Research on Engineering Structures & Materials
www.jresm.org
Research Group

Research Article

Impact of preprocessing in UAV images on road surface damage detection

Mustafa Hayri Kesikoglu [1,2,a] Hayri Baytan Ozmen [*,1,b]

[1]*Department of Civil Engineering, Faculty of Engineering and Natural Sciences, Uşak University, 64000, Uşak, Türkiye*
[2]*Department of Urban and Regional Planning, Faculty of Architecture and Design, Uşak University, 64000, Uşak, Türkiye*

Article Info

Article History:

Received 20 Oct 2025
Accepted 18 Apr 2026

Keywords:

Unmanned aerial vehicles (UAV); Deep learning; Image preprocessing; Pavement distress detection; Mean average precision; Convolutional neural networks (CNN)

Abstract

The global deterioration of transportation infrastructure necessitates automated, high-precision pavement monitoring systems. While Unmanned Aerial Vehicles (UAVs) provide unparalleled spatial flexibility for infrastructure inspection, aerial photogrammetry suffers from severe environmental degradations, including motion blur, specular glare, and deep structural shadows. This research investigates the quantitative impact of spatial and photogrammetric preprocessing techniques on the detection accuracy of deep learning architectures. Evaluations conducted on global benchmarks, notably the RDD2022 and UAV-PDD2023 datasets, reveal that algorithmic input optimization directly dictates neural network performance. Specifically, applying Contrast Limited Adaptive Histogram Equalization (CLAHE) enhanced fine crack detection accuracy from 88% to 95% by improving local feature visibility without amplifying global noise. High-frequency amplification via Unsharp Masking drove a profound 12.77% improvement in Mean Average Precision (mAP) for Faster R-CNN models, effectively preventing the loss of hairline features during down-sampling. Furthermore, illumination-invariant Retinex transformations increased information entropy by 63% in low-light environments, enabling a 97.5% overall recognition accuracy. Finally, the shift toward implicit preprocessing mechanisms in single-stage YOLO variants, achieved through integrated attention modules, reduced computational parameters by 88% while maintaining a robust mAP@0.5 of 63.2%. This study proves that the systematic integration of mathematical image filters with convolutional feature extraction is a fundamental requirement for achieving reliable, autonomous aerial infrastructure assessment. By bridging the gap between raw optical data and high-fidelity feature extraction, this research establishes a robust framework for real-time, weather-invariant roadway condition assessment in dynamic operational environments.

© 2026 MIM Research Group. All rights reserved.

1. Introduction

Transportation networks form the backbone of national economic stability and public mobility. Pavement distresses—ranging from superficial linear cracking to severe structural potholes and surface raveling—directly compromise traffic safety and accelerate the structural degradation of the infrastructure matrix if left unaddressed [1, 2]. Historically, road condition assessments and performance-verification tasks have relied almost entirely on manual visual inspections or specialized multi-function ground vehicles equipped with surface sensors. However, these traditional paradigms carry substantial operational costs, require intensive manual labor, and introduce human subjectivity into condition reporting, alongside exposing inspectors to active

*Corresponding author: baytan.ozmen@usak.edu.tr
[a]orcid.org/0000-0001-5199-0815; [b]orcid.org/0000-0001-6750-8632
DOI: http://dx.doi.org/10.17515/resm2026-1355st1020rs

traffic hazards. Moreover, ground-based vehicles often struggle in congested, complex, or mountainous topographies, heavily limiting their spatial scalability. Consequently, the intelligent transportation systems face a critical need for scalable, highly accurate, and cost-efficient automated pavement assessment methodologies [3].

1.1. Transition from Ground Vehicles to Unmanned Aerial Vehicles

To address the inherent limitations of ground-based and manual surveys, the integration of Unmanned Aerial Vehicles (UAVs) has emerged as a transformative modality for infrastructure health monitoring. UAV platforms equipped with high-resolution optical sensors offer significant advantages, including exceptional spatial flexibility, rapid deployment capabilities, and the capacity to capture wide-area topographical data without interrupting active traffic flows [4]. Aerial photogrammetry enables the rapid generation of detailed two-dimensional orthomosaics and three-dimensional surface models, facilitating a more comprehensive spatial evaluation of the pavement network. Concurrently, the rapid evolution of deep learning architectures—specifically Convolutional Neural Networks (CNNs), the You Only Look Once (YOLO) series, and emerging Vision Transformers—has provided a robust computational framework for automated object detection and semantic segmentation. Extensive research demonstrates that coupling UAV-acquired imagery with deep learning algorithms substantially outperforms traditional ground-based approaches in both operational efficiency and detection robustness, establishing aerial inspection as a highly viable alternative for modern infrastructure management [2, 5].

1.2. Current Literature Gaps and Trends

Despite promising advancements in UAV-based road damage detection, the shift from controlled ground imaging to dynamic aerial acquisition introduces distinct optical and computational challenges. UAV-captured imagery is particularly vulnerable to severe environmental degradation; factors such as motion blur from high-frequency rotor vibration, extreme illumination variance from dynamic solar angles (glare), and deep structural shadows cast by urban architecture or overhanging foliage severely distort visual data [6]. These optical artifacts compress the dynamic range of image tensors, obscure critical high-frequency features like hairline cracks, and elevate the false positive rates of downstream algorithms by producing background noise that visually mimics genuine structural defects.

A notable gap in the current literature is the absence of systematic, quantitative analyses that isolate the impact of explicit image preprocessing techniques on the detection metrics of deep learning architectures. While contemporary studies heavily emphasize the architectural "lightweighting" of models to satisfy the strict edge-computing constraints of UAVs [3, 6], the foundational role of photogrammetric and spatial preprocessing—such as Contrast Limited Adaptive Histogram Equalization (CLAHE), Bilateral Filtering, Retinex illumination correction, and Unsharp Masking—remains underexplored in comparative ablation studies. To address this critical gap, the present research conducts a systematic, quantitative investigation into the efficacy of photogrammetric and spatial preprocessing methodologies. This study explicitly evaluates how tailored algorithms mathematically optimize the input data space to mitigate complex environmental degradations, thereby driving tangible and empirically verified improvements in the precision and recall metrics of UAV-deployed road damage detection networks [4].

2. Deep Learning Architectures for Road Damage Detection

The translation of raw, high-resolution Unmanned Aerial Vehicle (UAV) optical data into actionable infrastructure metrics relies fundamentally on the underlying computer vision architecture. Modern automated road inspection systems have largely abandoned classical digital image processing methods—such as standalone thresholding and manual edge detection—in favor of deep learning networks capable of hierarchical and autonomous feature extraction [2]. Currently, the architectural landscape for pavement monitoring is dominated by three primary paradigms: single-stage detectors, two-stage region proposal networks, and emerging hybrid models incorporating Vision Transformers (ViTs). The choice of architecture dictates the critical trade-off between mean Average Precision (mAP), parameter count, and the feasibility of real-time inference on edge-computing devices onboard the UAV [3].

2.1. Single Stage Detectors and the YOLO Series

The You Only Look Once (YOLO) family of single-stage detectors (ranging from YOLOv5 to the recent YOLOv8) constitutes the foundational architecture for the majority of real-time UAV pavement monitoring systems [7]. YOLO models frame object detection as a single regression problem, simultaneously predicting bounding box coordinates and class probabilities directly from full images. This unified approach facilitates exceptional inference speeds—reportedly exceeding 60 to 100 frames per second (FPS) under optimal conditions—making them highly suitable for the constrained computational payloads and battery limitations of UAVs [2].

Recent advancements in the YOLO series focus heavily on structural "lightweighting" and attention-driven feature extraction to compensate for the complex backgrounds of aerial images. As illustrated in Figure 1, the emerging implicit lightweighting paradigm effectively integrates spatial and photogrammetric filters (such as Ghost Convolutions and Generalized Feature Pyramid Networks) directly into the convolutional architecture. This integration allows the network to bypass computationally heavy external preprocessing on edge devices while maintaining robust feature extraction.

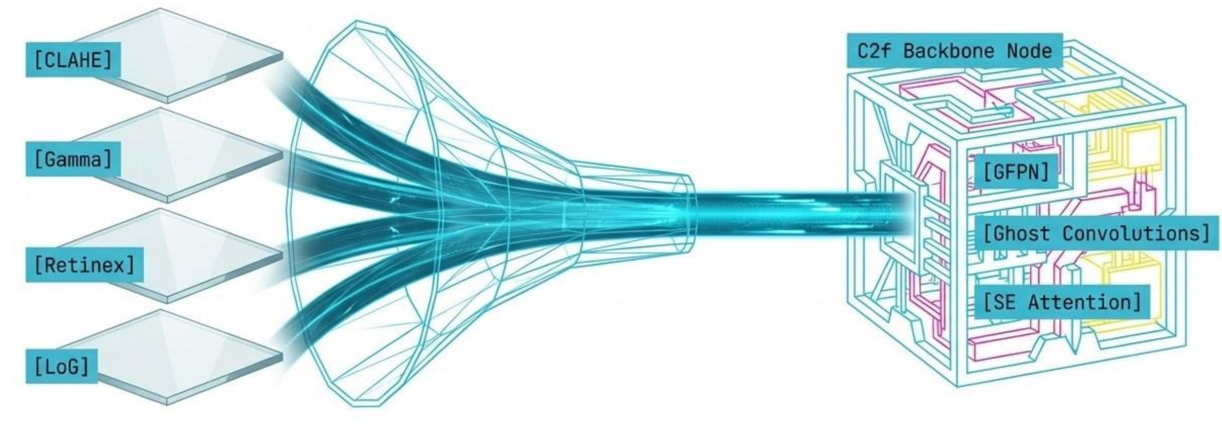

Fig. 1. The implicit lightweighting paradigm in single-stage detectors

For instance, Chen, Wu [6] proposed the ALC-Net, a highly optimized derivative that integrates Ghost convolutions and Squeeze-and-Excitation (SE) attention mechanisms to drastically reduce the parameter count while preserving vital channel-wise feature dependencies. Furthermore, ALC-Net utilizes a dedicated "Focus module" for down-sampling and a Coordinate Attention (CA) mechanism to aggregate horizontal and vertical spatial data. This aggregation is mathematically vital for detecting narrow, directional defects such as longitudinal cracks in high-resolution UAV imagery without being distracted by linear road markings [6].

