Full length article

Precise longitudinal crack detection via continuous texture reconstruction and deep segmentation

Peng An [a], Zhengru Ren [b], Long-Xiang Liu [a], Jia-Rui Lin [c], Yantao Yu [d], Yu-Tao Guo [a], Chao Hou [e], Zhen-Zhong Hu [a,*]

[a] Shenzhen International Graduate School, Tsinghua University, Shenzhen 518071, Guangdong, China
[b] School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[c] Department of Civil Engineering, Tsinghua University, Beijing 100084, China
[d] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, 999077, Hong Kong, China
[e] Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China

ARTICLE INFO

Keywords:
Automated pavement inspection
Longitudinal crack detection
Image stitching
Texture reconstruction
YOLOv8-seg
Morphological enhancement
Fine-grained segmentation

ABSTRACT

Pavement cracks are a major form of road infrastructure degradation, necessitating efficient and accurate detection for timely maintenance. Existing inspection methods rely either on labor-intensive manual surveys or automated systems constrained by high hardware costs and GPS dependency, limiting their flexibility for continuous surface assessment. This paper introduces a dual-channel crack detection model that integrates continuous pavement texture reconstruction with deep segmentation and high-precision boundary refinement algorithms, enabling on-site implementation and accuracy enhancement for longitudinal crack detection. A feature-based image stitching algorithm is developed to reconstruct continuous pavement textures from high-resolution images, enabling GPS-free crack localization. The proposed method further combines the YOLOv8-seg model with adaptive morphological operations to achieve pixel-level crack reconstruction. Comparative experiments reveal that the hybrid approach achieves superior segmentation performance with finer boundary delineation and improved branch recovery compared to the baseline model. The paper provides a practical solution for automated inspection and high-fidelity reconstruction of longitudinal cracks, effectively supporting pavement maintenance planning.

1. Introduction

The integrity and safety of transportation infrastructure are crucial to modern society, with road networks forming the backbone of the infrastructure. Cracks are one of the most common and critical issues among the various factors that can compromise pavement integrity. They not only affect the aesthetic appeal of pavements but also significantly impact the structural health and safety of the underlying pavements [1]. Therefore, the detection and localization of pavement cracks are of paramount importance for ensuring structural safety, estimating the remaining service life of pavements, and managing maintenance costs effectively [2]. Timely detection of these defects can prevent further deterioration and potential failures, making it a critical aspect of pavement maintenance and safety [3].

Current pavement distress surveys typically involve either low-cost manual visual inspection or automated mobile mapping systems (MMS). While MMS improves efficiency, it often incurs high costs and relies heavily on Global Navigation Satellite Systems (GNSS), limiting applicability in enclosed environments. Conversely, non-automated methods with fixed-point sampling remain subjective and hazardous. The accuracy of the inspections is often constrained by the inspector's skills and capabilities, which can be challenging, especially in cases of fine or weathered cracks [4]. The need for more efficient, accurate, and automated approaches to pavement crack detection has become increasingly urgent, particularly with the application of advanced technologies in infrastructure management [5] and data chain generation [6]. The integration reflects the growing emphasis on assessing the maturity of intelligent construction management systems in modern infrastructure operations [7]. Recent advancements in robotics and automation demonstrate that robotic systems can significantly reduce human error while improving inspection efficiency with lightweight methods [8]. AI-based modeling has proven to enhance time–history prediction accuracy and demonstrate its applicability across structural domains [9].

* Corresponding author.
E-mail address: huzhenzhong@tsinghua.edu.cn (Z.-Z. Hu).

In the realm of pavement inspection, various sensing technologies have been employed including multi-sensor systems [10] and optical solutions [11]. Optical methods, in particular, stand out due to their affordability and non-destructive operation [12]. The evolution of optical techniques has been significantly influenced by advancements in 3D reconstruction algorithms. Early synthesis methods relied on light field and structure-from-motion (SfM) technology, where the sparse point clouds is generated during camera calibration [13]. Subsequent developments in multi-view stereo (MVS) enabled full 3D reconstruction through geometric reprojection, though these methods often struggled with unconstructed areas or erroneous geometries [14]. Modern photogrammetry tools leverage GPU acceleration to produce high-quality 3D meshes from RGB images, though their processing times remain prohibitive for real-time applications [15]. Simultaneous localization and mapping (SLAM) techniques have been introduced to address the need for real-time mapping and localization, which enables the camera pose estimation and incremental scene reconstruction during movement. While SLAM frameworks such as ORB-SLAM3 provide robust localization, they are primarily optimized for 3D navigation and pose estimation rather than pixel-level pavement analysis. Their inability to generate textured models limits their utility in precise crack quantification [16].

Crack detection and characterization have emerged as the predominant applications of optical pavement inspection systems in real-world infrastructure maintenance, following a long trajectory of technological advancement. Early studies that utilizes single optical images for crack detection presents several challenges, including discontinuities in crack representation and difficulties in exactly localizing the exact position of cracks, leading to low efficiency and impracticality for engineering applications [17]. To address the fragmentation, image stitching becomes essential to generate continuous texture strip maps, ensuring the topological integrity of long-span cracks and providing a consistent coordinate system for subsequent analysis. Reconstruction-based methods, which include texture reconstruction, object reconstruction, and spatial reconstruction (both 2D and 3D), have shown promise in various fields such as automated vehicle perception, medical imaging and augmented reality, but have not yet been extensively applied to pavement crack detection [18]. Recent work by [19] addresses occlusion challenges in unmanned aerial vehicle(UAV)-based pavement imaging through innovative stitching algorithms, yet their vertical imaging constraints limit applicability in angled inspection scenarios.

Recent developments in 3D reconstruction and deep learning have significantly advanced the pavement distress analysis, offering more precise and scalable solutions. The SfM technique, when integrated with imagery from UAVs or vehicle-mounted systems, has been validated as a cost-effective alternative to laser scanning, producing sufficient dense point clouds for pavement surface modeling and damage analysis [20]. Extensions of SfM, such as PP-SfM, have supported downstream tasks like defect segmentation when integrated with transformer-based networks [21]. The methods enable accurate retrieval of depth profiles associated with surface deformations, potholes, and irregularities that are often missed by conventional techniques. In addition, multi-view stereo reconstruction based on deep learning has also enabled the efficient capture of pavement textures [22]. Low-cost hardware alternatives, such as Microsoft Kinect, have been explored, where the Kinect Fusion modules were developed to support dense surface reconstruction and facilitate integration with automated crack detection workflows [23]. These strategies collectively minimize data acquisition costs while preserving a high degree of geometric accuracy during the reconstruction of the surrounding environment. Parallel works have also focused on the reconstruction of macro-texture from monocular RGB images, aided by RGB-D datasets to improve accuracy in single-image depth estimation [24]. Bird's-eye view reconstruction further expand the potential of camera-based systems by providing low-cost and high-fidelity spatial representations for downstream navigation or inspection tasks [25].

Alongside improvements in 3D sensing, the rise of deep learning has driven a methodological shift in crack detection strategies. While traditional convolutional neural networks (CNNs) have shown reliable performance in damage detection [26], their high computational complexity and excessive memory consumption often limit real-time deployment. In contrast, the YOLO (You Only Look Once) series of detectors has gained attention for its ability to perform end-to-end detection through a single forward pass, significantly reducing inference time while maintaining competitive accuracy [27,28]. These detectors are particularly suitable for embedded and edge devices, where low latency is critical. The progression from region-based R-CNN frameworks [29,30] to streamlined architectures like YOLO and its variants has enabled the real-time detection of pavement cracks in complex environments, including urban roadways and pedestrian infrastructure [31]. For instance, the combination of PCGAN with YOLO-MF has proven effective in addressing data scarcity and supporting crack enumeration tasks, achieving accuracy rates exceeding 98% under controlled conditions [3]. Similarly, the deployment of YOLO-based models on UAV platforms has facilitated real-time inspection of multi-scale road crack, with improved accuracy and efficiency [32]. Furthermore, recent transformer-based segmentation and feature-fusion frameworks has achieved sub-pixel quantification accuracy and enhanced robustness under complex environments [33].

To further improve performance under resource constraints, recent research has introduced architectural enhancements within YOLO's detection head and neck modules. Lightweight variants of YOLO, such as YOLOv8, have been developed incorporating bidirectional feature pyramid structures and adaptive loss functions to balance detection precision and inference speed, ensuring real-time performance without the need for dedicated GPU acceleration [34]. Similarly, recent studies have introduced automatic lightweight architectures optimized via discrete particle swarm optimization (DPSO), achieving a better balance between detection accuracy and inference efficiency on resource-limited devices [35]. Hybrid frameworks have also emerged that bridge the gap between crack detection and pixel-level segmentation. For example, YOLOv5 integrated with a pyramid structure of dilated convolutions has shown promising results in object feature extraction and detection tasks [36]. In addition, integrating attention mechanisms has enabled more precise localization of fine-grained surface defects [37], yet it remains limited in enhancing the continuity of crack structures in complex or fragmented regions. In general, the above advancements ensure a clear trajectory towards more accurate, efficient, and deployable crack detection systems that are adaptable to a wide range of inspection scenarios.

Despite the advancements, there are challenges and limitations associated with current crack detection methods. Low-cost optical systems without high-end IMU integration for pavement detection still face issues such as discontinuities, localization difficulties, and angular misalignments. Recent work shows that post-processing refinement techniques can substantially improve continuity and topology preservation in thin-crack segmentation [38]. Image-based texture reconstruction methods, while promising, have not been widely adopted for the detection of continuous longitudinal cracks, and their effectiveness in real-world applications remains to be fully explored [39]. The reliance on dense, multi-view image acquisition, which typically involves time-consuming, manually guided data capture around the target region, presents a significant challenge for real-time implementation in mobile inspection settings. Deep learning methods, especially YOLO variants, have shown high accuracy but are often limited by the resolution and quality of annotated data and the computational resources required for training and deployment [40]. Furthermore, the deployment of detection networks in current inspection systems remains constrained by their reliance on high-precision GPS signals, especially in enclosed or prolonged signal-degraded environments such as tunnels and dense urban canyons [31]. Another persistent challenge lies in image stitching [41], especially due to non-linear distortions caused

by platform motion, which limits their performance in continuous surface reconstruction.

This paper presents an efficient and continuous method for the optical detection and localization of longitudinal pavement cracks (> 0.3 m length) [42], effectively addressing these limitations. Specifically, the proposed approach utilizes image stitching to generate continuous texture strip maps, which ensures the topological integrity of crack features to enhance segmentation accuracy while enabling GPS-free spatial localization. Compared with multifunctional inspection vehicles that typically rely on Distance Measuring Instruments (DMI) or high-precision GNSS for positioning, the proposed stitching and homography-based trajectory reconstruction offers a cost-effective alternative. The methodology provides unique value for inspecting narrow urban underpasses, tunnels, and localized maintenance sections where high-end hardware deployment is either cost-prohibitive or physically constrained. Unlike the UAV-based approaches, the ground-based system utilized in the research enables consistent imaging angles and reduces occlusion risks through controlled acquisition geometry.

The principal contributions of the research are summarized as

- (1) proposing a dual-channel framework that integrates real-time texture stitching with deep segmentation, enabling continuous modeling and localization of longitudinal cracks across sequential frames without reliance on GPS;
- (2) developing a hybrid crack refinement pipeline built upon YOLOv8-based segmentation, integrating adaptive sliding-window localization and morphological operations for high-fidelity crack boundary recovery;
- (3) deploying a YOLOv8-based segmentation module in an on-site processing workflow, validating real-time, high-accuracy identification and localization during inspection.

In essence, the research introduces a dual-channel framework that addresses the inherent conflict between semantic localization and pixel-level morphological fidelity in pavement distress analysis. The core methodological contribution lies in the transition from general-purpose navigation logic to pavement-specialized measurement logic. Unlike generic computer vision tasks, pavement inspection requires a high degree of topological continuity over long spatial scales.

The remainder of the paper is organized as follows. Section 2 provides the formulation of the problem, offering mathematical definitions and descriptions. Section 3 elaborates on the proposed method, encompassing the methodology framework and the detailed procedures. Section 4 presents a comprehensive demonstration of the effectiveness of the proposed method and its quantitative analysis results. Section 5 provides a comprehensive evaluation and discussion of the experimental results. Section 6 concludes the paper and proposes directions for future research.

2. Problem formulation

2.1. Scenario description

A mobile inspection platform is employed to capture pavement surface images during traversal, equipped with a central control system and an inspection camera for coordinated control and high-resolution image acquisition, as shown in Fig. 1. The inspection vehicle moves along a predetermined path, captures images of the pavement surface, and processes the data in real-time to generate continuous pavement textures and identify cracks on site.

It is assumed that the inspection platform moves at a known varying speed and heading throughout the time series, with its camera maintaining a fixed orientation with respect to the ground plane. The camera captures pavement surface images at regular intervals, generating a sequence of frames, which are subsequently stitched together to construct a continuous texture map that reconstructs the pavement surface.

2.2. System modeling

Reference frames. To describe the spatial relationships in the pavement reconstruction, five reference frames are defined as illustrated in Fig. 1, including the global frame $\{N\}$ fixed to the ground; the body frame $\{B\}$ attached to the vehicle; the camera frame $\{C\}$ and pixel frame $\{P\}$ for image acquisition; and the texture frame $\{U\}$ for the stitched mosaic. Specifically, the texture frame $\{U\}$ serves as the unified coordinate system where the longitudinal axis X_u aligns with the inspection path, enabling the correlation of pixel-level crack features with physical locations.

Kinematics model. The inspection platform is modeled as a rigid body moving on a planar surface. Its trajectory is parameterized by the position $\mathbf{t}_b^n(t) = [x^n(t), y^n(t), z^n(t)]^\top$ and heading $\phi(t)$ in the global frame N, which are obtained from the onboard navigation system. Consequently, the coordinate transformation from the body frame B to the global frame N is determined by the planar rotation $\mathbf{R}_b^n(\phi(t))$ and translation $\mathbf{t}_b^n(t)$.

$$\mathbf{x}^n = \mathbf{R}_b^n(\phi(t))\,\mathbf{x}^b + \mathbf{t}_b^n(t). \tag{1}$$

$$\mathbf{R}_b^n(\phi) = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

Camera imaging model. The monocular camera is mounted with a fixed downward tilt angle θ_0, which determines the static rotation $\mathbf{R}_c^b$ from the camera frame C to the body frame B. Given the calibrated camera intrinsic matrix K and the assumption of a planar pavement surface at a working distance d, the relationship between a pixel coordinate $\mathbf{x}^p$ and its corresponding spatial point follows the standard pinhole camera model [43].

The inverse projection from the pixel frame P to the global frame N, as shown in Fig. 2, is essentially an Inverse Perspective Mapping (IPM) process [44], formulated as $\mathbf{x}^n = \mathcal{F}(\mathbf{x}^p; K, \mathbf{R}_c^b, \mathbf{R}_b^n, d)$. This standard geometric mapping serves as the basis for the subsequent texture rectification.

$$\mathbf{R}_c^b = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_0 & -\sin\theta_0 \\ 0 & \sin\theta_0 & \cos\theta_0 \end{bmatrix}. \tag{3}$$

Pavement surface texture model. Let $\mathbf{x}^u = [x^u, y^u]^\top$ denote a point in the stitched texture frame U, and $\mathbf{x}^p$ be its corresponding point in the original pixel frame P. The alignment involves a rotation and translation to rectify the coordinate orientation, given by

To facilitate continuous crack tracking, the perspective views captured by the camera are registered into a unified Bird's Eye View (BEV) texture map defined in the texture frame U, as visually summarized in the transformation pipeline in Fig. 2. It requires a specific transformation to associate the stitched texture coordinates with the global physical position.

$$\mathbf{x}^u = \mathbf{R}_p^u\,\mathbf{x}^p + \mathbf{t}_p^u, \tag{4}$$

where $\mathbf{R}_p^u \in \mathbb{R}^{2\times 2}$ is the rotation matrix, and $\mathbf{t}_p^u$ is the translation vector aligning the pixel origin to its position in the texture map. Specifically,

$$\mathbf{R}_p^u = \mathbf{R}_u^p = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}. \tag{5}$$

For each stitched frame, the inverse projection from the texture coordinates to the global coordinates can be derived from

$$\begin{aligned} \mathbf{x}^n &= \mathbf{R}_b^n\,\mathbf{R}_c^b \left(d\,K^{-1} \begin{bmatrix} \mathbf{x}^p \\ 1 \end{bmatrix} \right) + \mathbf{t}_b^n \\ &= \mathbf{R}_b^n\,\mathbf{R}_c^b \left(d\,K^{-1} \begin{bmatrix} \mathbf{R}_u^p(\mathbf{x}^u - \mathbf{t}_p^u) \\ 1 \end{bmatrix} \right) + \mathbf{t}_b^n \\ &= \mathbf{R}_b^n\,\mathbf{R}_c^b \left(d\,K^{-1} \begin{bmatrix} \mathbf{R}_u^p\,\mathbf{x}^u \\ 1 \end{bmatrix} \right) + \mathbf{t}_u^n, \end{aligned} \tag{6}$$

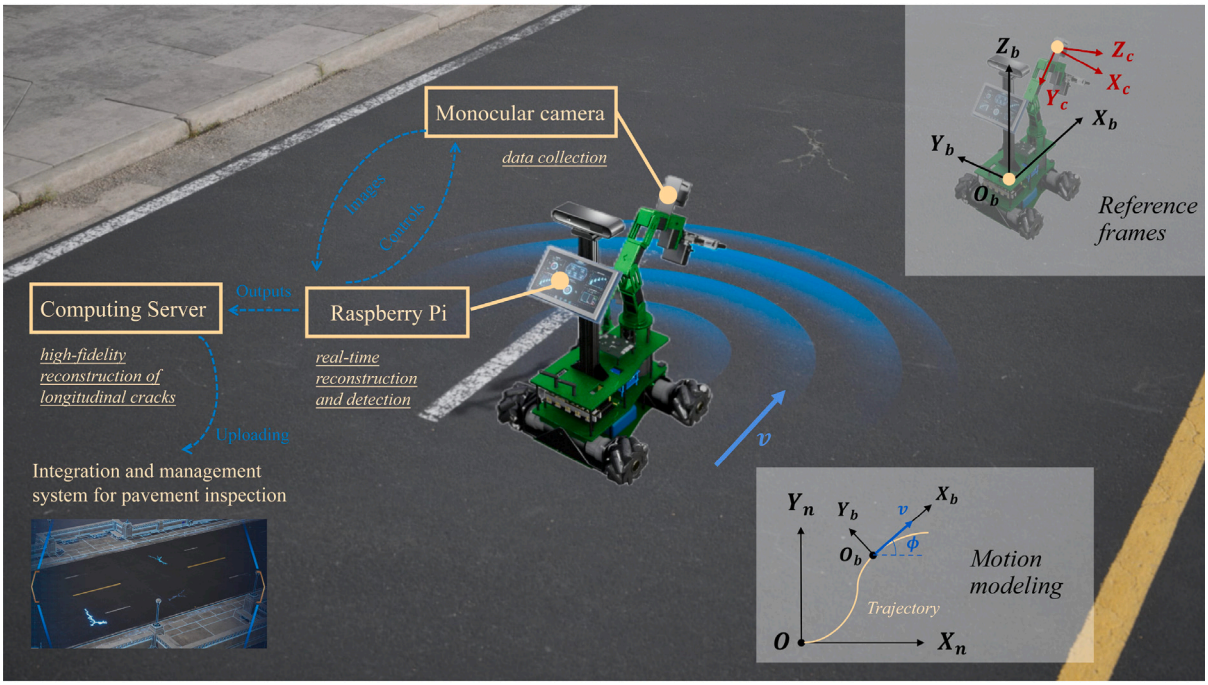


Fig. 1. Vehicle inspection scenario.

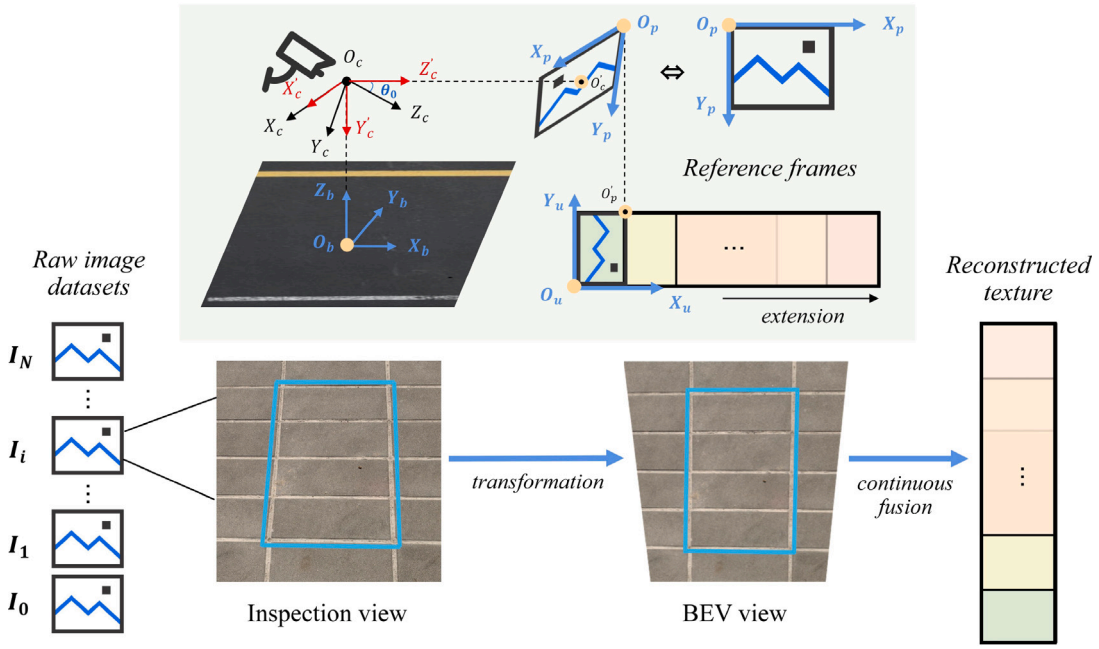


Fig. 2. Schematic of coordinate relationships.

where $\mathbf{R}_b^n = \mathbf{R}_b^n(\phi(t))$ is dependent on the vehicle's heading angle $\phi(t)$ at the time of image capture, and $\mathbf{t}_u^n$ denotes the translation from the origin of the frame U to the frame N.

