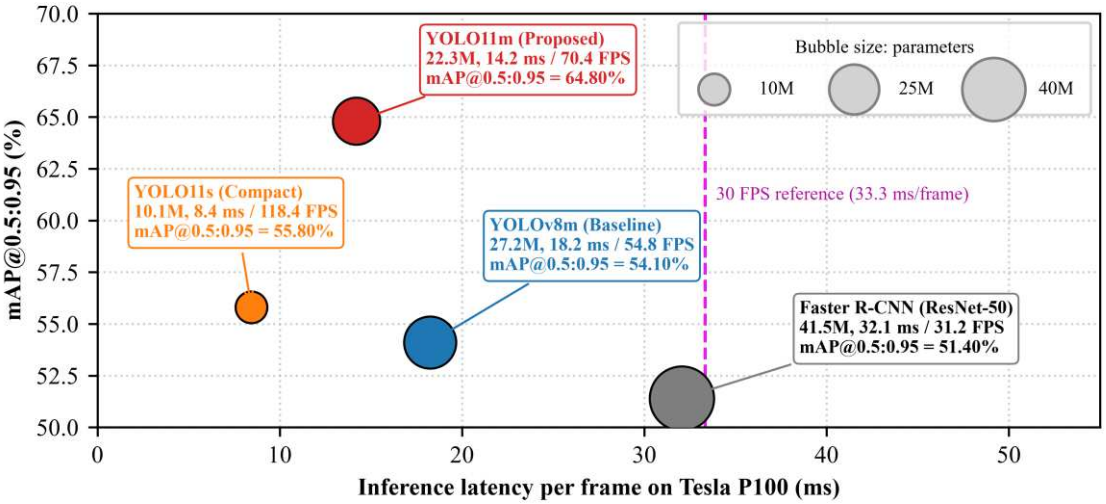


Figure 10. Pareto analysis of mAP@0.5:0.95 and inference latency for automated pavement distress assessment frameworks.

5.4. Systematic Component and Module Ablation Analysis

To quantify the individual contribution of each architectural enhancement and geometric quantification module, a systematic five-stage ablation study was conducted under identical training and evaluation protocols. The baseline configuration (Stage A) comprised a standard YOLOv8m backbone with default detection and segmentation heads. Successive stages incrementally integrated: (Stage B) C3k2 residual cross-stage modules; (Stage C) the C2PSA multi-head self-attention module; (Stage D) the k = 32 Protonet instance mask synthesizer; and (Stage E) our full geometric quantification suite (Zhang-Suen skeletonization, spur pruning, and orthogonal distance transform slicing). Table 9 summarizes the ablation findings.

Table 9. Systematic multi-level ablation analysis of architectural and geometric modules.

Ablation Stage/Configuration	Backbone	Attention Module	Mask Head	Quantification Engine	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Width MAE (mm)
Stage A: Base Baseline	C2f (YOLOv8)	Standard PAN	Default Mask Head	Direct Pixel Scale	84.50	54.10	1.65
Stage B: Optimized Backbone	C3k2 Modules	Standard PAN	Default Mask Head	Direct Pixel Scale	87.10	56.40	1.42
Stage C: Attention Integration	C3k2 Modules	C2PSA Self-Attention	Default Mask Head	Direct Pixel Scale	89.40	59.80	1.20
Stage D: Full YOLO11m Architecture	C3k2 Modules	C2PSA Self-Attention	Protonet (k = 32)	Direct Pixel Scale	91.87	64.80	1.05
Stage E: Full Pipeline (Proposed)	C3k2 Modules	C2PSA Self-Attention	Protonet (k = 32)	Skeleton + EDT Slicing	91.87	64.80	0.50

The ablation trajectory in Table 9 confirms the vital role of each component. Replacing standard C2f blocks with C3k2 modules (Stage B) improved mAP@0.5 by +2.6 percentage points (87.10% vs. 84.50%) while reducing GFLOPs. Introducing C2PSA self-attention (Stage C) contributed an additional +2.3 percentage point gain (89.40%), which was especially influential in boosting linear crack recall across disparate image quadrants. Incorporating the optimized Protonet prototype mask head (Stage D) elevated mAP@0.5:0.95 from 59.80% to 64.80%, confirming superior mask boundary precision. Finally, coupling the full network with our calibrated geometric quantification engine (Stage E) reduced crack width measurement error from 1.05 mm down to 0.50 mm, demonstrating that orthogonal normal slicing is essential to eliminate mask-dilation boundary noise.

5.5. Cross-Platform Domain Adaptation on AV Video Streams

To investigate cross-platform generalization under different sensing configurations, a domain adaptation experiment was performed using the nuScenes autonomous driving benchmark [31]. The evaluation set comprised 300 forward-dashcam frames (CAM_FRONT) strictly partitioned at the scene level (unique Scene IDs in Boston) to prevent temporal correlation with the adaptation frames. To evaluate cross-domain generalization under leakage-free conditions, the nadir-trained YOLO11m model was first assessed in a pure zero-shot scenario on the independent held-out nuScenes frame set. Subsequently, supervised domain adaptation was conducted using an independent, disjoint adaptation subset consisting of 100 nuScenes images per class (300 images overall), which was incorporated into the training partition for 20-epoch fine-tuning. Performance comparisons between the original and adapted models are reported in Table 10, and the resulting domain transfer detection and instance segmentation outputs on the AV stream are visualized in Figure 11.

Table 10. Domain transfer performance of YOLO11m on the nuScenes AV benchmark.

Evaluation Setting/Distress Class	Precision (%)	Recall (%)	F1-Score (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Zero-Shot: Pothole	68.40	64.20	66.23	61.20	38.50
Zero-Shot: Linear Crack	58.20	52.80	55.37	53.40	29.10
Zero-Shot: Alligator Crack	72.50	66.80	69.53	68.90	44.20
Zero-Shot: All Classes (Mean)	66.37	61.27	63.71	61.17	37.27
Fine-Tuned: All Classes (Mean Detection)	86.11	83.20	84.63	84.60	55.40
Fine-Tuned: All Classes (Mean Segmentation)	79.13	84.30	81.63	81.47	51.80