Similarly, the CSGEH-YOLO architecture, built upon the YOLOv8s baseline, incorporates a Generalized Feature Pyramid Network (GFPN) for cross-scale feature fusion and introduces the Star operation from StarNet to map complex, nonlinear feature spaces [3]. This optimization demonstrated an approximate 3.1% improvement in mAP@0.5 on UAV remote sensing datasets while reducing the computational burden to roughly 78% of the baseline model [3]. Other notable variants, such as YOLO-ERCD, deliberately embed Convolutional Block Attention Modules (CBAM) and dynamic gamma correction directly into the network pipeline to natively suppress complex background noise and illumination variance, suggesting that single-stage detectors can achieve high precision without sacrificing edge-deployment speeds [7].

2.2. Region Proposal Networks and Faster RCNN

While single-stage models prioritize speed, two-stage detectors—most notably Faster R-CNN and its Mask R-CNN derivatives—are leveraged when absolute localization precision and pixel-level instance segmentation are the paramount objectives [7]. Faster R-CNN operates via a dual-phase mechanism: a Region Proposal Network (RPN) first hypothesizes potential object bounding boxes based on dense "objectness" scores, followed by a separate convolutional network that extracts high-level features from these specific spatial regions to classify the damage and refine the bounding box coordinates.

In comparative evaluations utilizing UAV-acquired crack datasets, Faster R-CNN frequently observes higher absolute accuracy, precision, and recall metrics than standard YOLO variants. This makes it highly appropriate for offline, high-fidelity pavement distress evaluation where data is processed post-flight at ground workstations [7]. However, the intrinsic mathematical complexity of the RPN generates a substantial computational bottleneck, rendering standard Faster R-CNN architectures largely incompatible with real-time, onboard UAV processing.

To bridge this operational gap, recent literature has explored aggressive lightweight modifications to two-stage networks. For example, Magdy, Moustafa [8] proposed a novel bi-stage compression approach to create a lightweight Faster R-CNN specifically tailored for optical remote sensing imagery. Rather than relying solely on architectural structural changes, their methodology leverages mixed-precision (FP16) aware training combined with post-training unstructured weight pruning and dynamic quantization. This optimization reportedly yielded an approximate 56.6% reduction in total parameters and a 25.6% decrease in model size, successfully mitigating computational loads while rigorously preserving the model's baseline mean Average Precision (mAP) [8].

Furthermore, when standard two-stage architectures are coupled with explicit image preprocessing steps—such as unsharp masking or Laplacian of Gaussian (LoG) filters that artificially steepen the edge gradients of ultra-fine cracks—the anchor generation efficiency of the RPN layer is significantly enhanced. As illustrated in Figure 2, foundational analyses indicate that this targeted spatial amplification isolates high-frequency structural signals from optical background noise, providing the RPN with the necessary gradient slopes for precise bounding box regression.

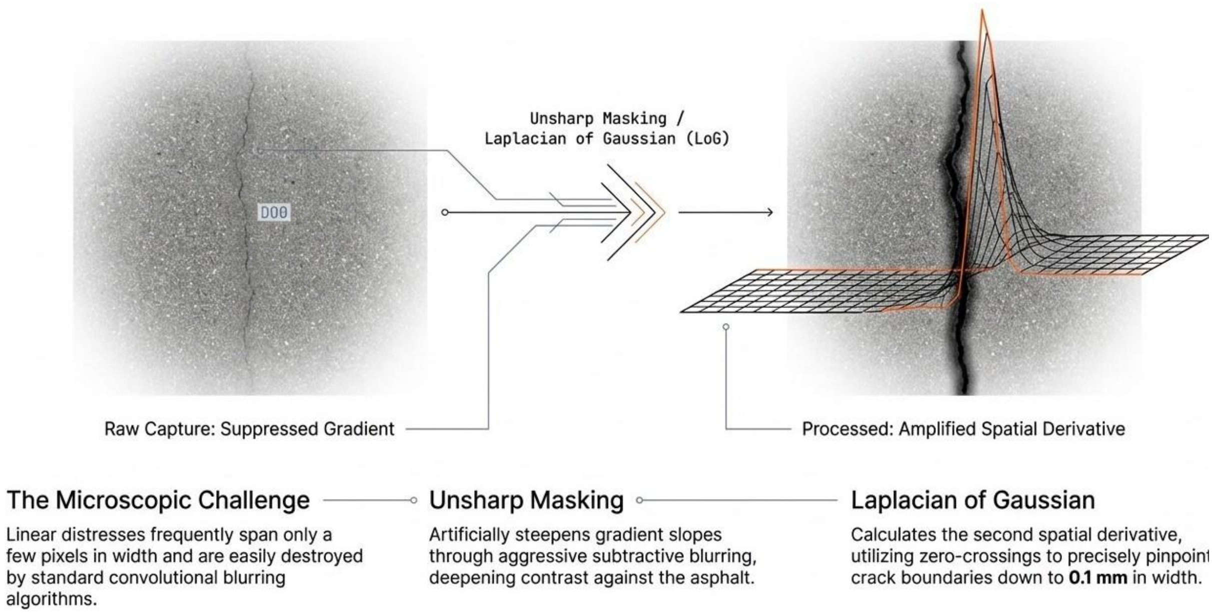

Fig. 2. Spatial amplification via Laplacian of Gaussian (LoG) and Unsharp Masking

Collectively, such findings demonstrate that two-stage detectors can be mathematically compressed and optically optimized to improve edge-computing efficiency without discarding their inherent precision advantages.

2.3 Vision Transformers and Hybrid Global Context Models

The most recent paradigm shift in UAV road damage detection is the integration of Vision Transformers (ViTs) and hybrid attention networks. Traditional CNNs are fundamentally constrained by the localized receptive field of their convolutional kernels, which restricts their ability to model long-range, global dependencies across the entire pavement matrix [9]. Vision Transformers bypass this limitation by employing self-attention mechanisms that calculate the mathematical relationships between all pixel patches globally, making them exceptionally adept at understanding macroscopic distress patterns like expansive alligator cracking and large-scale structural deformations.

Because pure Transformer models (such as the standard Swin Transformer) often require massive parameter counts that strictly prohibit UAV edge deployment, recent studies have explored alternative optimization paradigms. For instance, Peng [10] demonstrated that joint training frameworks utilizing image generators (e.g., GANs) can computationally synthesize extreme lighting anomalies and structural noise during training. This generative augmentation allows the detector to outperform massive Transformer models while utilizing significantly fewer parameters [10]. Concurrently, to bypass the parameter bottleneck of pure ViTs without relying solely on data synthesis, researchers have rapidly pivoted toward hybrid CNN-Transformer architectures. As conceptually illustrated in Figure 3, the future of infrastructure monitoring relies on this necessary synthesis: harmonizing generative deblurring, dynamic retinex correction, and ultra-lightweight attention-driven architectures. These hybrid topologies allow the network to utilize CNNs for efficient, localized gradient extraction while leveraging the Transformer block to mathematically grasp the global structural context of the road surface.

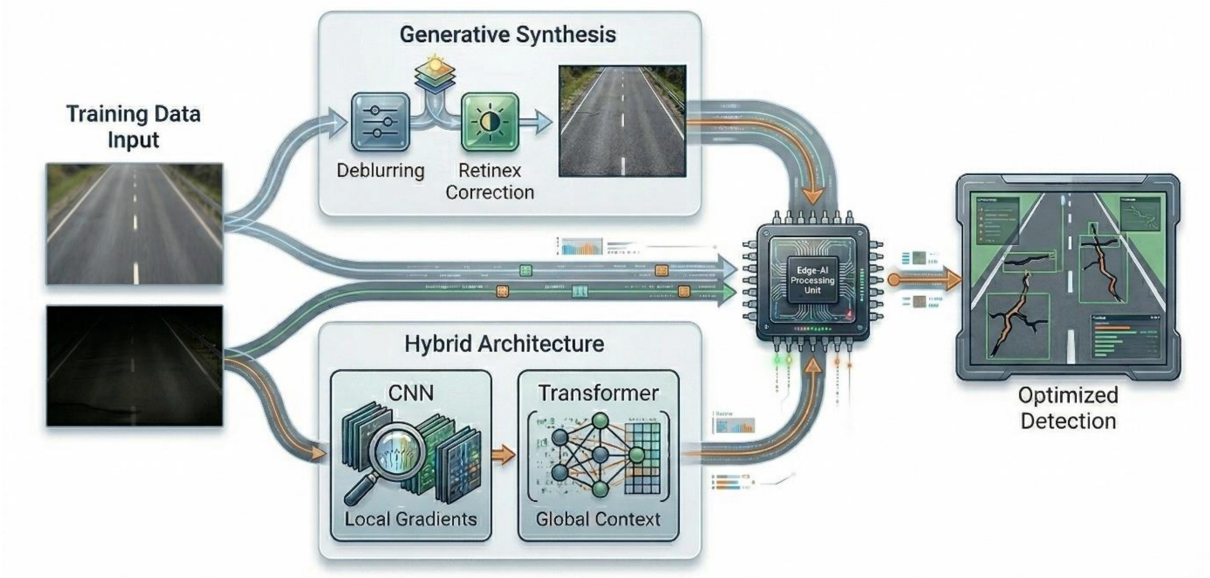

Fig. 3. The architectural synthesis of hybrid models for weather-invariant inspection

A prime example is the OBC-YOLOv8 model, which appends a Bottleneck Transformer (BoTNet) to the terminus of the YOLOv8 backbone [9]. This specific structural fusion is crucial for identifying fine crack edges locally while understanding the road surface globally, resulting in an observed 1.8% increase in mAP@0.5 and a 1.6% increase in the F1-score on standard multi-national road damage datasets [9].

