Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00117-8

Original Article

Automatic Assessment of Road Marking Surface Defects Using Computer Vision

Xiaotian Jiang [1, 2]

1. *Innovation Technology Department, Guizhou Expressway Industry Co. Ltd., Guiyang, CHN*
2. *Research Department, Guizhou Highway Transportation M&M Technology Innovation Center, Guiyang, CHN*

Received: March 04, 2026 | Review began: March 31, 2026 | Review ended: June 10, 2026 | Published: June 11, 2026

Abstract

Automated assessment of road marking condition is essential for data-driven pavement maintenance, yet existing methods either rely on large annotated datasets or lack interpretable degradation grading. This paper proposes a training-free framework that quantifies road marking deterioration through three physics-informed indicators: the Luminance Loss Index (LLI), which captures intensity attenuation; the Crack Index (CI), which measures surface fracturing; and the Spalling Area Ratio (SAR), which reflects structural loss relative to the original marking geometry. The three indicators are fused into a composite degradation score S via weighted linear combination, and a grid search procedure optimises the fusion weights and classification thresholds to assign each marking segment to one of three condition grades (Normal, Minor Damaged, or Damaged) without requiring annotated training data. Evaluated on a dataset of 1,672 line segment instances under 5-fold cross-validation, the proposed method achieves an accuracy of 0.7111 ± 0.0265 and a macro-F1 of 0.7175 ± 0.0247, outperforming an Otsu thresholding baseline by 7.7 percentage points on both metrics. Ablation experiments confirm that SAR provides the strongest discriminative signal, while CI improves Damaged class recall by 6.6 percentage points over the best two-indicator combination (LLI+SAR), reducing the risk of overlooking severely degraded markings. The framework requires no deep learning infrastructure and produces interpretable condition grades suitable for operational road safety management in scenarios where large annotated datasets are unavailable.

Categories: Image Processing and Analysis, Computer Vision, IoT Applications

Keywords: road marking inspection, surface defect assessment, image processing, interpretable framework, intelligent infrastructure inspection, computer vision

Introduction

Road markings serve as a fundamental component of traffic management infrastructure, providing critical visual guidance for vehicle navigation, lane discipline, and pedestrian safety [1,2]. Prolonged exposure to traffic loading, ultraviolet radiation, and mechanical abrasion causes progressive deterioration of marking materials, which is commonly reflected in reduced visibility, retroreflectivity loss, and surface degradation [3,4]. Failure to detect and remediate such defects in a timely manner may increase traffic risk, particularly under low-visibility conditions such as nighttime driving or adverse weather [5,6]. Systematic and accurate condition assessment of road markings has therefore become an increasingly important component of road asset management and maintenance programs [7,8].

Traditional inspection relies predominantly on manual field surveys, in which trained inspectors evaluate marking quality against standardized criteria [9,10]. While interpretable, these methods are labor-intensive, time-consuming, and subject to inter-rater variability, making large-scale or high-frequency monitoring impractical [11-13]. Instrument-based methods, including mobile retroreflectometers, offer greater objectivity but require specialized equipment and controlled measurement conditions, limiting deployment to periodic rather than continuous monitoring [14].

Digital image processing has opened new avenues for automated assessment. Early computer vision approaches applied global thresholding techniques such as Otsu binarization to estimate degradation based on darkened pixel ratios [15,16]. While computationally efficient, these methods treat luminance as the sole condition indicator and are unable to distinguish between different defect types or capture structural damage independently. Their sensitivity to illumination variation, shadow interference, and pavement color diversity further limits reliability in real-world deployment [17,18].

More recently, deep learning methods, particularly convolutional neural networks and encoder-decoder architectures such as U-Net, have demonstrated strong performance in pavement defect segmentation tasks [19,20]. However, their adoption in road marking assessment remains constrained by three practical limitations. Supervised models require large annotated datasets that are costly to compile for road marking defects, which exhibit high intra-class variability and imbalanced class distributions [21,22]. Trained models also show limited generalizability to new road environments or camera configurations without retraining [23]. Furthermore, the computational demands of deep network inference are incompatible with lightweight inspection systems commonly used in routine maintenance operations [24]. These constraints motivate the development of assessment frameworks that achieve reliable performance without dependence on large-scale training data or high-performance computing infrastructure.

This study proposes a training-free, physics-informed road marking condition assessment framework based on the fusion of three complementary defect indicators: the Luminance Loss Index (LLI), the Crack Index (CI), and the Spalling Area Ratio (SAR). Each indicator targets a distinct physical failure mode. LLI quantifies luminance degradation through relative intensity reduction within the marking region. CI characterizes structural cracking through morphological analysis of elongated dark edge components. SAR estimates coating spallation extent by computing the proportion of surface discontinuity within the marking region. The three indicators are normalized and fused into a composite deterioration score through weighted linear combination, with parameters jointly optimized to maximize multi-class classification performance. The framework requires no neural network training, operates on standard RGB imagery, and produces condition grades directly interpretable within existing road maintenance classification systems. The proposed approach offers a practical and deployable solution for routine road marking inspection in data-scarce maintenance environments.

Materials And Methods

Overview of the assessment framework

The proposed framework processes road marking images through six sequential stages to produce a standardized condition grade. First, raw RGB images are acquired from field inspection, with each image containing one or more annotated marking regions defined by polygon segmentation masks. Second, each marking region undergoes image preprocessing, including median filtering for noise suppression and perspective transformation for geometric correction, to ensure that subsequent feature measurements are not confounded by imaging artifacts or viewpoint distortion. Third, three defect indicators are independently extracted from each preprocessed region: the LLI quantifies brightness degradation, the CI characterizes structural cracking, and the SAR measures coating loss extent. Fourth, the three raw indicator values are normalized against optimized reference values and fused into a single composite deterioration score S through weighted linear combination. Fifth, the normalization reference values, fusion weights, and classification thresholds are jointly optimized on the training partition of each cross-validation fold by maximizing macro-averaged F1-score through grid search. Finally, each marking region is assigned a condition grade (Normal, Minor Damaged, or Damaged) by comparing its composite score S against two optimized thresholds. The following sections describe each stage in detail.

Image preprocessing

Raw inspection images frequently contain impulsive (salt-and-pepper) noise introduced by sensor characteristics and surface texture irregularities, which can interfere with edge detection and pixel-level intensity measurements in subsequent processing stages. To mitigate this, a median filter with a 5×5 kernel is applied to each input image prior to feature extraction. Median filtering effectively suppresses this noise while preserving edge sharpness, making it preferable to Gaussian smoothing for this application [25-27]. The filtered image is then converted to grayscale to reduce computational complexity and to provide a luminance representation suitable for all three defect indicators.

