

Article

Not peer-reviewed version

---

# Improving the Reliability of UAV-Based Crack Inspection of Port Quay Walls Using Anomaly Detection

---

Masachika Akage , [Daisuke Yoshida](#) \* , [Wakana Fujimoto](#)

Posted Date: 27 April 2026

doi: 10.20944/preprints202604.1845.v1

Keywords: UAV; port quay wall inspection; crack inspection; anomaly detection; false-positive reduction; Vision Transformer; infrastructure maintenance; aerial imagery



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Improving the Reliability of UAV-Based Crack Inspection of Port Quay Walls Using Anomaly Detection

Masachika Akage <sup>1</sup>, Daisuke Yoshida <sup>1,\*</sup> and Wakana Fujimoto <sup>2</sup>

<sup>1</sup> Graduate School of Informatics, Osaka Metropolitan University, Osaka 558-8585, Japan

<sup>2</sup> College of Sustainable System Sciences, Osaka Metropolitan University, Osaka 599-8531, Japan

\* Correspondence: daisuke@omu.ac.jp

## Abstract

UAV-based crack inspection of port quay walls is promising for efficient infrastructure maintenance, but its practical deployment remains hindered by frequent false positives caused by debris, stains, and irregular surface textures. This study proposes a false-positive reduction framework for a crack inspection system based on aerial images acquired by a small general-purpose UAV. The proposed method introduces anomaly detection after object detection so that detected crack candidate regions are re-evaluated based on their deviation from the learned feature distribution of crack images. A Vision Transformer (ViT)-based anomaly detection model is employed, and both standard-threshold and low-threshold object detection settings are investigated. Experimental validation across five verification areas showed that the combination of standard-threshold object detection and anomaly detection consistently improved F1 and F2 scores over the conventional baseline, demonstrating stable suppression of false positives while maintaining crack detectability. Under the low-threshold setting, Frangi filter-based preprocessing was more effective than grayscale-based preprocessing, achieving a favorable balance between broader crack extraction and false-positive suppression in some 5 m cases. However, this advantage decreased as image resolution deteriorated. Overall, the results indicate that the most robust configuration in the current framework is the combination of standard-threshold object detection and anomaly-based false-positive suppression. In contrast, the benefit of low-threshold operation depends strongly on image resolution. The findings also suggest that practical deployment requires calibration of the anomaly-detection threshold based on site conditions and GSD.

**Keywords:** UAV; port quay wall inspection; crack inspection; anomaly detection; false-positive reduction; Vision Transformer; infrastructure maintenance; aerial imagery

---

## 1. Introduction

Social infrastructure, including bridges, roads, and port facilities, constitutes an indispensable foundation for economic activity in modern society. According to the Global Infrastructure Outlook published by the Global Infrastructure Hub [1], substantial infrastructure investment demand will persist worldwide, while current investment trends remain insufficient to fully meet this demand. This situation underscores the growing importance of the maintenance, management, and renewal of existing infrastructure.

Among infrastructure assets, port facilities are particularly exposed to severe environmental conditions, including continuous exposure to waves, seawater spray, and corrosion-related deterioration. As a result, the need for labor-saving and efficient inspection is even greater than for many other types of concrete infrastructure. In Japan, the national policy of infrastructure digital transformation (DX) has also encouraged the active use of digital technologies, such as UAVs and AI, for the inspection of port facilities. This policy direction is also reflected in the official inspection

guidelines for port facilities issued by the Ministry of Land, Infrastructure, Transport and Tourism, which explicitly encourage the use of digital technologies, such as UAVs and AI, in periodic inspection activities [2].

At the same time, the concrete surfaces of port facilities often exhibit highly complex, non-uniform appearances due to attached debris and other types of environmental noise. Conventional crack-detection approaches, particularly those that rely primarily on object-detection models, frequently generate false positives by mistaking noise for cracks. This lack of reliability is a major barrier to practical deployment. From an infrastructure safety perspective, missed crack detections must be minimized, whereas excessive false positives increase the workload of inspectors and diminish the benefits of automation. Therefore, a crack inspection system for port facilities must simultaneously achieve high detection coverage and reliable false-positive suppression.