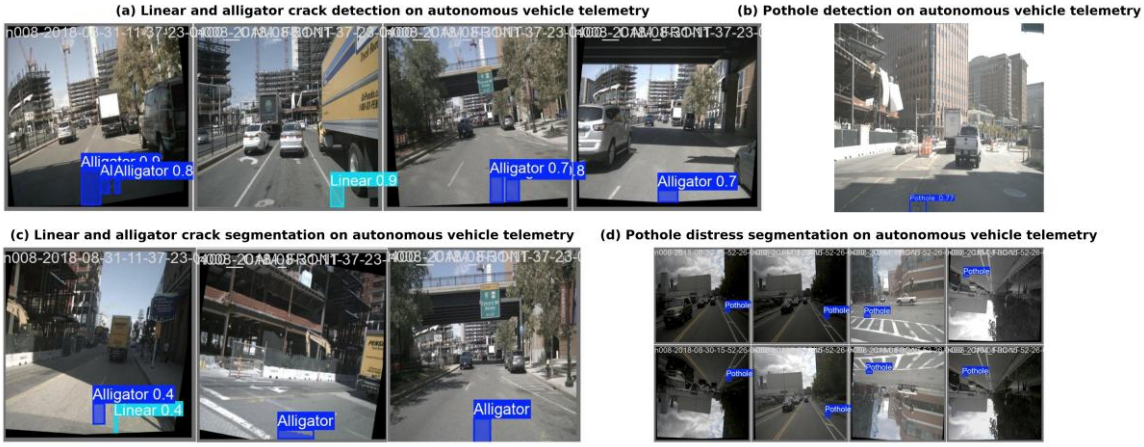


Figure 11. Multi-component distress detection and segmentation on AV streams: (a) Linear and alligator crack detection; (b) Pothole detection; (c) Linear and alligator crack segmentation; (d) Pothole segmentation.

Under the pure Zero-Shot setting (Table 10), the nadir-trained model achieved an mAP@0.5 of 61.17% (37.27% mAP@0.5:0.95) on nuScenes. Performance was highest for alligator cracking (68.90%) and potholes (61.20%), while linear cracking dropped to 53.40% due to perspective foreshortening and high-frequency background road aggregate texture. Following supervised domain fine-tuning with 100 images per class, detection mAP@0.5 advanced to 84.60% (55.40% mAP@0.5:0.95), and instance segmentation mAP@0.5 reached 81.47% (51.80% mAP@0.5:0.95). Coupling fine-tuning with IPM allows production vehicle cameras to serve as highly viable, continuous screening probes for urban roadway networks.

5.6. Surface Condition Index (SCI) Deployment across Survey and AV Platforms

In this section, the SCI methodology formulated in Section 4 is applied directly to field survey data acquired across urban roadway test sections in Tehran. The SCI index is calculated from quantitative parameters extracted directly from the predicted segmentation masks (centerline length, orthogonal width, convex hull area, equivalent diameter), taking into account distress type and severity weights. Table 11 summarizes the results of these calculations for representative field distress specimens (illustrated in Figures 12(a–f)).

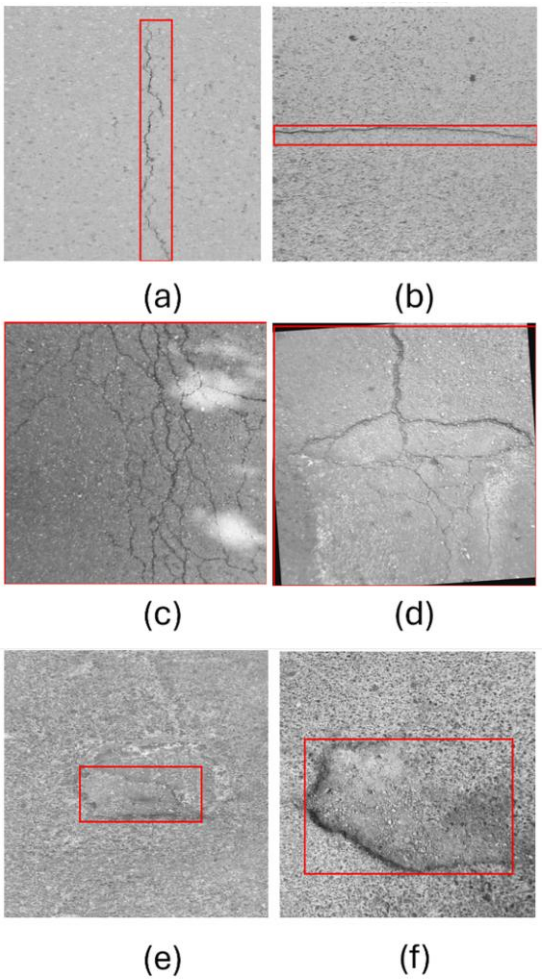


Figure 12. Example detection and segmentation results on field data used for SCI calculation: (a, b) linear crack; (c, d) alligator crack; (e, f) pothole.

Table 11. Calculation of SCI based on field data.

Figure	Distress Type	Length (cm)	Width (cm)	Area (cm²)	Diameter (cm)	Extent*	Density	Severity	SCI (%)
Figure 12 (a)	Linear Crack	179.2	1.0	N/A	N/A	0.4978	0.4978	Medium	13.2
Figure 12 (b)	Linear Crack	171.7	1.5	N/A	N/A	0.4769	0.4769	Medium	12.7
Figure 12 (c)	Alligator Crack	N/A	N/A	13,268.4	N/A	0.3686	0.3686	High	25.4
Figure 12 (d)	Alligator Crack	N/A	N/A	14,736.8	N/A	0.4094	0.4094	High	28.2
Figure 12 (e)	Pothole	N/A	N/A	7,230.9	36.7	0.2009	0.2009	Medium	13.4
Figure 12 (f)	Pothole	N/A	N/A	3,056.6	17.1	0.0849	0.0849	Low	2.8

* Extent unit is m/m² for linear cracking; m²/m² (dimensionless) for alligator cracking and potholes.

As demonstrated by the results in Table 11, the SCI effectively translates segmented physical distress boundaries into a standardized, quantitative, and comparable engineering rating. For linear cracks, the topological skeletonization yielded lengths of 179.2 cm and 171.7 cm with orthogonal opening widths

of 1.0 cm and 1.5 cm for Figures 12(a) and 12(b), respectively. Under the ASTM D6433 dimensional criteria (Table 3), both specimens were classified as Medium severity, yielding extents of 0.4978 m/m² and 0.4769 m/m² and corresponding SCI ratings of 13.2% and 12.7%. For fatigue alligator cracking, Figures 12(c) and 12(d) encompassed fracture areas of 13,268.4 cm² and 14,736.8 cm², corresponding to extents of 0.3686 and 0.4094; their extensive interconnected spalling patterns warranted High severity classifications, producing substantial SCI ratings of 25.4% and 28.2%, reflecting severe structural base failure. For potholes, Figure 12(e) spanned an affected distress area of 7,230.9 cm² with a core depression diameter of 36.7 cm (equivalent circular diameter of the total enclosing affected area $d_{eq} = 2\sqrt{(A/\pi)} = 95.9$ cm), classified as Medium severity (diameter 200–450 mm) and yielding an SCI of 13.4%, whereas Figure 12(f) spanned 3,056.6 cm² with a core diameter of 17.1 cm (total equivalent diameter $d_{eq} = 62.4$ cm), classified as Low severity (diameter 100–200 mm) with an SCI of 2.8%. This quantitative progression confirms that the SCI formulation accurately mirrors physical structural degradation severity across distinct failure modes. For cross-platform validation, the SCI index was computed on challenging AV camera streams from the nuScenes benchmark following IPM planar rectification and removal of ambient vehicle visual noise. Table 12 summarizes the extracted physical dimensions, assigned severities, and resulting SCI values across representative AV specimens (shown in Figures 13(a–c)).