Perspective distortion is a common artifact in road surface imagery acquired from vehicle-mounted or handheld cameras, where oblique viewing angles cause geometric deformation of the marking region and introduce systematic errors in area-based measurements [28]. Perspective transformation is applied to the grayscale image prior to feature extraction, correcting this distortion by mapping the marking region to a frontal orthographic view using a homographic transformation estimated from the boundary coordinates of the marking polygon [29,30]. This step ensures that pixel area ratios computed in SAR and crack density estimates in CI correspond accurately to physical surfae proportions. In the present experiments, the dataset images had been geometrically normalised during annotation, and perspective transformation was therefore not applied. The step is nonetheless described here for completeness, as it would be required when deploying the framework on uncorrected field imagery.

Defect feature extraction

Road marking deterioration manifests through three physically distinct failure modes: luminance degradation due to pigment fading, structural cracking due to material fracture, and surface spallation due to coating delamination. Each failure mode is captured by a dedicated indicator, and the three indicators are designed to target physically distinct degradation phenomena, ensuring that their fusion provides complementary rather than redundant information.

Luminance loss detection

As road marking paint ages, photooxidation and abrasion progressively deplete the reflective pigment and glass bead content of the coating surface, resulting in a measurable reduction in the mean luminance of the marking region [31]. The LLI quantifies this degradation by comparing the mean pixel intensity of the marking region against an upper reference value derived from the brightest pixels within the same region. LLI is defined as:

$$LLI = \frac{R_{ref} - R_{mean}}{R_{ref}} \tag{1}$$

where R_{ref} is the 95th percentile grayscale intensity within the marking polygon, used as an intra-region upper reference approximating the luminance of the least attenuated surface area, and R_{mean} is the mean grayscale intensity of all pixels within the polygon. By using an intra-region reference rather than a fixed absolute threshold, the formulation is robust to global illumination variation and camera exposure differences across images. A value of LLI = 0 indicates uniform luminance with no detectable degradation, while values approaching 1 indicate severe and widespread brightness reduction. To ensure numerical stability, pixels with $R_{ref} < 1$ are assigned LLI = 0 to avoid division instability in near-black regions. The resulting values are additionally clipped to the unit interval to prevent negative outputs caused by measurement noise.

Crack detection

Surface cracking represents a structurally distinct failure mode in which the marking coating fractures along linear trajectories under mechanical stress and thermal cycling [32]. Unlike luminance degradation, which affects the marking surface uniformly, cracks manifest as narrow elongated dark features embedded within an otherwise intact coating. Neither LLI nor SAR is sensitive to this pattern: LLI responds to mean brightness reduction rather than localized linear features, and SAR requires closed edge contours enclosing filled regions and is therefore insensitive to isolated linear edge fragments that do not form enclosed boundaries. CI is therefore introduced as a dedicated indicator for structural cracking.

CI is computed through a three-stage procedure. First, Canny edge detection is applied to the preprocessed grayscale image using adaptive thresholds derived from the Otsu binarization value T_{otsu} of the image, with the low and high thresholds set to $0.4 \cdot T_{otsu}$ and $0.85 \cdot T_{otsu}$, respectively. The detected edges are masked to retain only those falling within the marking polygon. Following edge detection, a morphological closing operation with a 3 × 3 kernel is applied to connect short discontinuities within crack edges. Second, connected component analysis is performed on the resulting binary edge map, and each component is evaluated against three criteria to determine whether it corresponds to a genuine crack: the component area must be no less than 30 pixels, the aspect ratio of its bounding box must be no less than 2.5, and the mean grayscale intensity of the component must be lower than 90% of the marking region mean or lower than an absolute threshold of 160, which was determined empirically based on the grayscale intensity distribution of the dataset. These criteria jointly enforce the elongated morphology and relative darkness characteristic of surface cracks while suppressing false detections from pavement texture and marking boundary edges. Third, the total area of all accepted crack components A_{crack} is divided by the total marking polygon area A_M to yield CI:

$$CI = \frac{A_{crack}}{A_M} \tag{2}$$

where A_{crack} is the cumulative pixel area of all connected components satisfying the crack criteria, and A_M is the total pixel area of the marking polygon. Higher values of CI indicate greater surface crack density within the marking region, while a value of CI = 0 indicates no detectable cracking.

Spalling area detection

The SAR quantifies the proportion of a road marking's surface area that has undergone large-scale coating detachment. Unlike LLI, which captures overall luminance attenuation, or CI, which targets fine linear cracks, SAR is specifically designed to detect macro-scale spalling, which represents the most structurally severe form of road marking degradation.

SAR is computed through a four-stage morphological pipeline applied within the segmented marking region $\mathcal{M}$:

Stage 1 - Adaptive edge detection: Canny edge detection is applied to the grayscale image with thresholds derived from the Otsu binarization value T_{otsu} of the marking region:

$$T_{low} = 0.3 \cdot T_{otsu} \tag{3}$$

$$T_{high} = 0.75 \cdot T_{otsu} \tag{4}$$

Using region-adaptive thresholds, rather than global fixed values, ensures consistent sensitivity across images with varying overall brightness. The lower thresholds relative to those used in CI reflect the larger spatial scale of spalled regions, which produce broader and more diffuse edge responses than fine linear cracks.

Stage 2 - Dilation: The resulting binary edge map is dilated with a 5 × 5 rectangular structuring element over two iterations. The 5 × 5 kernel size was selected to bridge gaps of up to two pixels between adjacent edge fragments while avoiding excessive merging of spatially distinct spalled regions. This step prevents a single large spall from being fragmented into multiple disconnected components.

Stage 3 - Morphological closing: A closing operation with the same kernel fills small holes within edge-enclosed regions, consolidating the detected spall boundaries into compact, filled blobs that approximate the true spatial extent of each spalled patch.

Stage 4 - Area ratio computation: Let $\tilde{E}$ denote the binary edge map after the dilation and closing operations described in Stages 2 and 3. The processed binary map is masked to the marking polygon $\mathcal{M}$, and SAR is defined as the fraction of mask pixels classified as spalled:

$$SAR = \frac{\left|\left\{p \in \mathcal{M} \mid \tilde{E}(p) > 0\right\}\right|}{|\mathcal{M}|} \tag{5}$$

Unlike LLI, which quantifies the overall reduction in reflective intensity across the marking surface, SAR operates on the structural geometry of the surface by detecting closed edge contours that enclose spalled patches. This edge-based formulation makes SAR sensitive to the spatial pattern of coating detachment rather than its photometric consequence, ensuring that SAR captures structural degradation aspects that are complementary to the photometric signal of LLI and the linear fracture patterns detected by CI.