This study aims to improve the reliability of UAV-based crack inspection of port quay walls by establishing a practical false-positive suppression framework in which anomaly detection is introduced as a post-processing step after object detection. The framework is designed to reduce false positives caused by debris and other external noise while preserving practical crack detectability under realistic inspection conditions. In addition, the study investigates whether a low-threshold object detection strategy combined with anomaly-based filtering can achieve both higher crack detectability and better false-positive suppression than conventional approaches. To clarify the effect of the proposed method itself, the evaluation is conducted at the single-image level, without incorporating orthophoto generation into the processing pipeline. The present study does not aim to propose a new backbone architecture for crack detection or anomaly detection. Rather, its contribution lies in establishing and validating a practical reliability-enhancement framework for UAV-based crack inspection of port quay walls under realistic geomatics and infrastructure-inspection conditions.

## 2. Related Work

Crack detection in concrete infrastructure is a key process in structural condition assessment and maintenance decision-making. In response to the growing demand for efficient inspection, numerous studies have explored automated crack detection using image processing, machine learning, and deep learning techniques, including methods based on UAV-acquired imagery. These approaches offer several advantages over conventional manual inspection, such as shorter inspection time, more quantitative evaluation, and reduced labor requirements [3,4].

Among these approaches, UAV-based crack inspection has attracted considerable attention because it enables efficient image acquisition over large or difficult-to-access structures [3–5]. Recent studies have reported promising results from combining UAV imagery with deep learning-based crack-detection models, including Transformer-based approaches for crack inspection of concrete infrastructure and water-related concrete structures [3,6,7]. Some of these studies have further incorporated photogrammetric products such as orthophotos and spatial coordinate systems, allowing detected cracks to be visualized and managed in a georeferenced framework [6]. Nevertheless, important limitations remain in practical inspection environments. Many existing studies, particularly those targeting high-precision inspection of port-related concrete infrastructure, still depend on high-end equipment or relatively well-controlled imaging conditions [8].

In contrast, port facilities are particularly challenging environments because concrete surfaces are often affected by attached debris, stains, uneven textures, and other environmental disturbances. Under such conditions, object detection-based methods frequently produce false positives by mistaking visually similar patterns for actual cracks. This problem is a major obstacle to real-world implementation, since excessive false detections increase the burden of manual verification and reduce confidence in automated inspection results [4]. Recent studies have also emphasized the importance of lightweight, practically deployable crack-detection models for complex inspection environments [9,10].

One intuitive way to reduce false positives is to introduce an additional image classification stage after object detection. However, such an approach is fundamentally constrained by the closed-set assumption, which requires all relevant classes to be predefined during training. In real port environments, non-crack patterns are highly diverse and cannot be exhaustively modeled in advance. Consequently, classification-based approaches may be less robust when confronted with previously unseen false-positive patterns [11].

In this context, anomaly detection provides a more appropriate framework for practical crack inspection. Rather than explicitly learning all non-crack categories, anomaly detection models characterize the feature distribution of crack images and identify samples that deviate from this distribution as anomalous [12–14]. This property makes anomaly detection particularly suitable for port facility inspection, where diverse and unknown sources of environmental noise are common [12,13]. Accordingly, combining object detection with anomaly detection is expected to improve inspection reliability by systematically suppressing false positives without requiring exhaustive annotation of all possible noise patterns [12]. Transformer-based approaches have recently shown strong performance in image-based crack detection and related visual inspection tasks, including semantic segmentation of concrete cracks and anomaly-aware structural image analysis [15,16].

Based on this perspective, the present study extends an existing UAV-based crack inspection framework for port quay walls and introduces anomaly detection to improve inspection reliability [4,12,17]. In addition, whereas previous studies have often emphasized georeferenced visualization through orthophoto generation and spatial projection, this study first examines fundamental discrimination performance at the single-image level, enabling the contribution of the proposed method to be evaluated independently of external factors associated with orthorectification and geometric transformation [4]. Accordingly, the present study should be understood as a domain-focused investigation of reliability enhancement for UAV-based port quay wall inspection, rather than as an attempt to establish universal performance across heterogeneous infrastructure classes.