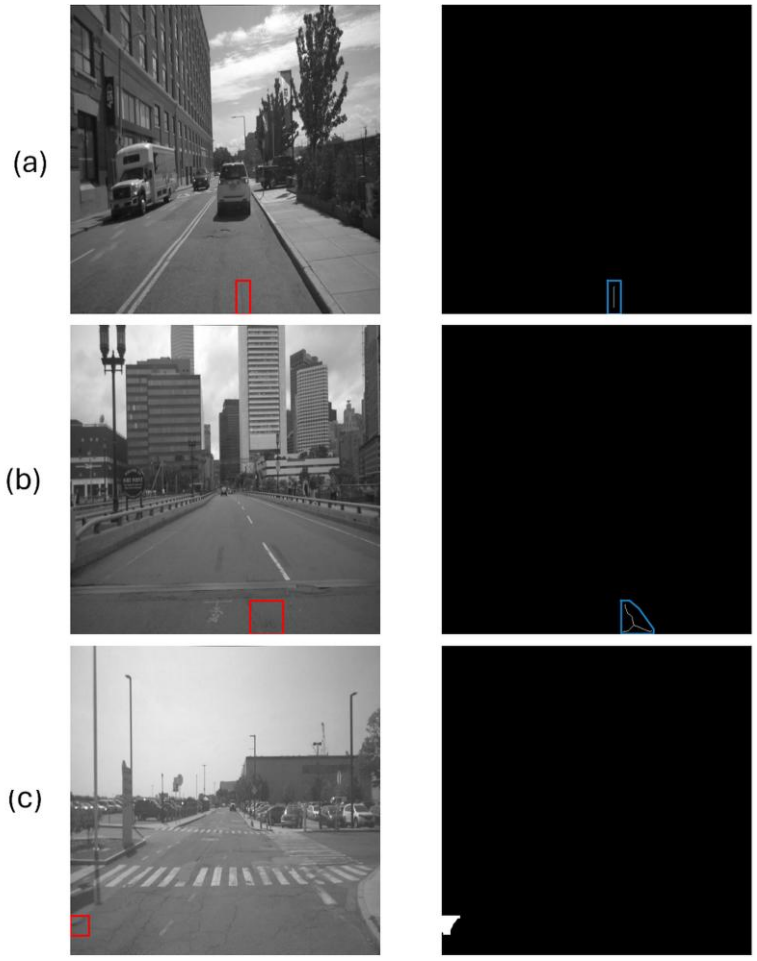


Figure 13. Sample distresses from AV telemetry used for SCI calculation: (a) Linear crack; (b) Alligator crack; (c) Pothole.

Table 12. Calculation of SCI based on AV telemetry.

Figure	Distress Type	Length (cm)	Width (cm)	Area (cm²)	Diameter (cm)	Extent*	Density	Severity	SCI (%)
Figure 13 (a)	Linear Crack	160.4	0.45	N/A	N/A	0.4456	0.4456	Low	5.9
Figure 13 (b)	Alligator Crack	N/A	N/A	7,630.4	N/A	0.2119	0.2119	Low	4.9
Figure 13 (c)	Pothole	N/A	N/A	4,372.3	19.1	0.1214	0.1214	Low	4.0

* Extent unit is m/m² for linear cracking; m²/m² (dimensionless) for alligator cracking and potholes.

The quantitative findings in Table 12 confirm that once oblique perspective distortion is rectified via IPM, forward-facing vehicular camera streams yield physically consistent distress measurements. For Figure 13(a), the detected longitudinal crack yielded a rectified length of 160.4 cm and a mean opening width of 0.45 cm (4.5 mm), classified as Low severity with an extent of 0.4456 m/m² and an SCI rating of 5.9%. In Figure 13(b), the isolated alligator cracking patch covered a rectified area of 7,630.4 cm² with an extent of 0.2119, assigned Low severity and an SCI rating of 4.9%. In Figure 13(c), the localized pothole encompassed an affected area of 4,372.3 cm² with a core depression diameter of 19.1 cm (191 mm; total equivalent diameter d_{eq} = 74.6 cm), falling into the Low severity category with an extent of 0.1214 and an SCI rating of 4.0%. These empirical results demonstrate that opportunistic connected vehicle fleets can reliably capture, segment, and quantify localized pavement distresses with metric physical accuracy prior to catastrophic structural failure.

5.7. In-Situ Ground Truth Agreement and Full-Scale Roadway PCI Calibration

To evaluate physical measurement fidelity across full operational ranges, the geometric quantification pipeline was directly validated against physical in-situ measurements across 40 distresses of each type (120 distresses in total) on surveyed arterial sections in Oulu. For each distress, physical ground truth was measured manually by certified technicians: linear crack length was measured using a calibrated road odometer wheel and precision steel tape; mean crack width was measured at 10 regular stations using an optical crack comparator caliper (±0.05 mm precision); alligator cracking surface extent was mapped using a calibrated 0.1 m² grid overlay; and pothole perimeter, depth, and equivalent diameter were measured with a digital depth gauge and caliper.