Visual illustration of defect indicators

To illustrate the detection capability of the three proposed indicators, Figure 1 presents a representative road marking sample alongside the corresponding detection results for LLI, CI, and SAR. The sample shown is a severely degraded marking exhibiting multiple co-occurring damage types. This sample was selected to simultaneously illustrate all three detection mechanisms; in practice, individual indicators respond selectively according to the dominant failure mode present.

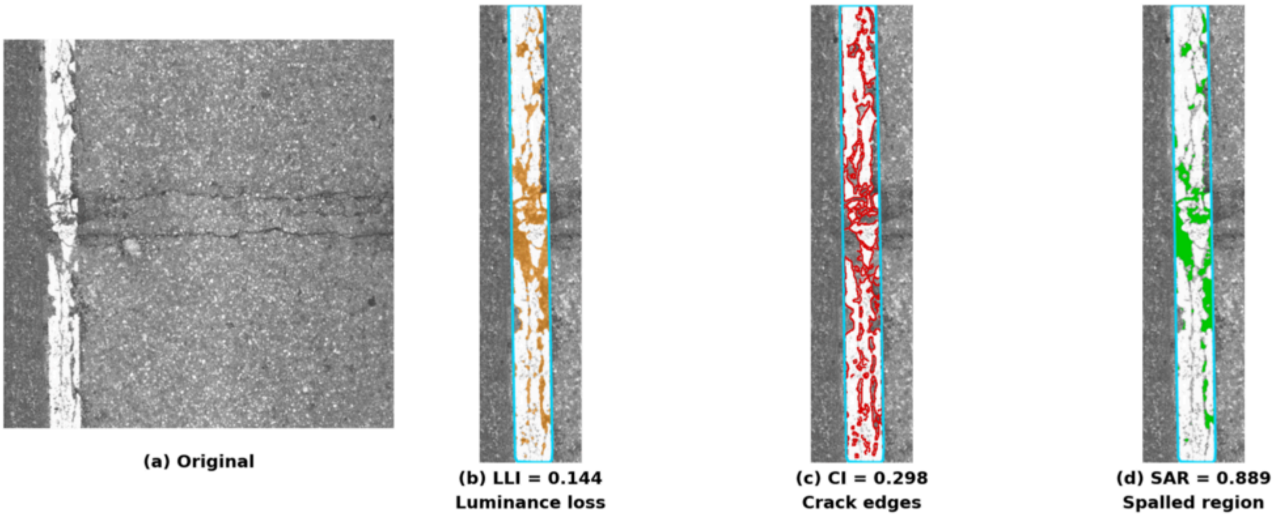

FIGURE 1: Illustration of the original road-marking sample (a) along with three defect indicators on road marking sample: (b) LLI, (c) CI, and (d) SAR

CI, Crack Index; LLI, Luminance Loss Index; SAR, Spalling Area Ratio

Figure 1(b) shows the pixel-level luminance deviation map used to visualise LLI, in which pixels whose intensity falls significantly below the 95th-percentile reference value are rendered in orange, illustrating the spatial distribution of brightness attenuation across the marking surface. Figure 1(c) shows the CI result, in which Canny-detected edge segments satisfying the elongation and darkness criteria are rendered in red, tracing the fine linear fractures distributed along the marking surface. Figure 1(d) presents the SAR result, in which the morphologically processed edge map captures closed contours enclosing spalled patches, highlighted in green. Unlike LLI, which measures overall luminance attenuation, SAR detects the spatial extent of discrete regions where the coating has physically separated from the substrate, leaving structurally missing patches regardless of their absolute brightness.

Although all three indicators respond to the same severely damaged sample, their spatial distributions are visually distinct: LLI captures diffuse global luminance attenuation, CI traces the fine linear crack network, and SAR localises discrete coating detachment patches through closed edge contours.

Indicator normalization and weighted fusion

Since LLI, CI, and SAR operate on different numerical scales, direct summation would bias the composite score toward the indicator with the largest absolute magnitude. Each indicator is therefore normalized by a reference value $R_{ref}^{(\cdot)}$ and clipped to the unit interval:

$$LLI_n = \min\left(\frac{LLI}{R_{ref}^{LLI}},\, 1\right) \tag{6}$$

$$CI_n = \min\left(\frac{CI}{R_{ref}^{CI}},\, 1\right) \tag{7}$$

$$SAR_n = \min\left(\frac{SAR}{R_{ref}^{SAR}},\, 1\right) \tag{8}$$

Each reference value $R_{ref}^{(\cdot)}$ defines the saturation threshold for its corresponding indicator: any raw value exceeding $R_{ref}^{(\cdot)}$ is treated as equivalent to the most severe degradation state and mapped to 1. The reference values R_{ref}^{LLI}, R_{ref}^{CI}, and R_{ref}^{SAR} are not set manually but are optimized jointly with the weighting coefficients through the grid search procedure described in the following section.

The three normalized indicators are then combined into a single composite score S through a weighted linear fusion:

$$S = w_1 \cdot LLI_n + w_2 \cdot CI_n + w_3 \cdot SAR_n \tag{9}$$

where w_1, w_2, $w_3 \geq 0$, and $w_1 + w_2 + w_3 = 1$. The composite score S serves as a unified measure of road marking degradation: a value approaching zero indicates a well-maintained marking with minimal detectable damage, while a value approaching one indicates severe degradation across multiple damage dimensions. The weighted formulation allows the relative contribution of each indicator to be calibrated to its discriminative capacity on the target dataset, rather than treating all three dimensions as equally informative. The optimal weighting reflects the relative discriminative contribution of each indicator, which is determined empirically through the parameter optimization procedure rather than assumed a priori.

Parameter optimization and condition grade classification

The composite score S depends on three sets of parameters: the reference values R_{ref}^{LLI}, R_{ref}^{CI}, and R_{ref}^{SAR} used for normalization; the fusion weights w_1, w_2, and w_3; and the classification thresholds n_{th} (the boundary between Normal and Minor Damaged) and m_{th} (the boundary between Minor Damaged and Damaged). Rather than setting these parameters manually, they are determined jointly through a grid search conducted on the training partition of each cross-validation fold, with Macro-F1 score as the optimization objective.