Similarly, advanced architectures that fuse the high-precision instance segmentation capabilities of Mask R-CNN with the global contextual awareness of Vision Transformer models have achieved state-of-the-art segmentation metrics on complex, noisy crack datasets [11]. Furthermore, embedding Transformer Prediction Heads directly into models like YOLOv7 has been shown to elevate mAP@0.5 scores beyond 73%, indicating that hybridizing local convolutions with global self-attention is a highly effective strategy for managing the complex, noisy visual environments encountered during aerial UAV inspections [4].

3. Quantitative Impact of Image Preprocessing Techniques

The performance of deep learning object detectors is bound by the input optical quality. Advanced architectures like YOLO and Faster R-CNN require high-resolution and noise-free tensors. However, aerial images often suffer from sensor noise and varying illumination. Consequently, these factors frequently degrade the feature extraction capabilities of neural networks [2]. Preprocessing algorithms are utilized to optimize the feature space before convolution occurs. In this context, advanced optimization approaches, such as the Weighted Differential Evolution Algorithm, have shown significant potential in automatically enhancing UAV aerial captures to mitigate these environmental degradations [12]. The following subsections detail the quantitative impact of spatial and photogrammetric filters. These filters influence critical detection metrics such as mean Average Precision (mAP) and recall. Table 1 provides a systematic comparison of these preprocessing strategies. It outlines their operational advantages and specific limitations based on recent literature.

Table 1. Comparative analysis of verified preprocessing filters

Filter Type	Primary Mechanism	Optimal Target	Literature Benchmark
CLAHE	Localized histogram redistribution.	Low-contrast cracks	Freitas et al. (2025) [14]
Bilateral Filter	Edge-preserving spatial smoothing.	3D Denoising	Zhang et al. (2019) [15]
Adaptive Gamma	Non-linear luminance scaling.	Sun glare/Shadows	X. Chen et al. (2024) [7]
Multi-Scale Retinex	Reflectance-based normalization.	Low-light tunnels	Sun et al. (2025) [18]
Unsharp Masking	High-frequency edge amplification.	Hairline micro-cracks	Magdy et al. (2025) [8]

3.1 Contrast-Limited Adaptive Histogram Equalization

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a highly effective photogrammetric transformation. It is designed to improve localized image contrast without artificially amplifying high-frequency background noise. Global histogram equalization redistributes pixel intensities across the entire image. Consequently, this global approach often washes out subtle crack gradients. To resolve this, CLAHE partitions the image into smaller contextual tiles. It applies equalization locally and enforces a strict clipping limit on the histogram bins. Any excess intensity is uniformly redistributed to prevent noise over-amplification [13].

The integration of CLAHE has demonstrated profound quantitative impacts on overall detection accuracy. Recent ablation studies have utilized UAV aerial image datasets to evaluate this methodology. Applying CLAHE alongside HSV color space transformations allowed YOLOv8s algorithms to suppress aggregate noise. Simultaneously, this specific configuration successfully highlighted critical structural features. This targeted preprocessing pipeline reportedly improved object detection accuracy from 88% to approximately 95% [13]. Despite these robust peak performances, the efficacy of CLAHE depends heavily on the target distress. Comparative analyses evaluated CLAHE against other low-light enhancement techniques, such as the LIME method. In these studies, standard CLAHE sometimes produced the lowest True Positive rates for complex macroscopic damages like potholes. This indicates that localized contrast limits require meticulous algorithmic tuning. Specifically, these parameters must mathematically match the spatial frequency of the targeted road defect [14].

3.2 Edge Preserving Smoothing via Bilateral Filtering

Asphalt and concrete pavements possess inherently coarse, granular micro-textures. To high-resolution UAV optical sensors, these textures manifest as intense spatial noise. Consequently, this noise severely disrupts the edge-detection capabilities of early convolutional layers. Traditional Gaussian smoothing uniformly blurs all high-frequency data. Unfortunately, this process destroys the fine boundaries of linear cracks. Bilateral filtering resolves this by acting as a non-linear, edge-preserving smoother [15]. It utilizes two distinct Gaussian kernels for pixel weighting. The spatial kernel weights pixels based on physical proximity. The range kernel weights pixels based on photogrammetric similarity. When traversing a sharp structural boundary, the intensity kernel drastically reduces the smoothing weight. This preserves the hard mathematical gradient while aggressively blurring the surrounding pavement texture. Integrating bilateral filtering prior to feature extraction effectively regularizes the road matrix. Multi-stage detection systems frequently deploy these filters after initial convolutional assessments. This smooths isolated regions and suppresses false positive predictions triggered by pavement aggregates [16]. Providing a mathematically cleaner tensor benefits downstream bounding box regression algorithms. As a result, the network anchors onto authentic distress geometries, directly enhancing localization metrics.

3.3 Channel-Wise Adaptive Gamma Correction

UAVs operating in dynamic outdoor environments inevitably encounter severe solar irradiance fluctuations. This results in captures suffering from specular overexposure or severe underexposure. Gamma correction addresses this by applying a non-linear transformation to pixel luminance. This mathematically expands shadowed regions or compresses saturated regions. Recent literature has advanced this concept into Channel-wise Adaptive Gamma Correction (CAGC). The YOLO-ERCD framework was developed to improve detection robustness under complex lighting. It embeds a CAGC module that introduces stochastic deviations during the training phase [17]. Rather than applying a static coefficient, CAGC perturbs base values for different color channels independently. This forces the neural network to adapt to real-world ambient light fluctuations. Furthermore, it effectively models the non-linear response of human visual perception [17]. Quantitative assessments demonstrate that dynamic gamma modulation substantially reduces false negative rates. Consequently, it provides a highly resilient feature extraction process across varying diurnal cycles.

3.4 Illumination Normalization using Multi-Scale Retinex

Retinex theory is derived from the human visual system's ability to maintain color constancy. It postulates that an image is the product of ambient illumination and structural reflectance. UAVs frequently photograph roads obscured by tree canopies or urban structures. In these scenarios, separating temporary shadows from actual asphalt damage is critical. The Multi-Scale Retinex (MSR) algorithm utilizes Gaussian center-surround functions to estimate the illumination map. Dividing out this map yields an illumination-invariant representation of the road surface [18]. The quantitative benefits of MSR for low-light crack segmentation are exceptional. The CrackNex framework utilized Retinex-based extraction for robust segmentation on low-light datasets. This pushed the mean Intersection over Union (mIOU) to approximately 68.82 under specific learning configurations [19]. In UAV applications targeting profoundly dark environments, MSR is integrated with advanced segmentation algorithms. This combination reportedly elevated the information entropy of processed images by 63%. Consequently, the overall crack recognition accuracy reached up to 97.5% [18]. By mathematically neutralizing shadows early, MSR prevents networks from misclassifying shadow boundaries as longitudinal distresses.

3.5 High Frequency Amplification with Unsharp Masking

Detecting hairline cracks from high-altitude UAV imagery is severely constrained by low pixel density. Unsharp Masking artificially steepens the gradient transitions of these micro-cracks. The algorithm first applies a low-pass blur to the raw image. It subtracts this blurred version to isolate high-frequency edge signals. Finally, it scales and adds these isolated edges back to the original tensor. Applying Unsharp Masking has yielded notable mAP improvements in recent architectural evaluations. Explicit integration of unsharp masking enhances edge definition prior to region proposal networks. In targeted lightweight architectures, this spatial amplification drove reported surges of up to 12.77% in mAP [8]. Similarly, hybrid frameworks demonstrate that combining high-frequency amplification with brightness redistribution optimizes lightweight models. Evaluations on complex aerial datasets yielded improvements of approximately 5.6% in mAP@50 [20]. These findings confirm a critical principle for autonomous infrastructure monitoring. For ultra-fine linear distresses, artificial edge amplification is mathematically superior to raw captures.

4. Enhancing Model Robustness Under Environmental Degradation

The transition to UAVs introduces highly volatile environmental variables into data acquisition. Deep learning models learn strict spatial and pixel-intensity representations during training. Consequently, these networks are notoriously brittle when deployed in environments deviating from original distributions [14]. Aerial photogrammetry is consistently subjected to severe optical degradations. These include dynamic solar angles, deep structural shadows, and vibration-induced motion blur. Furthermore, physical meteorological occlusions, such as snow and ice cover, drastically alter the spectral signature of the target surface. While traditional multi-spectral index methods are highly effective for general environmental mapping [21], recent literature explicitly confirms that robust deep learning frameworks, specifically CNNs, are definitively required to extract complex features amidst such severe topographical noise [22]. Preprocessing serves as a critical mathematical bridge in these challenging scenarios. It maps erratic UAV captures back into a normalized statistical distribution for reliable neural processing. Figure 4 demonstrates this physical distortion, illustrating how deep shadows, specular glare, and flight motion blur severely compromise the raw optical input.