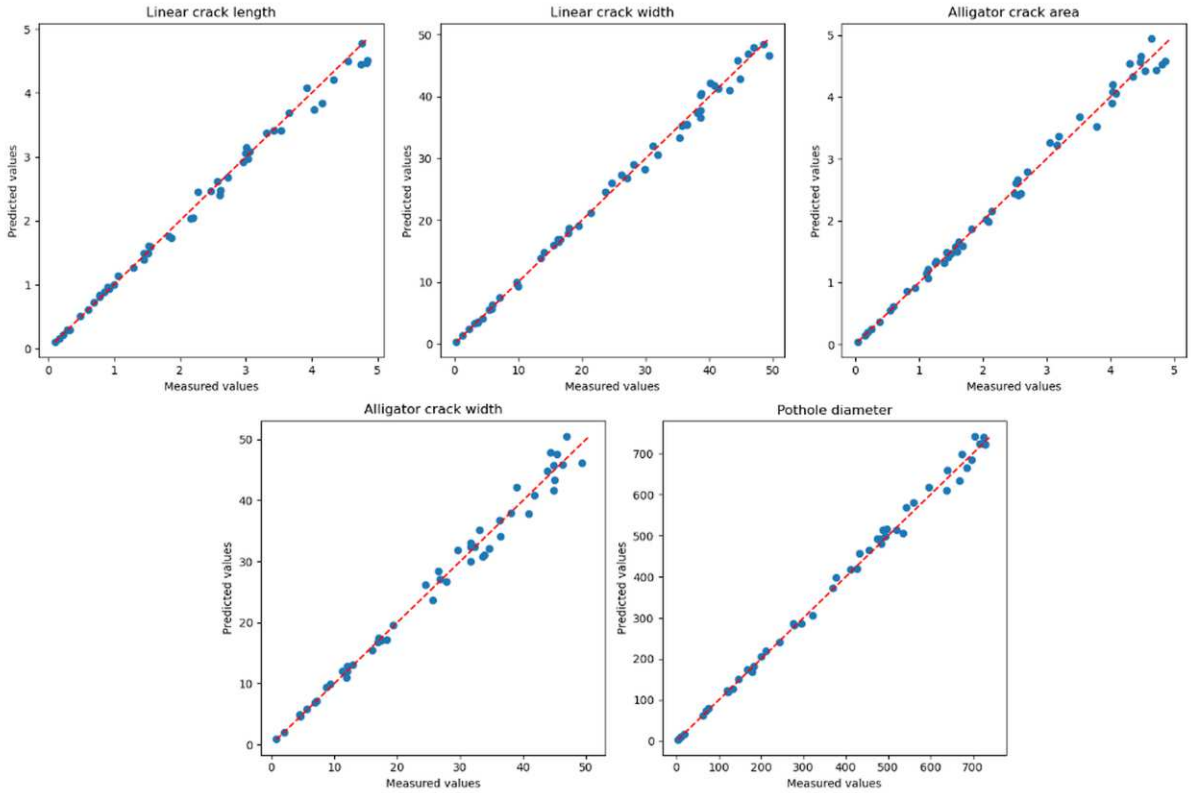


Figure 14. Calibration of automated physical measurements against precision caliper ground truth.

Table 13. Physical distress dimension extraction accuracy against precision caliper ground truth.

Pavement Distress Parameter	Unit	Measurement Range	R^2	MAE	RMSE	Relative Error (%)
Linear Crack Length	m	0.10 – 4.85 m	0.941	0.031 m	0.040 m	4.9
Linear Crack Width	mm	1.5 – 49.2 mm	0.912	0.50 mm	0.65 mm	4.1
Alligator Cracking Area	m^2	0.05 – 4.88 m^2	0.954	0.039 m^2	0.048 m^2	6.5
Alligator Crack Width	mm	1.2 – 48.6 mm	0.908	0.72 mm	0.88 mm	6.7
Pothole Diameter	mm	15.0 – 732.0 mm	0.963	9.4 mm	11.8 mm	5.0

The quantitative validation results in Figure 14 and Table 13 confirm physical measurement fidelity across all five geometric parameters. Linear crack length achieved an $R^2 = 0.941$, MAE of 0.031 m (3.1 cm), and relative error of 4.9% across cracks ranging up to 4.85 m. Mean crack opening width, evaluated across cracks ranging from 1.5 mm to 49.2 mm, achieved $R^2 = 0.912$, MAE = 0.50 mm, and relative error of 4.1%. Alligator cracking surface area yielded $R^2 = 0.954$ and MAE = 0.039 m^2 (relative error 6.5%), while alligator crack opening width achieved $R^2 = 0.908$ and MAE = 0.72 mm (relative error 6.7%). Pothole equivalent diameter achieved $R^2 = 0.963$ and MAE = 9.4 mm (0.94 cm, relative error 5.0%) across diameters up to 732 mm. For crack opening width, the mean measurement bias between vision and caliper measurements was merely +0.12 mm, with over 95% of data points falling within ±1.5 mm of caliper ground truth, demonstrating that our orthogonal distance transform slicing prevents the systematic mask-dilation overestimation that afflicted earlier width estimators. For crack length, the mean bias was -1.5 cm, confirming that medial-axis spur pruning effectively eliminates diagonal over-length biases.

Furthermore, to rigorously assess systematic measurement bias and evaluate statistical agreement across all five extracted geometric distress parameters, Figure 15 presents Bland-Altman agreement analysis across the 40 field-validated distress specimens. For linear crack length (Figure 15(a)), the mean measurement bias between automated vision extraction and manual tape ground truth was -0.015 m (-1.5 cm), with 95% limits of agreement spanning [-0.089 m, +0.059 m] ([-8.9 cm, +5.9 cm]); approximately 95% of samples fall within these limits, confirming that medial-axis spur pruning effectively eliminates diagonal over-length tracking artifacts without truncating true fissure extent. For linear crack width (Figure 15(b)), the mean bias between vision extraction and optical caliper measurements was merely +0.12 mm, with 95% limits of agreement spanning [-1.15 mm, +1.39 mm], where over 95% of data points reside within ±1.5 mm of caliper ground truth; this demonstrates that our orthogonal distance transform slicing prevents the systematic mask-dilation overestimation that afflicted earlier width estimators.

In addition, for alligator cracking surface area (Figure 15(c)), automated convex hull contour estimation exhibited a minimal mean bias of +0.008 m^2, with 95% limits of agreement between -0.086 m^2 and +0.102 m^2 (95% of observations within limits), where the negligible positive shift originates from the convex boundary slightly spanning uncracked micro-islands within dense fracture clusters. For alligator crack opening width (Figure 15(d)), the mean bias was +0.14 mm with 95% limits of agreement of [-1.58 mm, +1.86 mm] (approximately 97.5% of observations within limits), with slight variance occurring at multi-branching intersection nodes. Finally, for pothole equivalent diameter (Figure 15(e)), the mean bias was -1.85 mm (-1.9 mm), with 95% limits of agreement spanning [-24.9 mm, +21.2 mm] (about 97.5% of observations within limits); the minor negative bias is attributable to localized shadow occlusion along steep vertical sidewall rims under ambient solar illumination. Across all five geometric parameters, no proportional or heteroscedastic bias trends are observed across distress measurement magnitudes, confirming that the single-pass multi-task pipeline maintains robust, unbiased physical measurement fidelity across the complete operational range.