The search proceeds in two stages. In the first stage, a coarse grid is constructed by enumerating candidate values for each reference value and each weight pair (w_1, w_2), with $w_3 = 1 - w_1 - w_2$ determined implicitly by the unit-sum constraint. Candidate values for each reference value are drawn from a predefined range covering the observed distribution of each raw indicator, and weights are sampled from a uniform grid over the simplex $(w_1, w_2) : w_1, w_2 \geq 0,\ w_1 + w_2 \leq 1$. For each parameter combination, the composite score S is computed for all training samples, and the classification thresholds n_{th} and m_{th} are selected by exhaustive search over all candidate threshold pairs drawn from the sorted score values of the training samples, subject to the constraint $n_{th} < m_{th}$. In the second stage, a fine-grained threshold search is applied around the coarse optimum to refine n_{th} and m_{th} at higher resolution. This two-stage design reduces the computational cost of the joint parameter search while achieving threshold precision comparable to a full fine-grained enumeration. The combination yielding the highest Macro-F1 on the training partition is retained as the parameter set for that fold.

A road marking is assigned to one of three condition grades based on the following rules, where $0 \leq n_{th} < m_{th} \leq 1$ is enforced throughout the search:

$S < n_{th}$: Normal

$n_{th} \leq S < m_{th}$: Minor Damaged

$S \geq m_{th}$: Damaged

This classification scheme ensures that grade boundaries are learned from data rather than prescribed by domain heuristics, while remaining fully interpretable as threshold comparisons on a physically meaningful continuous score.

Evaluation metrics

Classification performance is evaluated using overall accuracy, Macro-F1 score, and per-class precision, recall, and F1 score. Overall accuracy measures the proportion of correctly classified samples across all three condition grades. Per-class precision and recall quantify, respectively, the fraction of predicted positives that are true positives and the fraction of true positives that are retrieved, for each grade individually. The per-class F1 score is the harmonic mean of precision and recall, providing a balanced measure that is sensitive to both false positives and false negatives within each class.

Macro-F1 is adopted as the primary performance metric, computed as the unweighted mean of per-class F1 scores:

$$\text{Macro-F1} = \frac{1}{3}\sum_{c=1}^{3} F1_c \tag{10}$$

where $F1_c$ denotes the F1 score of class c as defined above. This formulation treats each condition grade equally regardless of its frequency in the dataset, which is preferable to Micro-F1 in the presence of class imbalance. Micro-F1 aggregates contributions proportionally to class size and therefore tends to be dominated by the majority class, potentially concealing poor performance on minority classes such as Minor Damaged.

A confusion matrix is additionally reported to provide a complete picture of inter-class misclassification patterns, including the most safety-critical error type of predicting Damaged markings as Normal, which would result in severely deteriorated markings being left without maintenance intervention.

All metrics are reported as mean ± standard deviation computed across five stratified cross-validation folds. Stratification ensures that the class distribution within each fold reflects that of the full dataset, preventing folds from being dominated by a single condition grade. This ensures that performance estimates are not dependent on a single data partition and that the variability of the method across different training and test splits is made explicit.

Dataset and experimental setup

The dataset used in this study is the Road-marking Computer Vision Model, publicly available via Roboflow Universe [33]. The original dataset contains 3,360 images across three splits (train: 2,488; valid: 581; test: 291), with instance-level polygon segmentation annotations stored in YOLOv8 format, assigned to five condition categories: Normal_line, Minor_Damaged_line, Damaged_line, Faded_line, and Damaged_Arrow. The images consist of overhead daytime photographs of white road markings captured from a top-down perspective, and have been pre-processed with perspective correction prior to release, so the rectification step is not applied in these experiments. This pre-processing step is a characteristic of the dataset rather than a requirement of the proposed method; in deployment scenarios where perspective correction has not been applied, this step would need to be performed prior to indicator extraction.

This study focuses exclusively on straight-line marking instances and retains three condition grades: Normal_line, Minor_Damaged_line, and Damaged_line. Faded_line annotations are excluded because fading represents a luminance attenuation phenomenon that is subsumed by the LLI indicator, and these instances can be accommodated within the Normal or Minor Damaged grades depending on the degree of attenuation. Damaged_Arrow instances are excluded as their irregular geometry introduces confounding shape-related edge responses that are not representative of linear marking degradation. After filtering, the working dataset comprises 1,672 annotated straight-line instances, distributed across three condition grades, as reported in Table 1.

Condition Grade	Instances	Proportion
Normal	601	35.9%
Minor Damaged	657	39.3%
Damaged	414	24.8%
Total	1,672	100%

TABLE 1: Dataset composition and class distribution of straight-line marking instances used in this study

All experiments are conducted under a stratified 5-fold cross-validation protocol (implemented using scikit-learn's StratifiedKFold), which ensures that the class distribution within each training and test fold mirrors that of the full dataset. Parameter optimization follows the grid search procedure described above and is performed independently on each training fold to prevent any leakage of test-set information into the parameter selection process. Performance metrics are reported as mean ± standard deviation aggregated across the five test folds.

Results And Discussion

This section presents the experimental results and discussion in four parts. First, the discriminative capacity of the three proposed indicators is examined through a feature separability analysis. Second, an ablation study evaluates the contribution of each indicator combination to overall classification performance. Third, the proposed method is compared against a traditional Otsu thresholding baseline, and the design choice of using three indicators over two is justified from a safety-oriented perspective. Finally, the distribution of composite scores and the validity of the learned classification thresholds are analyzed.

Feature separability analysis

Before evaluating the full classification pipeline, the discriminative capacity of each individual indicator is examined by analyzing its statistical distribution across the three condition grades. Figure 2 shows the box plots of LLI, CI, and SAR for Normal, Minor Damaged, and Damaged marking instances.