Given the inverse projection Eq. (6), it can be observed that the transformation is constructed through a composition of known rotation and translation matrices, along with the intrinsic camera model and depth scaling. When the depth d is known or can be reasonably approximated (e.g., for flat ground assumption), the transformation becomes deterministic and bijective for each frame, implying that a one-to-one correspondence between the coordinates in U and N can be established.

Nevertheless, over a complete inspection sequence, small deviations due to motion estimation noise, sensor timing, and depth approximation may lead to accumulated scale drift in the stitched texture. A global linear scaling correction can be introduced to solve the problem based on the known kinematics of the inspection platform. Define the two-dimensional path of the inspection platform by a differentiable vector $\mathbf{P} \in \mathbb{R}^2$. By integrating the speed over time from the starting point to a given time t, the length of the path traveled in the frame N can be expressed as

$$L^n(t) = \|\mathbf{P}(t)\| = \int_0^t \|v(\tau)\| \, \mathrm{d}\tau, \tag{7}$$

The total length of the inspection path in the texture frame U is defined as the accumulated Euclidean distance between consecutive points along the central axis of the stitched texture. Let $\{\mathbf{x}_k^u = [x_k^u, y_k^u]^\top \mid k = 0, 1, 2, \ldots, \mathrm{K}\}$ denote a sequence of points sampled along the centerline of the texture map at time t, where K is the total number of the sampled points. The total path length in U is given by

$$L^u(t) = \sum_{k=0}^{\mathrm{K}-1} \left\| \mathbf{x}_{k+1}^u - \mathbf{x}_k^u \right\| = \sum_{k=0}^{\mathrm{K}-1} \sqrt{(x_{k+1}^u - x_k^u)^2 + (y_{k+1}^u - y_k^u)^2}. \tag{8}$$

Accordingly, a time-dependent scaling factor s is introduced as the ratio between the physical path length $L^n(t)$ and the corresponding texture length $L^u(t)$, expressed as

$$s(t) = \frac{L^n(t)}{L^u(t)}. \tag{9}$$

The scale factor enables the refined mapping of texture coordinates $[x^u, y^u]^\top$ in the texture to global coordinates $[x^n, y^n]^\top$, expressed as

$$\begin{bmatrix} x^n \\ y^n \end{bmatrix} = s(t) \begin{bmatrix} x^u \\ y^u \end{bmatrix} + \mathbf{t}_u^n. \tag{10}$$

The correction may compensate for accumulated drift and ensures spatial consistency between the reconstructed texture and the actual path traversed during inspection.

2.3. Problem statement

Consider a sequence of pavement surface images $\mathcal{I} = \{I_i \mid i = 1, 2, \ldots, N\}$ captured at regular intervals by a vehicle-mounted monocular camera. Each image $I_i \in \mathcal{I}$ is defined over a set of pixels $\mathcal{P}_i$.

The transformation pipeline begins with the raw image I_i, which is undistorted to obtain $\tilde{I}_i$, and then projected onto the pavement plane via homography to generate the BEV image $\hat{I}_i$. Let K denote the intrinsic camera matrix and (R_i, t_i) the extrinsic parameters at frame I_i. The camera calibration is required to estimate the corrected intrinsic matrix $\hat{K}$, distortion coefficients $(k_1, k_2, p_1, p_2, k_3, k_4, k_5, k_6)$, and the distortion model $\mathcal{D}(\cdot)$ to get $\tilde{I}_i$, which compensates for the deformation induced by lens and manufacturing inaccuracies. The obtained images are then mapped to the BEV plane via the calculation of homography matrix H_i.

The BEV images $\hat{I}_i$ are required to be spatially aligned and fused into a global long-strip texture map $\mathcal{T}$ defined over U, formulated as

$$\mathcal{T} = \bigcup_{i=1}^{N} \mathcal{T}_i, \quad \mathcal{T}_i = \mathcal{W}_p^u(\hat{I}_i, \mathbf{x}^p, \phi(t)), \tag{11}$$

where $\mathcal{W}_p^u(\cdot)$ denotes the warping function that maps each pixel $\mathbf{x}^p = [x^p, y^p]^\top$ in $\hat{I}_i$ to its corresponding position $\mathbf{x}^u = [x^u, y^u]^\top$ in the frame U, forming the texture $\mathcal{T}(x^n, y^n)$. As a result of Eq. (6), $\mathbf{x}^u$ is further corresponded to its global coordinate $[x^n, y^n]^\top$.

The construction of $\mathcal{T}$ is the foundation for high-fidelity crack detection. As the vehicle moves, the texture map $\mathcal{T}$ and crack set $\mathcal{C} = \{C_m \mid m = 1, 2, \ldots, M\}$ are incrementally updated, where M denotes the total number of the detected cracks. At each new frame I_{i+1}, the updates are described as

$$\mathcal{T} \leftarrow \mathcal{T} \cup \mathcal{T}_{i+1}, \quad \mathcal{C} \leftarrow \{C_m\} \cup \{C_{M+1}, \ldots, C_{M'}\}, \tag{12}$$

ensuring real-time responsiveness and continuity of crack localization. Within the unified map $\mathcal{T}$, each crack region $C_m \subset \mathrm{dom}(\mathcal{T}) \subset U$ is identified by segmenting a connected set of pixels that exhibit crack-like characteristics based on a predefined detection function $f_{\mathrm{det}}(\cdot)$, as defined by

$$C_m = \left\{ (x^u, y^u) \in \mathrm{dom}(\mathcal{T}) \;\middle|\; f_{\mathrm{det}}(x^u, y^u) = 1 \right\}. \tag{13}$$

Each crack C_m is then assigned a spatial footprint $\Omega_m \subset \mathbb{R}^2$ in the N, which enables real-time, spatially consistent detection and mapping of pavement surface cracks, in particular facilitating the high-resolution analysis of long longitudinal cracks across multiple frames.

Given the vehicle's motion parameters $v(t)$ and $\phi(t)$ under the assumption of straight-line motion, the research aims to (1) reconstruct a continuous and real-time long-strip texture representation of the pavement surface accumulated along the driving trajectory, and (2) enable accurate and efficient detection and localization of the longitudinal cracks $\mathcal{C} = \{C_m\}$ with high spatial precision, where $C_m \cap C_{m'} = \emptyset$ for $\forall m \neq m'$, with $m, m' \in \{1, 2, \ldots, M\}$.

3. Texture-aligned framework for continuous modeling of longitudinal cracks

3.1. Overview

A comprehensive framework for real-time pavement texture reconstruction and crack detection is developed to enhance pavement inspection efficiency and accuracy. The overall structure establishes a dual-channel processing pipeline as shown in Fig. 3, including (i) a feature-based fusion pipeline for continuous pavement texture reconstruction, and (ii) an enhanced YOLOv8-seg detector integrated with adaptive morphological processing for precise longitudinal crack identification.

The first module focuses on the continuous reconstruction of pavement texture. To address geometric deformation and environmental interference, the preprocessing pipeline employs standardized protocols for edge sharpening, homography-based perspective normalization, and radial distortion compensation. Feature detection and matching techniques are then applied to correlate key points across consecutive images. Subsequently, the stitched frames are integrated into a coherent reconstruction of the pavement surface.

The second module focuses on real-time crack detection and high-precision reconstruction. It begins with the instance segmentation via the YOLOv8-seg model. After initial segmentation, a hybrid refinement strategy is employed, consisting of a sliding window scan with watershed correction and adaptive morphological operations. The local morphological operations sharpen crack boundaries, mitigate errors in both geometry and extent, and further restore the continuity of fine cracks (< 3 mm width). Consequently, a high-resolution crack map is generated to ensure topological continuity, allowing for the precise delineation of local edges and the extension of crack branches. The seamless integration of these two core components forms the cornerstone of the proposed framework. The reconstructed pavement texture offers the necessary contextual information for georeferenced detection of longitudinal cracks.

3.2. Feature-based continuous reconstruction of pavement texture

The continuous reconstruction of pavement texture requires a sequence of techniques for image enhancement, geometric rectification, feature-based image stitching, and the reconstruction of continuous textures over time.

3.2.1. Image enhancement and rectification

To enhance edge features while mitigating noise interference, a hybrid spatial filter is applied to the raw images, combining Gaussian smoothing to suppress environmental noise with a Laplacian operator to amplify intensity gradients, as illustrated in Fig. 4.

To ensure geometric fidelity, the enhanced images undergo radial distortion correction followed by a perspective transformation to a BEV view, thereby eliminating perspective foreshortening, as shown in Fig. 5.

The rectification is achieved by a homography matrix H^{rect} that maps the coordinates to the unified plane as

$$\begin{bmatrix} \hat{x}^p \\ \hat{y}^p \\ 1 \end{bmatrix} = H^{\mathrm{rect}} \begin{bmatrix} \tilde{x}^p \\ \tilde{y}^p \\ 1 \end{bmatrix}, \tag{14}$$

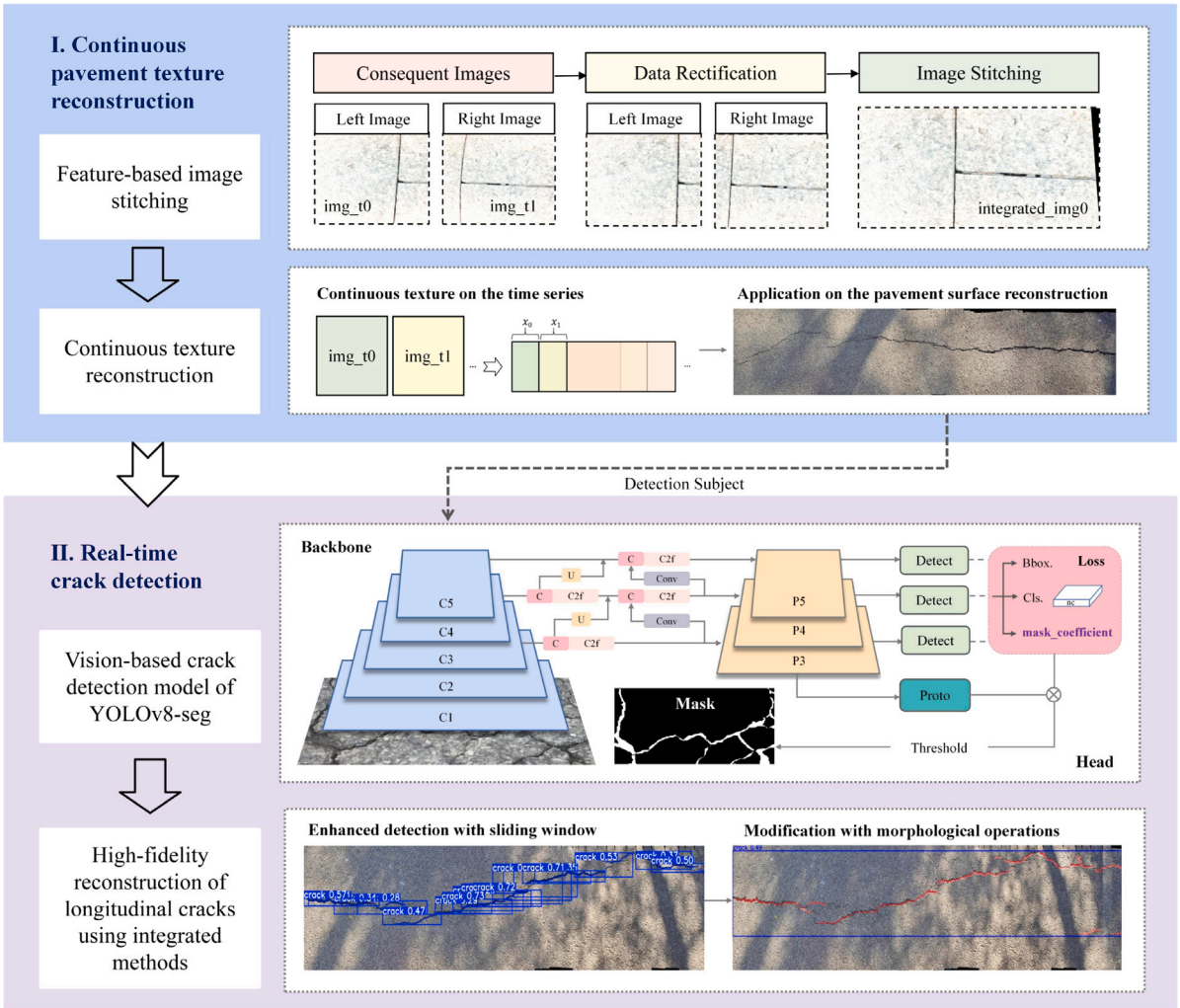


Fig. 3. Dual-channel framework for feature-based continuous modeling of longitudinal cracks.

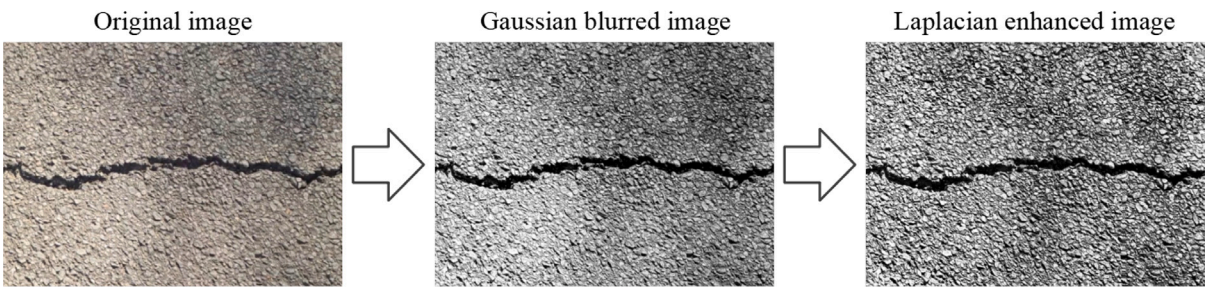


Fig. 4. An example of image enhancement.

where $[\hat{x}^p, \hat{y}^p]^\top$ denotes the coordinates in the BEV view. The matrix H^{rect} is computed by identifying corresponding points between the original image and the desired top-down projection through pre-calculated calibration parameters. The rectified image yields a top-down, undistorted pavement view consistent with the physical geometry.

3.2.2. Feature-based image stitching

Feature-based image stitching merges consecutive pavement images into a unified texture map by detecting, matching, and geometrically aligning key points across images. The Scale-Invariant Feature Transform (SIFT) [45] is employed to extract robust keypoints and generate descriptors from the rectified images, using the difference of Gaussians (DoG). The detected keypoints are refined through quadratic interpolation in scale-space to achieve pixel-level localization accuracy and are assigned a dominant orientation based on the local gradient magnitude and direction, after which a feature descriptor is constructed for each keypoint by sampling local gradient values within a fixed neighborhood, organizing them into a regular grid, and normalizing them to form a 128-dimensional vector $\mathbf{f} = f(Q) \in \mathbb{R}^{128}$, where $f(\cdot)$ encodes the local gradient structure around the keypoint $Q \in \hat{I}$.

Feature matching is then performed across consecutive images based on descriptor similarity, where the geometric transformation is modeled by a homography matrix H^{mos}. This matrix is estimated using the DLT method, with RANSAC [46] employed to suppress outliers.

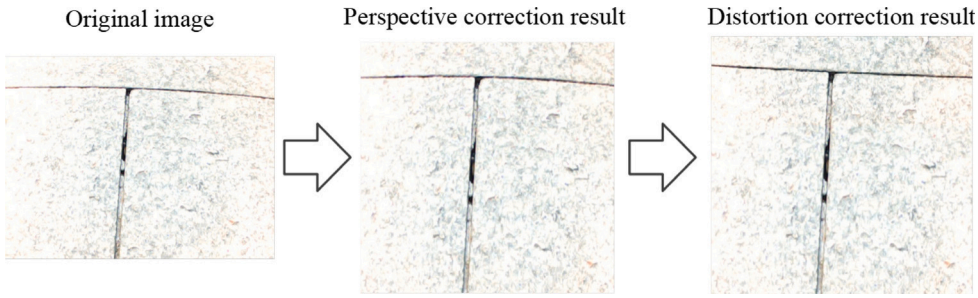


Fig. 5. An example of image rectification.

Once estimated, $\hat{I}_2$ is warped and blended with $\hat{I}_1$ to form a seamless mosaic, denoted as

$$I_2^{\text{mos}}(x_1^p, y_1^p) = H^{\text{mos}}(\hat{I}_2(x_2^p, y_2^p)), \tag{15}$$

where $[x_1^p, y_1^p]^\top$ are the corresponding points of $[x_2^p, y_2^p]^\top$ under the pixel frame of $\hat{I}_1$. Overlapping regions are integrated using Laplacian pyramid blending [47] to ensure smooth transitions between frames.

3.2.3. Continuous texture reconstruction

The continuous pavement texture reconstruction is achieved through incremental stitching of time-series images $\mathcal{I}$ via a three-phase integration: local feature alignment, overlap blending, and global consistency verification.

To integrate frame I_i into the current map $\mathcal{T}_{i-1}$, an adaptive alpha blending weight α is utilized to ensure spatial coherence at the stitching frontier, given by

$$\alpha(x^u, y^u) = \exp\left(-\frac{\|(x^u, y^u) - \partial\mathcal{T}_{i-1}\|^2}{2\sigma_g^2}\right), \tag{16}$$

where $\partial\mathcal{T}_{i-1}$ represents the boundary of the current texture map.

To maintain global consistency and mitigate cumulative alignment errors over long sequences, bundle adjustment [48] is periodically applied to jointly optimize the homography sequence $\{H_i^{\text{mos}}\}$:

$$\min_{\{H_i^{\text{mos}}\}} \sum_{i=2}^{N} \sum_{j=i-k}^{i-1} \sum_{(p,q)\in\mathcal{M}_{ij}} \|p - (H_i^{\text{mos}})^{-1} H_j^{\text{mos}} q\|^2 + \lambda \sum_{i=2}^{N} \|\log(H_i^{\text{mos}}(H_{i-1}^{\text{mos}})^{-1})\|_F^2, \tag{17}$$

where the first term minimizes reprojection errors, while the second term regularizes homography variations to maintain motion continuity. Finally, boundary refinement is performed via content-aware cropping to eliminate boundary inconsistencies, resulting in a visually coherent surface reconstruction.

3.3. Longitudinal crack detection and high-fidelity reconstruction

3.3.1. Real-time crack detection and segmentation

Within the scope of real-time pavement crack detection, the YOLOv8-seg model employs a specialized architecture integrating detection and segmentation via prototype-based mask generation.

an enhanced YOLOv8-seg detector integrated with adaptive morphological processing for precise longitudinal crack identification.

To enhance segmentation accuracy for fine crack structures, a crack-width weighted Dice coefficient is employed as the segmentation loss. $\mathcal{L}_{\text{mask}}$ is expressed as

$$\mathcal{L}_{\text{mask}} = 1 - \frac{2\sum_p w(p)\,(\mathbf{M}_k)_p\,(\hat{\mathbf{M}}_k)_p}{\sum_p w(p)\,(\mathbf{M}_k)_p + \sum_p w(p)\,(\hat{\mathbf{M}}_k)_p}, \quad w(p) = 1 + \exp(-\alpha D_T(p)), \tag{18}$$

where $(\mathbf{M}_k)_p$ and $(\hat{\mathbf{M}}_k)_p$ denote the predicted and ground-truth mask values at pixel p, respectively. $D_T(p)$ represents the distance from pixel p to the crack centerline, and $w(p)$ assigns higher importance to pixels near the centerline, controlled by parameter α.