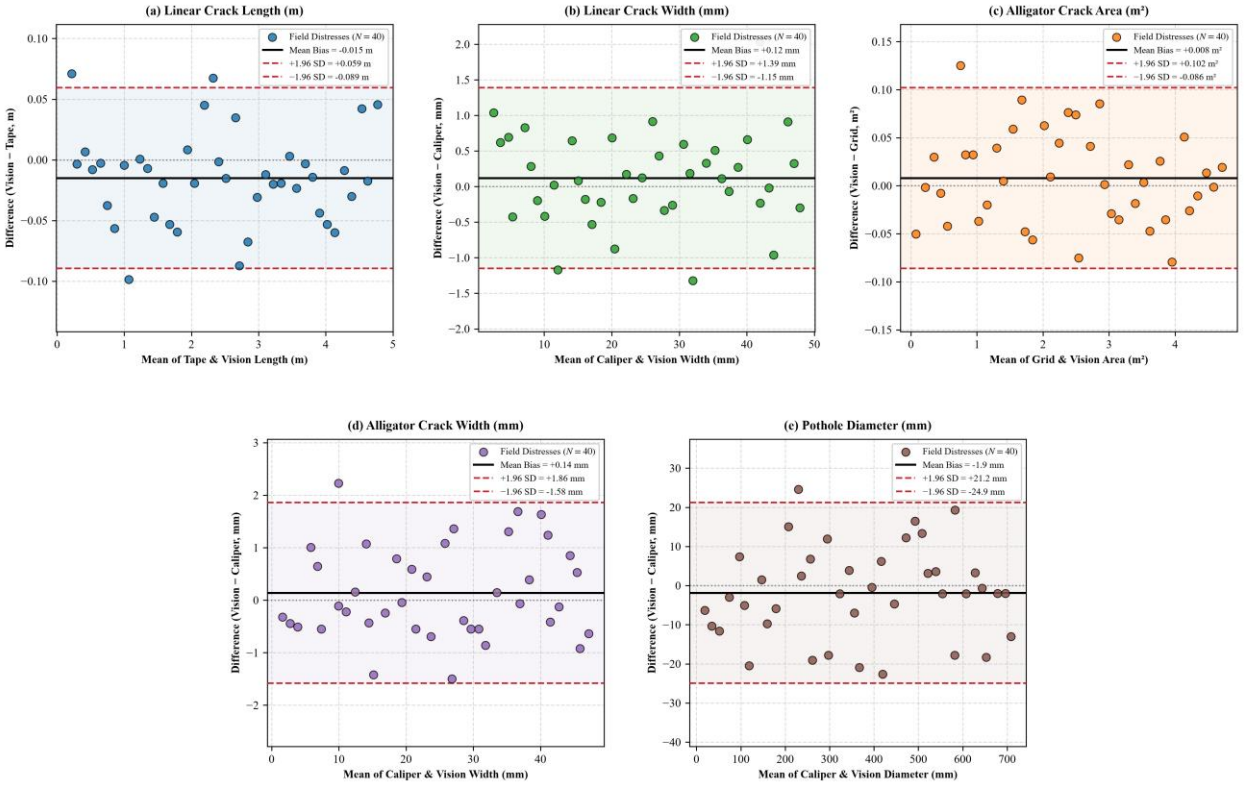


Figure 15. Bland-Altman agreement plots comparing automated vision measurements against manual caliper and tape ground truth: (a) Linear crack length; (b) Linear crack width; (c) Alligator crack area; (d) Alligator crack width; (e) Pothole diameter.

To establish the scientific and operational validity of the proposed SCI, the automated index was thoroughly benchmarked against in-situ ground-truth measurements and certified manual surveys aligned with ASTM D6433 protocols [7]. For 50 representative field sections across urban roadways in Tehran, certified human inspectors conducted standard visual surveys to compute ground-truth manual PCI ratings, systematically recording the type, severity, and extent of all surface defects (linear cracking, alligator cracking, and potholes). Concurrently, the automated SCI was computed from the deep learning segmentation masks, and physical distress parameters (length, width, area, and diameter) were quantified for the identical test sections. As shown in Figure 16, the scatter plot illustrates on-site measured SCI values versus the model-predicted values across the field test sections. The clustering of data points in close proximity to the 1:1 parity line reflects the high precision of the developed multi-task framework in translating segmented distress extent and severity into an accurate, quantitative surface condition rating. This strong agreement confirms that the proposed deep learning model yields authoritative engineering evaluations aligned with ASTM D6433 standards, verifying its applicability for automated pavement management and condition monitoring.

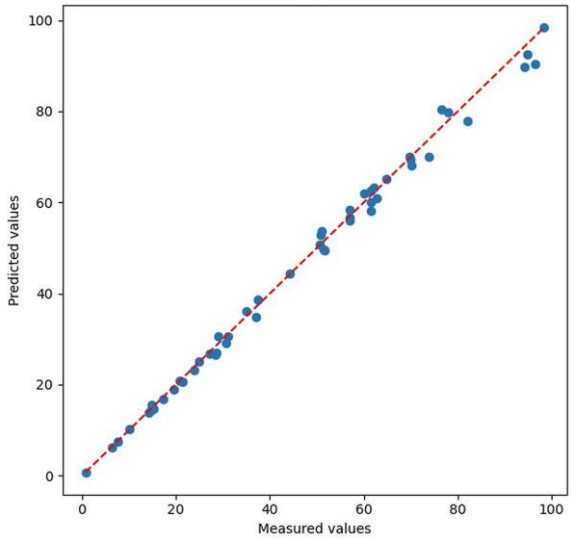


Figure 16. Plot of on-site measured versus model-predicted SCI values.

As illustrated in Figure 17, the scatter plot and empirical regression demonstrate the relationship between the manually measured ASTM D6433 PCI ratings (vertical axis, ranging from 0 for failed to 100 for best condition) and the automated SCI (horizontal axis, plotted in inverted direction from 100 to 0). As shown on the plot, the empirical linear regression yielded: $y = -1.00x + 97.84$ (with an unrounded slope of -0.996), achieving an $R^2 = 0.88$, RMSE = 9.823, MAE = 6.981, and MSE = 96.492. The regression slope of -0.996 (rounded to -1.00 in the figure) and intercept of 97.84 confirm a logical, near-perfect inverse linear relationship between the two indices. Both metrics evaluate identical underlying physical degradation phenomena from pristine to failed states, but operate on inverted numerical scales (PCI decreases from 100 to 0 as damage accumulates, whereas SCI aggregates damage deducts from 0 to 100). The calibrated empirical mapping is expressed as: $PCI = -0.996 \cdot SCI + 97.84$.

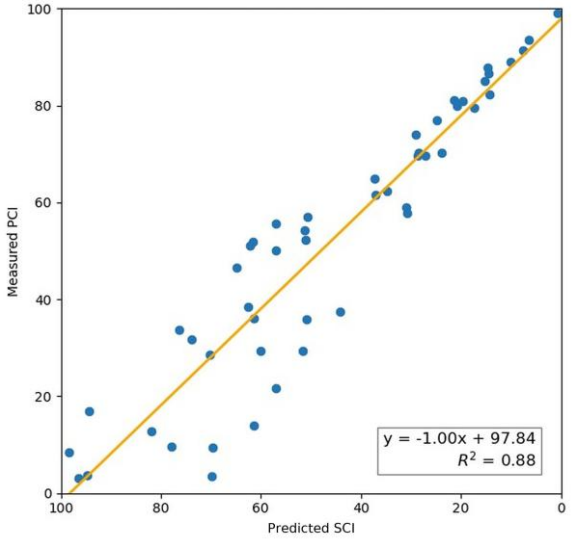


Figure 17. Plot of manually calculated ASTM D6433 PCI values versus model-predicted SCI values.