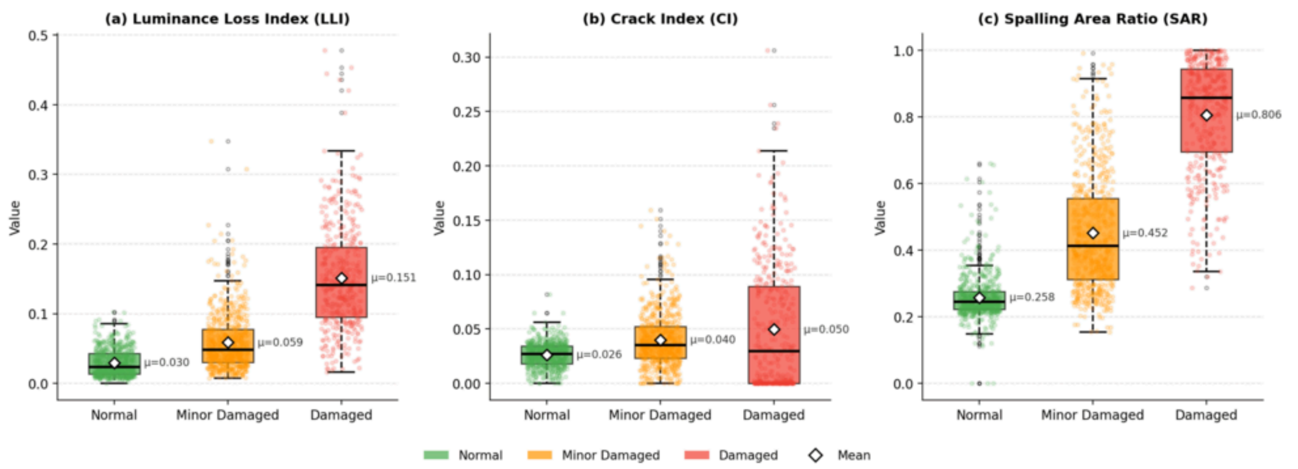

FIGURE 2: Box plots of (a) LLI, (b) CI, and (c) SAR across three road-marking condition grades

SAR exhibits the strongest class separability among the three indicators. The mean SAR values increase monotonically from 0.258 (Normal) to 0.452 (Minor Damaged) and 0.806 (Damaged), with minimal overlap between the Normal and Damaged distributions. This sharp gradient reflects the structural nature of spalling: as coating detachment progresses, an increasing fraction of the marking surface is replaced by dark voids, producing a consistently large spalling area ratio in severely degraded samples.

LLI also shows a clear monotonic trend, with mean values of 0.030, 0.059, and 0.151 for Normal, Minor Damaged, and Damaged grades, respectively. The separation between Normal and Damaged is substantial, though the Minor Damaged distribution overlaps with both adjacent grades, indicating that luminance attenuation alone is insufficient to reliably distinguish early-stage degradation from well-maintained markings.

CI presents the weakest separability, with mean values of 0.026, 0.040, and 0.050 across the three grades. The inter-class differences are small relative to within-class variance, particularly for the Damaged grade where the standard deviation (0.057) exceeds the mean (0.050). This behaviour is physically consistent: in severely damaged markings, large-scale coating detachment eliminates the continuous surface that would otherwise produce elongated crack edges, causing CI to respond weakly precisely where damage is most severe. Despite its limited individual discriminative power, CI provides information that is largely orthogonal to SAR and LLI, and its contribution to separating Normal from Minor Damaged instances motivates its inclusion in the multi-indicator fusion.

The optimized parameter values across all five folds are reported in Table 2. The fusion weights remain entirely stable ($w_1 = 0.20$, $w_2 = 0.20$, and $w_3 = 0.60$) across all folds, confirming that the grid search consistently identifies SAR as the dominant contributor, consistent with its superior separability observed in Figure 2.

Fold	R_{LLI}	R_{CI}	R_{SAR}	w_1	w_2	w_3	n_{th}	m_{th}
1	0.26	0.38	0.29	0.2	0.2	0.6	0.6043	0.7087
2	0.1	0.38	0.29	0.2	0.2	0.6	0.6517	0.7926
3	0.18	0.32	0.29	0.2	0.2	0.6	0.6353	0.7429
4	0.26	0.38	0.29	0.2	0.2	0.6	0.6158	0.7083
5	0.26	0.38	0.29	0.2	0.2	0.6	0.6222	0.7125
Mean	0.212	0.368	0.29	0.2	0.2	0.6	0.6259	0.733

TABLE 2: Optimized parameter values across five cross-validation folds

Composite score S distribution and threshold analysis

The composite score S is computed for each instance using the mean optimized parameters reported in Table *2*, and fold-specific parameters are used during the actual classification evaluation. The per-grade statistics of S are summarised in Table *3*.

Condition Grade	Mean S	Std S
Normal	0.543	0.089
Minor Damaged	0.647	0.069
Damaged	0.739	0.053

TABLE 3: Composite score S statistics by condition grade

The distributions exhibit a clear monotonic ordering across the three grades, confirming that the weighted fusion of LLI, CI, and SAR produces a well-ordered and interpretable composite degradation signal. The Normal-Minor Damaged mean gap (0.104) is comparable in magnitude to the Minor Damaged-Damaged gap (0.092), but the substantially larger within-class standard deviation of Normal ($\sigma = 0.089$) relative to Damaged ($\sigma = 0.053$) indicates that classification errors are expected to concentrate at the Normal-Minor Damaged boundary rather than the Minor Damaged-Damaged boundary. This distributional spread is an inherent property of gradual pavement marking degradation, where the transition from intact to lightly worn is continuous rather than discrete.

The grid search optimization yields mean thresholds of $n_{th} = 0.622 \pm 0.019$ and $m_{th} = 0.713 \pm 0.032$ across the five folds (Table 2), partitioning the S axis into three contiguous decision regions corresponding to Normal ($S < n_{th}$), Minor Damaged ($n_{th} \leq S < m_{th}$), and Damaged ($S \geq m_{th}$). The cross-fold standard deviations of n_{th} (± 0.019) and m_{th} (± 0.032) are small relative to the inter-class score separation, indicating that the decision boundaries are insensitive to the specific training partition and that the composite score S generalises consistently across different subsets of the dataset.

Ablation study on feature combinations

With the composite score S and its decision boundaries established in the preceding section, an ablation study is conducted to evaluate the individual and combined contributions of the three indicators. Seven feature combinations and one traditional baseline are compared under the same 5-fold cross-validation protocol: three single-indicator methods (LLI, CI, SAR), three two-indicator combinations (LLI+CI, LLI+SAR, CI+SAR), the proposed three-indicator combination (LLI+CI+SAR), and an Otsu thresholding baseline for reference. The complete results are reported in Table 4, and the performance profiles and aggregated confusion matrix for the proposed method are shown in Figure 3.