Furthermore, a prototype loss $\mathcal{L}_{\text{proto}}$ is introduced to maintain spatial consistency, utilizing a directional filter $\mathbf{M}_{\text{orient}}$ to emphasize linear crack morphologies, presented as

$$\mathcal{L}_{\text{proto}} = \sum_{j=1}^{N} \|\mathbf{P}_j \odot \mathbf{M}_{\text{orient}} - \hat{\mathbf{P}}_j\|_F, \tag{19}$$

where $\mathbf{M}_{\text{orient}}$ is the directional filter emphasizing linear crack structures, $\hat{\mathbf{P}}_j$ the ground-truth prototype mask for the jth prototype, and $\odot$ denotes the Hadamard product.

3.3.2. High-fidelity crack reconstruction

To overcome resolution discrepancies from the deep learning model, a hybrid refinement framework is developed, executing a three-stage strategy: (i) sliding-window localized optimization, (ii) two-phase morphological conditioning, and (iii) marker-controlled watershed refinement, as visualized in Fig. 6.

The process begins by synthesizing longitudinal crack seeds through confidence-weighted aggregation, as

$$C^{\text{merged}} = \frac{\sum_{v=1}^{V} A_v\, C_v}{\sum_{v=1}^{V} A_v}. \tag{20}$$

where C_v and A_v denote the confidence score and area of the vth crack segment, respectively, and V is the total number of the merged segments. Based on the merged segments, raw binary masks of the longitudinal cracks are constructed for the texture $\mathcal{T}$, forming the initial pixel-level crack regions as

$$C_m^{(0)} = \left\{(x^u, y^u) \in \mathcal{T} \;\middle|\; \mathbf{M}_m(x^u, y^u) = 1\right\}, \quad m \in \{1, 2, \ldots, M\}. \tag{21}$$

where $\mathbf{M}_m$ denotes the mth instance mask defined on $\mathcal{T}$, and M is the total number of detected longitudinal cracks.

Subsequently, a localized optimization of sliding windows B_{w} is performed to maximize congruence with gradient-aligned crack boundaries, given by

$$\hat{B}_{\text{w}} = \underset{B_{\text{w}}\in\text{W}}{\text{argmax}} \left(\frac{|\nabla\mathcal{T} \cap B_{\text{w}}|}{|B_{\text{w}}|}\right). \tag{22}$$

where $(u_{\text{w}}, v_{\text{w}})$ denotes the top-left coordinates of window B, and $w_{\text{w}}, h_{\text{w}}$ are its width and height. $\hat{B}_{\text{w}}$ is the optimized window. The detection windows are geometrically aligned with the underlying crack structures, as illustrated in Fig. 6.

Morphological conditioning then refines boundaries using a pixel-wise membership probability p_b based on grayscale continuity Λ and intensity conformity Ψ as

$$p_b(x^u, y^u, \lambda) = (\Lambda(x^u, y^u))^{\lambda} \cdot (\Psi(x^u, y^u))^{1-\lambda}, \tag{23}$$

where $\Lambda(x^u, y^u)$ measures the local grayscale continuity along the crack direction, $\Psi(x^u, y^u)$ is the normalized membership of pixel intensity within an adaptive crack intensity range $[\eta_{\min}, \eta_{\max}]$, and $\lambda \in [0, 1]$ balances the contribution of continuity and intensity conformity.

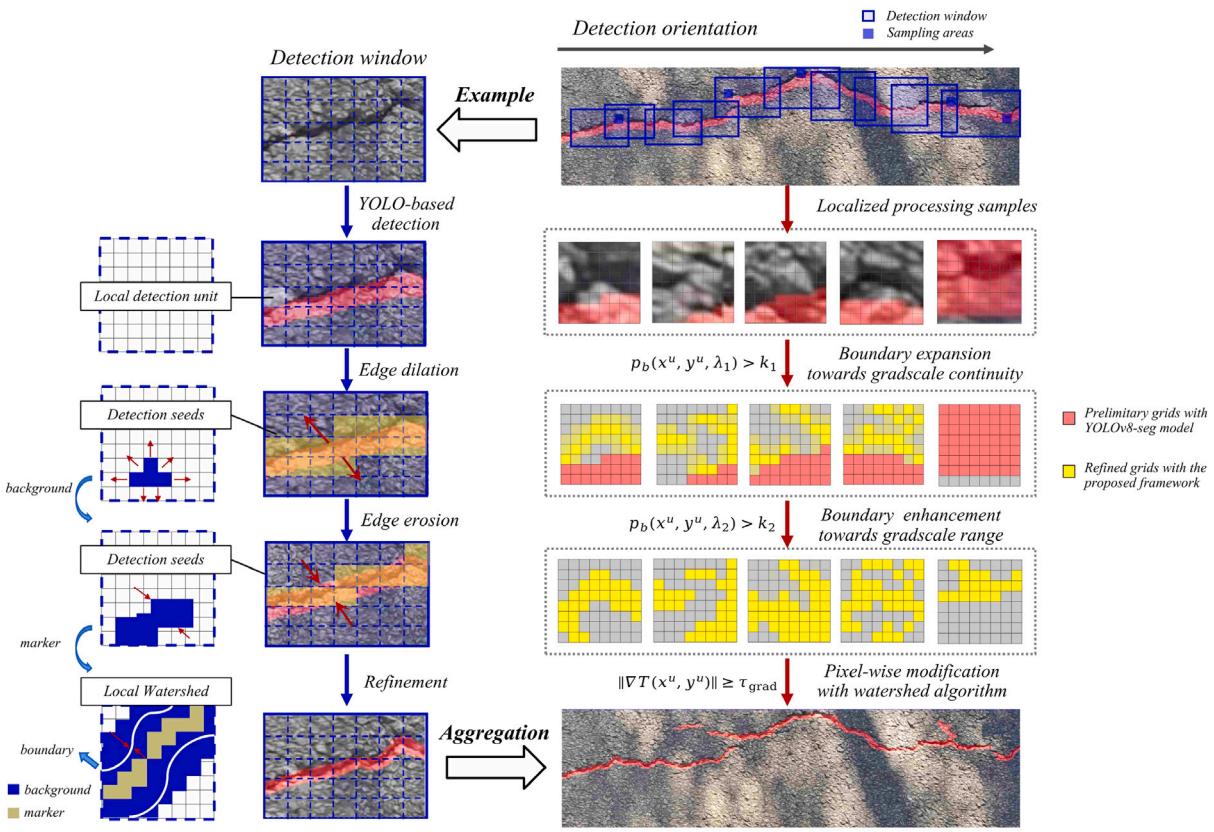


Fig. 6. High-fidelity crack reconstruction pipeline using sliding window and morphological operations.

The continuity term is instantiated as an exponential decay with respect to the local grayscale deviation

$$\Lambda(x^u, y^u) = \exp\left(-\frac{|\Phi(x^u, y^u) - \bar{\Phi}|}{\sigma_\Phi}\right), \tag{24}$$

where $\bar{\Phi}, \sigma_\Phi$ are the mean and standard deviation of local grayscale values along the crack seed.

In the first phase of the morphological operations, the edge dilation process expands the crack boundary with each detection seed serving as a growth origin. The pixel-wise membership probability is thresholded as $p_b(x^u, y^u, \lambda_1) > k_1$ to probabilistically propagate the boundary into adjacent low-contrast regions with continuous grayscale profiles, where k_1 is the probabilistic threshold controlling boundary growth. The weighting coefficient $\lambda = \lambda_1$ for the dilation phase is set to a relatively high value to emphasize the role of grayscale continuity. The updated crack region that recovers subtle crack branches is expressed as

$$C_m^{(1)} = \{(x_u, y_u) \in C_m^{(0)} \mid p_b(x_u, y_u, \lambda_1) > k_1\}. \tag{25}$$

In the second phase, precision boundary alignment is achieved by edge erosion process that prunes low-probability pixels. A stricter probabilistic threshold $k_2 > k_1$ is applied in the edge enhancement to delineate crack boundaries with higher confidence, and the weighting coefficient $\lambda = \lambda_2$ is set to a lower value to emphasize pixel intensity. The updated crack region is expressed as

$$C_m^{(2)} = \{(x_u, y_u) \in C_m^{(1)} \mid p_b(x_u, y_u, \lambda_2) > k_2\}, \quad \lambda_2 < \lambda_1. \tag{26}$$

The weighting coefficients are adaptively determined by local image statistics. Specifically, the weighting factor λ_1 is dynamically governed by the local structural coherence C_{coh} derived from the eigenvalues of the structure tensor as

$$C_{coh} = \left(\frac{\mu_1 - \mu_2}{\mu_1 + \mu_2}\right)^2, \tag{27}$$

which prioritizes geometric continuity in anisotropic regions where linear crack extensions are statistically significant. Conversely, the intensity conformity weight λ_2 is modulated by the local coefficient of variation, scaling the reliance on grayscale contrast inversely to the background texture complexity.

The probabilistic thresholds are determined by an offline calibration protocol linked to the imaging system's sensitivity. The dilation threshold k_1 is calibrated to the sensor's noise equivalence level to maximize topological recall, while the erosion threshold k_2 is established as a texture discrimination level to strictly filter aggregate artifacts and delineate high-confidence crack markers. The process is further regulated by a feedback loop that tightens the pruning criteria if the enhancement entropy exceeds a stability index, indicating chaotic pixel inclusion, to ensure pixel-wise precision.

Built upon the probabilistically enhanced boundaries, a marker-controlled watershed algorithm is implemented to the crack region $C_m^{(2)}$ and $C_m^{(1)}$ for precision augmentation. The flooding process is constrained by both gradient information from the morphological operations, presented as

$$\nabla T(x^u, y^u) = \left(\frac{\partial T}{\partial x^u}, \frac{\partial T}{\partial y^u}\right) = (\mathcal{T} * S_\theta, \mathcal{T} * S_{\theta+\pi/2}), \tag{28}$$

where $\nabla \mathcal{T}(x^u, y^u)$ is the gradient field, and S_θ denotes a directional Sobel kernel rotated by θ. The flooding sources are automatically generated from $C_m^{(1)}$, which provides high recall but relatively low accuracy, while the markers are derived from the connected components of $C_m^{(2)}$, characterized by low recall but very high accuracy. The flooding process ceases upon gradient peaks, which are the boundary of the cracks. The watershed boundaries $\partial\mathcal{R}$ are derived by minimizing the functional

$$\begin{aligned} \partial\mathcal{R} &= \arg\min_{\partial\Omega}\left(\int_{\partial\Omega} \|\nabla\mathcal{T}(x^u, y^u)\|^{-1} ds + \alpha \cdot \text{dist}(x^u, y^u; C_m^{(2)})\right), \\ &\text{s.t. } \|\nabla T(x^u, y^u)\| \geq \tau_{\text{grad}}, \end{aligned} \tag{29}$$

where the second term penalizes deviations from the probabilistic boundary prior. The formulation ensures pixel-wise precision and topological consistency. As visualized in the detection window of Fig. 6, the watershed process is initialized using the eroded region as high-confidence markers and the dilated region as background seeds. The flooding propagates from the seeds along the gradient field and terminates where the fronts collide, defining the precise crack boundary. The final set of the longitudinal cracks C_m is defined as the union of watershed basins $\{\Omega_\ell\}$, whose boundaries are constrained by the optimized watershed contour $\partial\mathcal{R}$, as shown in Eq. (30).

$$C_m = \bigcup_{\ell\,:\,\Omega_\ell \cap C_m^{(2)} \neq \varnothing} \Omega_\ell, \qquad \text{with } \partial\Omega_\ell \subseteq \partial\mathcal{R}. \tag{30}$$

4. Experiments and results

4.1. Experimental setup

The experimental setup was designed to investigate pavement inspection and crack detection using a miniature mobile vehicle equipped with a Raspberry Pi-controlled platform and an inspection camera. The setup simulates an autonomous pavement inspection system operating in real-world conditions.

Pavement inspection. The experiments were conducted in Beijing (Tsinghua University) and Shenzhen (University Town of Nanshan District), where we collected a total of 13 inspection datasets spanning diverse pavement scenarios, including different weather conditions, pavement types, lighting conditions and crack morphologies. Each dataset consisted of road sections ranging from 10 to 30 m in length, ensuring that the evaluation covered a wide variety of real-world conditions. During the experiments, the inspection platform was programmed to traverse a predetermined path, with a tilted monocular camera capturing pavement surface images continuously. The captured data were processed by the platform in real time to generate pavement textures and detect longitudinal cracks, and simultaneously transmitted to the computing server for high-fidelity crack reconstruction.

To simplify the motion model, the platform was constrained to move along a straight-line path with a constant heading ($\phi = \phi_0$), accounting for minor disturbances while neglecting the effects of turning on motion control. The moving speed was fixed at $v = 0.5$ m/s, and the camera was mounted with a fixed orientation relative to the ground($\theta = \theta_0$). A custom control script deployed on the Raspberry Pi managed vehicle motion and data acquisition, while simultaneously handling data synchronization and transmission. The script also ensured precise timestamping by aligning image capture with the vehicle's speed and position. Optimized control parameters including camera frame rate and sampling intervals further guaranteed reliable and reproducible data quality. The setup of the inspection platform enabled consistent image acquisition during movement.

The primary data consisted of continuous pavement texture images, captured as a sequence of overlapping frames. These frames were complemented by auxiliary datasets, including real-time speed logs, pose angles from the robotic arm, and the total runtime of the inspection process. The collected dataset samples included:

- **Image Dataset**: A sequence of high-resolution pavement surface images captured at a predefined frame rate.
- **Speed logs**: Real-time speed measurements sampled at regular intervals.
- **Runtime Logs**: Total operation time for each inspection run.
- **Pose Data**: Recorded angles and positions of the robotic arm equipped with the camera at each timestamp.

YOLOv8-based model training. To ensure the generalization capability of the deep segmentation model, a large-scale composite dataset, referred to as CRACK10000, was constructed by aggregating several public benchmark datasets, including CRACK500 [49], DeepCrack [50], CFD [51], the Concrete Crack Dataset [52], and GAPs [53]. The repository covers asphalt and concrete pavements, and interference factors such as shadows, oil stains, and water stains.

From the aggregated pool of over 10,000 images, a high-quality subset of 6800 images was selected for this study, as detailed in Table 1. The selection criteria prioritized image clarity and the presence of representative longitudinal crack features. The dataset was randomly partitioned into three non-overlapping subsets: a training set (5780 images, 85%), a validation set (680 images, 10%), and a testing set (340 images, 5%). The split ensures sufficient data for feature learning while reserving a representative set for hyperparameter tuning and final evaluation.

Despite the large number of samples in the datasets, the quality of their original annotations is inconsistent; for example, CRACK500 and DeepCrack often have coarse or imprecise boundaries in complex scenes. To guarantee annotation quality, strictly pixel-level ground truth labeling was performed. The annotations from public datasets were manually inspected and corrected using the LabelMe tool to fix disconnected segments and coarse boundaries. Data augmentation techniques, including random flipping, rotation, and brightness adjustment, were applied to the training set to prevent overfitting and enhance the model's robustness against environmental variations.

Prior to training, images were resized to 640 × 640 pixels and augmented using random flipping, rotation and brightness adjustments to improve model generalization. The training was performed on an NVIDIA RTX 4090 GPU, and the trained model was deployed on the platform for real-time inference and crack segmentation during inspection.

The selected model is capable of real-time detection due to its ability to process images in a single pass. During the experiments, the network was trained for 150 epochs using stochastic gradient descent (SGD) with an initial learning rate $\eta = 0.01$ and momentum $\mu = 0.937$, reaching convergence by the end of the training process. To ensure robust segmentation, a prototype update threshold and a smoothing factor were adopted for controlling mask activation and temporal consistency during inference.

4.2. Evaluation metrics

To evaluate the performance of the trained YOLOv8-seg model for crack detection and segmentation, several key metrics are employed to provide a comprehensive assessment of model accuracy and robustness, including precision, recall, and mean average precision (mAP), frames Per second (FPS), as well as three self-defined geometric descriptors designed to quantify crack integrity, fragmentation resistance, and topological complexity.

- **Precision:** measures the proportion of true positive crack detections out of all detections made by the model, defined as

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{31}$$

  where TP denotes the number of true positives (correct detections), and FP denotes the number of false positives (incorrect detections).
- **Recall:** evaluates the model's ability to detect all actual cracks in the image, defined as

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{32}$$

  where FN represents the number of false negatives (missed cracks).
- **Mean average precision (mAP):** an overall performance metric that evaluates the precision of the model across different IoU thresholds and multiple classes. It aggregates the precision scores at different recall levels.

Table 1
Composition and characteristics of the CRACK10000 subset.

Dataset name	Source	Quantity	Characteristics
CRACK500	Pavement	500	Crack images with pixel-level annotations, featuring variable illumination and complex backgrounds.
DeepCrack	Pavement	537	Multi-scale cracks with challenging non-crack background textures.
CFD (Crack Forest)	Pavement	118	Urban road surfaces with noise such as shadows and manhole covers.
Concrete Crack	Concrete	~3,700	High-resolution images of concrete surfaces, primarily designed for classification tasks.
GAPs	Pavement	1969	High-resolution asphalt pavement images with diverse distress types and high quality.

- **Intersection over Union (IoU):** measures the extent of overlap between the predicted segmentation mask and the ground truth mask. It is a strict metric for segmentation accuracy, defined as

$$\text{IoU} = \frac{TP}{TP + FP + FN}, \tag{33}$$

- **F1-Score:** the harmonic mean of Precision and Recall, providing a comprehensive evaluation that balances both false positives and false negatives. It is defined as

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{34}$$

- **Frames Per Second (FPS):** quantifies the computational efficiency of the proposed method by measuring the average number of images processed per second during inference. A higher FPS indicates better suitability for real-time or near-real-time inspection scenarios. FPS is defined as

$$\text{FPS} = \frac{N}{T}, \tag{35}$$

where N denotes the total number of processed images and T denotes the total inference time in seconds.

- **Continuity (average continuous length):** a self-defined geometric descriptor introduced in this study to quantify the completeness of detected cracks, defined as

$$\text{Continuity} = \frac{\text{Total Length}}{\text{Logical Count}}. \tag{36}$$

The metric represents the average length of each independent crack segment. For the detection of a specific area, larger Continuity value indicates more complete and less fragmented crack detection results.

- **Spatial Connectivity Index (SCI):** another self-defined geometric descriptor designed to evaluate the spatial connectivity and fragmentation resistance of crack networks, defined as

$$\text{SCI} = \frac{L_{\max}}{\text{Total Length}}, \quad \text{SCI} \in [0, 1], \tag{37}$$

where $L_{\max}$ denotes the length of the largest connected crack component. An SCI value close to 1 indicates that the crack pattern is dominated by a single, well-connected main crack, while a low SCI value implies highly scattered and fragmented crack structures.

- **Branches:** a topological metric introduced to evaluate the model's ability to recover fine, intricate crack networks. It is calculated by extracting the one-pixel-wide morphological skeleton of the segmented crack mask and counting the number of branching nodes within the skeletonized structure.

4.3. Results

The results presented in Section 4.3 aim to demonstrate the feasibility of the proposed pipeline and the effectiveness of each processing stage. To clearly illustrate the workflow and intermediate outcomes, representative samples are selected from the self-constructed real-world pavement dataset, and the corresponding results at different stages of the algorithm are visualized and analyzed. The selected examples are primarily used to showcase the behavior of the proposed method under realistic operating conditions and to qualitatively validate the functionality of each module. Quantitative evaluations on public benchmark datasets with ground-truth annotations, as well as comprehensive validation on the full real-world pavement dataset, are reported in Section 5.

4.3.1. Pavement texture reconstruction

The image alignment procedures, which illustrate the intermediate steps and results, are presented in Fig. 7. Initially, a pair of sequential frames from the original image stream is presented, which preserves temporal continuity and consistent spatial alignment over time. Subsequent image pairs depict the results of feature matching, where the keypoints are detected and matched across the images, with corresponding points highlighted by colored lines. The density and accuracy of these connections reflect the quality of feature matching. For each image pair, a homography matrix is calculated based on the matched points, transforming coordinates from the right image to the left image.

The final stitched result after image alignment is shown in Fig. 7. The original misalignment is corrected, and the images are seamlessly combined into a continuous texture map, demonstrating the effectiveness of the feature-based approach for accurate image alignment.

The stitching quality was evaluated on UDIS-D [54], a standard test dataset for seamless stitching, which comprises 1106 image pairs. The stitching error was quantified by computing the pixel-wise difference between the corresponding regions of the stitched results and the benchmark images. The average stitching error was found to be 0.85 pixels, indicating a high level of alignment accuracy. The feature matching success rate reached 96.4%, where a keypoint match was considered correct if the SSIM (structural similarity index) between the corresponding regions exceeded 0.60, further confirming the reliability of the feature matching process.

To further evaluate the continuous pavement texture reconstruction, a comparison was conducted between the proposed method and an existing technique, Patch-NetVLAD [55]. Patch-NetVLAD has shown to be effective for large-scale image matching tasks, utilizing a deep learning-based architecture. It employing a patch-based approach combined with a VLAD (vector of locally aggregated descriptors) representation to match and stitch image patches. However, the method may struggle in some pavement texture scenarios, particularly with complex pavement surfaces and variations in lighting or weather conditions.