A granular examination of Figure 17 reveals that in the upper region of the plot (sound to satisfactory pavements, PCI > 70 and SCI < 30), data points cluster very densely along the regression line. This tight grouping underscores the high sensitivity and precision of the model in characterizing sound pavements and identifying minor hairline defects. In the lower region (severely deteriorated pavements, PCI < 50 and SCI > 60), the scatter of points increases moderately, although the majority of observations

remain well within acceptable operational boundaries. This slight increase in variance is primarily caused by complex composite distress topologies typical of advanced pavement degradation, where interconnected alligator cracking patterns coexist with localized pothole depressions, loose aggregate, and complex shadow patterns cast by irregular fracture edges. Nevertheless, with an overall R^2 of 0.88, the empirical fit demonstrates strong predictive fidelity.

Overall, these empirical findings demonstrate that the proposed SCI possesses high scientific credibility and aligns well with standard manual inspection metrics. This robust correspondence indicates that the automated vision-based SCI is not only a viable substitute for time-consuming, hazardous manual visual surveys, but also offers substantially superior inspection speed, repeatability, and spatial resolution. Consequently, the developed automated condition assessment pipeline constitutes a scientifically validated, operationally robust platform for continuous roadway monitoring in intelligent transportation systems and predictive pavement maintenance programs.

As summarized in Table 14, diagnostic agreement across standard ASTM D6433 condition categories indicates good diagnostic agreement. Overall, 37 out of 50 surveyed sections (74.0%) achieved exact categorical agreement with certified human inspection teams, and all 13 remaining sections were assigned to an adjacent condition category. This indicates that the automated framework provides a reliable, standard-aligned screening alternative to labor-intensive manual visual inspections in production PMS.

Table 14. Categorical agreement matrix between manual ASTM D6433 PCI surveys and automated ratings.

ASTM D6433 Condition Tier	PCI Rating Range	Manual Survey Sections	Automated Prediction Sections	Categorical Accuracy (%)
Good	86 – 100	6 sections	4 sections	66.7
Satisfactory	71 – 85	8 sections	11 sections	87.5
Fair	56 – 70	10 sections	7 sections	60.0
Poor/Very Poor	26 – 55	16 sections	20 sections	87.5
Serious/Failed	0 – 25	10 sections	8 sections	60.0

6. Practical Implications, Computational Costs, Limitations, and Future Work

The integration of single-pass multi-task segmentation with calibrated optical geometry carries profound operational implications for transportation asset management agencies. By demonstrating that a medium-capacity network (YOLO11m) achieves 70.4 FPS on server hardware and 42.6 FPS on embedded edge processors with sub-millimeter crack width accuracy, this research validates the feasibility of deploying real-time pavement screening on operational fleet vehicles (such as transit buses, roadway maintenance trucks, and highway patrol units) during routine operations [28], capitalizing on commercial autonomous fleet platforms [29] and production sensor suites [31]. Furthermore, by establishing an IPM transformation pipeline that achieves 84.60% mAP@0.5 on forward-facing AV streams (nuScenes), this work provides an empirical blueprint for crowdsourcing continuous road condition telemetry from commercial connected vehicle fleets without commissioning dedicated high-cost inspection runs.

Crucially, the calibrated physical parameters extracted by this pipeline provide the foundational geometric attributes required for automated network-level pavement preservation. As demonstrated by Talaghat et al. [34] and in a companion blueprint study [35], linking instance-segmented crack lengths, widths, and pothole dimensions to spatial survey coordinates enables dynamic condition mapping and

automated maintenance prioritization. Moreover, as highlighted by Sedighian-Fard et al. [36], accurate and standardized initial distress quantification is the fundamental prerequisite for subsequent performance tracking and predictive pavement preservation programs.

From an operational and deployment perspective, the proposed methodology demonstrates remarkable adaptability and engineering resilience under diverse field operating scenarios. The unified single-pass formulation effectively resolves the latency and synchronization bottlenecks that historically impeded edge deployment, enabling robust execution across commodity mobile processors and automotive edge modules alike. Looking forward toward continuous network-wide scaling, several promising avenues for technological extension emerge. While planar homography and calibrated GSD deliver sub-millimeter metric fidelity across standardized dry asphalt conditions, expanding the optical pipeline to accommodate dynamic environmental variations, such as transient surface moisture or rapid illumination shifts under dense tree canopies, presents an exciting opportunity for multi-modal sensor fusion. Future extensions could seamlessly couple optical RGB feeds with low-cost thermal infrared channels, explore temporal frame-to-frame feature aggregation from vehicle trajectory sequences, and incorporate generative synthetic data augmentation via advanced generative adversarial networks [59] and WGAN-GP architectures [60] to further enrich training distributions for rare distress morphologies. In doing so, this framework establishes a versatile foundation for continuous, crowdsourced roadway health monitoring across next-generation intelligent transportation ecosystems.

7. Conclusions

The paper presented a holistic and integrated framework for autonomous assessment of pavement condition from the stage of distress detection to the development of a quantitative index. The main goal was to develop and validate deep learning-based models that can accurately and efficiently assess key pavement distresses in real-world scenarios. Key findings of this study are as follows:

1. During the controlled architectural evaluation, the YOLO11m-seg model achieved mAP values of 91.87% for detection and 89.23% for segmentation on the field test dataset. Compared with the YOLOv8m-seg baseline (84.50% detection mAP and 81.70% segmentation mAP), the proposed architecture demonstrated consistent performance improvements, confirming the effectiveness of YOLO11 architectural enhancements for complex pavement distress patterns;
2. The proposed framework demonstrated effective cross-domain generalization to challenging AV imagery. Despite substantial differences in imaging geometry (horizontal perspective versus BEV imagery) and environmental variations, the model maintained robust performance after supervised domain adaptation. Using a separate AV adaptation subset (100 images per class, 300 images in total) for fine-tuning, the adapted model achieved 84.60% detection mAP and 81.47% segmentation mAP on the independent AV evaluation dataset. The limited performance reduction compared with field imagery confirms the effectiveness of the proposed adaptation strategy in mitigating domain shift;
3. This research was not limited to distress identification alone. Utilizing the accurate outputs of the segmentation model, the physical parameters of the distress, such as length, width, area, and diameter, were successfully extracted. Based on these parameters, the SCI was developed that is able to integrate the type, severity, and density of distress into a single quantitative measure;
4. The SCI was effectively applied to both field data and AV data. The results showed that the index is able to distinguish between low-severity distresses and severe distresses. Furthermore, empirical regression against certified manual ASTM D6433 PCI surveys on 50 full-scale test sections demonstrated strong linear predictive fidelity (PCI = 97.84 - 0.996 × SCI, R^2 = 0.88,

MAE = 6.98). Furthermore, Bland-Altman analysis of field-validated specimens demonstrated minimal measurement bias, high agreement with ground-truth measurements, and consistent accuracy across all five geometric distress parameters;

5. Moreover, this research developed a comprehensive pipeline that connects raw visual observations with actionable, data-driven pavement management decisions. Through the integration of a high-performance segmentation model (YOLO11m) and a quantitative condition assessment index (SCI), the proposed approach enables objective, scalable, and efficient continuous monitoring of road infrastructure.