Method	Accuracy	Macro-F1	F1-Normal	F1-Minor	F1-Damaged
Baseline (Otsu)	0.6340±0.034	0.6404±0.033	0.6781±0.039	0.5025±0.065	0.7405±0.016
LLI only	0.6304±0.017	0.6446±0.015	0.6064±0.038	0.5793±0.032	0.7480±0.016
CI only	0.5401±0.021	0.5123±0.025	0.6029±0.017	0.4954±0.039	0.4387±0.052
SAR only	0.5120±0.015	0.4693±0.016	0.7386±0.039	0.1096±0.017	0.5597±0.009
LLI + CI	0.6286±0.015	0.6446±0.013	0.6179±0.032	0.5604±0.020	0.7554±0.015
LLI + SAR	0.7249±0.024	0.7290±0.020	0.7830±0.038	0.6702±0.031	0.7338±0.018
CI + SAR	0.6453±0.025	0.6155±0.023	0.7745±0.029	0.6258±0.033	0.4462±0.052
LLI+CI+SAR (Proposed)	0.7111±0.027	0.7175±0.025	0.7629±0.029	0.6448±0.034	0.7448±0.016

TABLE 4: Ablation study results: performance of eight feature combinations under five-fold cross-validation (mean ± SD)

CI, Crack Index; LLI, Luminance Loss Index; SAR, Spalling Area Ratio

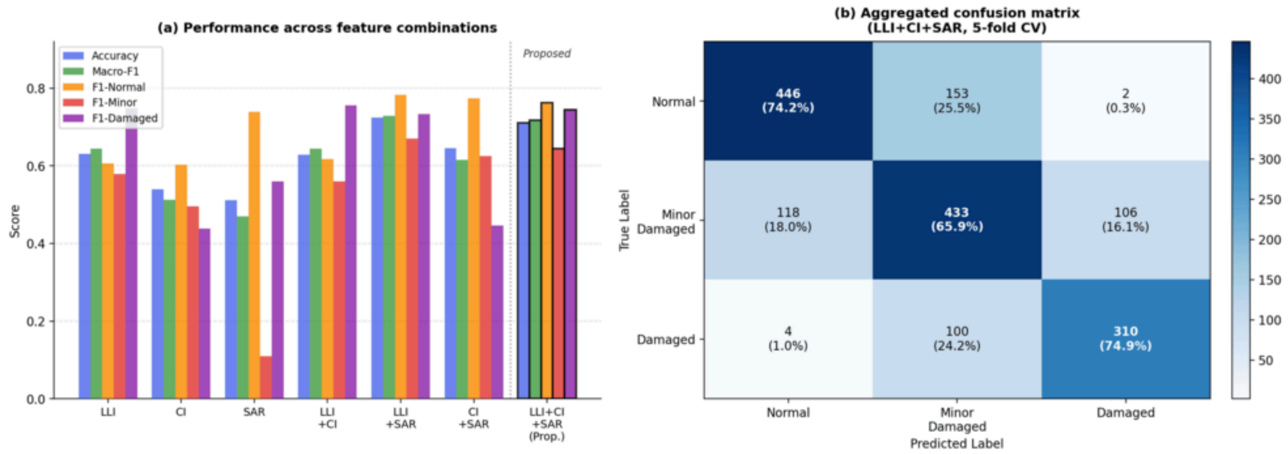

FIGURE 3: Ablation study results: (a) performance across feature combinations and (b) aggregated confusion matrix

CI, Crack Index; LLI, Luminance Loss Index; SAR, Spalling Area Ratio

Among single-indicator methods, LLI alone achieves the highest Macro-F1, outperforming CI and SAR. The poor performance of SAR alone is primarily attributable to its near-zero. Although SAR separates Normal from Damaged effectively, it provides almost no signal for identifying Minor Damaged instances, confirming that a single structural indicator is insufficient for three-class discrimination.

Among two-indicator combinations, LLI+SAR achieves the highest overall performance with Accuracy and Macro-F1, substantially outperforming all single-indicator methods. Adding CI to form the proposed LLI+CI+SAR combination slightly reduces overall Accuracy and Macro-F1 relative to LLI+SAR, but improves by 1.1 percentage points and, more importantly, raises Damaged recall from 0.683 to 0.749, reducing the rate at which severely degraded markings are missed. This trade-off is discussed further in the following section from a safety-oriented perspective.

The aggregated confusion matrix in Figure 3(b) reveals that out of 414 Damaged instances, only 4 are misclassified as Normal, representing the most safety-critical error type. The majority of misclassifications occur at the Normal-Minor Damaged boundary (153 Normal instances predicted as Minor Damaged; 118 Minor Damaged predicted as Normal), reflecting the inherently subtle visual differences between intact and lightly worn markings, which remain the primary source of misclassification even in the full three-indicator system.

Comparison with existing methods

Comparison With Baseline (Otsu Thresholding)

The proposed method is first compared against the Otsu thresholding baseline, which applies a single globally optimized intensity threshold to segment road markings without any composite indicator fusion. As reported in Table 4, the proposed LLI+CI+SAR method achieves an Accuracy of 0.7111 ± 0.0265 and a Macro-F1 of 0.7175 ± 0.0247, compared to 0.6340 ± 0.0336 and 0.6404 ± 0.0326 for the baseline, representing improvements of 7.7 percentage points in both metrics.

The most pronounced gains appear in F1-Minor, where the proposed method improves from 0.5025 ± 0.0648 to 0.6448 ± 0.0336, and F1-Normal (from 0.6781 to 0.7629). The baseline's weakness on Minor Damaged instances is attributable to its reliance on a single global intensity threshold, which cannot capture the multi-dimensional degradation signals, namely luminance attenuation, surface cracking, and coating detachment, that distinguish lightly worn markings from intact ones. The proposed method integrates all three complementary cues through weighted fusion, enabling more robust discrimination across all degradation grades.

For the safety-critical Damaged class, the improvement is comparatively modest, as severely degraded markings tend to exhibit unambiguous low-intensity responses that even a simple threshold can detect reliably. The proposed method maintains comparable fold stability, indicating that the additional indicators do not introduce instability in the Damaged class. Overall, the results confirm that composite indicator fusion provides consistent gains over single-threshold approaches, particularly for the intermediate degradation grade that is most difficult to classify and most prevalent within the dataset.

Safety-oriented justification for three-indicator fusion

Among all feature combinations evaluated in the ablation study, LLI+SAR achieves the highest overall Accuracy and Macro-F1, making it the strongest two-indicator baseline. The proposed LLI+CI+SAR method records slightly lower values on both overall metrics, a difference that warrants closer examination from a safety perspective.