The qualitative results of the comparison are presented in Fig. 8. While the proposed method matches fewer keypoints, the matches are highly accurate with minimal noise, ensuring stable and precise alignment. In contrast, Patch-NetVLAD typically matches a larger number of keypoints, but in areas with abrupt changes in texture or lighting

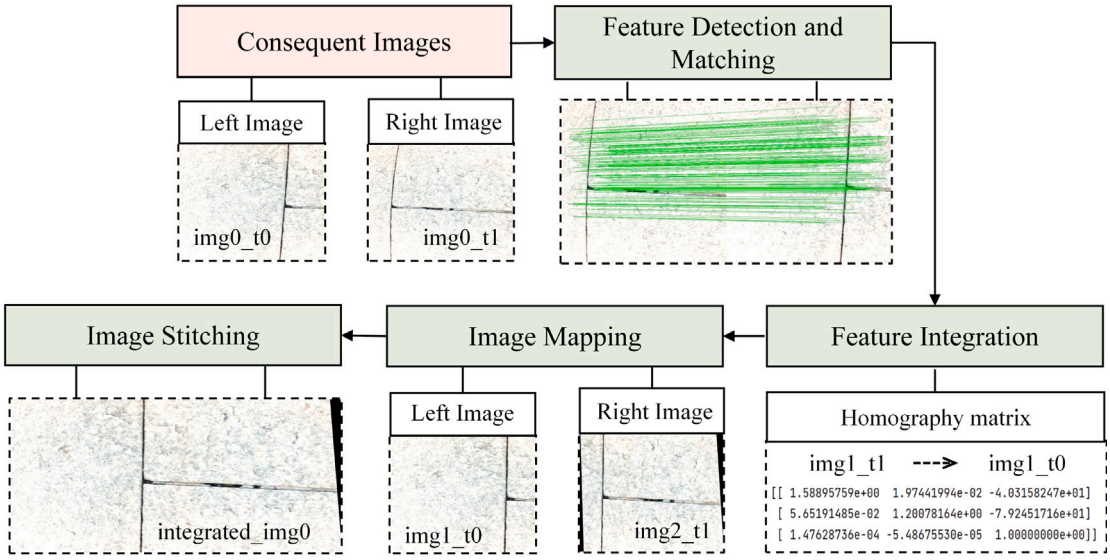
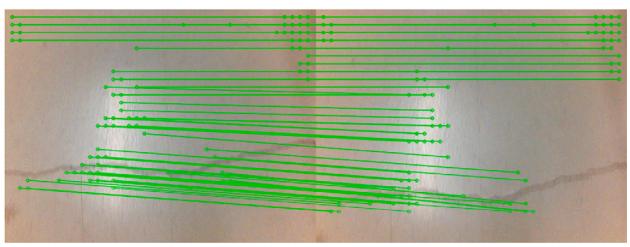
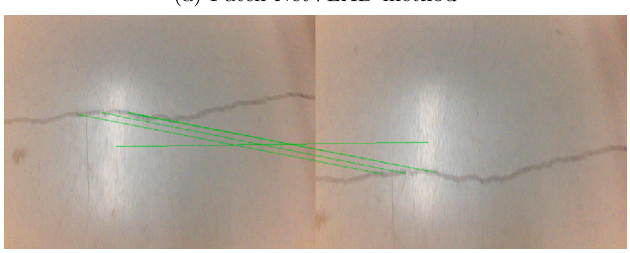


Fig. 7. Pipeline of feature-based image stitching illustrating intermediate steps and results.



(a) Patch-NetVLAD method



(b) The proposed method

Fig. 8. Qualitative comparison between Patch-NetVLAD and the proposed method.

conditions, it tends to generate a significant number of misalignments and noisy correspondences.

In terms of numerical results, the evaluation metrics were computed using both methods on the same dataset, as shown in Table 2. The Patch-NetVLAD method produced an error of 1.12 pixels on the same dataset, which is higher than the proposed method. Meanwhile, Patch-NetVLAD achieved a feature matching success rate of approximately 88.7%, lower than the proposed approach. The quantitative difference highlights the superior precision of the proposed stitching approach, resulting in better-quality texture maps.

Both qualitative and quantitative results confirm the effectiveness of the proposed image stitching method, which is further substantiated through its application in subsequent process. For continuous texture reconstruction over time series, the proposed method progressively integrates frames captured at different time steps to generate a coherent, high-fidelity pavement texture map.

The selected sample for presentation corresponds to a representative longitudinal crack captured under typical daytime conditions, with uneven illumination and dappled shadows caused by surrounding objects. Such lighting variations are frequently encountered in real-world pavement inspections, making this example representative for illustrating the stitching and detection performance of the proposed method. As shown in Fig. 9, the selected image patches are presented at three representative time instances, respectively $t = 39$ s, $t = 50$ s, and $t = 75$ s. The lower part of the figure displays the final result of a continuous texture map reconstructed from the temporally ordered patches. The individual segments, originally separated in time, are seamlessly aligned to form a visually and structurally consistent representation of the selected pavement section.

The time-series results highlight the robustness of the proposed framework in handling continuous input streams. Each new frame is contextually anchored to previous observations, thereby preserving both geometric alignment and texture continuity. The temporal integration of local features, such as crack edges, surface patterns, and illumination transitions, plays a critical role in achieving a high-quality composite. Such continuous representation is crucial for reliable pavement condition analysis, especially in scenarios involving subtle surface degradation or branching cracks.

4.3.2. *Baseline crack detection using YOLOv8-seg*

The convergence behavior of the trained YOLOv8-seg model is depicted in Fig. 10, which presents the evolution of loss components for both the training and validation sets. Specifically, four loss terms are reported, including $\mathcal{L}_{box}$ (box_loss), $\mathcal{L}_{seg}$ (seg_loss), $\mathcal{L}_{cls}$ (cls_loss), and $\mathcal{L}_{dfl}$ (dfl_loss). The consistent downward trends in both training and validation losses during the first 140 epochs indicate effective learning. Notably, a sharp drop in the training losses (e.g., $\mathcal{L}_{box}$ and $\mathcal{L}_{seg}$) is observed at epoch 140, which is an expected behavior resulting from the default training strategy of YOLOv8, where Mosaic data augmentation is automatically disabled during the final 10 epochs. As a result, the model experiences an overfitting tendency on the simpler training images, leading to a slight increase in the validation segmentation loss. For the validation sets, while $\mathcal{L}_{box}$ and $\mathcal{L}_{cls}$ decrease substantially, converging to values below 1.0, the more gradual decline of $\mathcal{L}_{dfl}$ and the relatively elevated value of $\mathcal{L}_{seg}$ after 150 epochs suggest persistent difficulties in accurately modeling fine crack boundaries, particularly in thin or low-contrast regions.

Table 2
Quantitative comparison between Patch-NetVLAD and the proposed method.

Metric	Proposed method	Patch-NetVLAD	Difference
Mean alignment error (pixels)	0.85	1.12	−0.27
Feature matching success rate	96.4%	88.7%	+7.7%

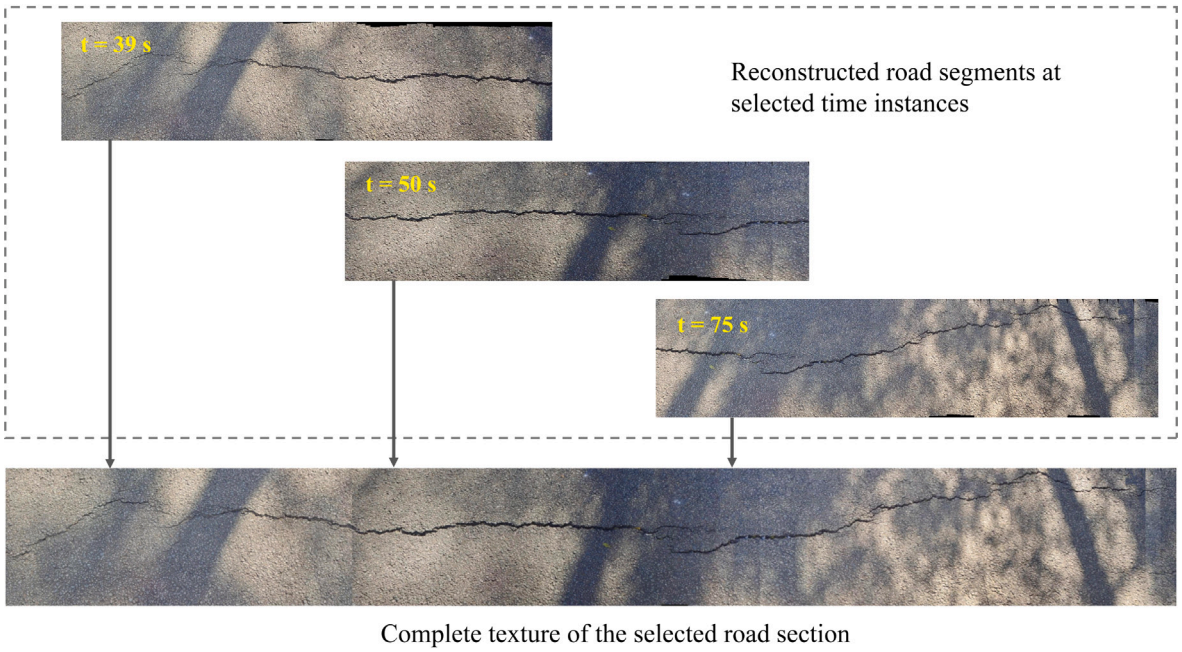


Fig. 9. Texture reconstruction of the detected pavement section.

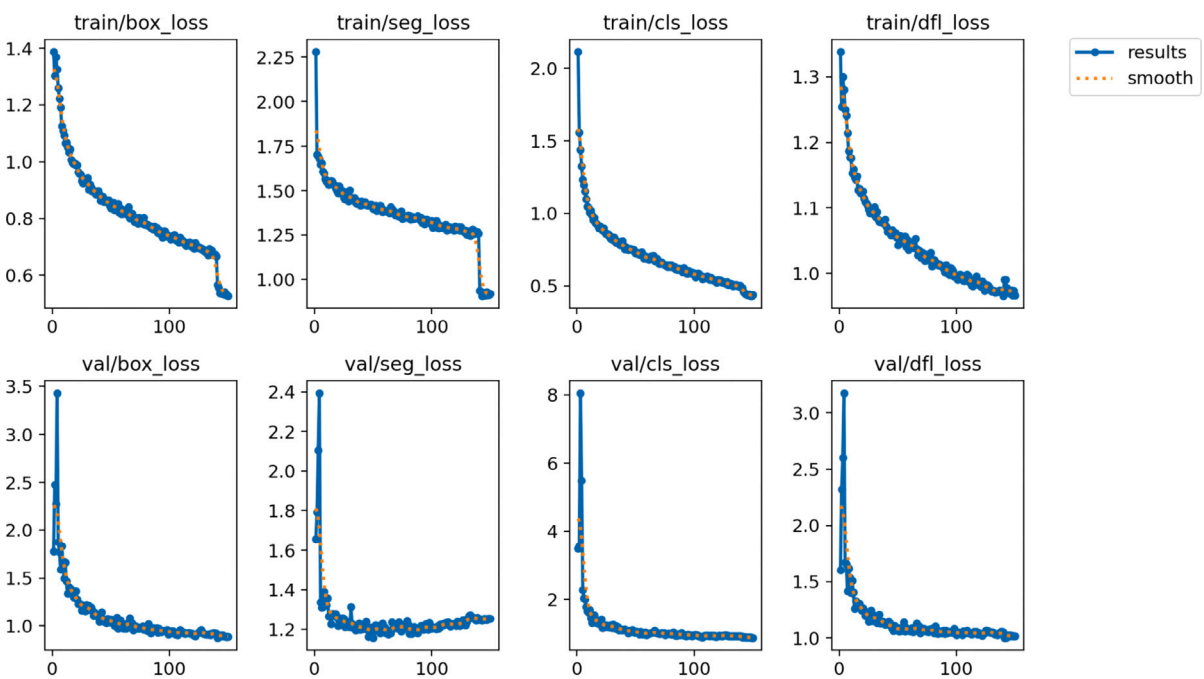


Fig. 10. Training loss curves of the detection model.

The model's performance on bounding boxes(B) and masks(M) is evaluated using precision, recall, and mAP at IoU thresholds of 0.50 and 0.50–0.95, as shown in Fig. 11. For bounding box detection, precision(B) and recall(B) increase rapidly within the first 20 epochs and then stabilize around 0.80, indicating reliable crack detection with few false positives. For segmentation masks, precision(M) and

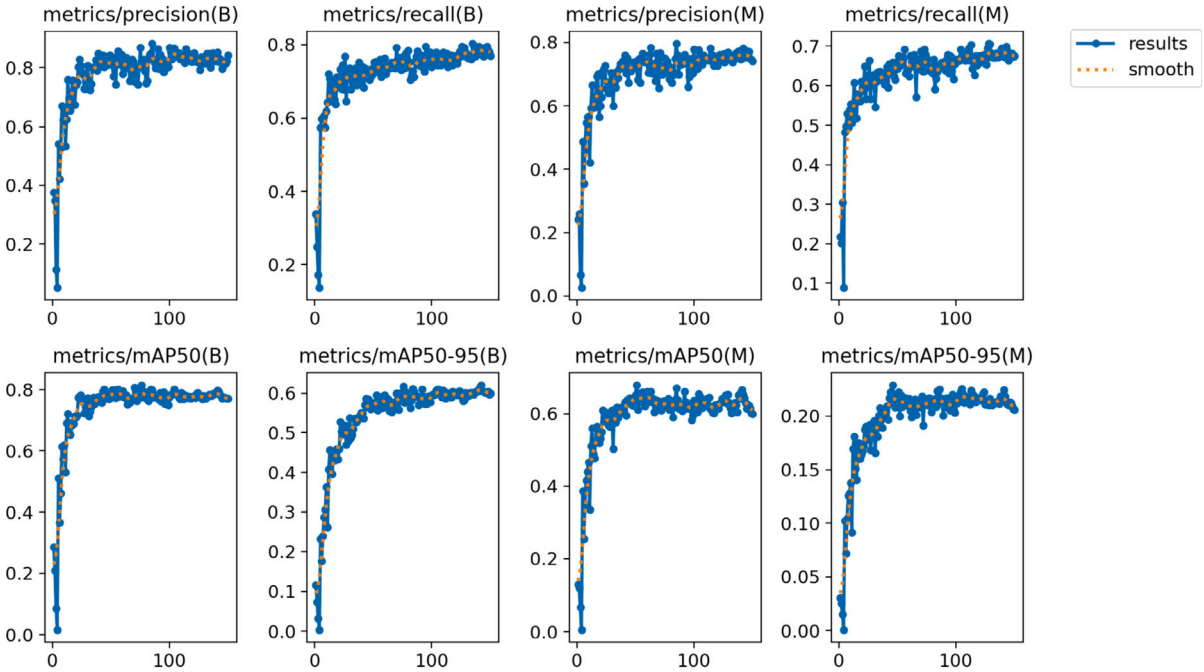


Fig. 11. Progression of the performance metrics during training.

recall(M) also exhibit stable convergence, though recall(M) remains slightly lower than recall(B) due to the complex and irregular crack geometries in the test set. The mAP50 scores further confirm model accuracy, stabilizing at 0.78 for bounding boxes and 0.63 for masks. At stricter evaluation thresholds, the mAP50–95 scores decrease to 0.58 for bounding boxes and 0.23 for masks, indicating that while the model performs well in coarse detection, fine-grained segmentation of thin or small crack structures remains challenging. The discrepancy aligns with the observed deficiencies in mAP50-95(M), suggesting that the quality of segmentation is the primary constraint on performance. Nevertheless, while the late-stage overfitting affects the fine boundary modeling at the very end, the overall detection metrics remained stable, ensuring the baseline's availability.

A selection of testing samples from the trained detection model is shown in Fig. 12. In general, the numerical and visual results suggest that the trained YOLOv8-seg model exhibits robust generalization performance across both the detection and segmentation tasks.

In the raw detection output of the inspection road section shown in Fig. 13, it is evident that the YOLOv8-based model consistently locates the cracks but struggles to define precise crack boundaries, as shown across the sample segments at $t = 39$ s, $t = 50$ s, and $t = 75$ s. Specifically, the rectangular bounding boxes generated by YOLOv8 are relatively coarse, and minor interruptions in crack continuity often lead to misidentification. The limitation is particularly evident in areas with complex textures or shadows, where the model fails to capture thinner or partially occluded cracks, resulting in incomplete or inaccurate representations.

The limitations stem from the model's training dataset. The YOLOv8 was trained primarily on lower-resolution images, which makes it ill-suited for handling the high-resolution textures and fine details present in the pavement images used in the study. As a result, cracks in high-resolution textures are not captured with sufficient precision, particularly in terms of boundary detection where fine details are critical.

Although the YOLOv8-based model delivers rapid results suitable for real-time crack detection, its outputs are inadequate for high-precision boundary segmentation, especially for longitudinal cracks in high-resolution imagery. Therefore, additional refinement techniques are necessary to improve the accuracy of the crack boundaries.

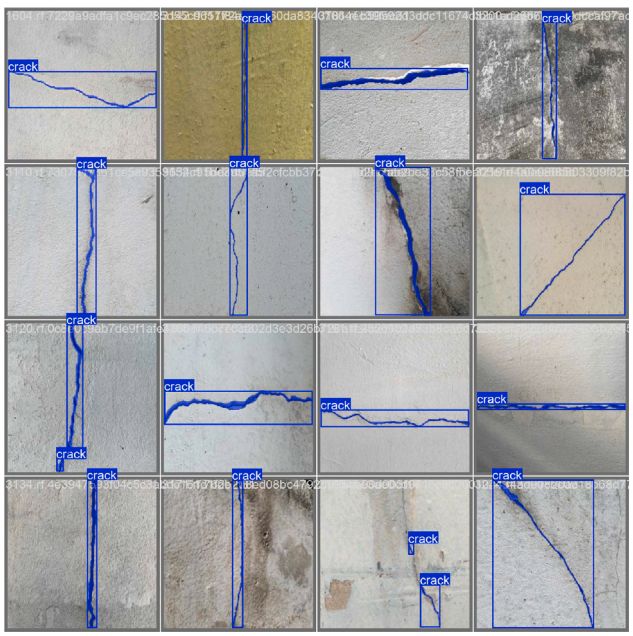
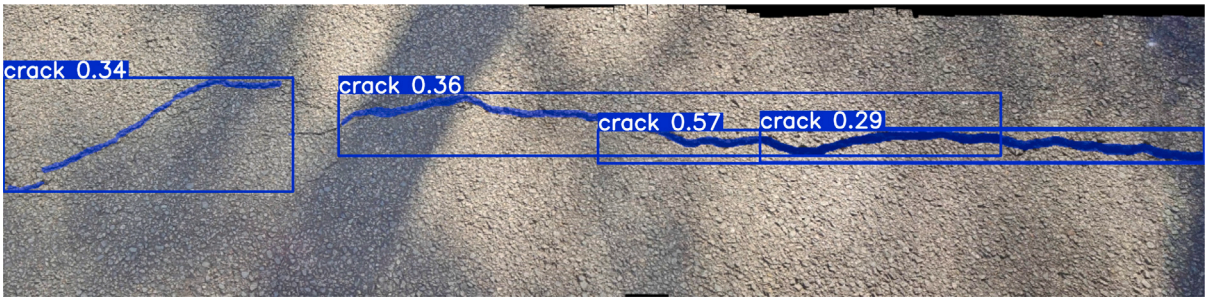


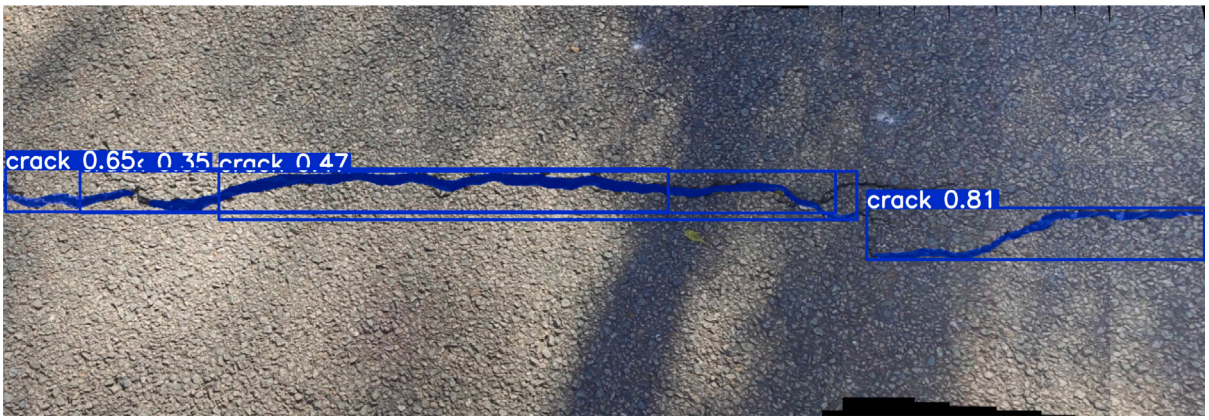
Fig. 12. The testing samples of the trained detection model.