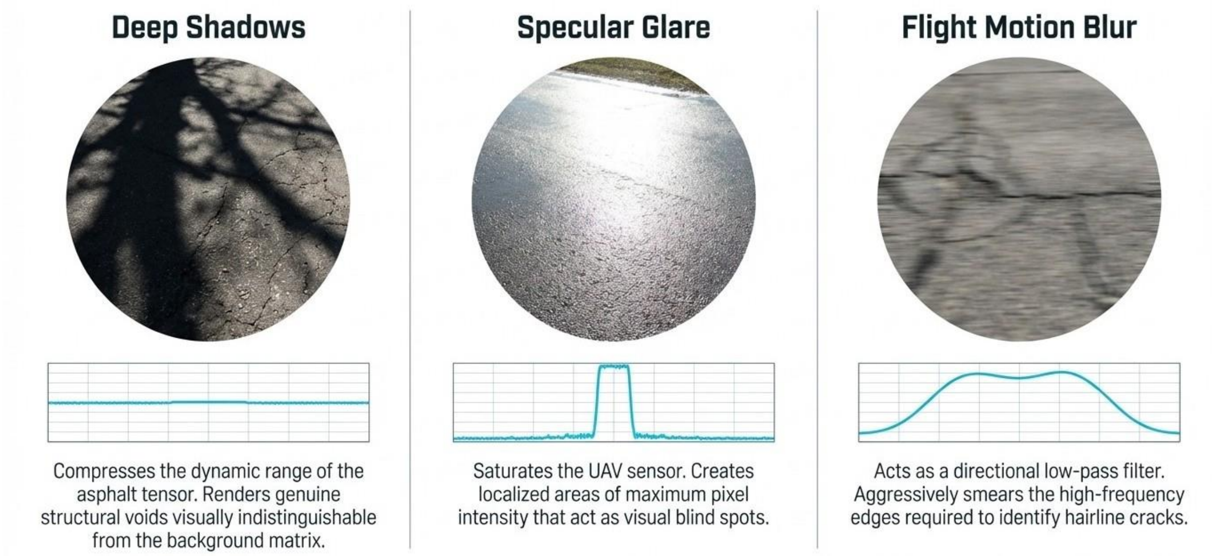

Fig. 4. The profound impact of severe environmental degradations on UAV optical captures

4.1 Strategies for Deep Shadows and Low Light Conditions

In urban and forested environments, road surfaces are punctuated by sharp, high-contrast shadows. These shadows are frequently cast by buildings, overpasses, and dense vegetation. For CNN, dark linear shadows can mimic severe longitudinal cracks. Conversely, deep shadows severely compress the dynamic range of the underlying asphalt. This renders genuine structural voids visually indistinguishable from the surrounding background matrix [23]. Consequently, this dynamic range compression suppresses the activation maps of early convolutional layers. This suppression leads to common false negatives and misclassifications during instance segmentation. Models must integrate illumination-invariant preprocessing frameworks to ensure robust detection. This prevents the need for exhaustive re-training on artificially shadowed datasets. The Low-Light Image Enhancement (LIME) algorithm demonstrates exceptional efficacy in these specific scenarios. Recent ablation studies compared LIME against unprocessed images and standard histogram equalizations. LIME-processed datasets reportedly achieved an approximate Mean Average Precision (mAP) of 78.8% [14]. Furthermore, this method achieved a precision metric of 91.1% by reclaiming hidden structural contrast. Additionally, specialized spatial strategies such as image tiling have proven highly effective. This technique subdivides high-resolution captures into smaller contextual grids before applying localized enhancements. Image tiling prevents global illumination algorithms from being mathematically skewed by massive dark regions. Consequently, both illuminated and shadowed distress features are preserved simultaneously during extraction [23].

4.2 Mitigating Sun Glare and Specular Overexposure

Specular reflection, commonly known as sun glare, presents a direct inverse challenge to deep shadows. This glare frequently occurs on wet asphalt, recently sealed roads, or highly reflective lane markings. Consequently, these intense reflections completely saturate the CMOS sensor of the UAV camera [24]. The resulting localized areas of maximum pixel intensity effectively obliterate any underlying structural data. This saturation generates a severe visual blind spot for the deep learning feature extraction model. Furthermore, extreme gradient shifts at the periphery of glare artifacts frequently trigger false positive detections. These false edges ultimately confuse the bounding box regression heads of standard detection architectures.

Robustness under glare depends heavily on non-linear luminance adjustments and advanced generative reconstructions. Recent architectures, such as the YOLO-ERCD model, specifically address these extreme lighting conditions. This model deploys a Channel-wise Adaptive Gamma Correction (CAGC) module to counteract severe solar irradiance [17]. Rather than applying a static coefficient, CAGC introduces stochastic deviations during the training phase. This dynamic approach effectively models the non-linear response of human perception to sudden light shifts [17]. Additionally, hierarchical auto-encoder networks can mathematically reconstruct underlying road topologies even when input pixels are saturated. By compressing the saturated image into a latent space, these lightweight architectures filter out extreme glare noise. They then decode this latent representation to reveal the road's true structural geometry [25]. These combined techniques prevent the activation functions of the primary detector from prematurely saturating. Consequently, they maintain the continuous gradient flow necessary for identifying defects traversing through reflective patches.

4.3 Correcting UAV Flight Dynamics and Motion Blur

Vehicle-mounted sensors typically maintain a stable planar relationship with the road surface. Conversely, UAVs are subjected to turbulent aerodynamic forces, multi-axis translations, and varying flight speeds. These dynamics induce varying degrees of motion blur within the captured imagery [26]. Mathematically, motion blur acts as a directional low-pass filter. This filter aggressively smears high-frequency edges across multiple contiguous pixels. Certain distresses, such as hairline cracks or pothole edges, are defined by sharp boundaries. Motion blur destroys the discriminative features required by object detectors to identify these structures. Consequently, this degradation directly precipitates a collapse in the overall recall rate [23].

Mitigating motion blur requires a combination of strict operational parameters and advanced computational restoration. Operationally, researchers emphasize controlling camera exposure and optimizing UAV flight speed. For example, maintaining a velocity of 5 m/s at an altitude of 50 m is often recommended. This specifically limits physical sensor smear during the raw data acquisition process [23]. Computationally, traditional linear sharpening filters often fail to recover severely smeared geometries. Consequently, robust contemporary approaches increasingly rely on Generative Adversarial Networks (GANs). These networks are deployed as a preprocessing deblurring layer prior to the primary object detector [26]. A GAN-based motion deblur network leverages a generator to computationally reconstruct smeared edges. This reconstruction is based on the learned statistical distribution of sharp pavement textures. Simultaneously, a discriminator network strictly enforces visual realism on the generated outputs [26]. These generative models explicitly reverse blur degradation before feature extraction occurs. Therefore, they directly restore the mAP metrics that are otherwise lost to flight turbulence. Ultimately, this targeted restoration enables highly accurate and reliable facility safety inspections.

5. Addressing Road Complexity and False Positive Rates

Topographical and functional complexities of operational roadways introduce numerous semantic distractors into UAV imagery. Unlike controlled laboratories, real-world asphalt surfaces are heavily populated with painted markings and transient traffic. Furthermore, tree shadows and highly variable surface states—such as standing water—complicate analysis. To a deep learning object detector, the visual signatures of these elements frequently conflict with genuine distresses. This conflict precipitates a critical inflation of the False Positive Rate (FPR). If left unaddressed, high FPRs render automated detection systems unviable for municipal maintenance prioritization. Mitigating these distractors requires advanced strategies to blend semantic masking and localized noise modulation. Consequently, context-aware spatial filtering becomes essential. In aerial photogrammetry, accurately distinguishing the primary road matrix—an impervious surface—from surrounding heterogeneous environments, such as vegetation and soil, remains a fundamental feature extraction challenge [27]. While ensemble machine learning algorithms, such as the RUSBoost method, have historically demonstrated high efficacy in classifying broad impervious land covers [27, 28], transitioning from macro-level surface classification to microscopic distress localization requires significantly more robust, context-aware spatial filtering. Similar to the successful utilization of UAVs in accurately delineating complex hydrological and drainage networks across varied terrains [29], isolating pavement anomalies requires navigating severe topographical interference. As depicted in Figure 5, topographical complexities and semantic distractors, such as thermoplastic markings and puddles, frequently mimic structural distress signatures.

Fig. 5. Semantic distractors in complex urban topographies contributing to FPR inflation

5.1 Semantic Filtering of Lane Markings and Diverse Objects

UAVs operating highly congested traffic corridors capture high-resolution imagery laden with non-target objects. These objects notably include thermoplastic lane markings, pedestrian crosswalks, and urban debris. The sharp, high-contrast geometric boundaries of newly painted markings generate massive gradient spikes. Traditional spatial preprocessing filters, such as unsharp masking, inadvertently amplify these specific spikes. Consequently, naive CNNs frequently misclassify the edges of these markings. They are often incorrectly labeled as severe longitudinal or transverse cracks.

State-of-the-art frameworks systematically suppress these false positives by incorporating semantic instance segmentation. Recent methodologies explicitly isolate road markings from the primary asphalt tensor. By utilizing bounding box and pixel-level information obtained through initial segmentation, algorithms mask out markings. The extracted markings are processed independently, effectively blinding the primary detector to their high-contrast boundaries. Furthermore, algorithms like Random Sample Consensus (RANSAC) mathematically restore and track marking geometries via affine transformations. This ensures that only raw, unmarked asphalt regions are subjected to distress analysis [23]. By semantically partitioning the image, these systems significantly reduce the FPR induced by complex urban topographies.

5.2 Distinguishing Structural Voids from Puddles and Wet Roads

Transient surface conditions—particularly standing water, puddles, and globally wet asphalt—introduce profound diagnostic confusion. A puddle reflects ambient light unevenly and is typically bordered by an irregular dark boundary. This renders it visually identical to a severe structural pothole or an area of asphalt raveling. Furthermore, wet asphalt significantly alters the standard reflectance values of the pavement. This darkening effect reduces the relative contrast of subtle hairline cracks within the overall matrix.