Thus, the novelty of the present study lies not in introducing anomaly detection itself as an abstract concept, but in validating a practical false-positive suppression framework for UAV-based port quay wall inspection under visually challenging field conditions.

### 3. Methodology

This section describes the baseline crack-detection system based on aerial images acquired by a small, general-purpose UAV, as well as the false-positive reduction method proposed in this study. The baseline framework is derived from a previously developed crack detection system intended for practical use with commercially available small UAV platforms. In contrast, the proposed method introduces an anomaly-detection stage to improve reliability under visually complex conditions during port inspections. In this study, analysis is conducted at the single-image level rather than on orthophotos in order to exclude external factors, such as subtle geometric distortion and image-quality degradation caused by orthorectification, thereby enabling a more rigorous evaluation of the proposed algorithm. This design choice does not imply that the proposed framework is limited to non-georeferenced image analysis. Rather, the single-image setting was adopted in the present study to isolate the intrinsic discrimination capability of the false-positive reduction framework before integration with orthophoto-based and georeferenced inspection workflows.

#### 3.1. Baseline UAV-Based Crack Detection System

The baseline system adopted in this study is based on an existing crack-detection framework developed for practical implementation using small general-purpose UAVs. The original framework was designed to support cost-effective crack inspection with aerial images captured by commercially available UAV platforms, thereby reducing the economic and operational barriers to practical deployment in local governments and infrastructure maintenance agencies.

In the baseline workflow, aerial images are first acquired over the inspection site using a small UAV. The acquired images are then divided into smaller patches by applying overlap-based splitting

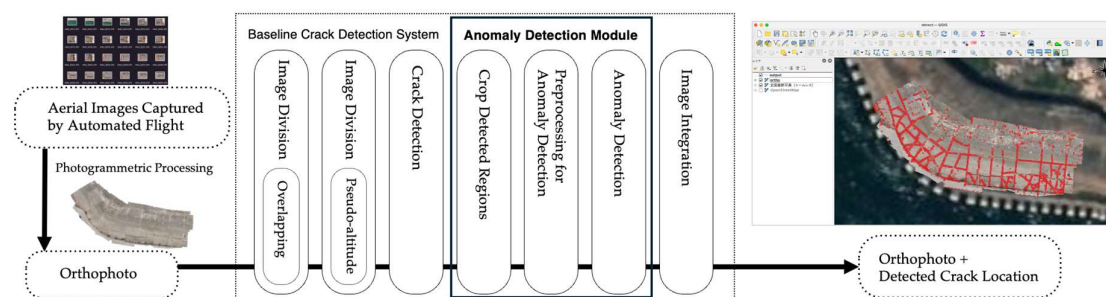
and pseudo-altitude splitting, both of which were introduced in the previous study. Crack candidate regions are subsequently detected from the segmented images using a deep learning-based object detection model, and the detected regions are represented by bounding boxes on the input images. Each bounding box retains both positional information and a confidence value indicating the likelihood that the region contains a crack.

Although the original system includes orthophoto generation, the present study intentionally omits orthorectification and instead performs all analyses on single aerial images. This design choice was made to eliminate external factors that may influence detection accuracy, such as subtle geometric distortion and image-quality changes introduced during orthophoto generation. As a result, the intrinsic discrimination capability of the proposed false-positive reduction algorithm can be evaluated more rigorously, particularly with respect to disturbances such as debris and surface noise commonly observed in port facilities.

The baseline detector used in this study is the pretrained YOLOR-based model inherited from the previous system [18]. In the object detection stage, two operational policies are considered: a standard threshold setting as in the previous study and a lower threshold setting intended to extract crack candidates more comprehensively. The former tends to provide stable detection for clear cracks, whereas the latter is designed to reduce missed detections at the cost of increased false positives. The proposed method addresses this trade-off by filtering the detected candidates in a subsequent anomaly-detection stage.