4.3.3. High-fidelity crack reconstruction with precision augmentation

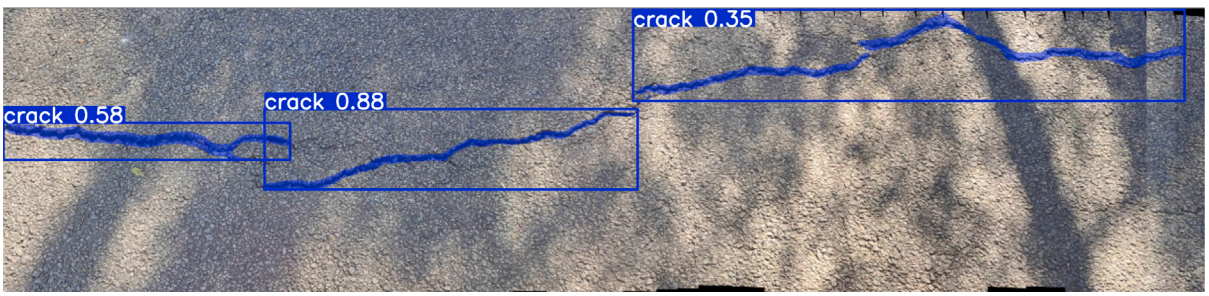
The preliminary reconstruction generates coarse longitudinal cracks by aggregating temporally consistent crack segments using a minimum confidence threshold of $C_0 = 0.3$. In Fig. 14, the results of the integrated methods are presented at the time instances $t = 39$ s, $t = 50$ s, and $t = 75$ s. As the time series progresses, the refinement process progressively enhances the delineation of crack contours and the recovery of fine structural details. The sliding window adaptively targets uncertain or



(a) Raw detection at $t = 39\,\mathrm{s}$



(b) Raw detection at $t = 50\,\mathrm{s}$



(c) Raw detection at $t = 75\,\mathrm{s}$

Fig. 13. Raw output of YOLOv8-based crack detection on the time series.

fragmented regions, where sequential morphological operations and watershed segmentation are performed to connect disjointed segments and smooth noisy edges.

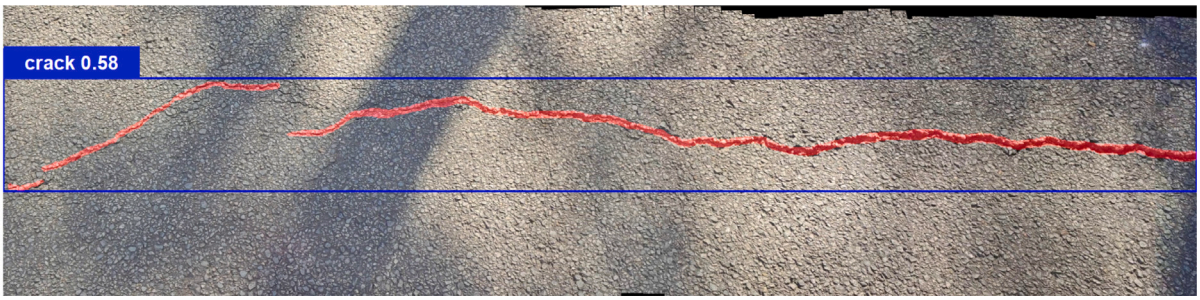
In terms of segmentation accuracy, YOLOv8 effectively localizes cracks but struggles to delineate crack boundaries precisely, particularly when cracks are thin, fragmented, or partially occluded. The refined method, which integrates YOLOv8 with the morphological and watershed techniques, demonstrates a marked improvement in segmentation accuracy. The statistical results of the generalization assessment were given in Section 5.2.

For a more intuitive comparison, Fig. 15 provides a detailed view that contrasts the raw YOLOv8 output with the refined detection results for the given sample. Based on the reconstructed pavement texture, the segmentation results obtained by YOLOv8 and by the proposed method are rendered in different colors for visual comparison. In general, the proposed method generates crack edges that are narrower and more continuous compared to the raw YOLOv8 output. Furthermore, two local regions are cropped and enlarged on the left and right sides for detailed local visualization.

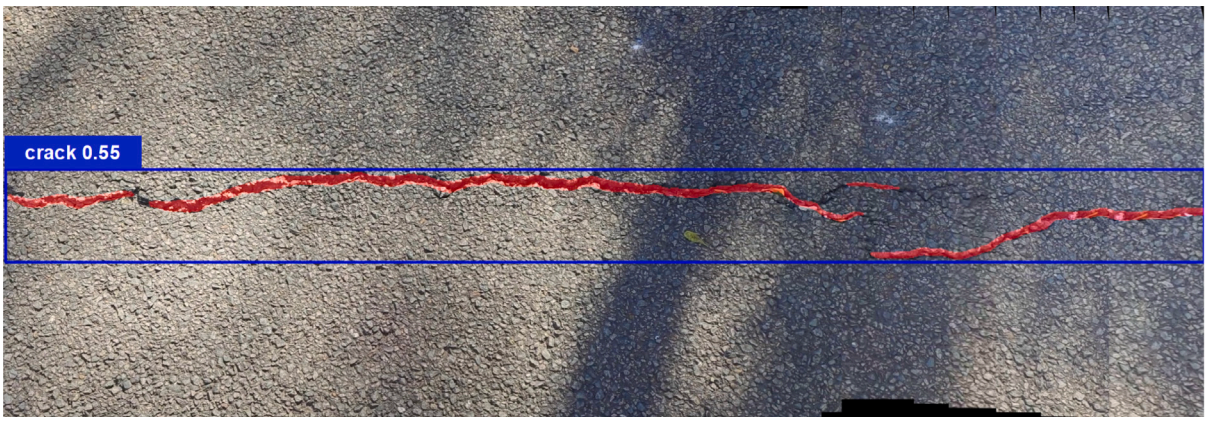
The left zoom highlights the effect of edge erosion. The raw YOLOv8 output exhibits coarse and blurred crack boundaries with edges that are over smoothed and dilated, leading to exaggerated crack widths and a loss of fine details. By contrast, the proposed method sharpens the boundaries via erosion and watershed segmentation, producing high-resolution results consistent with the physical structure of longitudinal cracks, thereby achieving refined segmentation with improved edge clarity.

The right zoom highlights the effect of edge dilation. The raw YOLOv8 output fails to capture numerous fine, low-contrast branches, resulting in the omission of discontinuous and fragmented crack segments. In contrast, the proposed method employs dilation and watershed refinement to longitudinally extend the crack boundaries, enhancing continuity and restoring subtle branches that were previously undetected. The local extension enables a more accurate characterization of crack propagation, and provides a more complete representation of the longitudinal cracks.

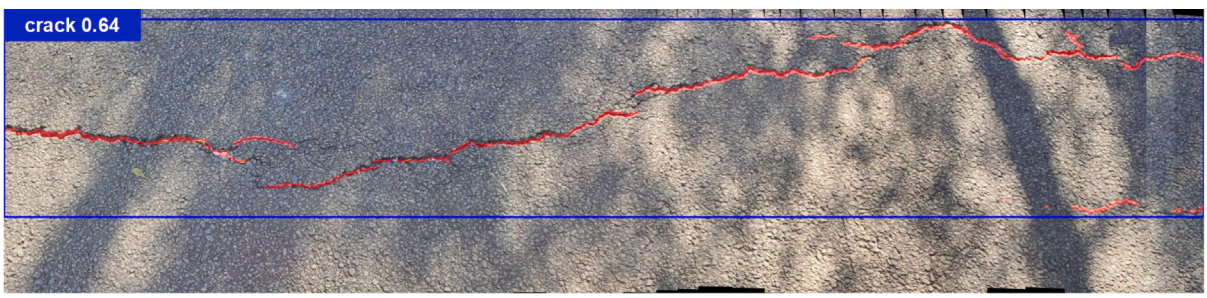
Collectively, the local comparisons highlight the complementary effects of the refinement process. The hybrid refinement framework



(a) Refined detection at $t = 39\,\mathrm{s}$



(b) Refined detection at $t = 50\,\mathrm{s}$



(c) Refined detection at $t = 75\,\mathrm{s}$

Fig. 14. Refined output of crack detection based on the proposed method.

not only corrects boundary misalignments and sharpens the edges, but also recovers missed micro-structures, resulting in more continuous, complete, and precise segmentation than raw YOLOv8 outputs.

5. Evaluation and discussion

While Section 4 presents the intermediate and final results of each module in the proposed pipeline, Section 5 provides a comprehensive evaluation of the proposed method framework against state-of-the-art approaches, including quantitative comparison, ablation analysis, and generalization assessment.

5.1. Quantitative comparison with state-of-the-art methods

To comprehensively evaluate the performance of the proposed method, we conducted a comparative study against three representative state-of-the-art(SOTA) methods, including the classic **U-Net** baseline, the specialized crack detection model **DeepCrack**, the recent transformer model **SAM 3**, and the SegFormer-based segmentation architecture **CrackSegFormer**. Specifically, U-Net serves as a classic encoder–decoder baseline widely adopted in medical and industrial segmentation tasks due to its effective feature fusion via skip connections. DeepCrack is a specialized model explicitly designed for crack detection, leveraging hierarchical feature learning to capture crack structures at multiple scales. SAM 3 is a transformer-based architecture adapted from the "Segment Anything" paradigm to handle diverse segmentation targets. CrackSegFormer exploits self-attention mechanisms to capture long-range contextual dependencies, enabling more accurate delineation of complex crack topologies. Our proposed method combines a custom-trained YOLOv8-seg model with a high-fidelity reconstruction (HFR) pipeline, referred to as **YOLOv8+HFR**, for comparison against the aforementioned methods.

The experiments were conducted on the strictly held-out testing set containing 340 images sourced from CFD and DeepCrack datasets (as partitioned in Section 4.1), each annotated with pixel-wise ground truth masks. The dataset covers diverse crack types, including numerous fine and shallow crack branches. All models were evaluated on the same testing set under identical experimental settings.

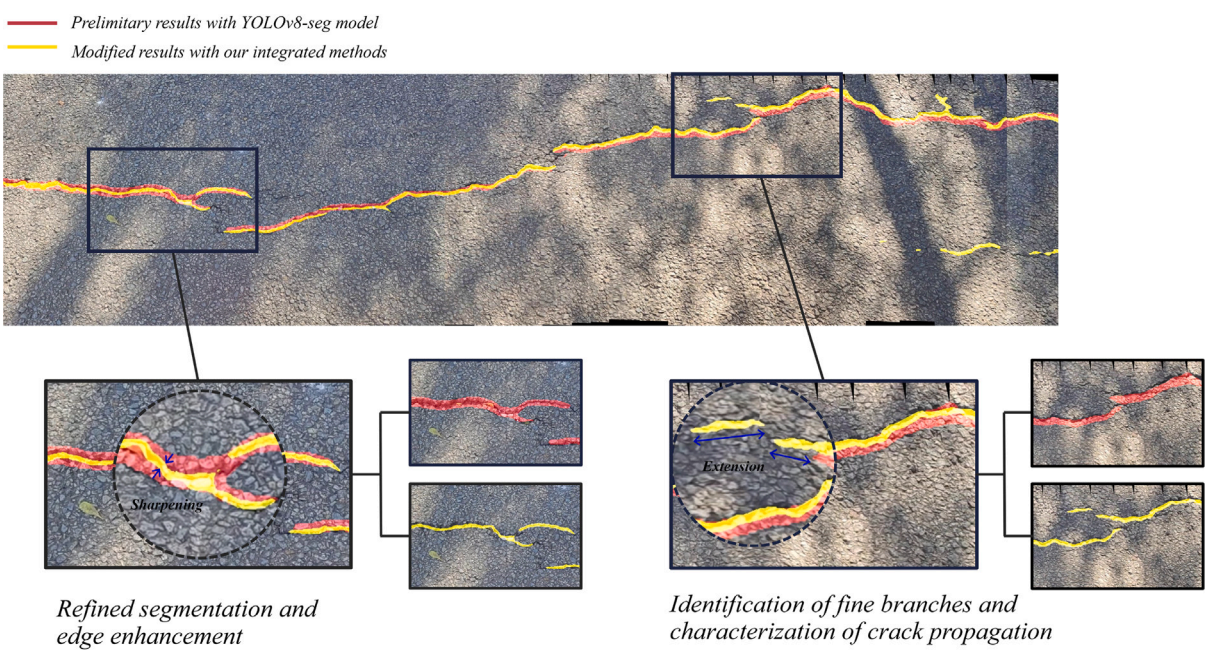


Fig. 15. Local comparison between raw and refined output of longitudinal crack detection.

Table 3
Quantitative comparison with state-of-the-art methods on the testing set.

Method	IoU	Precision	Recall	F1-Score	Time (ms)	FPS
U-Net	0.5123	0.7114	0.6426	0.6752	254.72	3.97
DeepCrack	0.2728	0.4986	0.3924	0.4390	344.18	2.94
SAM 3	0.4400	0.5073	0.7680	0.6110	2413.81	0.43
CrackSegFormer	0.5452	0.6915	0.7152	0.7000	258.76	3.92
Ours	**0.6128**	**0.7489**	**0.7781**	**0.7632**	569.84	1.71

The input image resolution was set to 640 × 640, and all models performed inference at 512 × 512 to ensure uniformity. The optimal parameter configurations for each model were used. Following detection, the performance metrics IoU, Precision, Recall, F1-score, and computational efficiency (Time in ms and FPS) were calculated across the testing set, as shown in Table 3. Additionally, six representative images capturing different crack characteristics were selected to visualize the segmentation results of all methods, for qualitative comparison between the SOTA methods and ours, as shown in Fig. 16. The corresponding Error Maps for our method are also presented to highlight the effectiveness of our approach.

Quantitative analysis. As shown in Table 3, our method outperforms the comparison methods across the key metrics. U-Net acts as a conservative predictor, achieving relatively high Precision (0.7114) but moderate Recall (0.6426), indicating it tends to miss fine crack details. DeepCrack shows limited generalization capability on the dataset, with the lowest IoU of 0.2728. SAM 3, leveraging the powerful generalization of foundation models, achieves a high Recall (0.7680) comparable to ours; however, its low Precision (0.5073) suggests a tendency towards over-segmentation and false positives. CrackSegFormer emerges as the most competitive baseline, securing a significantly higher Recall (0.7152) than U-Net while maintaining a much higher Precision (0.6915) than SAM 3. In contrast, our method achieves the best balance between Precision (0.7489) and Recall (0.7781). Notably, while maintaining a high Precision level superior to CrackSegFormer, our method significantly boosts the Recall rate, resulting in a 9.0% increase in F1-score (0.7632) and a 12.4% improvement in IoU(0.6128) compared to the second-best model CrackSegFormer. The balanced performance indicates that the proposed method is particularly well suited for practical pavement inspection scenarios, where both detection accuracy and robustness are critical.

Computational efficiency. In addition to accuracy metrics, the proposed method exhibits acceptable computational efficiency. As indicated in Table 3, our method runs at 1.71 FPS with an average processing time of 569.84 ms per image, which is slower than U-Net (3.97 FPS), CrackSegFormer (3.92 FPS), and DeepCrack (2.94 FPS), but significantly faster than SAM 3 (0.43 FPS). Considering the substantial performance gains in IoU and F1-score, these results indicate that the proposed pipeline achieves a reasonable trade-off between segmentation accuracy and computational cost. Notably, when only the YOLOv8-seg model is employed, the average inference time on the same testing set is reduced to 65.16 ms per image, achieving a throughput of 15.5 FPS. This demonstrates the deployment feasibility of the proposed framework under the intended inspection scenario, where real-time crack localization can be achieved using the lightweight YOLOv8-seg model on the inspection platform, while the computationally intensive HFR module can be executed on a computing server for high-fidelity reconstruction.

Qualitative analysis. To further analyze the robustness of the algorithms, we visualized the detection results on six representative samples in Fig. 16. These samples characterize diverse pavement conditions and crack types, including complex textures with markings(Sample 2), varying lighting (Sample 4), complex topology with branching cracks (Samples 3 and 5), and alligator cracking patterns (Samples 1 and

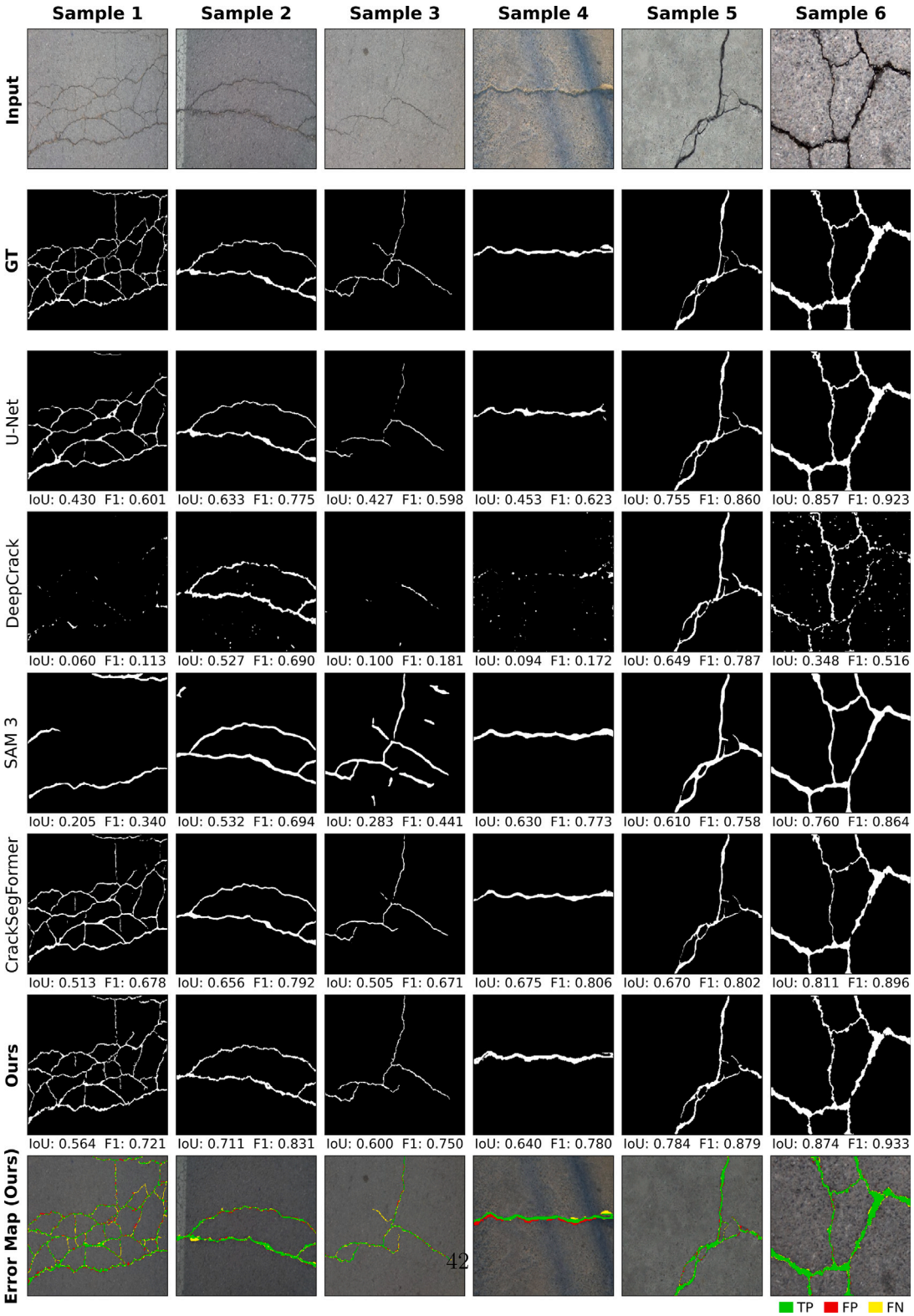


Fig. 16. Qualitative comparison on six representative samples.

6). Collectively, these samples cover a wide range of crack combinations in terms of depth, width, and severity, making them suitable for qualitative comparison across different methods.

- **Noise robustness:** Our method demonstrates strong adaptability to complex environmental interferences, achieving a balance between high precision and robust generalization across various crack types. In comparison, the SOTA methods exhibit limitations. DeepCrack is highly susceptible to background noise, resulting in a low F1-score (e.g. Samples 4 and 6), whereas SAM 3 tends to over-segment boundaries (e.g., Samples 2 and 3), leading to reduced precision. CrackSegFormer exhibits excellent noise

robustness, particularly under varying illumination (e.g., Sample 4), where it correctly avoids false positives; however, it occasionally exhibits slight over-segmentation on complex textures (e.g., Samples 1 and 2). In contrast, our method effectively suppresses background noise while preserving structural integrity, particularly for fine and shallow cracks with complex patterns (Samples 1–3), where precision consistently exceeds 0.7. For coarse and deep cracks (Sample 6), precision can even reach 0.961. Although U-Net and CrackSegFormer produce accurate overall shapes and can handle small or uneven cracks, they are less robust when dealing with fragmented cracks or irregular edges (e.g., Sample 1), resulting in lower precision in these cases. These results indicate that our method not only preserves fine crack details but also maintains stable performance across diverse crack geometries. Consequently, the Error Maps are dominated by green (True Positive) regions with minimal red (False Positive), confirming accurate delineation without interference from shadows or pavement textures.