The key distinction lies in the Damaged class, where false negatives carry the greatest operational consequence. The proposed method achieves a Damaged Recall of 0.7490 ± 0.0512, compared to 0.6834 ± 0.0371 for LLI+SAR, a gain of +6.6 percentage points. This improvement means that the proposed method correctly identifies substantially more severely degraded markings that would otherwise be overlooked by LLI+SAR. The corresponding F1-Damaged improves from 0.7338 to 0.7448, confirming that the recall gain outweighs the associated precision reduction on the F1 metric.

The trade-off is a moderate reduction in Damaged Precision, indicating that CI introduces some false positives for the Damaged class. This is consistent with CI's relatively low inter-class separability. The crack detection indicator occasionally assigns elevated degradation scores to markings with surface texture patterns that resemble crack edges, leading to marginal over-prediction of the Damaged grade. In road marking maintenance, however, a false positive that triggers an unnecessary inspection incurs far lower cost than a false negative that allows a severely degraded marking to remain unaddressed. The recall-oriented improvement conferred by CI therefore justifies its inclusion in the final fusion model, even at the cost of a modest reduction in overall Accuracy and Macro-F1.

Limitations and generalizability

The proposed method is designed for scenarios where annotated training data are unavailable, relying solely on physics-informed indicators and parameter optimisation on the available labelled data through cross-validation. While this design choice enables deployment in data-scarce contexts, it also defines the boundary conditions under which the method has been validated. The current dataset consists primarily of images acquired under normal daylight illumination, and the performance of the composite score S under adverse conditions such as night-time imaging, rainfall, and cast shadow has not been evaluated. In such conditions, the luminance-based indicator LLI and the crack-based indicator CI are particularly susceptible to systematic bias, as ambient light variations directly affect pixel intensity and edge contrast distributions. Future work should assess the robustness of the proposed indicators under diverse lighting and weather conditions, and consider incorporating illumination normalisation as a preprocessing step.

The method is not directly compared against deep learning-based approaches in this study, as standard supervised methods typically require substantially larger annotated datasets than are assumed available in the target deployment scenario. This is an intentional scope constraint rather than an oversight: the proposed framework is positioned as a training-free alternative for authorities and operators who lack the resources to collect and annotate large-scale datasets. Nevertheless, a systematic comparison with supervised methods on a larger, multi-source dataset would be valuable to quantify the performance gap and identify the conditions under which the additional cost of data annotation is justified.

The current validation is based on a single dataset collected from one geographic region and road network, which limits the assessment of cross-scenario generalisation. Variations in road marking materials, paint types, and pavement surface textures across different regions may affect the absolute values of LLI, CI, and SAR, potentially requiring recalibration of the fusion weights and classification thresholds. Although the cross-fold stability of the optimised parameters in Table *2* suggests that the learned parameters are robust to data partitioning within this dataset, external validation on independent datasets from different regions and road types remains an important direction for future work.

Conclusions

This paper presents a training-free framework for automated road marking condition assessment that integrates three complementary physics-informed indicators: LLI, CI, and SAR. The three indicators are fused into a single composite degradation score S through a weighted linear combination, and a grid search optimisation procedure determines the fusion weights and classification thresholds that partition S into three condition grades, namely Normal, Minor Damaged, and Damaged, without requiring large-scale annotated training data, relying only on the available labelled samples for parameter optimisation.

Under 5-fold cross-validation on a dataset of 1,672 line segment instances, the proposed method achieves an overall Accuracy of 0.7111 ± 0.0265 and a Macro-F1 of 0.7175 ± 0.0247, representing improvements of 7.7 percentage points over the Otsu thresholding baseline on both metrics. Ablation experiments confirm that all three indicators contribute to classification performance: SAR provides the strongest individual discriminative signal, while CI confers a targeted improvement in Damaged class recall, reducing the rate at which severely degraded markings are overlooked. The optimised fusion weights and thresholds remain stable across folds, indicating that the learned parameters are robust to data partitioning within the evaluated dataset. The proposed framework requires no deep learning infrastructure, requiring only a modest labelled dataset for parameter optimisation rather than large-scale annotation, and no specialised hardware beyond a standard road survey camera, making it suitable for lightweight deployment in routine pavement inspection workflows. The condition grades produced are directly interpretable in terms of maintenance priority, supporting evidence-based prioritisation of road marking maintenance interventions. Future work will focus on extending validation to multi-source datasets covering diverse road marking types, pavement materials, and imaging conditions, as well as exploring adaptive indicator weighting strategies that can accommodate scene-specific variations without manual recalibration.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Xiaotian Jiang

Acquisition, analysis, or interpretation of data: Xiaotian Jiang

Drafting of the manuscript: Xiaotian Jiang

Critical review of the manuscript for important intellectual content: Xiaotian Jiang

Supervision: Xiaotian Jiang

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** This work was funded by the Science and Technology Project of Shuicheng Operation Management Center, Guizhou Expressway Group Co., Ltd. (Grant No. SCZX2025092), for the project titled "Research and Application of High-Performance Thermoplastic Road Marking Paints". **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

Acknowledgements

This work was supported by the Science and Technology Project of Shuicheng Operation Management Center, Guizhou Expressway Group Co., Ltd. (Grant No. SCZX2025092), for the project titled "Research and Application of High-Performance Thermoplastic Road Marking Paints".