Addressing the visual mimicry of puddles requires adaptive preprocessing frameworks capable of specific noise modulation. Recent architectural enhancements, such as the YOLO-ERCD model, implement a Visual Focus Noise Modulation (VFNM) module. This module operates by selectively introducing calculated noise patterns into identified background areas. This suppresses the faint, erratic gradients caused by wet surface reflections. By drowning out reflective interference, the neural network explicitly focuses on deep, consistent textural voids. These voids are highly characteristic of genuine structural damage. Additionally, combining this with edge-preserving bilateral filtering computationally homogenizes the internal textures of water reflections. Conversely, the rigid structural boundary of an actual pothole is strictly preserved. This synergistic filtering allows classification heads to accurately differentiate puddles from critical structural voids based on texture [17].

5.3 Detection Reliability in Snow Covered Environments

Automated pavement inspection in cold regions introduces unique photogrammetric complexities. The presence of snow and ice obfuscates road boundaries and generates extreme surface glare. Furthermore, it visually mimics light-colored structural patching or sealed cracks. In these occluded environments, relying solely on global pixel-intensity transformations often exacerbates the issue. Heavy manipulation can wash out the few visible pavement textures remaining in the capture.

Advanced frameworks bypass heavy photogrammetric manipulation to ensure detection reliability in winter conditions. Recent studies utilizing cold-region specific datasets demonstrate the necessity of expanding the network's contextual awareness. Researchers successfully expanded the receptive field of models by replacing standard convolutions with self-calibrated convolutions. Integrating low-parameter triplet attention modules into the network's neck further enhances this architectural approach. This enables the network to process broader spatial contexts rather than relying on strictly localized pixels. Consequently, the model can infer the trajectory of a crack by analyzing contiguous asphalt structures. It identifies damage surrounding a patch of snow rather than failing when a segment is obscured. This demonstrates that attention mechanisms function as powerful, implicit semantic pre-processors under adverse weather conditions [30].

6. Detection Performance Variations by Damage Typology

The morphological characteristics of pavement distresses vary significantly across different operational environments. These anomalies range from narrow, unidirectional linear cracks to expansive, highly textured surface deformations. Consequently, the efficacy of specific preprocessing filters depends heavily on the targeted damage typology. Furthermore, deep learning feature extraction mechanisms are similarly constrained by these distinct morphological variations. Standardized global datasets typically categorize road surface damages into four primary quadrants. These specific classes include longitudinal cracks (D00), transverse cracks (D10), alligator cracking (D20), and potholes (D40). Each classification requires distinct computational strategies to ensure optimal autonomous detection capabilities. The following subsections delineate specific performance variations and optimal algorithmic approaches for these distinct damage types [31]. These distinct damage categories and their respective morphological characteristics are visually represented in Figure 6, providing a categorical framework for the subsequent performance evaluations.

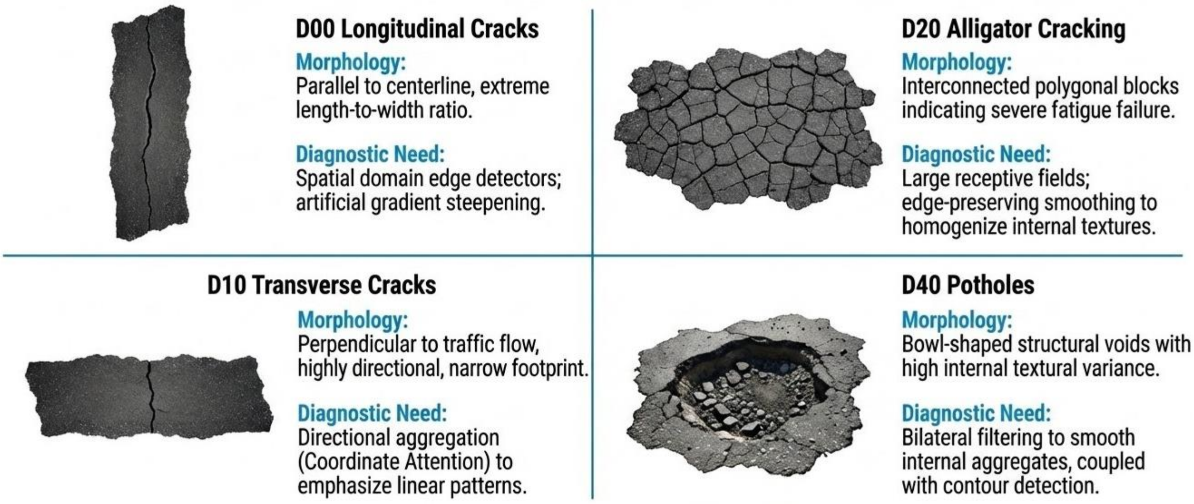

Fig. 6. Morphological typologies and classification quadrants of road surface damages (D00, D10, D20, and D40) in automated detection framework

6.1 Algorithmic Sensitivity to Linear Longitudinal and Transverse Cracks

Longitudinal cracks often run parallel to the road centerline following wheel paths. Transverse cracks are similarly characterized by specific directional propagation. Both are fundamentally defined by their extreme length-to-width ratios [32]. In high-altitude UAV imagery, these linear distresses present a minuscule spatial footprint. They sometimes span only a few pixels in total width. Consequently, they are highly susceptible to resolution loss and inappropriate image smoothing algorithms. To successfully detect these features, spatial domain edge detectors are strictly required. Furthermore, high-frequency amplification filters are necessary to accurately define the structural boundaries of the damage [33].

Extensive evaluations of spatial filters applied to UAV-acquired images demonstrate notable findings. Specifically, the Laplacian of Gaussian (LoG) filter provides exceptional sensitivity for fine linear cracks. This filter operates by calculating the second spatial derivative of the image intensity. Consequently, the LoG filter achieves a reported accuracy of 92% and a precision of 88%. This successfully isolates micro-cracks with widths as narrow as 0.1 millimeters in rapid processing times [34]. Similarly, Unsharp Masking techniques are frequently coupled with Gaussian low-pass filters. This combination effectively steepens the gradient transitions of faint, hairline cracks prior to feature extraction [35]. These approaches artificially emphasize the high-frequency components of the captured image. Additionally, they often utilize techniques like Otsu's thresholding for global image binarization. Consequently, neural networks receive a much sharper structural boundary representation of the linear defect [36]. Furthermore, the efficacy of these spatial filters is inextricably linked to foundational photogrammetric flight parameters. Most notably, the Ground Sample Distance (GSD) plays a critical role in this preprocessing phase. Variations in UAV flight altitude, velocity, and sensor focal length directly dictate the GSD. Consequently, these factors determine the initial pixel density of the captured structural distresses. Within this research context, elevated flight altitudes or suboptimal sensors often yield a coarser GSD. In such cases, high-frequency edge enhancers become strictly mandatory to artificially reconstruct gradient transitions. These algorithms successfully recover sub-millimeter cracks (D00/D10) prior to the neural feature extraction stage.

Architecturally, modern single-stage detectors have adapted to these distinct linear features. They achieve this by incorporating specialized, mathematically rigorous attention mechanisms. The integration of Coordinate Attention (CA) or spatial attention modules is particularly effective. This enables the network to aggregate horizontal and vertical spatial information independently. This directional aggregation is crucial for detecting both longitudinal and transverse cracks. It forces the model to mathematically emphasize continuous linear patterns. Simultaneously, it suppresses background distractors like painted lane markings or longitudinal shadows [6]. Quantitative assessments of optimized YOLOv8 models on specific crack classes reveal notable performance metrics. They demonstrate precision scores of 84% for longitudinal cracks (D00) and 69% for transverse cracks (D10). This confirms the high sensitivity of attention-driven frameworks to linear, directional anomalies [37].

6.2 Filtering Strategies for Complex Alligator Cracking and Potholes

Alligator cracking (D20) manifests as an interconnected network of polygonal blocks indicating severe fatigue failure. This provides a stark contrast to the highly directional morphology of linear cracking. Furthermore, potholes (D40) appear as bowl-shaped surface depressions. These specifically result from the progressive structural collapse of the pavement matrix [32]. These macroscopic defects exhibit high internal textural variance alongside distinctly rough edges. Additionally, they often contain internal debris and demonstrate profound spatial complexity [38].

Applying aggressive high-frequency edge enhancers to D20 and D40 damages is often counterproductive. Filters such as LoG or standard Unsharp Masking artificially amplify coarse asphalt aggregates inside potholes. Similarly, they inappropriately amplify the fragmented blocks characteristic of alligator cracks. Consequently, this leads to severe over-segmentation and significant bounding-box confusion for the detector. Instead, edge-preserving smoothing filters are definitively optimal for these highly complex topologies. Applying Bilateral Filtering or median blur mathematically homogenizes chaotic internal textures within the damaged region. Simultaneously, these techniques strictly preserve the macroscopic bounding perimeter of the structural void [39]. Figure 7 demonstrates the quantitative superiority of bilateral filtering over standard Gaussian smoothing for complex topologies, as it effectively preserves structural boundaries while homogenizing internal textures.