- **Continuity of fine cracks:** Recovering the topological continuity of fine, intricate cracks remains a significant challenge. In the representative samples, our method shows a substantial advantage in Recall and connectivity. In scenarios characterized by alligator cracking (Sample 1) and fine branching (Sample 3), baseline methods like U-Net often yield fragmented and disconnected segments, resulting in low Recall rates (e.g., 0.485 for Sample 3). DeepCrack exhibits even poorer recovery, with generally low Recall values and missing many crack components; it only performs adequately when background noise is minimal and crack edges are clear (e.g., Sample 5), making it overall unsuitable for fine cracks. SAM 3 achieves very high overall Recall, sometimes exceeding that of our method, indicating that it captures most crack regions; however, this comes at the cost of precise edge delineation, and its Recall on fine, mesh-like cracks (Sample 1) remains low, revealing limitations in generalization for small-scale structures. CrackSegFormer mitigates fragmentation significantly, effectively bridging gaps in complex topologies (e.g., Samples 3 and 5); however, it still falls slightly short in capturing the most subtle, hair-like crack branches (Samples 1-3) when compared to our method. In contrast, our approach successfully restores these subtle structures, boosting Recall to 0.724 in Sample 3 and 0.732 in Sample 1. This improvement is visually evident in the Error Maps, where continuous green lines trace the intricate crack networks, few yellow (False Negative) gaps. These observations demonstrate our model's capability to preserve the topological completeness of fine cracks, which is critical for accurate pavement condition assessment.

Overall, the visual comparison aligns with the quantitative data. the proposed method demonstrates superior applicability across varying crack depths and widths, successfully recovering fine, continuous structures while minimizing false positives in complex backgrounds.

5.2. Generalization assessment in real-world pavement scenarios

5.2.1. Robustness verification under complex real-world conditions

To verify the robustness and generalization capability of the proposed pipeline in practical engineering applications, we conducted a comprehensive generalization assessment using a total of 13 real-world pavement inspection datasets collected in field experiments. These datasets span diverse scenarios, including varying weather conditions, pavement types, lighting environments, and crack morphologies. The diversity of experiment conditions ensures that the evaluation comprehensively covers a wide range of real-world conditions.

The specific characteristics of the 13 tested scenarios are detailed in Table 4. The datasets were meticulously classified to verify the method effectiveness under realistic and demanding conditions. In terms of weather, the majority of data (11 sets) were collected under sunny conditions. This selection was intentional to capture a wide range of lighting conditions, particularly scenarios with complex light-shadow interplay. Notably, two datasets (Samples 6 and 10) were collected under snowy conditions after snowfall, with residual de-icing agents on the pavement surface, introducing unique texture interference for the detection algorithm.

Regarding lighting conditions, we specifically prioritized "mottled" lighting environments (e.g., under tree canopies). These scenarios are typical in urban roads but highly challenging due to the irregular high-contrast shadows that can be easily confused with cracks. The pavement materials include both asphalt and concrete, featuring diverse distress types such as longitudinal cracks, alligator cracking, transverse cracks, and fine fissures. Additional disturbances, including road markings, spalling, joints, debris, and uneven aging, further increase the complexity of the detection task. The comprehensive setup effectively validates the method's generalization and robustness across distinct real-world domains.

The detection targets in the experiment were the continuous pavement texture maps reconstructed by the proposed stitching algorithm. Each texture map contains one or multiple continuous crack instances spanning the entire road section. To quantitatively evaluate the detection performance, we calculated five key metrics for each of the 13 scenarios, as presented in Table 5.

The metrics listed in the table serve as indicators of different crack characteristics. Total Length reflects the overall scale of the distress; Avg Width and Logical Count describe the morphological complexity and the number of independent crack segments; Continuity measures average continuous length, serving as a critical indicator of crack integrity, where higher values imply smoother and less fragmented detection; and SCI measures the dominance of the main crack, with values closer to 1 indicating strong resistance to fragmentation.

To facilitate a more intuitive assessment of the algorithm performance under complex real-world conditions, representative detection results are visualized in Fig. 17. Based on the quantitative metrics and detailed manual inspection, the detection outcomes were categorized into four states. Specifically, a result was labeled as *Success* when the Continuity metric exceeded 1000 pixels and manual verification confirmed that more than 60% of the actual crack regions were correctly reconstructed as continuous structures. Cases with crack coverage between 20% and 60% were classified as *Omission*, while *Noise* state is defined when the false positive area exceeds 10% of the detected region. Those with more than 90% of crack regions missed were considered *Failure*. According to this criterion, the proposed method successfully detected valid cracks in 12 out of the 13 tested scenarios, achieving an overall success rate of 92.3%. The detailed performance characteristics are analyzed below.

- **Overall integrity and topological preservation.** From an overall perspective, the quantitative results in Table 5 demonstrate that the proposed method preserves crack integrity and topology effectively across diverse pavement conditions. The average Continuity reaches 3732.42 pixels, indicating that detected cracks are generally reconstructed as long and coherent structures rather than fragmented segments. Meanwhile, the average SCI of 0.70 suggests that, in most scenarios, the dominant crack component accounts for the majority of the total crack length, reflecting strong resistance to structural fragmentation and stable topological reconstruction performance.
- **Robustness under complex lighting conditions.** Under complex lighting conditions with strong light–shadow contrast, the proposed method exhibits notable robustness. In mottled illumination environments, which are widely regarded as one of the most challenging scenarios for visual pavement inspection, the algorithm consistently identifies crack features in both illuminated and shadowed regions. As illustrated by Sample 11 (Fig.

Table 4
Details of the 13 real-world pavement inspection datasets covering diverse environmental conditions.

No.	Surface type	Weather	Lighting	Crack type	Surface condition
1	Asphalt	Sunny	Strong	Longitudinal/Fine	Aging surface
2	Asphalt	Sunny	Strong	Longitudinal/Fine	Aging surface
3	Asphalt	Sunny	Strong	Transverse	Markings, Strong shadows
4	Asphalt	Sunny	Strong	Alligator	Spalling
5	Asphalt	Sunny	Mottled	Longitudinal	Strong shadows
6	Asphalt	Snowy	Weak	Longitudinal	De-icing agent
7	Asphalt	Sunny	Mottled	Longitudinal	Debris, Strong shadows
8	Asphalt	Sunny	Mottled	Mixed	Strong shadows
9	Concrete	Sunny	Mottled	Alligator/Fine	Joints
10	Concrete	Snowy	Mottled	Longitudinal	De-icing agent
11	Concrete	Sunny	Mottled	Longitudinal penetrating	Strong shadows
12	Concrete	Sunny	Mottled	Alligator	Strong shadows
13	Asphalt	Sunny	Mottled	Alligator	Debris, Strong shadows

Table 5
Quantitative detection results of the proposed method on 13 real-world datasets.

No.	Count	Total length (px)	Avg width (px)	SCI (–)	Continuity (px)	State
1	3	14,050	15.75	0.5951	4683.33	Success
2	2	7830	15.92	0.5476	3915.00	Success
3	2	10,769	13.79	0.7980	5384.50	Success
4	1	6776	13.18	1.0000	6776.00	Success
5	3	5912	14.02	0.5338	1970.67	Success
6	4	7873	13.79	0.6671	1968.25	Success
7	3	8155	15.04	0.6959	2718.33	Success
8	2	6113	15.78	0.8078	3056.50	Success
9	17	55,120	13.34	0.8327	3242.35	Success
10	4	6693	15.82	0.4018	1673.25	Success
11	1	8591	15.96	1.0000	8591.00	Success
12	2	7812	15.63	0.6457	3906.00	Success
13	9	5726	15.28	0.5138	636.22	Omission
Avg	**5**	**11,647**	**14.87**	**0.70**	**3732.42**	**SR: 92.3%**

17a), a longitudinal penetrating crack on a concrete surface is reconstructed as a single continuous structure, achieving an SCI of 1.0000 and a logical count of 1, indicating complete preservation of crack continuity. A similar phenomenon is observed in Sample 5, where fine crack branches in both bright and shaded areas are successfully detected. The relatively lower Continuity value in this case is attributed to the presence of inherently fragmented micro-branches in shadowed regions, which further confirms the algorithm's sensitivity to fine-scale crack structures. Consistent situations can also be observed in Sample 9 and Sample 6, where numerous fragmented branches are faithfully reconstructed under mottled lighting, demonstrating strong capability in detecting complex crack morphology without being affected by illumination-induced noise.

In contrast, extremely strong illumination represents another challenging condition, as excessive brightness may cause surface whitening and reduce crack-to-background contrast. Samples 2 and 3 show that the proposed method maintains good generalization under such conditions. Specifically, in Sample 2, shallow and thin cracks are still detected with high completeness, although strong illumination slightly reduces edge contrast and introduces minor discontinuities. Sample 3 further demonstrates that the algorithm maintains robust and stable performance even under multiple challenges including strong illumination, road markings and localized shadows, achieving a high SCI of 0.7980 while accurately reconstructing both the main crack and its fine branches. Overall, these results confirm that the proposed method is highly robust to diverse lighting conditions and exhibits strong immunity to surface noise and illumination interference.

- **Adaptability to weather variations and surface residues.** To further evaluate robustness under varying weather conditions, special attention was given to post-snowfall scenarios, specified

Table 6
Quantitative comparison of average metrics and success rates among different methods on the real-world datasets.

Method	Count	Total length (px)	Avg width (px)	SCI (–)	Continuity (px)	Rate breakdown
U-Net	4	6723	12.59	0.68	2948.13	Success: 53.8%; Omission: 30.8% Noise: 7.7%; Failure: 7.7%
DeepCrack	21	6982	15.24	0.43	781.12	Success: 15.4%; Omission: 23.1% Noise: 23.1%; Failure: 38.5%
SAM 3	6	10,621	17.11	0.57	2609.37	Success: 84.6%; Omission: 15.4% Noise: 0%; Failure: 0%
CrackSegFormer	7	8042	12.48	0.69	1874.92	Success: 53.8%; Omission: 7.7% Noise: 7.7%; Failure: 30.8%
Ours	**5**	**11,647**	**14.87**	**0.70**	**3732.42**	**Success: 92.3%**; Omission: 7.7% Noise: 0%; Failure: 0%

as Sample 6 and 10. In such cases, temperature variations and residual de-icing agents significantly alter pavement appearance, often producing locally whitened regions and irregular textures that interfere with visual crack detection. As shown in Sample 6 (Fig. 17f), a fine crack located in the upper-left region is partially missed due to heavy coverage of de-icing material. In Sample 10, the SCI drops to 0.4018, and manual inspection reveals three noticeable discontinuities along the main crack, resulting in an increased logical count of 4. The results indicate that the variation in surface conditions may cause some degree of fragmentation.

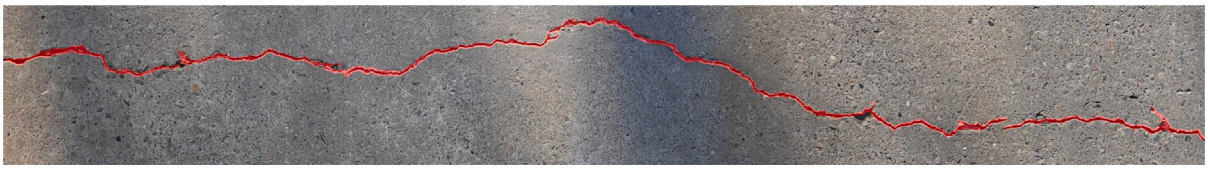
- **Large-scale crack reconstruction capability.** The capability of large-scale crack reconstruction is highlighted by Sample 9, which represents a highly complex concrete pavement with extensive distress. Despite the large spatial scale, with a total crack length exceeding 55,000 pixels and a logical count of 17, the algorithm maintains a high SCI of 0.8327. The result in Fig. 17b visually and quantitatively demonstrates the advantage of the proposed framework in stitching long pavement textures and reconstructing extensive crack networks spanning tens of meters without severe fragmentation.
- **Omission and limitation analysis.** The only omission occurs in Sample 13, which features a shallow, net-like alligator crack pattern under mottled lighting with surface debris. In this case, the appearance of low crack depth, weak contrast, and high-frequency background noise jointly degrade the visual saliency of crack features on the illuminated side. The fragmented responses fail to satisfy the structural coherence requirements of the YOLOv8-based post-processing logic, leading to omission.

5.2.2. Comparative validation against SOTA methods

To rigorously evaluate the effectiveness of the proposed YOLOv8+HFR framework in complex real-world environments, we compared it with the four state-of-the-art methods in Section 5.1. The quantitative results are summarized in Table 6.

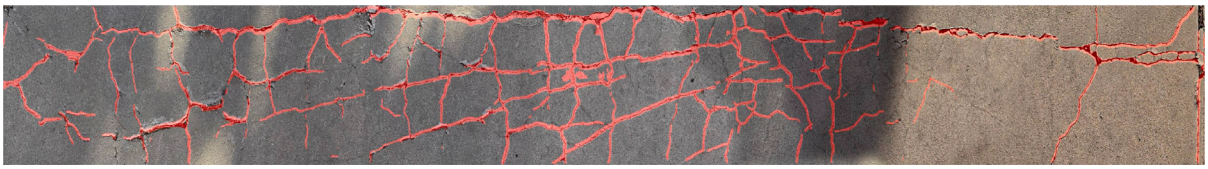
The qualitative visualizations in Fig. 18 corroborate the statistical data, revealing distinct behaviors under complex environmental conditions. A common deficiency observed among the SOTA methods is their limited capability in handling complex topological structures, particularly for alligator cracking and fine mesh patterns. In Sample 9 and 13, all baseline algorithms exhibited poor detection performance, characterized by significant omissions or total failures. Beyond the commonalities, the methods displayed distinct morphological tendencies and performance trade-offs:

- **U-Net (High Precision, Low Recall):** U-Net tends to predict thinner crack widths, with an average crack width of 12.59 pixels, which is significantly smaller than that of the other methods. While it produces clean results with minimal noise, its conservative prediction strategy leads to a high Omission rate (30.8%). Quantitatively, although its SCI (0.68) and Continuity are second only to our method, its total crack length is significantly lower, indicating that while U-Net maintains topological smoothness in the areas it successfully localizes, it fails to perceive substantial portions of the crack network. As visualized in Samples 3 and 5, U-Net is particularly sensitive to illumination variations, and crack segments located in shadowed or mottled lighting regions are prone to omission, which accounts for a significant portion of the performance discrepancy.
- **SAM 3 (Good Generalization, Rough Edges):** SAM 3 demonstrates impressive generalization capabilities, achieving a high success rate of 84.6% without any "Noise" or "Failure" cases. It is robust to lighting variations and captures most crack structures. However, its segmentation boundaries are notably rough and dilated, resulting in the highest average width (17.11 px). Still, it struggles with complex topological structures. In the alligator cracking scenario (Sample 9, Fig. 18(d)), SAM 3 missed the fine, uneven mesh details, yielding a very low SCI of 0.2178.
- **DeepCrack (High Fragmentation, Poor Robustness):** DeepCrack performed poorly in these challenging scenarios, with the highest Failure rate (38.5%) and Noise rate (23.1%). It is highly sensitive to environmental interference and fails to detect fine cracks. Even in the most successful case (Sample 11, Fig. 18(e)), significant false positives caused by shadow interference are visible, and the detected boundaries are considerably wider than the actual cracks. The overall performance of DeepCrack in these scenarios is markedly inferior to the other evaluated algorithms.
- **CrackSegFormer (Coarse Segmentation and Poor Semantic Selectivity):** Despite its strong performance on the strictly controlled testing set, CrackSegFormer struggles significantly with generalization in complex real-world conditions. Quantitatively, it achieves a competitive average SCI (0.69) and total length (8042 px); however, it suffers from severe topological fragmentation across the evaluation scenarios, as evidenced by a low average Continuity (1874.92 px). While it yields the lowest average crack width (12.48 px) among all methods, it often fails to preserve the topological details of the actual distress. As shown in Sample 7 (Fig. 18(f)), the model captures the main crack relatively completely, albeit with coarse boundaries. The abnormal Logical Count (17) and severely degraded Continuity (568.12) in this case are mainly attributed to misclassifying black padding regions from panoramic stitching as fragmented cracks. This indicates that the model is highly sensitive to general strong gradient changes but lacks the semantic selectivity required to effectively distinguish domain-specific features. Consequently, its



(a) Sample 11: concrete, mottled, longitudinal penetrating crack

Count: 1, SCI: 1.0000, Continuity: 8591.00

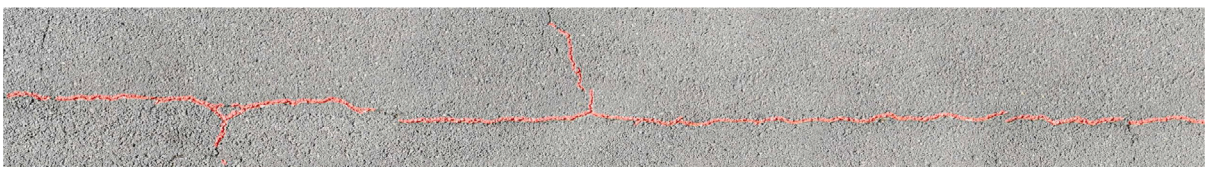


(b) Sample 9: concrete, mottled, alligator/fine crack

Count: 17, SCI: 0.8327, Continuity: 3242.35

(c) Sample 5: asphalt, mottled, longitudinal crack

Count: 3, SCI: 0.5338, Continuity: 1970.67

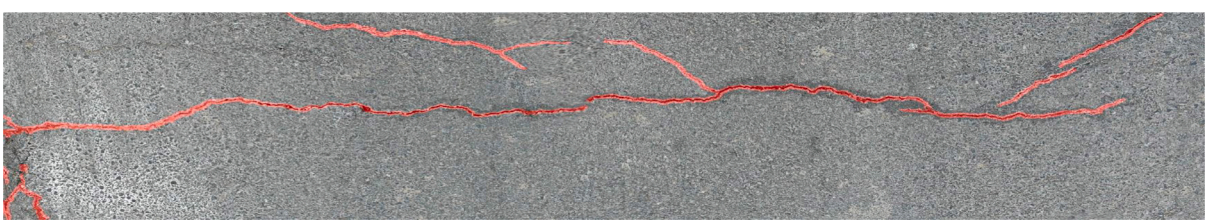


(d) Sample 2: asphalt, strong lighting, longitudinal/fine crack

Count: 2, SCI: 0.5476, Continuity: 3915.00

(e) Sample 3: asphalt, strong lighting, transverse crack, with markings

Count: 2, SCI: 0.7980, Continuity: 5384.50



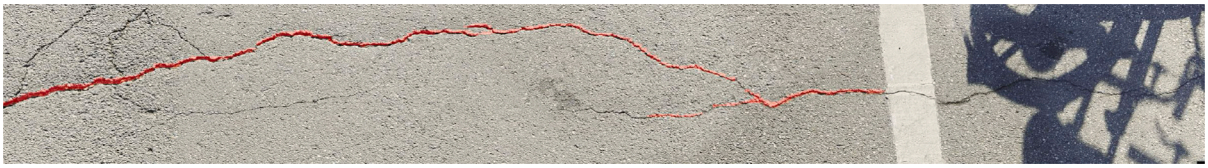
(f) Sample 6: asphalt, weak lighting, longitudinal crack, with de-icing agent

Count: 4, SCI: 0.6671, Continuity: 1968.25

Fig. 17. Visualization of detection results on six representative real-world samples.

susceptibility to real-world noise and artifacts directly contributes to a high Failure rate (30.8%) in practical applications.

- **The Proposed Method (Balanced and Robust):** In contrast, our method achieves the best balance. It maintains the highest Success Rate (92.3%) and Continuity (3732.42 px). Unlike U-Net, it recovers complete crack structures; unlike SAM 3, it preserves fine geometric details (width 14.87 px) and handles complex topologies (SCI 0.70) effectively. This confirms that the proposed YOLOv8+HFR framework is best suited for precise and continuous pavement inspection.

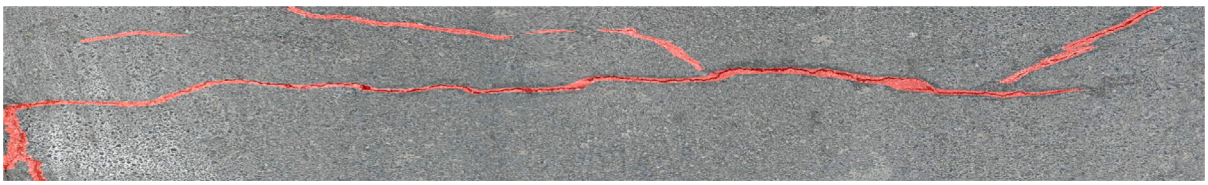


(a) Unet detection, Sample 3

Count: 3, SCI: 0.7787, Continuity: 2214.67

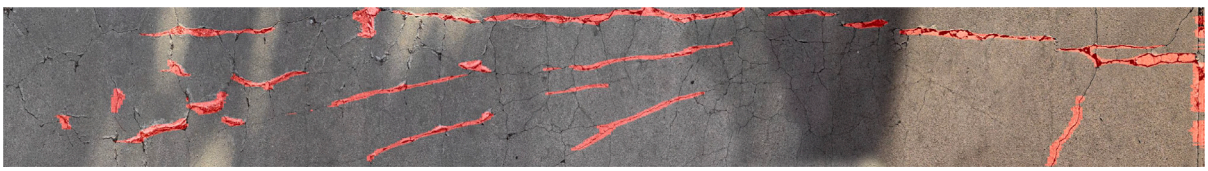
(b) Unet detection, Sample 5

Count: 4, SCI: 0.7573, Continuity: 1926.25



(c) SAM 3 detection, Sample 6

Count: 4, SCI: 0.7573, Continuity: 1926.25



(d) SAM 3 detection, Sample 9

Count: 24, SCI: 0.2178, Continuity: 439.46

(e) DeepCrack detection, Sample 11

Count: 3, SCI: 0.9866, Continuity: 3771.33

(f) CrackSegFormer detection, Sample 7

Count: 17, SCI: 0.5663, Continuity: 568.12

Fig. 18. Visualized results of SOTA methods for real-world crack detection.