### 3.2. Proposed False-Positive Reduction Method

Figure 1 illustrates the workflow of the proposed method. The proposed framework introduces an anomaly-detection stage after the baseline crack-detection process in order to determine whether each detected region corresponds to a crack. More specifically, the regions enclosed by the detected bounding boxes are extracted from the original image, preprocessed, and then passed to an anomaly-detection module. Through this two-stage design, the proposed method aims to distinguish true cracks from visually similar disturbances, including debris and other environmental noise, that are difficult to reject using object detection alone.



**Figure 1.** Workflow of the proposed false-positive reduction framework. The black-framed block indicates the anomaly-detection module newly introduced in the proposed method, while the baseline crack detection workflow is adapted from [4].

A key concept of the proposed framework is that crack candidates can first be broadly collected during the object detection stage, even under a relatively low detection threshold, and then systematically screened in the second stage. This design is intended to improve inspection reliability by balancing high crack detectability with effective false-positive suppression in practical port inspection environments.

#### 3.2.1. Region Extraction from Detected Bounding Boxes

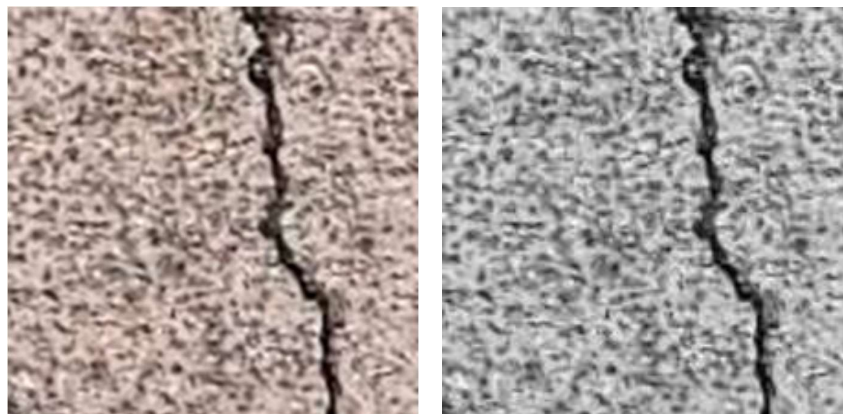
In the proposed method, each region detected by the baseline object detector is extracted from the original aerial image using the bounding box coordinates. These cropped regions are treated as crack candidates and serve as input units for the subsequent anomaly-detection stage. Specifically, the bounding-box regions were cropped from the original aerial images using a script-based procedure based on the label files generated in the preceding crack-detection stage. The resulting cropped regions were then preprocessed according to the experimental condition and forwarded to the subsequent ViT-based anomaly detection module. The purpose of this step is to isolate local regions that have already been identified as potentially crack-related and to re-evaluate them in greater detail using a different decision framework.

This region-based processing enables the proposed method to function as a second-stage verifier. While the object detector is responsible for efficiently locating possible crack regions in large aerial images, the anomaly-detection module focuses on whether the local visual characteristics of each cropped region are consistent with those of actual cracks. As a result, regions corresponding to debris, stains, or irregular concrete textures can be selectively removed from the final detection results.

### 3.2.2. Preprocessing for Anomaly Detection

Before anomaly detection, the cropped candidate regions are further processed to emphasize crack-relevant features while suppressing background variations and irrelevant appearance cues. In this study, preprocessing is not merely normalization, but an important step for enhancing the visual properties that help distinguish cracks from non-crack disturbances in port environments.