References

1. Babić D, Babić D, Fiolic M, Ferko M: Road markings and signs in road safety. Encyclopedia. 2022, 2:1738-1752. 10.3390/encyclopedia2040119
2. Babić D, Fiolić M, Babić D, Gates T: Road markings and their impact on driver behaviour and road safety: A systematic review of current findings. Journal of Advanced Transportation. 2020, 2020:7843743. 10.1155/2020/7843743
3. Soilán M, González-Aguilera D, Del-Campo-Sanchez A, Hernández-López D, Del Pozo S: Road marking degradation analysis using 3D point cloud data acquired with a low-cost Mobile Mapping System. Automation in Construction. 2022, 141:104446. 10.1016/j.autcon.2022.104446
4. Kim D: Development of predictive and evaluation models for the retro reflectivity performance of pavement lane markings. F1000Research. 2025, 14:907. 10.12688/f1000research.169127.2
5. Kučina I, Kosovec B, Jezidžić F, Ferko M, Babić D: The influence of road marking visibility on lateral vehicle position and driving speed. Acta Polytechnica CTU Proceedings. 2025, 52:69-76. 10.14311/app.2025.52.0069
6. Smadi O, Souleyrette RR, Ormand DJ, Hawkins N: Pavement marking retroreflectivity: Analysis of safety effectiveness. Transportation Research Record: Journal of the Transportation Research Board. 2008, 2056:17-24. 10.3141/2056-03
7. Xu S, Wang J, Wu P, Shou W, Wang X, Chen M: Vision-based pavement marking detection and condition assessment—A case study. Applied Sciences. 2021, 11:3152. 10.3390/app11073152
8. He H, Xu A, Han X, Wang H, Wang L, Su W: LiDAR perception and evaluation method for road traffic marking retroreflection. Transportation Research Record: Journal of the Transportation Research Board. 2023, 2677:258-279. 10.1177/03611981221145135
9. Abu-Samra S, Zayed T, Tabra W: Pavement condition rating using multiattribute utility theory. Journal of Transportation Engineering, Part B: Pavements. 2017, 143:04017011. 10.1061/jpeodx.0000011
10. Huang Y, Xu B: Automatic inspection of pavement cracking distress. Journal of Electronic Imaging. 2006, 15:013017. 10.1117/1.2177650
11. Yang X, Zhang J, Liu W, Jing J, Zheng H, Xu W: Automation in road distress detection, diagnosis and treatment. Journal of Road Engineering. 2024, 4:1-26. 10.1016/j.jreng.2024.01.005
12. Janouskova K, Rigotti M, Giurgiu I, Malossi C: Model-assisted labeling via explainability for visual inspection of civil infrastructures. Computer Vision - ECCV 2022 Workshops. Karlinsky L, Michaeli T, Nishino K (ed): Springer, Cham; 2023. 13807:244-257. 10.1007/978-3-031-25082-8_16
13. Yu J, Jiang J, Fichera S, Paoletti P, Layzell L, Mehta D, Luo S: Road surface defect detection—From image-based to non-image-based: A survey. IEEE Transactions on Intelligent Transportation Systems. 2024, 25:10581-10603. 10.1109/tits.2024.3382837
14. Lee S, Cho BH: Evaluating pavement lane markings in metropolitan road networks with a vehicle-mounted retroreflectometer and AI-based image processing techniques. Remote Sensing. 2023, 15:1812. 10.3390/rs15071812
15. Han H, Deng H, Dong Q, Gu X, Zhang T, Wang Y: An advanced Otsu method integrated with edge detection and decision tree for crack detection in highway transportation infrastructure. Advances in Materials Science and Engineering. 2021, 2021:9205509. 10.1155/2021/9205509
16. Safaei N, Smadi O, Masoud A, Safaei B: An automatic image processing algorithm based on crack pixel density for pavement crack detection and classification. International Journal of Pavement Research and Technology. 2021, 15:159-172. 10.1007/s42947-021-00006-4

17. Yu H, Zhu L, Li D, Wang Q, Liu X, Shen C: Comparative study on concrete crack detection of tunnel based on different deep learning algorithms. Frontiers in Earth Science. 2022, 9:817785. 10.3389/feart.2021.817785

18. Bang S, Park S, Kim H, Kim H: Encoder-decoder network for pixel-level road crack detection in black-box images. Computer-Aided Civil and Infrastructure Engineering. 2019, 34:713-727. 10.1111/mice.12440

19. Jia S: Semantic segmentation of pavement cracks based on an improved U-Net. Journal of Computing and Electronic Information Management. 2023, 10:6-10. 10.54097/jceim.v10i3.8672

20. Botezatu A-P, Burlacu A, Orhei C: A review of deep learning advancements in road analysis for autonomous driving. Applied Sciences. 2024, 14:4705. 10.3390/app14114705

21. Benlliure-Jimenez C, Penate-Sanchez A, Lorenzo-Navarro J, Castrillón-Santana M, Hernandez-Tejera FM: An annotation assistant for monitoring the electrical grid using aerial images. Knowledge-Based Systems. 2026, 336:115355. 10.1016/j.knosys.2026.115355

22. Chen Q, Fu S: Continuous pavement crack detection using ECA-enhanced instance segmentation of video images. Construction and Building Materials. 2025, 465:140247. 10.1016/j.conbuildmat.2025.140247

23. Zhou W, Zhan Y, Zhang H, Zhao L, Wang C: Road defect detection from on-board cameras with scarce and cross-domain data. Automation in Construction. 2022, 144:104628. 10.1016/j.autcon.2022.104628

24. Yang H, Ma T, Tong Z, Wang H, Wang N, Cheng H: Deployment strategies for lightweight pavement defect detection using deep learning and inverse perspective mapping. Automation in Construction. 2024, 167:105682. 10.1016/j.autcon.2024.105682

25. Huang T, Yang G, Tang G: A fast two-dimensional median filtering algorithm. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1979, 27:13-18. 10.1109/tassp.1979.1163188

26. Shah A, Bangash JI, Khan AW, Ahmed I, Khan A, Khan A, Khan A: Comparative analysis of median filter and its variants for removal of impulse noise from gray scale images. Journal of King Saud University Computer and Information Sciences. 2022, 34:505-519. 10.1016/j.jksuci.2020.03.007

27. Limbong H, Harahap LS, Gading RA: Comparison of median filter and Gaussian filter performance in removing salt and pepper noise. Journal of Artificial Intelligence and Engineering Applications (JAIEA). 2025, 4:1849-1854. 10.59934/jaiea.v4i3.1033

28. Li P, Zhou B, Wang C, Hu G, Yan Y, Guo R, Xia H: CNN-based pavement defects detection using grey and depth images. Automation in Construction. 2024, 158:105192. 10.1016/j.autcon.2023.105192

29. Chou CP, Leong KW, Chen AC, Lee YX: Road marking retroreflectivity study via a visual algorithm. International Journal of Pavement Research and Technology. 2020, 13:614-620. 10.1007/s42947-020-6001-x

30. Nie Y, Wu W, Shan J, Peng H, Guo F, Liu Y, Xiao J: Road marking distress detection and assessment based on UAV imagery. Materials. 2026, 19:992. 10.3390/ma19050992

31. Mohamed M, Skinner A, Abdel-Rahim A, Kassem E, Chang K: Deterioration characteristics of waterborne pavement markings subjected to different operating conditions. Journal of Transportation Engineering, Part B: Pavements. 2019, 145:04019003. 10.1061/jpeodx.0000101

32. Wei Z, Luo Z, Lei M, Li Y, Ma H, Peng J: Towards failure mechanism of hot melt road marking coatings. Transportation Research Procedia. 2025, 82:1280-1292. 10.1016/j.trpro.2024.12.125

33. markdamages: Road-marking Computer Vision Model. Accessed: June 08, 2026: https://universe.roboflow.com/markdamages/road-marking-zvq3d/dataset/7.