5.3. Ablation study

5.3.1. Benchmark-oriented ablation

To rigorously validate the contribution of each component within the proposed YOLOv8+HFR framework, we conducted a comprehensive ablation study. The core motivation of the method pipeline is to address the inherent trade-off between semantic localization and pixel-level boundary precision in crack detection. By isolating the effects of the YOLOv8 baseline, the standalone HFR algorithm, and specific morphological operations including morphological dilation and erosion, we analyzed how each module influences the geometric metrics (IoU, Precision, Recall, F1-Score) and the topological quality of the detected cracks. The quantitative results of the ablation study are summarized

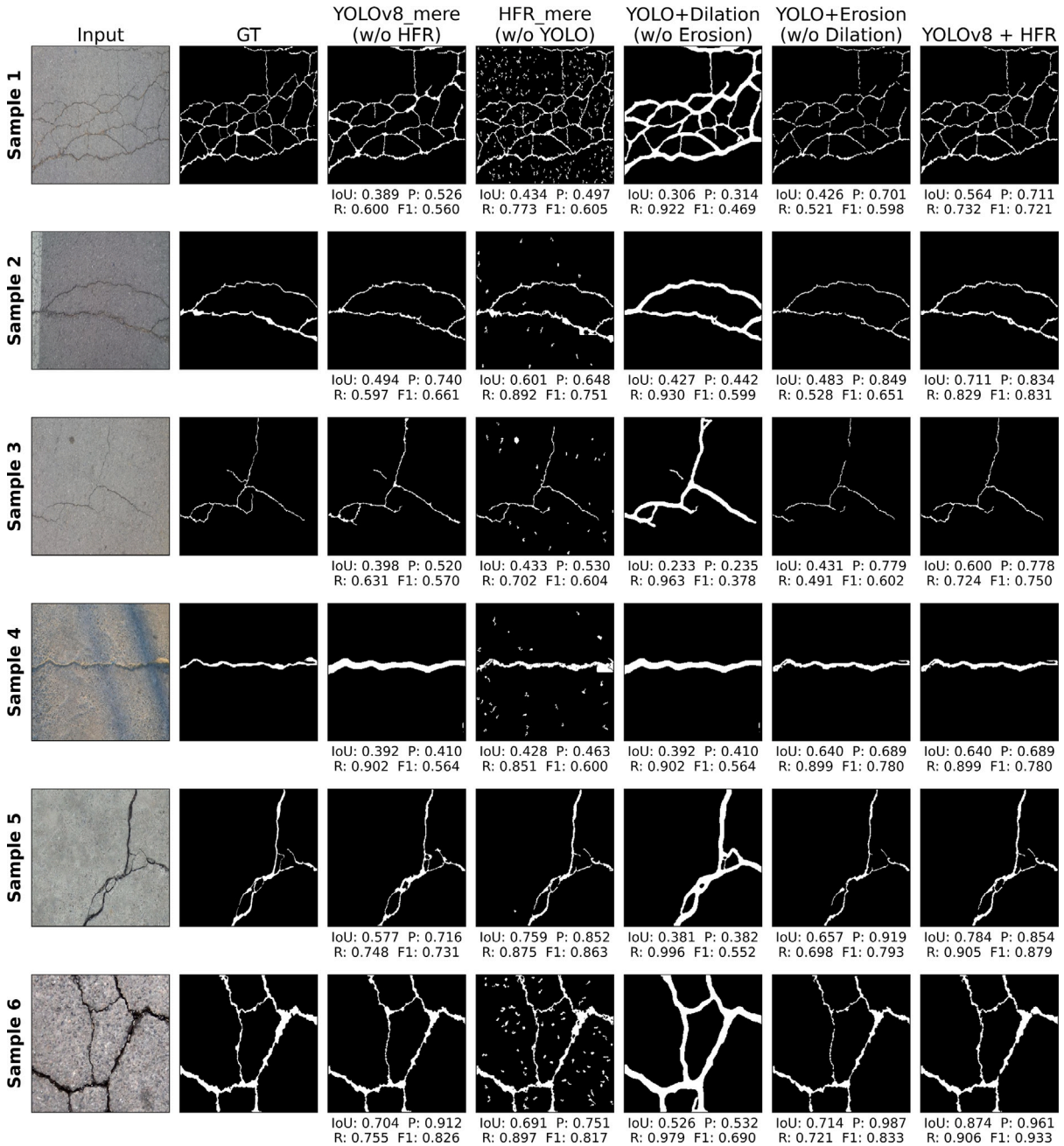


Fig. 19. Benchmark-oriented ablation study of the proposed pipeline.

Table 7
Ablation study on benchmark dataset.

Ablation variant	IoU	Precision	Recall	F1-score
YOLOv8 (Baseline)	0.4328	0.5592	0.6645	0.6073
HFR Only	0.4789	0.5810	0.7380	0.6502
YOLOv8 + Dilation	0.3802	0.3950	**0.9180**	0.5524
YOLOv8 + Erosion	0.4651	**0.7955**	0.5320	0.6376
YOLOv8 + HFR	**0.6128**	0.7489	0.7781	**0.7632**

in Table 7, and the corresponding qualitative visualizations on six representative samples are presented in Fig. 19.

The baseline model, denoted as "YOLOv8-mere", employs the standard YOLOv8-seg model trained in the study to process high-resolution pavement images. Quantitatively, the baseline achieves an F1-Score of 0.6073 and an IoU of 0.4328 (Table 7). The Recall (0.6645) and Precision (0.5592) are acceptable. As shown in the third column of Fig. 19, although the baseline results appear visually reasonable, they still suffer from local omissions, as well as noticeable boundary ambiguity. For instance, in the representative case of alligator cracking (Sample 1), while the main crack structure is roughly localized, the internal mesh topology is only partially recovered, with many fine interconnections missing. This indicates that the raw outputs of the CNN decoder are insufficient for capturing the high-frequency spatial details required for engineering-grade crack inspection. For relatively wide cracks (Sample 4), the model tends to over-recall the crack region, which leads to imprecise boundary localization. Overall, the YOLOv8-seg model achieves a reasonable balance between precision and recall, providing

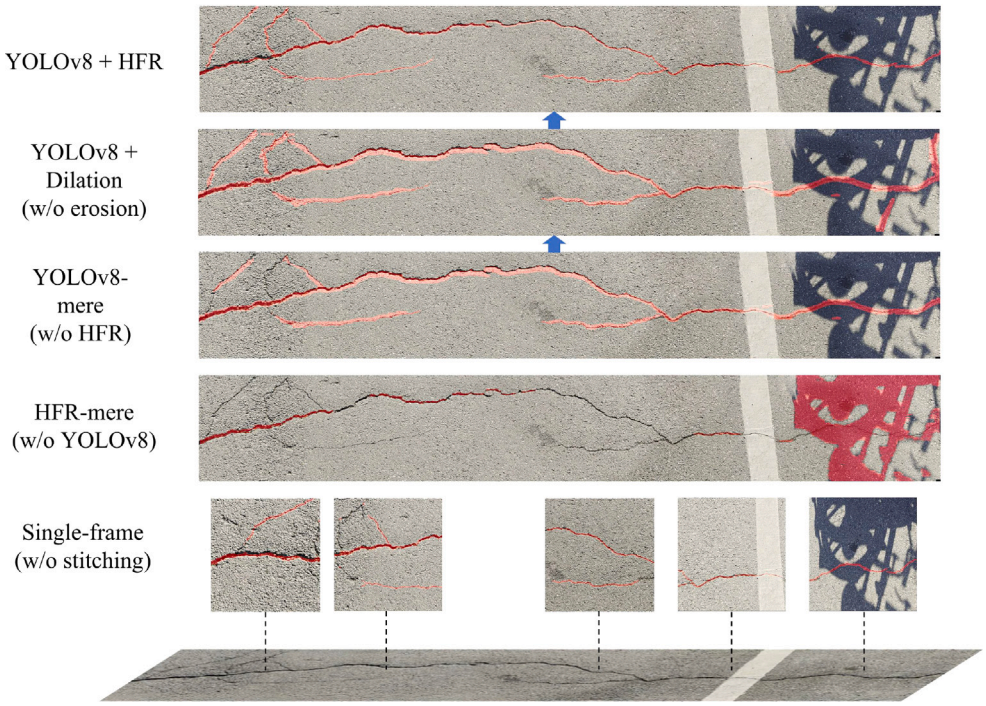


Fig. 20. Visual ablation comparison in a real-world deployment scenario.

a stable lower-bound performance for crack detection. Although it does not exhibit particularly high precision or recall compared with specialized models, its robust predictions and high computational efficiency make it well suited as a baseline model.

To investigate the contribution of the proposed refinement pipeline within the proposed framework, we evaluated the "HFR-mere" variant, which relies solely on the high-fidelity reconstruction (HFR) module without the semantic guidance of YOLOv8. As reported in Table 7, the HFR-only approach exhibits a characteristically high Recall of 0.7380, indicating strong sensitivity to texture variations. However, the advantage comes at the expense of Precision, which drops to 0.5810. Qualitative results in Fig. 19 reveal that HFR behaves similarly to a high-pass filter, amplifying not only crack structures but also background noise, surface roughness, and non-crack anomalies. In Samples 2 and 4, the output is dominated by scattered noise responses, making it difficult to distinguish true structural distress from environmental interference. The observation confirms that, without semantic constraints, purely morphological methods lack sufficient contextual awareness for robust crack discrimination.

To further analyze the role of individual morphological operations, we isolated the effects of dilation and erosion when applied to YOLOv8 probability maps. The YOLOv8 + Dilation variant aggressively expands detected regions to improve connectivity between fragmented segments, achieving the highest Recall of 0.9180 among all configurations. However, this strategy severely degrades Precision to 0.3950 and results in the lowest IoU of 0.3802. Visual inspection shows significant crack thickening and frequent merging of adjacent structures, particularly in alligator cracking scenarios, which leads to substantial loss of topological fidelity. In contrast, the YOLOv8 + Erosion variant focuses on suppressing noise by shrinking detected regions. While this approach yields the highest Precision of 0.7955, it substantially compromises continuity, with Recall dropping to 0.5320. Fine crack branches are frequently eliminated by the erosion operation (e.g. Sample 3), causing the detection results to degenerate into sparse and disconnected fragments. The phenomenon is partly attributed to minor boundary inaccuracies and the fragmentation of subtle branches in the YOLOv8 predictions, which are amplified by the dilation process. As a result, erosion-only strategies are inherently inadequate for preserving thin, continuous crack geometries.

The proposed YOLOv8 + HFR framework effectively integrates the complementary strengths of these components through a "locate-then-refine" strategy. YOLOv8 first establishes a reliable semantic region of interest, within which the HFR module performs constrained morphological reconstruction. Quantitatively, the synergy leads to the best overall performance, achieving the highest F1-Score of 0.7632 and IoU of 0.6128, with Recall and Precision simultaneously maintained at 0.7781 and 0.7489, respectively. Compared to the baseline, the IoU improves by nearly 18%, while the F1-Score surpasses the high-recall dilation variant by over 22%. Qualitatively, the proposed method demonstrates superior morphological fidelity: it suppresses background noise that dominates HFR-only results, preserves fine crack branches that are lost under erosion, and avoids the excessive thickening induced by dilation. Across multiple challenging samples, crack skeletons remain continuous while boundaries are accurately aligned with true physical edges. The results collectively validate that the proposed framework achieves a well-balanced trade-off between noise suppression, continuity preservation, and geometric accuracy, which is critical for engineering-grade pavement crack inspection.

5.3.2. *Deployment-oriented ablation study*

To validate the practical efficacy of the proposed pipeline under realistic engineering constraints, we conducted a deployment-oriented ablation study on the 13 diverse pavement scenarios described in Section 5.2. Unlike pixel-level comparisons against ground truth under controlled conditions, the analysis focuses on the transition from discrete frame-level detection to continuous topological reconstruction in real-world deployment, with particular attention to the impact of ablating key algorithmic components such as the stitching module, YOLOv8, and the two-step HFR processing on the final detection performance. The quantitative results reported in Table 8 represent the metrics

Table 8
Quantitative performance comparison of ablation variants averaged over the 13 validation samples.

Method	Logical count (count)	Total length (px)	SCI (–)	Avg. width (px)	Branches (count)	Continuity (score)	FPS (Hz)
YOLOv8 + HFR	4.08	11,647	0.695	14.87	685	3732.42	1.09
YOLOv8 + Erosion	5.77	10,907	0.560	14.85	647	2047.84	1.75
YOLOv8 + Dilation	3.85	11,589	0.730	21.35	283	5027.05	2.36
YOLOv8-mere	4.46	9272	0.640	19.24	239	2884.73	10.60

averaged over all 13 validation samples, and Table 9 summarizes the mean incremental changes in crack length and width computed from the results in Table 8. Due to extensive overlap in consecutive frames and abundant background noise in HFR-mere, the quantitative results of these two ablations are limited; thus, the tables focus on the two key steps of the HFR module to highlight their contributions. For intuitive illustration, Fig. 20 visualizes the step-by-step ablation process of the stitching and detection pipeline on a representative sample featuring a long-span longitudinal crack with dense branching and strong shadow interference.

Impact of continuous stitching versus single-frame processing. The bottom row of Fig. 20 highlights the fundamental limitation of traditional single-frame detection in deployment scenarios. When processed independently, long-span cracks are physically truncated by image boundaries, resulting in disconnected segments that lack global topological context. As illustrated in the magnified regions, faint crack traces entering low-contrast areas are frequently missed without the support of spatial continuity. By introducing the stitching module, the crack is reconstructed within a unified spatial domain as a continuous entity. Such long-range continuity is fundamentally unattainable under discrete frame processing, confirming stitching as a prerequisite for subsequent high-fidelity reconstruction.

Role of semantic guidance from YOLOv8. The necessity of deep semantic guidance is demonstrated by comparing the HFR-mere variant with the YOLOv8-mere baseline. Without semantic constraints, the HFR-mere approach exhibits strong sensitivity to texture but lacks contextual discrimination. As shown in Fig. 20, dark shadows and pavement markings are frequently misidentified as cracks. In contrast, the YOLOv8-mere model effectively suppresses these non-crack artifacts and focuses on genuine distress regions. However, quantitative results in Table 8 reveal that the raw YOLOv8 output still suffers from fragmentation, achieving an average Continuity score of 2884.73 with a Logical Count of 4.46. The limitation is also evident in the example of YOLOv8-mere detection, where the detected cracks exhibit excessively coarse edges, inflated widths, and a local absence of numerous fine branches, demonstrating that YOLOv8 alone is still insufficient for high-fidelity crack reconstruction.

"Extend-then-prune" mechanism in HFR. The core contribution of the proposed framework lies in the sequential application of morphological dilation and erosion. The dilation stage has aggressively recovered fragmented segments, achieving the lowest Logical Count and highest Continuity score, but at the cost of boundary inflation, as visualized in Fig. 20. Subsequent erosion corrects the geometric bias by thinning the structure while preserving the newly established connections. As detailed in Table 9, the erosion stage removes approximately 30.35% of the excessive width introduced by dilation, yielding crack geometries that are both topologically coherent and physically realistic. The combination of the two stages enables the HFR pipeline to achieve an overall increase of 11.02% in crack length and a 22.71% reduction in edge width relative to the YOLOv8 baseline, producing crack geometries that are both topologically coherent and physically realistic.

Topological complexity and branch recovery. An important observation from Table 8 concerns the Branches metric. While dilation improves connectivity, it also tends to merge fine parallel branches into coarse blobs, resulting in a slight increase in the total branch count. The overall HFR framework increases the average branch count from 239 to 685, demonstrating that the erosion step not only refines boundaries but also excavates internal crack structures, recovering secondary branches previously buried within coarse masks. The resulting tree-like morphology confirms the method's superiority in preserving the intrinsic fractal characteristics of pavement distress.

Computational cost and deployment feasibility. It is worth noting that the YOLOv8-mere processing speed recorded in this deployment study (10.60 Hz) is lower than the 15.5 FPS reported in Section 5.1. The discrepancy arises from the different evaluation conditions. While Section 5.1 measures the pure inference speed on standard discrete images, this deployment-oriented evaluation processes the large-scale, continuous pavement scenarios, which inherently introduces additional memory handling and image tiling operations.

Furthermore, the introduction of HFR reduces this processing speed from 10.60 Hz to 1.09 Hz due to CPU-bound skeletonization and sliding-window analysis. Nevertheless, a processing rate of approximately 1 Hz remains acceptable for offline inspection or slow-moving survey platforms. The computational trade-off yields a 29.4% increase in Continuity and nearly a threefold improvement in branch recovery, justifying the overhead for high-precision engineering assessment.

5.4. Discussion

5.4.1. Advantages

The proposed dual-channel framework for pavement texture reconstruction and crack detection demonstrates both real-time capability and high-fidelity performance, with strong applicability to longitudinal cracks and practical potential for on-site deployment in modern pavement maintenance systems.

Balanced Precision-Recall optimization and morphological fidelity. A primary advantage of the framework is its superior balance between Precision and Recall. As demonstrated in the comparative experiments with SOTA models (Section 5.1), the framework maintains a high precision level superior to both classic semantic segmentation models (e.g., U-Net) and modern Transformer baselines (e.g., CrackSegFormer), while simultaneously delivering a notable improvement in recall. The performance gain stems from the strategic decoupling of the detection and refinement objectives. YOLOv8-seg is utilized as the baseline for its environmental robustness; however, to mitigate the omission of subtle cracks, our pipeline incorporates an "extend-then-prune" optimization strategy to fully capture fine-grained fragments within candidate regions. As validated in the ablation study (Section 5.3), the HFR module effectively compensates for the limitations of the baseline in resolving complex edge details via micro-scale sliding-window refinement. The design allows the system to retain the macro-scale environmental robustness of YOLOv8 while executing micro-scale refinement, effectively resolving issues of edge coarseness and omission

Table 9
Statistical analysis of morphological evolution during the HFR process.

Process stage	Length change (px)	Proportion	Width change (px)	Proportion
Dilation phase	+1,344	+12.81%	2.11	+10.97%
Erosion phase	−188	−1.59%	−6.48	−30.35%
Overall net change	**+1,156**	**+11.02%**	**−4.37**	**-22.71%**

in low-contrast areas. Consequently, the dual-channel framework ensures that the final segmentation is not only topologically complete but also morphologically accurate.

Robust generalization on continuous topologies. Another key advantage of the framework is its robust applicability for practical inspection and reconstruction of longitudinal cracks. Pavement texture reconstruction serves as a crucial prior step in the dual-channel framework, where the feature-based approach eliminates the need for high-precision positioning modules and ensures adaptability in GPS-denied environments. It offers essential support for accurate, in-situ identification of common longitudinal cracks, representing a substantial advancement from conventional fixed-point crack detection. Beyond its theoretical adaptability, the system's algorithmic resilience is empirically substantiated by the extensive generalization assessment conducted on 13 complex real-world datasets in Section 5.2. Unlike traditional thresholding methods sensitive to environmental noise, the proposed framework demonstrates exceptional robustness across varying illumination and surface conditions, achieving an overall success rate of 92.3%, significantly outperforming the SOTA baselines evaluated under identical conditions. The adaptive weighting mechanism (λ_1, λ_2) allows the model to generalize well to extensive crack networks where crack branches are tortuous but physically linked. By dynamically modulating the erosion and dilation criteria based on local structure tensors, the system effectively suppresses aggregate noise while preserving the geometric continuity of complex, net-like crack structures. This adaptability ensures consistent segmentation performance even when transitioning between asphalt surfaces with varying aggregate roughness, validating its generalization potential for continuous pavement distress patterns.

Moreover, distinctive engineering advantage of the proposed framework is its high degree of specialization for high-fidelity pavement surface reconstruction. While SLAM or Visual Odometry (VO) systems are highly effective for camera trajectory estimation and robotic localization, they are not primarily optimized for the continuous and dense reconstruction of surface textures. Standard SLAM frameworks typically prioritize pose consistency over the morphological fidelity of the environment, resulting in sparse or low-resolution maps that are insufficient for pixel-level crack quantification. In contrast, our approach leverages homography-based constraints specifically tailored to the planar nature of road surfaces to generate stable, high-resolution texture strip maps. The specialization ensures that the reconstructed "digital strip map" maintains superior metric accuracy and morphological integrity, bridging the gap between basic visual localization and the rigorous demands of infrastructure distress assessment in GNSS-denied environments.