One preprocessing strategy adopted in this study is grayscale transformation. The cropped candidate regions are converted into grayscale using the weighted averaging scheme defined in ITU-R BT.601 [19]. The purpose of this transformation is to eliminate the influence of color information, which can vary substantially across sites, and to allow the anomaly detection model to focus on geometric and luminance-based features of cracks. In port facilities, the color of concrete surfaces varies with material composition, age, weathering, and moisture conditions. If color information is preserved, the anomaly detection model may overfit to specific background colors contained in the training data and become less robust when applied to other sites. By converting the images to grayscale, the proposed method suppresses this site-dependent color bias and instead emphasizes brightness contrast and morphological characteristics, thereby improving generalization across different port environments. An example of a grayscale transformation applied to a cropped crack candidate region is shown in Figure 2.



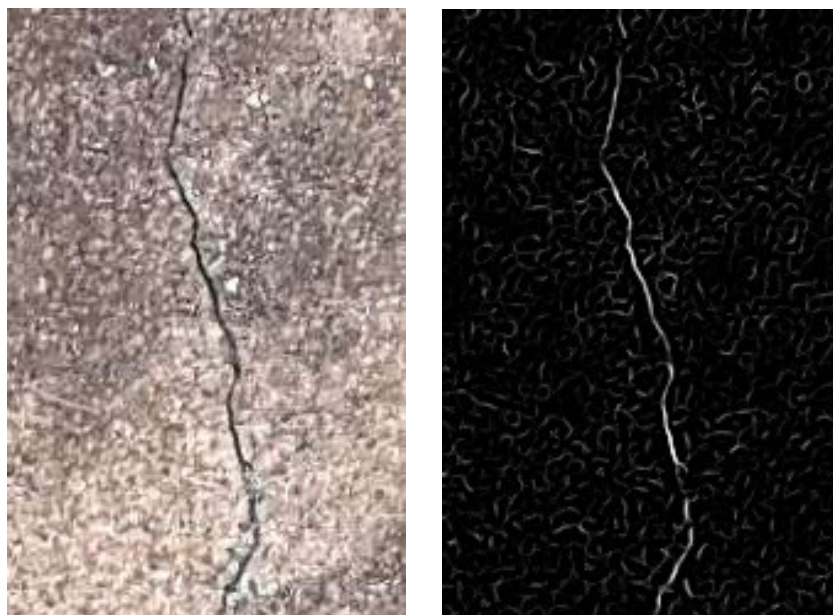
**Figure 2.** Example of grayscale transformation used for preprocessing of a cropped crack candidate region. Left: original image; right: grayscale image.

In addition, Frangi filtering is applied in some experimental settings to selectively enhance line-like structures in the cropped candidate regions. Originally proposed for vessel enhancement in medical image analysis [20], the Frangi filter has also been used in other fields to emphasize elongated

structures [21]. In this study, the Frangi filter is applied to highlight crack-like linear continuity while suppressing irregular background patterns on concrete surfaces. The motivation for using Frangi filtering is particularly strong when the object detection threshold is lowered. Under such settings, the detector tends to respond not only to cracks but also to small surface unevenness and texture patterns on concrete. Since cracks can be regarded as elongated and directionally continuous structures, Frangi filtering is expected to emphasize crack-like geometry while attenuating blob-like or irregular noise. In this study, the filter is implemented using the scikit-image library [22]. In the implementation used in this study, the Frangi filtering step was applied using the `frangi` function from the `skimage.filters` module, without explicit parameter overriding in the script. The mathematical formulation of the Frangi filter used in this study is summarized in Equations (1) and (2), and an example of Frangi filtering applied to a cropped crack candidate region is shown in Figure 3.

$$H = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix} \quad (1)$$

$$V(s) = \begin{cases} 0 & \text{if } \lambda_2 > 0 \\ \exp\left(-\frac{R_B^2}{2\beta^2}\right) \left(1 - \exp\left(-\frac{S^2}{2c^2}\right)\right) & \text{otherwise} \end{cases} \quad (2)$$