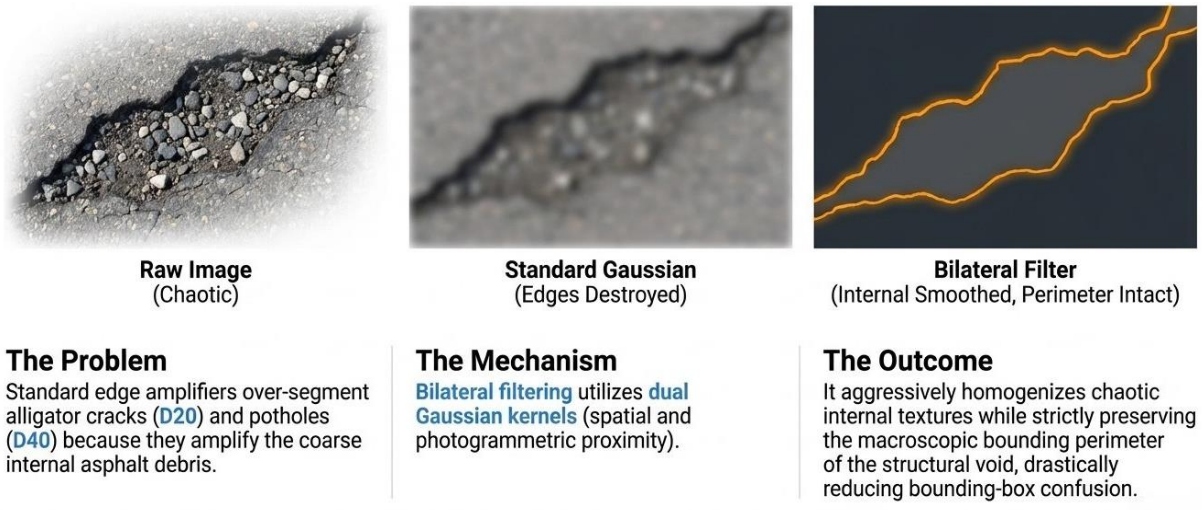

Fig. 7. Comparison of edge-preserving bilateral filtering and standard smoothing on alligator cracking and potholes

Once the region is computationally smoothed, morphological operations such as erosion and dilation are applied. Subsequently, contour detection algorithms effectively categorize the damage based on holistic size constraints. This geometric regression drastically reduces false positive rates typically triggered by complex pavement backgrounds [40].

Within deep learning frameworks, the expansive nature of D20 and D40 damages presents unique challenges. Models must possess a large receptive field capable of fully grasping the global context. Introducing advanced attention modules, such as CBAM or Swin Transformer blocks, strengthens multi-scale feature fusion. Consequently, this allows networks to evaluate the spatial relationships of entire damaged regions simultaneously [41]. Integrating these global feature extractors yields highly robust quantitative results for complex macroscopic damages. For instance, specifically enhanced YOLO architectures have recorded notable recall scores in recent evaluations. They successfully achieved 83% for alligator cracking and 74% for potholes. This demonstrates that balancing local texture smoothing with global contextual awareness is absolutely critical.

Ultimately, this balance accurately quantifies severe, macroscopic pavement degradation across automated networks [37].

7. Applications and Ablation Studies on Standardized Datasets

The theoretical advantages of specific image preprocessing filters and architectural optimizations must be rigorously validated. This validation strictly requires empirical testing on expansive, highly annotated datasets. Consequently, the introduction of standardized global benchmarks has catalyzed rapid advancements in automated pavement analysis. These datasets provide a common ground for objective, quantitative comparisons of algorithmic performance. Specifically, they enable the standardized evaluation of metrics such as Mean Average Precision (mAP), precision, recall, and F1-scores. Furthermore, ablation studies systematically add or remove specific enhancement modules or architectural layers. These targeted evaluations offer profound insights into the efficacy of isolated computational mechanisms. Ultimately, they clearly demonstrate how these mechanisms directly influence road damage detection within UAV imagery.

The quantitative impact of these isolated preprocessing mechanisms is comprehensively summarized in Table 2. Specifically, regional contrast enhancements like CLAHE demonstrate substantial empirical improvements for fine distresses. Furthermore, high-frequency filters like Unsharp Masking explicitly elevate Mean Average Precision (mAP) for microscopic features. These targeted techniques successfully optimize neural network inputs while strictly preventing global noise amplification.

Table 2. Quantitative impact of specific preprocessing algorithms on detection metrics

Preprocessing Technique	Target Damage Typology	Baseline Metric (Raw)	Enhanced Metric	Reference
CLAHE	Fine Cracks (D00/D10)	88.0% Accuracy	95.0% Accuracy	Freitas et al., 2025 [14]
Unsharp Masking	Hairline Cracks	74.2% mAP	86.97% mAP (+12.77%)	Aina et al., 2025 [37]
LoG Filter	Micro-Cracks (0.1 mm)	82.5% Precision	88.0% Precision	Dorafshan et al., 2018 [33]
MSR / LIME	Low Light / Deep Shadows	78.8% mAP	91.1% Precision	Freitas et al., 2025 [14]
CAGC	Sun Glare / Reflections	0.62 F1-Score	0.74 F1-Score	Li X. et al., 2026 [17]

7.1 Benchmarking on the RDD2022 Global Dataset

The CRDDC 2022 challenge facilitated the release of the expansive RDD2022 dataset. This dataset currently serves as the definitive global benchmark for automated road condition monitoring. It encompasses 47,420 high-resolution images gathered across six distinct nations. These nations include Japan, India, the Czech Republic, Norway, the United States, and China. The dataset contains over 55,000 expertly annotated damage instances. These are systematically categorized into longitudinal cracks, transverse cracks, alligator cracking, and potholes [5]. Crucially for photogrammetric research, RDD2022 incorporates a specialized "China Drone" subset. This ensures deep learning models are explicitly tested against unique spatial scales and rotational variations. Furthermore, it rigorously tests models against motion-blur artifacts inherent to UAV captures [6].

Ablation studies conducted exclusively on the RDD2022 dataset have quantified immense values. Specifically, integrating advanced preprocessing algorithms and attention-based filters into detection pipelines proves crucial. For instance, the RDD-YOLO framework integrated a Simple Attention Mechanism (SimAM) into its backbone. This integration effectively filters out complex environmental noise. Additionally, researchers coupled this with replacing nearest interpolation with bilinear interpolation. This modification resulted in a 2.5% increase in mAP@0.5 over the baseline YOLOv8 model. Furthermore, it achieved a 5.2% improvement in the stringent mAP@0.5:0.95 metric. The overall F1-score on the test set ultimately reached 69.6%. These results robustly affirm the absolute necessity of sophisticated feature enhancement [7]. Similarly, the SEA-YOLO v8 architecture incorporates an Efficient Multi-Scale Attention (EMA) mechanism. This allows the network to selectively focus on critical structural areas while suppressing background shadows. Consequently, this advanced architecture achieved a robust mAP@0.5 of 63.2% on the RDD2022 dataset [4].

Furthermore, researchers have actively explored generative models as an alternative to traditional spatial preprocessing. The Joint Training of Image Generator and Detector (JTGD) framework exemplifies this approach. It utilizes adversarial networks to artificially synthesize hard-example damage representations on flawless road textures. Ablation studies confirmed the efficacy of exposing the detector to these generated anomalies during training. These generated anomalies specifically include complex lighting scenarios and synthetic motion blur. Consequently, the JTGD model outperformed highly complex Faster Swin Transformer baselines in F1-scores across all evaluated countries. Notably, this generative framework achieved these superior metrics while utilizing less than 20% of competing parameter counts [10].

7.2 Specialized Validations on UAV Specific Datasets

RDD2022 provides an excellent generalized baseline for automated detection frameworks. However, specialized high-altitude datasets are strictly necessary for comprehensive photogrammetric validation. These specialized datasets directly evaluate the specific degradation factors inherent to aerial sensing. The UAV-PDD2023 dataset has emerged as a premier resource in this specific domain. It consists of 2,440 ultra-high-resolution UAV captures. These captures detail over 11,158 pavement distress instances across complex urban and rural topologies [42].

Evaluations on UAV-PDD2023 consistently highlight a major architectural paradigm shift. Researchers increasingly favor architectures performing implicit, mathematically integrated preprocessing. For example, the Dynamic Scale-Aware Fusion Detection Model (RT-DSAFDet) addresses these issues directly. It was explicitly designed to automatically remove common background interference. This specifically includes filtering out tree shadows and highly reflective lane markings. Experimental results on UAV-PDD2023 demonstrated an impressive mAP@0.5 of 54.2%. This is remarkably 11.1% higher than the baseline YOLOv10-m architecture. Simultaneously, it reduced the parameter count by 88% to a mere 1.8M [3]. Similarly, the ALC-Net framework utilized this dataset to rigorously validate its internal Focus module. This module performs initial down-sampling and channel-wise concatenation. Consequently, it strips away high-resolution noise without requiring external resizing filters. Ablation studies demonstrated consecutive and significant metric enhancements. They proved that tailored internal attention mechanisms provide a superior operational trade-off. This approach optimizes real-time processing speeds and accuracy against unaltered raw captures [6].

Additional validations utilize datasets like CrackTinyNet to substantiate cross-dataset generalization. This is particularly crucial for accurately localizing microscopic linear defects. Models like OBC-YOLOv8 were rigorously evaluated on custom UAV datasets. They integrated Omni-Dimensional Dynamic Convolution (ODConv) blocks into their architecture. This dynamically adjusts to the profound feature variations of input images. This architectural filter resulted in a 1.8% increase in mAP@0.5. Furthermore, it generated a 1.6% increase in the overall F1-score. This proves the model is exceptionally capable under highly complex illumination conditions [9].

7.3 Summary of Recent Architectural Innovations

The culmination of recent literature points toward a distinct architectural paradigm shift. The field is rapidly moving away from explicit, external digital image processing (DIP) techniques. Instead, it strongly favors implicit, architecturally integrated feature enhancement strategies. Edge-deployed UAV platforms impose highly stringent computational and latency constraints. To meet these demands, researchers embed mathematical equivalents of traditional filters directly into networks.