Computational efficiency and deployment strategy. a critical consideration for real-world deployment is the balance between detection latency and segmentation precision. The lightweight, vision-based acquisition module mounted on the mobile inspection platform enables continuous on-site monitoring. As detailed in the ablation study (Table 8), while the integration of the HFR module introduces computational overhead and reduces FPS to 1.09 Hz, the YOLOv8-mere baseline maintains a rapid inference rate of 10.60 Hz. The performance demonstrates the deployment feasibility of the proposed framework under the intended inspection scenario in Fig. 1. To address the trade-off in practical engineering, we utilize a decoupled edge-cloud deployment strategy. In practice, the lightweight YOLOv8-seg backbone is deployed on the mobile edge device, specified as the detection platform, to perform real-time coarse detection with FPS $>$ 10 Hz. It ensures immediate feedback to the operator regarding the presence and approximate location of cracks, facilitating timely on-site marking. Simultaneously, the raw high-resolution imagery and bounding box coordinates are streamed to a backend server, where the computationally intensive HFR refinement is executed asynchronously. The distributed approach effectively leverages the real-time responsiveness of the detection model while reserving the high-fidelity morphological reconstruction for centralized processing, thereby validating the system's engineering applicability for rigorous pavement maintenance management.

Interpretation of continuity metrics. The two continuity-related indicators, Continuity and SCI, are specifically defined to characterize the structural coherence of reconstructed longitudinal cracks. The effectiveness of the proposed method has been systematically validated through both comparative experiments and ablation studies, based on the two indicators. Nevertheless, these metrics are intended as scenario-specific, relative measures rather than absolute physical standards. Their main role is to provide a consistent basis for comparing continuity preservation and relative performance trends.

While maximizing continuity is not universally optimal – particularly for naturally discontinuous fractures – it serves as the primary determinant of reconstruction success in the present study cases, given the specific focus on recovering long-span, continuous longitudinal cracks. Empirical verification confirms that continuity values below 1000 pixels typically indicate significant segmentation failures or extensive fragmentation. It should be noted that neither Continuity nor the auxiliary SCI metric functions as a standalone criterion. To preclude regional omissions, these shape-based descriptors must be cross-referenced with Total Length and a predefined threshold for logical crack count. Consequently, a multi-dimensional assessment combining these metrics is essential to accurately classify the detection state.

To sum up, the proposed framework's compatibility with modern digital infrastructure makes it a promising component of future intelligent pavement management systems. By providing high-fidelity pavement condition data, it enables the transition from reactive to predictive maintenance strategies, which can lower lifecycle costs, minimize service interruptions, and improve overall pavement safety and longevity. Its modular design also supports integration with cloud-based platforms, GIS databases, or digital twin frameworks, enhancing its adaptability to various deployment scenarios.

5.4.2. Limitations

Despite the promising results, several limitations must be considered when evaluating the system's applicability in practical pavement condition assessment.

Operational dependencies and scanning trajectory. The deployment is constrained by the acquisition stability and the scanning mode. High-resolution stitching imposes strict requirements on image clarity. Engineering trials indicate that motion blur becomes a detrimental factor when the acquisition speed exceeds 0.8 m/s, causing feature mismatching. Thus, the current system is strictly limited to low-speed

operation ($0.2 \sim 0.5$ m/s) to guarantee the pixel-level fidelity required for micro-crack detection.

While vehicle-mounted road inspection systems typically operate at speeds ranging from 30 km/h to 50 km/h to prioritize macroscopic, network-level coverage, such high-speed acquisition relies heavily on extremely expensive hardware (e.g., line-scan cameras and high-precision IMUs) and often compromises the capture of fine structural details. In contrast, the proposed method is specifically positioned within the paradigm of *low-speed, high-precision inspection* utilizing cost-effective visual sensors. Given the frame rate limitations of standard cameras and the high overlap ratio required for seamless texture stitching, sacrificing acquisition speed becomes a necessary engineering trade-off, which ensures the micro-scale morphological fidelity necessary for precise crack quantification and localized maintenance planning. Consequently, the system is particularly suitable for unmanned ground vehicles or autonomous robotic platforms operating in targeted road segments, pedestrian infrastructures, or enclosed environments (e.g., tunnels) where high-speed vehicles are restricted.

Furthermore, while the edge-cloud strategy effectively addresses onboard computational limits, it inherently relies on stable data transmission bandwidth. In practical engineering, the uplink latency for transmitting uncompressed high-fidelity data becomes a bottleneck. The latency is particularly critical in remote inspection areas with unstable network coverage, where data buffering may restrict the real-time synchronization of high-precision results despite the efficient edge-side processing.

In addition, unlike Ground Penetrating Radar (GPR), the optical system lacks penetrative capabilities and is sensitive to surface occlusions. Physical obstructions such as accumulated debris, standing water, or road obstacles can block the visual line of sight, inevitably leading to segmentation discontinuities in uncleaned areas.

Moreover, the current deployment configuration utilizes a single-pass trajectory along the pavement extension direction, optimizing for longitudinal through-cracks with potential branches. It does not currently account for transverse reciprocating scanning with lateral overlap. Consequently, while the system effectively characterizes longitudinal defects, its field of view and reconstruction logic exhibit limitations when inspecting wide-span transverse cracks or disconnected defects that extend beyond the single-pass capture width.

Generalization constraints on discontinuous and volumetric defects.

While the method demonstrates robust performance on connected topologies, it exhibits specific limitations when handling spatially disjointed or volumetric defects due to its inherent algorithmic dependencies. The performance of the HFR module is coupled with the YOLOv8 baseline, as the refinement process operates on the premise of topological connectivity and relies on the identified skeleton as a growing seed. Consequently, for highly fragmented cracks, especially isolated and fine-grained alligator cracking segments that are not physically connected to the main body, the HFR module lacks a valid propagation path if the baseline model fails to capture the initial seed. The dependency renders the method less effective for detecting discrete crack clusters that lack a connective structural backbone, as evidenced by the omission case in Sample 13.

Beyond the topological constraints, the system is also susceptible to interference from volumetric surface defects due to its reliance on grayscale intensity, particularly under strong and complex illumination. Volumetric depressions, such as potholes or shallow ruts, often mimic the low-intensity profile of cracks yet lack linear continuity, occasionally triggering false positives in our tests. Furthermore, under high-contrast lighting, the adaptive thresholds, while robust to general pavement texture, struggle to recover the boundaries of wide and shallow cracks from regions exhibiting sharp shadow gradients. The limitation leads to performance degradation in complex environments, as evidenced by the crack omission in Sample 13 and the interference caused by potholes in Sample 4.

Synthesizing the discussed algorithmic and operational constraints, the proposed framework is optimally defined for the high-precision detection of longitudinal through-cracks that align with the inspection vehicle's trajectory, along with their associated fine branches. Within the defined operational envelope, it demonstrates exceptional stability and precision across varying illumination and weather conditions, offering a robust solution for automated pavement maintenance.

6. Conclusion and future works

The study presents an integrated framework for pavement texture reconstruction and high-fidelity crack segmentation, demonstrating significant advancements in both accuracy and efficiency. By integrating high-resolution imagery, robust image stitching techniques, a YOLOv8-based crack detection model, and the proposed hybrid pipeline for precision augmentation, the system has shown promising results in reconstructing pavement textures and identifying longitudinal cracks with high precision. Particularly, the quantitative and qualitative results confirm the effectiveness of the dual-channel architecture, especially in achieving high-fidelity, continuous modeling of longitudinal pavement cracks. By combining feature-based stitching, adaptive alignment, and time-aware integration, the framework produces high-resolution, coherent textures that serve as a reliable foundation for subsequent crack reconstruction across sequential frames. These results are further refined through a hybrid framework designed for pixel-level accuracy in crack boundary delineation. In general, the approach effectively addresses key challenges in automated pavement condition inspection, offering a compelling solution for intelligent, data-driven maintenance.

Besides the achievements, several operational constraints still impact the system's real-world scalability, including motion blur induced by vehicle speed and visual occlusions caused by surface obstacles or adverse illumination. The reliance on high-resolution imagery introduces computational and storage challenges, particularly for large-scale applications. Additionally, the sensitivity of the image stitching process to environmental factors, such as vibration and camera misalignment, may lead to errors in the texture reconstruction. The crack segmentation pipeline, while significantly improved, still faces challenges in identifying crack branches within partially occluded or high-contrast regions, particularly for fragmented and spatially discontinuous cracks. The factors may affect the stability and timeliness of high-fidelity reconstruction in real-world inspection scenarios.

Future work will focus on mitigating the identified deployment constraints in order to enhance robustness and scalability in real-world applications. To bridge the speed gap for broader engineering applications, future iterations will incorporate advanced hardware with shorter exposure times and motion-aware deblurring algorithms to elevate the speed ceiling without sacrificing reconstruction fidelity. To mitigate the impact of surface occlusions and adverse illumination, future extensions will explore multi-view acquisition, temporal redundancy across consecutive frames, as well as complementary sensing modalities to improve visual continuity under challenging conditions.

From a system perspective, the computational and transmission burden induced by continuous high-resolution data streaming is planned to be addressed by investigating adaptive compression schemes, selective region-of-interest transmission, and more efficient edge-cloud collaboration strategies. In addition, lightweight refinement alternatives and incremental reconstruction mechanisms are expected to be explored to reduce reliance on full-resolution data while preserving reconstruction fidelity. The efforts will further improve the practicality of the framework for large-scale and long-duration pavement inspection tasks.

CRediT authorship contribution statement

Peng An: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zhengru Ren:** Supervision, Resources, Project administration, Conceptualization. **Long-Xiang Liu:** Investigation, Data curation. **Jia-Rui Lin:** Supervision. **Yantao Yu:** Supervision. **Yu-Tao Guo:** Supervision. **Chao Hou:** Supervision. **Zhen-Zhong Hu:** Supervision, Funding acquisition.

Funding

This work was supported by the National Key Research and Development Program of China [grant number 2022YFC3801100]; and Shenzhen Science and Technology Program [grant number SGDX 20240115110503006].

Declaration of competing interest

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

Data availability

Data will be made available on request.

References

[1] S. Bhat, S. Naik, M. Gaonkar, P. Sawant, S. Aswale, P.R. Shetgaonkar, A survey on road crack detection techniques, in: 2020 International Conference on Emerging Trends in Information Technology and Engineering, Ic-ETITE, 2020, pp. 1–6.

[2] N. Kheradmandi, V. Mehranfar, A critical review and comparative study on image segmentation-based techniques for pavement crack detection, Constr. Build. Mater. 321 (2022) 126162.

[3] D. Ma, H. Fang, N. Wang, C. Zhang, J. Dong, H. Hu, Automatic detection and counting system for pavement cracks based on PCGAN and YOLO-MF, IEEE Trans. Intell. Transp. Syst. 23 (11) (2022) 22166–22178.

[4] Q. Qiu, D. Lau, Real-time detection of cracks in tiled sidewalks using YOLO-based method applied to unmanned aerial vehicle (UAV) images, Autom. Constr. 147 (2023) 104745.

[5] L.S. Furtado, I.S. Bessa, N.d. Gurjão, J.B. Soares, Integrating smart city technologies for sustainable pavement infrastructure, Can. J. Civ. Eng. 52 (5) (2025) 569–582.

[6] Y.-Q. Xiao, S.-W. Li, Z.-Z. Hu, Automatically generating a MEP logic chain from building information models with identification rules, Appl. Sci. 9 (11) (2019) 2204.

[7] C. Lin, Z.-Z. Hu, C. Yang, Y.-C. Deng, W. Zheng, J.-R. Lin, Maturity assessment of intelligent construction management, Buildings 12 (10) (2022) 1742.

[8] C. Xing, G. Zheng, Y. Zhang, H. Deng, M. Li, L. Zhang, Y. Tan, A lightweight detection method of pavement potholes based on binocular stereo vision and deep learning, Constr. Build. Mater. 436 (2024) 136733.

[9] K. Ge, Y. Guo, C. Wang, Z.-Z. Hu, AI-based prediction of seismic time-history responses of RC frame structures considering varied structural parameters, J. Build. Eng. 106 (2025) 112643.

[10] A. Alrajhi, K. Roy, L. Qingge, J. Kribs, Detection of road condition defects using multiple sensors and IoT technology: A review, IEEE Open J. Intell. Transp. Syst. 4 (2023) 372–392.

[11] G. Yang, K.C.P. Wang, J.Q. Li, Y. Fei, Y. Liu, K.C. Mahboub, A.A. Zhang, Automatic pavement type recognition for image-based pavement condition survey using convolutional neural network, J. Comput. Civ. Eng. 35 (1) (2021) 04020060.

[12] Q. Mei, M. Gül, A cost effective solution for pavement crack inspection using cameras and deep neural networks, Constr. Build. Mater. 256 (2020) 119397.

[13] N. Snavely, S.M. Seitz, R. Szeliski, Modeling the world from internet photo collections, Int. J. Comput. Vis. 80 (2) (2008) 189–210.

[14] C. Griwodz, S. Gasparini, L. Calvet, P. Gurdjos, F. Castan, B. Maujean, G. De Lillo, Y. Lanthony, AliceVision meshroom: An open-source 3D reconstruction pipeline, in: Proceedings of the 12th ACM Multimedia Systems Conference, Association for Computing Machinery, 2021, pp. 241–247.

[15] Agisoft LLC, Agisoft Metashape: professional photogrammetry software, 2024, https://www.agisoft.com/. (Accessed 14 July 2025).

[16] R. Bajaj, Z.A. Al-Sabbag, C.M. Yeum, S. Narasimhan, 3D dense reconstruction for structural defect quantification, ASCE OPEN: Multidiscip. J. Civ. Eng. 2 (1) (2024) 04024001.

[17] L. Nie, C. Lin, K. Liao, S. Liu, Y. Zhao, Deep rectangling for image stitching: A learning baseline, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 5730–5738.

[18] Y.-F. Liu, S. Cho, B.F. Spencer, J.-S. Fan, Concrete crack assessment using digital image processing and 3D scene reconstruction, J. Comput. Civ. Eng. 30 (1) (2016) 04014124.

[19] J. Shan, W. Jiang, Y. Huang, D. Yuan, Y. Liu, Unmanned aerial vehicle (UAV)-based pavement image stitching without occlusion, crack semantic segmentation, and quantification, IEEE Trans. Intell. Transp. Syst. 25 (11) (2024) 17038–17053.

[20] L. Inzerillo, G. Di Mino, R. Roberts, Image-based 3D reconstruction using traditional and UAV datasets for analysis of road pavement distress, Autom. Constr. 96 (2018) 457–469.

[21] N. Wang, J. Dong, H. Fang, B. Li, K. Zhai, D. Ma, Y. Shen, H. Hu, 3D reconstruction and segmentation system for pavement potholes based on improved structure-from-motion (SFM) and deep learning, Constr. Build. Mater. 398 (2023) 132499.

[22] H.-C. Dan, B. Lu, M. Li, Evaluation of asphalt pavement texture using multiview stereo reconstruction based on deep learning, Constr. Build. Mater. 412 (2024) 134837.

[23] Y. Zhang, C. Chen, Q. Wu, Q. Lu, S. Zhang, G. Zhang, Y. Yang, A kinect-based approach for 3D pavement surface reconstruction and cracking recognition, IEEE Trans. Intell. Transp. Syst. 19 (12) (2018) 3935–3946.

[24] S. Dong, S. Han, C. Wu, O. Xu, H. Kong, Asphalt pavement macrotexture reconstruction from monocular image based on deep convolutional neural network, Computer-Aided Civ. Infrastruct. Eng. 37 (2022) 1754–1768.

[25] T. Zhao, L. Yang, Y. Xie, M. Ding, M. Tomizuka, Y. Wei, RoadBEV: Road surface reconstruction in bird's eye view, IEEE Trans. Intell. Transp. Syst. 25 (11) (2024) 19088–19099.

[26] W. Wang, C. Su, G. Han, H. Zhang, A lightweight crack segmentation network based on knowledge distillation, J. Build. Eng. 76 (2023) 107200.

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

[28] S. Halder, K. Afsari, Robots in inspection and monitoring of buildings and infrastructure: A systematic review, Appl. Sci. 13 (4) (2023) 2304.

[29] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2014, pp. 580–587.

[30] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1137–1149.

[31] R. Popli, I. Kansal, J. Verma, V. Khullar, R. Kumar, A. Sharma, ROAD: Robotics-assisted onsite data collection and deep learning enabled robotic vision system for identification of cracks on diverse surfaces, Sustainability 15 (12) (2023) 9314.

[32] Y. Zhang, Y. Lu, Z. Huo, J. Li, Y. Sun, H. Huang, USSC-YOLO: Enhanced multi-scale road crack object detection algorithm for UAV image, Sensors 24 (17) (2024) 5586.

[33] M. Ren, Y. Li, T. Hussain, Y. Wu, J. Li, Pixel-level concrete crack quantification through super resolution reconstruction and multi-modality fusion, Adv. Eng. Informatics 69 (2026) 103807.

[34] Ultralytics, YOLOv8 documentation, 2023, https://docs.ultralytics.com.

[35] G. Zhu, S.-L. Shen, J. Yao, M. Wang, J. Zhuang, Z. Fan, Automatic lightweight networks for real-time road crack detection with DPSO, Adv. Eng. Informatics 68 (2025) 103610.

[36] J. Wu, Z. Deng, Improved YOLOv5 model based on the pyramid structure of dilated convolutions, in: 2023 International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR, 2023, pp. 98–102.

[37] J. Li, H. Li, J. Tu, Z. Liu, J. Yao, L. Li, Pavement crack detection based on local enhancement attention mechanism and deep semantic-guided multi-feature fusion, J. Electron. Imaging 33 (6) (2024) 063027.

[38] W. Hou, J. He, C. Cui, F. Zhong, X. Jiang, L. Lu, J. Zhang, C. Tu, Segmentation refinement of thin cracks with minimum strip cuts, Adv. Eng. Informatics 65 (2025) 103249.

[39] S. Wang, X. Chen, Q. Dong, Detection of asphalt pavement cracks based on vision transformer improved YOLO V5, J. Transp. Eng. PART B-PAVEMENTS 149 (2) (2023).

[40] F.-J. Du, S.-J. Jiao, Improvement of lightweight convolutional neural network model based on YOLO algorithm and its research in pavement defect detection, Sensors 22 (9) (2022) 3537.

[41] S. Mathavan, K. Kamal, M. Rahman, A review of three-dimensional imaging technologies for pavement distress detection and measurements, IEEE Trans. Intell. Transp. Syst. 16 (5) (2015) 2353–2362.

[42] A.A. of State Highway, T. Officials, Standard Practice for Quantifying Cracks in Asphalt Pavement Surfaces from Collected Images Utilizing Automated Methods (AASHTO PP 67-14), Tech. Rep., American Association of State Highway and Transportation Officials, 2014, (Accessed 12 August 2025).

[43] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, second ed., Cambridge University Press, 2003.

[44] M. Bertozzi, A. Broggi, GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection, IEEE Trans. Image Process. 7 (1) (1998) 62–81.

[45] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[46] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.
[47] P. Burt, E. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540.
[48] B. Triggs, P.F. McLauchlan, R.I. Hartley, A.W. Fitzgibbon, Bundle adjustment—a modern synthesis, in: International Workshop on Vision Algorithms, Springer, 1999, pp. 298–372.
[49] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, H. Ling, Feature pyramid and hierarchical boosting network for pavement crack detection, IEEE Trans. Intell. Transp. Syst. 21 (4) (2019) 1525–1535.
[50] Y. Liu, J. Yao, X. Lu, R. Xie, L. Li, DeepCrack: A deep hierarchical feature learning architecture for crack segmentation, Neurocomputing 338 (2019) 51–63.
[51] Y. Shi, L. Cui, Z. Qi, F. Meng, Z. Chen, Automatic road crack detection using random structured forests, IEEE Trans. Intell. Transp. Syst. 17 (12) (2016) 3434–3445.
[52] Ç.F. Özgenel, A.G. Sorguç, Performance comparison of pretrained convolutional neural networks for concrete crack detection, in: ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, vol. 35, IAARC Publications, 2018, pp. 1–8.
[53] M. Eisenbach, R. Sims, R. Stricker, D. Milz, K. Lein, N. Geissler, K. Debes, H.-M. Gross, How to get pavement distress detection ready for deep learning? A systematic approach, in: 2017 International Joint Conference on Neural Networks, IJCNN, IEEE, 2017, pp. 2039–2047.
[54] L. Nie, C. Lin, K. Liao, S. Liu, Y. Zhao, Unsupervised deep image stitching: Reconstructing stitched features to images, IEEE Trans. Image Process. 30 (2021) 6184–6197.
[55] S. Hausler, S. Garg, M. Xu, M. Milford, T. Fischer, Patch-netvlad: Multi-scale fusion of locally-global descriptors for place recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2021, pp. 14141–14152.