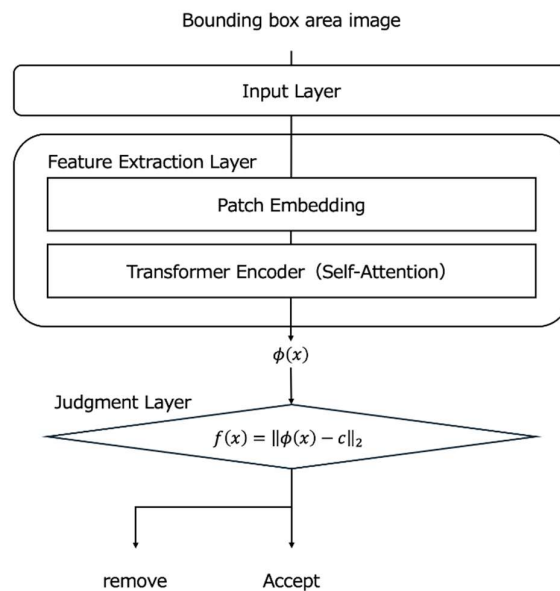
**Figure 3.** Example of Frangi filtering used to enhance crack-like linear structures in a cropped candidate region. Left: original image; right: Frangi-filtered image.

Here,  $\lambda_1$  and  $\lambda_2$  denote the eigenvalues of the Hessian matrix,  $R_B = \frac{|\lambda_1|}{|\lambda_2|}$  represents the degree of structural anisotropy, and  $S = \sqrt{\lambda_1^2 + \lambda_2^2}$  represents the strength of the local second-order structure, while  $\beta$  and  $c$  are constants controlling the sensitivity of the filter. As shown in Figure 3, this filtering enhances crack-like linear continuity while suppressing irregular surface patterns that often cause false positives in low-threshold detection.

### 3.2.3. ViT-Based Anomaly Detection Using a Crack Feature Center

After preprocessing, each cropped candidate region is evaluated by an anomaly-detection module in order to determine whether it is consistent with the learned feature distribution of crack images. In this study, a Vision Transformer (ViT) is adopted as the feature extractor [17]. Unlike conventional convolutional neural networks, which primarily capture local patterns via convolutional filters, Vision Transformer models capture relationships among image patches via self-

attention [17]. This property is well-suited to crack inspection because cracks often exhibit elongated and spatially continuous structures across the image. The training and inference workflow of the anomaly-detection module is illustrated in Figure 4.



**Figure 4.** Training and inference workflow of the ViT-based anomaly detection module using a crack feature center.

This design is particularly suitable for practical inspection environments in which all possible false-positive patterns cannot be predefined in advance. In such cases, unsupervised anomaly detection offers greater flexibility than conventional closed-set classification, as it assesses deviation from the learned crack distribution rather than relying on exhaustive prior labeling of non-crack categories.

More specifically, before being input to the ViT, each cropped candidate region was resized to  $224 \times 224$  pixels and converted into a three-channel representation. For both training and inference, the input tensor was normalized with mean = [0.5, 0.5, 0.5] and standard deviation = [0.5, 0.5, 0.5]. The resulting input image was then divided into  $16 \times 16$  pixel patches, and the pretrained ViT generates a 768-dimensional feature vector for each candidate region. The model is initialized with weights pretrained on ImageNet [23], and no additional fine-tuning of the ViT weights is performed in this study. Instead, 2445 crack-only images collected from the SDNET2018 dataset [24] are used to construct the crack feature center. Although SDNET2018 consists of close-range concrete crack images and therefore differs from the UAV-acquired port quay wall images used in the present verification, it was adopted in this study as a practical source of crack-only samples for constructing the crack feature center. The rationale is that the anomaly detection module was intended to model crack-relevant visual characteristics, particularly elongated and spatially continuous crack morphology, rather than site-specific background appearance. From this perspective, the SDNET2018-based feature center was expected to provide a useful reference for distinguishing crack-like structures from non-crack disturbances, even under the domain differences between ground-based crack images and UAV imagery. After applying the same preprocessing used during inference, feature vectors are extracted from the training crack images and averaged to construct a center feature vector in the embedding space. The center feature vector  $\mathbf{c}$  is defined as