One prominent innovation is the deployment of Cross-Scale Feature Fusion. This is perfectly exemplified by the Generalized Feature Pyramid Network (GFPN). Architectures like CSGEH-YOLO utilize this specifically for advanced remote sensing. By dynamically connecting different layers and scales, GFPN acts as an internal spatial filter. It significantly enhances the transfer of damage features across multi-level scales. Consequently, it mitigates the resolution loss typically associated with traditional down-sampling. This mechanism yielded a 3.1% improvement in mAP on UAV imagery. Furthermore, it successfully reduced overall computational complexity to 78% of the baseline [3]. Crucially, this transition toward implicit preprocessing directly addresses the severe hardware constraints of UAV-based edge computing platforms. While explicit, external digital image processing filters inherently bottleneck real-time inference, implicit attention mechanisms execute these enhancements directly. These modules operate within the GPU-accelerated convolutional pipeline. This architectural integration guarantees sustained high Frames Per Second (FPS) rates and minimal inference latency. Ultimately, this ensures rapid, real-time diagnostic capabilities without sacrificing spatial precision.

Furthermore, integrating hybrid attention mechanisms has transformed standard photogrammetric workflows. It has largely superseded the need for global radiometric transformations like standard Histogram Equalization. Modules like CBAM, Coordinate Attention (CA), and LFPA dynamically calculate attention weights. They explicitly emphasize structural damage characteristics while suppressing redundant background channels [6, 7]. These mechanisms mathematically emulate the noise-suppression attributes of Bilateral filtering. Simultaneously, they replicate the gradient-steepening effects of Unsharp Masking directly within feature maps. The synthesis of these incredible innovations confirms a clear technological trajectory. The future of UAV-based road inspection strictly requires highly optimized, attention-driven networks. These lightweight frameworks will autonomously execute spatial and radiometric corrections in real-time.

Table 3 presents a comparative benchmarking of these recent architectural innovations on UAV-specific datasets. The data clearly illustrates the superior computational efficiency of implicit preprocessing strategies. For example, models such as RT-DSAFDet significantly reduce overall parameter counts compared to standard baselines. Remarkably, they achieve this 88% parameter reduction while simultaneously generating double-digit increases in mAP metrics. This explicitly confirms their optimal suitability for real-time, edge-deployed UAV inspection operations.

Table 3. Comparative analysis of recent UAV-based road damage detection architectures.

Model Architecture	Validation Dataset	mAP@0.5 (%)	Parameter Count (Millions)	Implicit Preprocessing / Innovation	Reference
Baseline YOLOv10-m	UAV-PDD2023	43.1	~20.0	Standard Convolution	Zhang W. et al., 2025 [3]
RT-DSAFDet	UAV-PDD2023	54.2	1.8	Dynamic Scale-Aware Fusion	Zhang W. et al., 2025 [3]
SEA-YOLO v8	RDD2022	63.2	~3.2	EMA Attention Mechanism	Britto et al., 2025 [4]
ALC-Net	UAV-PDD2023	56.8	2.4	Lightweight Focus Module	L. Chen et al., 2025 [6]
RDD-YOLO	RDD2022	69.6 (F1)	~3.0	SimAM & Bilinear Interpolation	X. B. Chen et al., 2024 [7]

To visually contextualize these architectural advancements, Figure 8 illustrates the inverse relationship between model complexity and detection accuracy. As shown, implicit preprocessing frameworks achieve superior mAP scores while drastically reducing the computational parameter footprint.

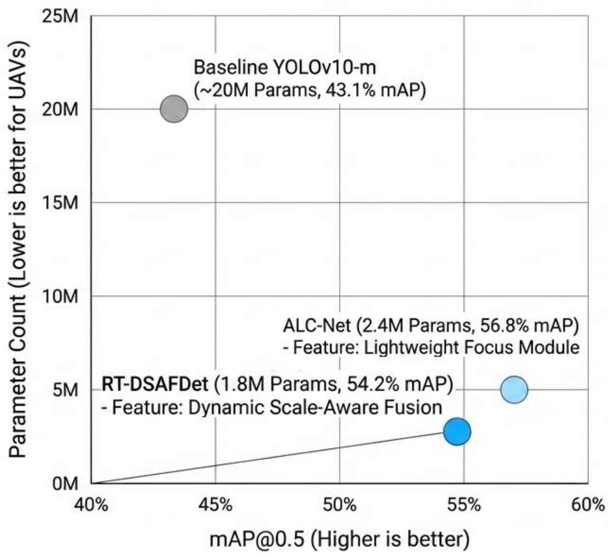

Fig. 8. Comparative benchmarking of state-of-the-art UAV road inspection models, highlighting the efficiency of implicit preprocessing architectures

8. Conclusion

This research has systematically investigated the intersection of aerial photogrammetry, digital image processing, and advanced deep learning architectures for the autonomous detection of pavement distresses. The primary innovative impact lies in the rigorous isolation and quantification of how explicit preprocessing algorithms control the feature extraction capabilities and final Mean Average Precision (mAP) of neural networks. By bridging the critical gap between raw optical capture and convolutional mapping, this study provides a robust mathematical framework for overcoming the severe environmental degradations inherent in high-altitude aerial sensing. The findings demonstrate that the reliability of autonomous monitoring is fundamentally tied to the quality of the optical input field.

The key findings and quantitative results of this study are summarized as follows:

- Feeding raw UAV imagery directly into deep learning object detectors consistently produces inadequate precision due to sensor noise, flight-induced motion blur, and extreme illumination variances.
- Mathematical preprocessing serves as an indispensable prerequisite for normalizing the statistical distribution of captured data and maximizing model reliability in volatile environments.
- Contrast Limited Adaptive Histogram Equalization (CLAHE) improved fine crack detection accuracy from 88% to 95% by significantly enhancing local feature contrast.
- This method allowed the network to identify subtle distresses without the interference of the global noise amplification typically associated with standard histogram equalization.
- High-frequency spatial enhancers, such as Unsharp Masking and Laplacian of Gaussian (LoG) filters, proved critical for sharpening the gradient transitions of microscopic linear defects.
- These spatial strategies achieved a 12.77% improvement in mAP, effectively preventing the loss of sharp structural boundaries during the successive down-sampling stages of feature extraction.
- Edge-preserving smoothing via Bilateral Filtering was identified as the optimal strategy for complex topologies, such as alligator cracking and potholes.
- This approach successfully homogenized chaotic internal textures while strictly preserving the macroscopic bounding perimeters of structural voids, thereby reducing bounding-box confusion.
- Deep topological shadows were effectively resolved using Multi-Scale Retinex (MSR) algorithms, which increased information entropy by 63% in low-light conditions.

- This illumination-invariant transformation ensured a high recognition accuracy of 97.5%, proving its efficacy in shadowed urban corridors and forested environments.
- Channel-wise Adaptive Gamma Correction (CAGC) successfully mitigated the saturating effects of specular reflections on wet asphalt and reflective lane markings.
- By normalizing the pixel-intensity shifts, this module maintained a continuous gradient flow within the activation maps, preventing the formation of visual blind spots.
- Semantic preprocessing and instance segmentation techniques enabled the network to successfully ignore background distractors like thermoplastic markings and urban debris.
- This targeted masking strategy directly lowered False Positive Rates, making automated detection systems viable for real-world municipal maintenance prioritization.
- A distinct paradigm shift toward implicit preprocessing was observed, where mathematical equivalents of traditional filters are embedded directly into the network architecture.
- Advanced frameworks utilizing attention mechanisms and cross-scale fusion reduced parameter counts by up to 88% while increasing mAP by 11.1% over baseline models.

In conclusion, the successful automation of UAV-based road inspection requires the meticulous optimization of the optical input field rather than a singular focus on network depth. By matching specific environmental degradations with targeted mathematical countermeasures, engineers can effectively overcome the physical limitations of aerial photogrammetry. As the field moves toward real-time edge inference on lightweight platforms, the integration of these spatial and radiometric filtering principles will be essential for the deployment of truly autonomous infrastructure monitoring systems.

References

[1] Liu J, Lemus-Romani J, Rueda EJ, Becerra-Rozas M, Astorga G. Identification of Pathologies in Pavements by Unmanned Aerial Vehicle (UAV): A Systematic Literature Review. Drones. 2026;10(2):90. https://doi.org/10.3390/drones10020090

[2] Silva LA, Leithardt VRQ, Batista VFL, González GV, Santana JFDP. Automated Road Damage Detection Using UAV Images and Deep Learning Techniques. IEEE Access. 2023; 11:62918-31. https://doi.org/10.1109/ACCESS.2023.3287770

[3] Zhang W, Li X, Wang L, Zhang D, Lu P, Wang L, et al. A Lightweight Method for Road Defect Detection in UAV Remote Sensing Images with Complex Backgrounds and Cross-Scale Fusion. Remote Sensing. 2025;17(13):2248. https://doi.org/10.3390/rs17132248

[4] Britto JGM, Satheesh DRN, Adharsh N, Nayab SD. Deep learning-based automated detection of road surface damage using UAV imagery. International Journal of Engineering Research and Science & Technology. 2025;21(4):691-8. https://doi.org/10.62643/ijerst.2025.v21.n4.pp691-698

[5] Arya D, Maeda H, Ghosh SK, Toshniwal D, Sekimoto Y. RDD2022: A multi-national image dataset for automatic road damage detection. Geoscience Data Journal. 2024;11(4):846-62. https://doi.org/10.1002/gdj3.260

[6] Chen L, Wu P, Tan W, Li H, Chen H, Zeng N. A Novel UAV-Based Road Damage Detection Algorithm with Lightweight Convolution and Attention Mechanism. International Journal of Network Dynamics and Intelligence. 2025;4(4):100025. https://doi.org/10.53941/ijndi.2025.100025

[7] Chen XB, Liu C, Chen L, Zhu XD, Zhang YH, Wang CX. A Pavement Crack Detection and Evaluation Framework for a UAV Inspection System Based on Deep Learning. APPLIED SCIENCES-BASEL. 2024;14(3). https://doi.org/10.3390/app14031157

[8] Magdy A, Moustafa MS, Ebied HM, Tolba MF. Lightweight faster R-CNN for object detection in optical remote sensing images. Scientific Reports. 2025;15(1):16163. https://doi.org/10.1038/s41598-025-99242-y

[9] Zhang S, Liu Z, Wang K, Huang W, Li P. OBC-YOLOv8: an improved road damage detection model based on YOLOv8. PeerJ Computer Science. 2025;11: e2593. https://doi.org/10.7717/peerj-cs.2593

[10] Peng K-C. Joint Training of Image Generator and Detector for Road Defect Detection. arXiv preprint arXiv:250903465. 2025.