Declaration of competing interest

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

Data availability

The datasets collected in Tehran, Iran, and Oulu, Finland, and used and/or analysed during the current study are available from the corresponding author on reasonable request. The nuScenes dataset used in this study is publicly available at https://www.nuscenes.org/.

References

[1] World Health Organization (WHO), Global status report on road safety 2023, World Health Organization, Geneva, Switzerland, 2023.

[2] AASHTO, Pavement Management Guide, 2nd Edition, AASHTO, Washington, D.C., 2012.

[3] C. Queiroz, H. Kerali, A review of institutional arrangements for road asset management, The World Bank, Washington, D.C., 2010.

[4] K.C. Sinha, S. Labi, B.R.D.K. Agbelie, Transportation infrastructure asset management in the new millennium: continuing issues, and emerging challenges and opportunities, Transportmetrica A: Transport Science 13 (7) (2017) 591–606. https://doi.org/10.1080/23249935.2017.1308977

[5] R. Haas, W.R. Hudson, J.P. Zaniewski, Modern Pavement Management, Krieger Publishing Company, Malabar, FL, 1994.

[6] M.Y. Shahin, Pavement Management for Airports, Roads, and Parking Lots: Second Edition, Springer, New York, NY, 2005. https://doi.org/10.1007/b101538

[7] ASTM D6433-20, Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys, ASTM International, West Conshohocken, PA, 2020. https://doi.org/10.1520/D6433-20

[8] D.H. Timm, J.M. McQueen, A study of manual vs. automated pavement condition surveys, Highway Research Center, Auburn University, Report IR-04-01, 2004.

[9] V. Edmondson, J. Woodward, M. Lim, M. Kane, J. Martin, I. Shyha, Improved non-contact 3D field and processing techniques to achieve macrotexture characterisation of pavements, Construction and Building Materials 227 (2019) 116693. https://doi.org/10.1016/j.conbuildmat.2019.116693

[10] S. Mathavan, K. Kamal, M. Rahman, A review of three-dimensional imaging technologies for pavement distress detection and measurements, IEEE Transactions on Intelligent Transportation Systems 16 (5) (2015) 2353–2362. https://doi.org/10.1109/TITS.2015.2428655

[11] C.C. Hsieh, Y.C. Tsai, Machine learning for crack detection: review and model performance comparison, Journal of Computing in Civil Engineering 34 (5) (2020) 04020038. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000918

[12] R. Roberts, L. Inzerillo, G. Di Mino, Exploiting low-cost 3D imagery for the purposes of detecting and analyzing pavement distresses, Infrastructures 5 (1) (2020) 6. https://doi.org/10.3390/infrastructures5010006

[13] Z. Du, J. Yuan, F. Xiao, C. Hettiarachchi, Application of image technology on pavement distress detection: A review, Measurement 184 (2021) 109900. https://doi.org/10.1016/j.measurement.2021.109900

[14] A.M. Saad, K.N. Tahar, Identification of rut and pothole by using multirotor unmanned aerial vehicle (UAV), Measurement 137 (2019) 647–654. https://doi.org/10.1016/j.measurement.2019.01.093

[15] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, H. Omata, Road damage detection and classification using deep neural networks with smartphone images, Computer-Aided Civil and Infrastructure Engineering 33 (12) (2018) 1127–1141. https://doi.org/10.1111/mice.12387

[16] D. Arya, H. Maeda, S.K. Ghosh, D. Toshniwal, A. Mraz, T. Kashiyama, Y. Sekimoto, Deep learning-based road damage detection and classification for multiple countries, Automation in Construction 132 (2021) 103935. https://doi.org/10.1016/j.autcon.2021.103935

[17] J. Guan, X. Yang, L. Ding, X. Cheng, V.C.S. Lee, C. Jin, Automated pixel-level pavement distress detection based on stereo vision and deep learning, Automation in Construction 129 (2021) 103788. https://doi.org/10.1016/j.autcon.2021.103788

[18] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91

[19] J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517–6525. https://doi.org/10.1109/CVPR.2017.690

[20] J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, arXiv preprint arXiv:1804.02767, 2018.

[21] A. Bochkovskiy, C.Y. Wang, H.Y.M. Liao, YOLOv4: Optimal speed and accuracy of object detection, arXiv preprint arXiv:2004.10934, 2020.

[22] G. Jocher, A. Chaurasia, J. Qiu, Ultralytics YOLOv8, version 8.0.0, 2023. https://github.com/ultralytics/ultralytics

[23] G. Jocher, J. Qiu, Ultralytics YOLO11, version 11.0.0, 2024. https://github.com/ultralytics/ultralytics

[24] H. Majidifard, Y. Adu-Gyamfi, W.G. Buttlar, Deep machine learning approach to develop a new asphalt pavement condition index, Construction and Building Materials 247 (2020) 118513. https://doi.org/10.1016/j.conbuildmat.2020.118513

[25] L.K. Suong, J. Kwon, Detection of potholes using a deep convolutional neural network, Journal of Universal Computer Science 24 (9) (2018) 1244–1257.

[26] S. Tighe, N. Li, L. Cowe Falls, R. Haas, Incorporating road safety into pavement management, Transportation Research Record 1699 (1) (2000) 1–10. https://doi.org/10.3141/1699-01

[27] E. Ibragimov, Y. Kim, J.H. Lee, J. Cho, J.J. Lee, Automated Pavement Condition Index Assessment with Deep Learning and Image Analysis: An End-to-End Approach, Sensors 24 (7) (2024) 2333. https://doi.org/10.3390/s24072333

[28] H. Bello-Salau, A.M. Aibinu, A.J. Onumanyi, E.N. Onwuka, J.J. Dukiya, H. Ohize, New road anomaly detection and characterization algorithm for autonomous vehicles, Applied Computing and Informatics 16 (1/2) (2020) 223–239. https://doi.org/10.1016/j.aci.2018.05.002

[29] M. Maurer, J.C. Gerdes, B. Lenz, H. Winner (Eds.), Autonomous Driving: Technical, Legal and Social Aspects, Springer, Berlin, Germany, 2016. https://doi.org/10.1007/978-3-662-48847-8

[30] C. Katrakazas, M. Quddus, W.H. Chen, L. Deka, Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions, Transportation Research Part C: Emerging Technologies 60 (2015) 416–442. https://doi.org/10.1016/j.trc.2015.09.011

[31] H. Caesar, V. Bankiti, A.H. Lang, S. Vora, V.E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, O. Beijbom, nuScenes: A multimodal dataset for autonomous driving, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11618–11628. https://doi.org/10.1109/CVPR42600.2020.01164

[32] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Second Edition, Cambridge University Press, Cambridge, UK, 2004. https://doi.org/10.1017/CBO9780511811685

[33] M. Nieto, L. Salgado, F. Jaureguizar, J.M. Cabrera, Stabilization of inverse perspective mapping images based on robust vanishing point estimation, in: 2007 IEEE Intelligent Vehicles Symposium, 2007, pp. 315–320. https://doi.org/10.1109/IVS.2007.4290133

[34] M.A. Talaghat, A. Golroo, V. Shahhosseini, M. Rasti, Fully automated pavement crack quantification and PCI estimation via a digital twin-ready vision-based deep learning pipeline, Results in Engineering 32 (2026) 111659. https://doi.org/10.1016/j.rineng.2026.111659

[35] M.A. Talaghat, A. Golroo, M. Rasti, V. Shahhosseini, Road Pavement Digital Twin: A Blueprint with a Case Study, ASCE Journal of Computing in Civil Engineering 40 (3) (2026) 05026001. https://doi.org/10.1061/JCCEE5.CPENG-7187

[36] M. Sedighian-Fard, A. Golroo, M. Javanmardi, A. Alahi, H. Dehghan Tezerjani, M.A. Talaghat, V. Shahhosseini, M. Rasti, Explainable AI-enhanced pavement deterioration modeling using hybrid classification-regression deep learning architecture, Scientific Reports (2026), published online 31 August 2026. https://doi.org/10.1038/s41598-026-67842-x

[37] J. Cohen, A coefficient of agreement for nominal scales, Educational and Psychological Measurement 20 (1) (1960) 37–46. https://doi.org/10.1177/001316446002000104

[38] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (6) (2017) 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

[39] J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, The Lancet 327 (8476) (1986) 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

[40] U.S. Army, Pavement Maintenance Management, Technical Manual TM 5-623, Headquarters, Department of the Army, Washington, D.C., 1982.

[41] N. Wong, C. Chambers, K. Stol, R. Halkyard, Autonomous vehicle following using a robotic driver, in: 2008 15th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), 2008, pp. 115–120. https://doi.org/10.1109/MMVIP.2008.4749517

[42] G. Chandan, A. Jain, H. Jain, Mohana, Real time object detection and tracking using deep learning and OpenCV, in: Proceedings of the 2018 International Conference on Inventive Research in

Computing Applications (ICIRCA), 2018, pp. 1305–1308. https://doi.org/10.1109/ICIRCA.2018.8597266

[43] R. Kavitha, S. Nivetha, Pothole and object detection for an autonomous vehicle using YOLO, in: Proceedings of the 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1585–1589. https://doi.org/10.1109/ICICCS51141.2021.9432186

[44] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (6) (1986) 679–698. https://doi.org/10.1109/TPAMI.1986.4767851

[45] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics 9 (1) (1979) 62–66. https://doi.org/10.1109/TSMC.1979.4310076

[46] S. Chambon, J.M. Moliard, Automatic road pavement assessment with image processing: Review and comparison, International Journal of Geophysics 2011 (2011) 989354. https://doi.org/10.1155/2011/989354

[47] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, P. Fieguth, A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure, Advanced Engineering Informatics 29 (2) (2015) 196–210. https://doi.org/10.1016/j.aei.2015.01.008

[48] P. Shi, C. Zhao, Review on deep based object detection, in: 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), 2020, pp. 372–377. https://doi.org/10.1109/ICHCI51889.2020.00085

[49] Y.J. Cha, W. Choi, O. Büyüköztürk, Deep learning-based crack damage detection using convolutional neural networks, Computer-Aided Civil and Infrastructure Engineering 32 (5) (2017) 361–378. https://doi.org/10.1111/mice.12263

[50] C.V. Dung, L.D. Anh, Autonomous concrete crack detection using deep fully convolutional neural network, Automation in Construction 99 (2019) 52–58. https://doi.org/10.1016/j.autcon.2018.11.028

[51] R. Fan, U. Ozgunalp, B. Hosking, M. Liu, I. Pitas, Pothole detection based on disparity transformation and road surface modeling, IEEE Transactions on Image Processing 29 (2020) 897–908. https://doi.org/10.1109/TIP.2019.2933750

[52] H. Liu, X. Miao, C. Mertz, C. Xu, H. Kong, CrackFormer: Transformer network for fine-grained crack detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 3783–3792.

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

[54] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, Cham, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[55] L.C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, arXiv preprint arXiv:1706.05587, 2017.

[56] J. Huyan, W. Li, S. Tighe, Z. Xu, J. Zhai, CrackU-net: A novel deep convolutional neural network for pixelwise pavement crack detection, Structural Control and Health Monitoring 27 (8) (2020) e2551. https://doi.org/10.1002/stc.2551

[57] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: Proceedings of the 34th International Conference on Machine Learning (ICML), 2017, pp. 214–223.

[58] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of Wasserstein GANs, Advances in Neural Information Processing Systems (NeurIPS) 30 (2017) 5767–5777.

[59] C. Han, T. Ma, J. Huyan, Z. Tong, H. Yang, Y. Yang, Multi-stage generative adversarial networks for generating pavement crack images, Engineering Applications of Artificial Intelligence 131 (2024) 107767. https://doi.org/10.1016/j.engappai.2023.107767

[60] M. Sedighian-Fard, A. Golroo, Automatic assessment of road pavement condition using a generative adversarial network model with gradient penalty and U-Net-based segmentation, Amirkabir Journal of Civil Engineering 57 (9) (2025) 1667–1696. https://doi.org/10.22060/ceej.2026.24715.8341

[61] D. Zhang, Q. Zou, H. Lin, X. Xu, L. He, R. Gui, Q. Li, Automatic pavement defect detection using 3D laser profiling technology, Automation in Construction 96 (2018) 350–365. https://doi.org/10.1016/j.autcon.2018.09.019

[62] S.C. Radopoulou, I. Brilakis, Automated detection of multiple pavement defects, Journal of Computing in Civil Engineering 31 (2) (2017) 04016057. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000623

[63] J. Ha, D. Kim, M. Kim, Assessing severity of road cracks using deep learning-based segmentation and detection, The Journal of Supercomputing 78 (16) (2022) 17721–17735. https://doi.org/10.1007/s11227-022-04560-x

[64] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, Generalized intersection over union: A metric and a loss for bounding box regression, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 658–666. https://doi.org/10.1109/CVPR.2019.00075

[65] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, Proceedings of the AAAI Conference on Artificial Intelligence 34 (07) (2020) 12993–13000. https://doi.org/10.1609/aaai.v34i07.6999

[66] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2999–3007. https://doi.org/10.1109/ICCV.2017.324

[67] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, J. Yang, Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection, Advances in Neural Information Processing Systems (NeurIPS) 33 (2020) 21002–21012.

[68] T.Y. Zhang, C.Y. Suen, A fast parallel algorithm for thinning digital patterns, Communications of the ACM 27 (3) (1984) 236–239. https://doi.org/10.1145/357994.358023