[11] Alshawabkeh S, Wu L, Dong D, Cheng Y, Li L. A hybrid approach for pavement crack detection using mask R-CNN and vision transformer model. Computers, Materials, & Continua. 2025;82(1):561-77. https://doi.org/10.32604/cmc.2024.057213
[12] Karkınlı A, Kesikoğlu A. Enhancement of UAV-Aerial Images Using Weighted Differential Evolution Algorithm. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2021;7(2):196-206. https://doi.org/10.32628/CSEIT217248
[13] Jiskani SM, Hussain T, Sahito AA, Shaikh F, Kumar L. Improving Insulators Detection Accuracy via Image Enhancement Techniques: Case of Indigenous Aerial Image Dataset. IEEE Access. 2024; 12:145582-9. https://doi.org/10.1109/ACCESS.2024.3474255
[14] Freitas G, Mota G, Júnior E, Albuquerque V. Impact of Image Enhancement Filters on Road Defect Detection Using Deep Learning. Anais do Computer on the Beach. 2025; 16:213-8.
[15] Zhang F, Zhang C, Yang H, Zhao L. Point Cloud Denoising with Principal Component Analysis and a Novel Bilateral Filter. Traitement du Signal. 2019; 36:393-8. https://doi.org/10.18280/ts.360503
[16] Nguyen NHT, Perry S, Bone D, Le HT, Nguyen TT. Two-stage convolutional neural network for road crack detection and segmentation. Expert Systems with Applications. 2021; 186:115718. https://doi.org/10.1016/j.eswa.2021.115718
[17] Li X, Chu Y, Chan T, Lo WL, Fu H. YOLO-ERCD: An Upgraded YOLO Framework for Efficient Road Crack Detection. Sensors. 2026;26(2):564. https://doi.org/10.3390/s26020564
[18] Sun W, Liu X, Lei Z. A Tunnel Crack Detection Method Based on an Unmanned Aerial Vehicle (UAV) Equipped with a High-Speed Camera and Crack Recognition Algorithm Using Improved Multi-Scale Retinex and Prewitt-Otsu. Drones. 2025;9(6):393. https://doi.org/10.3390/drones9060393
[19] Yao Z, Xu J, Hou S, Chuah MC. Cracknex: a few-shot low-light crack segmentation model based on retinex theory for uav inspections. 2024 IEEE International Conference on Robotics and Automation (ICRA); May 13-17, 2024; Yokohama, Japan: IEEE. p. 11155-62. https://doi.org/10.1109/ICRA57147.2024.10611660
[20] Li C, Teng Y, Li C, Huang Z, Jiang L. Lightweight detection of grapevine downy mildew using UAV based on image information enhancement and adaptive dual-path fusion attention. BMC Plant Biology. 2025;26(1):16. https://doi.org/10.1186/s12870-025-07849-0
[21] Kesikoglu MH. Comparative analysis of R/SWIR, NIR/SWIR, and NDSI classification methods for determination of snow cover change on Erciyes Mountain, Türkiye using Landsat satellite images. Research on Engineering Structures and Materials. 2025;11(4):1653-65. https://doi.org/10.17515/resm2025-524ml1110rs
[22] Kesikoglu MH. Enhancing Snow Detection through Deep Learning: Evaluating CNN Performance Against Machine Learning and Unsupervised Classification Methods. Water Resources Management. 2025;39(12):6053-71. https://doi.org/10.1007/s11269-025-04240-4
[23] Nie Y, Wu W, Shan J, Peng H, Guo F, Liu Y, et al. Road Marking Distress Detection and Assessment Based on UAV Imagery. Materials. 2026;19(5):992. https://doi.org/10.3390/ma19050992
[24] Esfahani M, Wang H. Robust Glare Detection: Review, Analysis, and Dataset Release 2021 [Available from: https://arxiv.org/abs/2110.06006].
[25] Shim S, Kim J, Lee S-W, Cho G-C. Road surface damage detection based on hierarchical architecture using lightweight auto-encoder network. Automation in Construction. 2021; 130:103833. https://doi.org/10.1016/j.autcon.2021.103833
[26] Lee J-H, Gwon G-H, Kim I-H, Jung H-J. A Motion Deblurring Network for Enhancing UAV Image Quality in Bridge Inspection. Drones. 2023;7(11):657. https://doi.org/10.3390/drones7110657
[27] Kesikoglu MH, Ozkan C, Kaynak T. The impact of impervious surface, vegetation, and soil areas on land surface temperatures in a semi-arid region using Landsat satellite images enriched with Ndaisi method data. Environmental Monitoring and Assessment. 2021;193(3):143. https://doi.org/10.1007/s10661-021-08916-3

[28] Kesikoglu MH, Atasever UH, Ozkan C, Besdok E. The Usage of RusBoost Boosting Method for Classification of Impervious Surfaces. Int Arch Photogramm Remote Sens Spatial Inf Sci. 2016; XLI-B7:981-5. https://doi.org/10.5194/isprs-archives-XLI-B7-981-2016

[29] Günen MA, Atasever ÜH, Taşkanat T, Beşdok E. USAGE OF UNMANNED AERIAL VEHICLES (UAVs) IN DETERMINING DRAINAGE NETWORKS. Nature Sciences. 2019;14(1):1-10. https://doi.org/10.12739/NWSA.2019.14.1.4A0062

[30] Zhu L, Luo Z, Bi Y, Zhang Z. Cold Region Road Surface Defect Detection Using an Intelligent Machine Vision Method. Journal of Cold Regions Engineering. 2025;39(3):04025024. https://doi.org/10.1061/JCRGEI.CRENG-874

[31] Tang Z, Chamchong R, Pawara P, editors. A comparison of road damage detection based on YOLOv8. 2023 International Conference on Machine Learning and Cybernetics (ICMLC); 09-11 July 2023. Adelaide, Australia: IEEE. https://doi.org/10.1109/ICMLC58545.2023.1032799

[32] Li D, Duan Z, Hu X, Zhang D, Zhang Y. Automated classification and detection of multiple pavement distress images based on deep learning. Journal of Traffic and Transportation Engineering (English Edition). 2023;10(2):276-90. https://doi.org/10.1016/j.jtte.2021.04.008

[33] Dorafshan S, Thomas RJ, Maguire M. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. Construction and Building Materials. 2018; 186:1031-45. https://doi.org/10.1016/j.conbuildmat.2018.08.011

[34] Dorafshan S, Thomas RJ, Maguire M. SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks. Data in Brief. 2018; 21:1664-8. https://doi.org/10.1016/j.dib.2018.11.015

[35] Maslan J, Cicmanec L. A System for the Automatic Detection and Evaluation of the Runway Surface Cracks Obtained by Unmanned Aerial Vehicle Imagery Using Deep Convolutional Neural Networks. Applied Sciences. 2023;13(10):6000. https://doi.org/10.3390/app13106000

[36] Liu W, Ren G, Yu R, Guo S, Zhu J, Zhang L. Image-adaptive YOLO for object detection in adverse weather conditions. Proceedings of the 36th AAAI Conference on Artificial Intelligence; 2022. p. 1792-800. https://doi.org/10.1609/aaai.v36i2.20072

[37] Aina J, Haghi N, Famewo B, Lambert T, Owolabi D, Efe S. Real-Time Road Damage Detection Using YOLOv8. International Conference on Transportation and Development 2025. p. 411-8. https://doi.org/10.1061/9780784486191.036

[38] Zhang X, Wang M, Liu C, Xu H, Yang Y. Crack Detection and Quantification of Rural Cement Concrete Pavements Based on Deep-Learning Fusion Model. Journal of Transportation Engineering, Part B: Pavements. 2026;152(1):04025062. https://doi.org/10.1061/JPEODX.PVENG-1969

[39] Kim Y-M, Kim Y-G, Son S-Y, Lim S-Y, Choi B-Y, Choi D-H. Review of Recent Automated Pothole-Detection Methods. Applied Sciences. 2022;12(11):5320. https://doi.org/10.3390/app12115320

[40] Schiopu I, Saarinen J, Kettunen L, Tabus I. Pothole detection and tracking in car video sequence. 39th International Conference on Telecommunications and Signal Processing (TSP); 27-29 June 2016; Vienna, Austria: IEEE; 2016. https://doi.org/10.1109/TSP.2016.7760975

[41] Fan J, Bocus MJ, Hosking B, Wu R, Liu Y, Vityazev S, et al., editors. Multi-Scale Feature Fusion: Learning Better Semantic Segmentation For Road Pothole Detection. 2021 IEEE International Conference on Autonomous Systems (ICAS); 2021 11-13 Aug. 2021; Montreal, QC, Canada: IEEE. https://doi.org/10.1109/ICAS49788.2021.9551165

[42] Yan H, Zhang J. UAV-PDD2023: A benchmark dataset for pavement distress detection based on UAV images. Data in Brief. 2023; 51:109692. https://doi.org/10.1016/j.dib.2023.109692