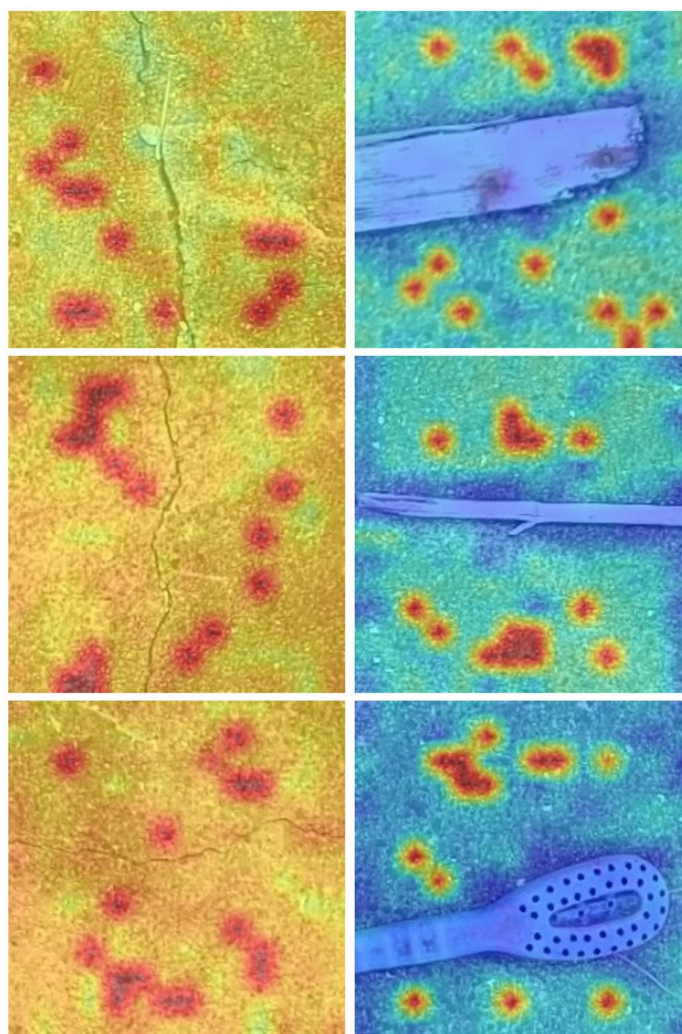
$$\mathbf{c} = \frac{1}{N} \sum_{i=1}^N \phi(x_i) \quad (3)$$

where  $N$  is the number of crack images used for training and  $\phi(x_i)$  denotes the feature vector extracted from training image  $i$ .

This center vector represents the centroid of normal crack patterns in terms of texture and shape. During inference, each preprocessed candidate region extracted from the object detection stage is passed through the same ImageNet-pretrained ViT-based feature extractor to obtain its feature vector in the same embedding space. An anomaly score is then computed as the Euclidean distance between the feature vector of the input region and the center feature vector obtained during training. The anomaly score  $f(x)$  is computed as

$$f(x) = \|\phi(x) - \mathbf{c}\|_2 \quad (4)$$

where  $\phi(x)$  denotes the feature vector of the input region, and  $\mathbf{c}$  is the center feature vector obtained during training. If the input region truly corresponds to a crack, its feature representation is expected to lie close to the learned crack distribution, resulting in a relatively small anomaly score. In contrast, if the input corresponds to debris or another non-crack disturbance, the extracted feature vector tends to deviate from the learned crack center, yielding a larger anomaly score. Regions whose anomaly scores exceed a predefined threshold are rejected as false positives. Examples of anomaly score distributions computed by the anomaly-detection module are shown in Figure 5. To enable relative comparison across test images, the heatmaps were globally normalized using the minimum and maximum anomaly scores observed across all images. This normalization was used solely for visualization, whereas the actual rejection of false positives was based on the original anomaly scores and the predefined threshold. These heatmaps visually support the above interpretation that true crack regions lie closer to the learned crack feature distribution, whereas false-positive regions tend to deviate from it and yield higher anomaly scores.



**Figure 5.** Comparison of anomaly score distributions computed by the anomaly detection model (left column: crack images; right column: false-positive images). Warm colors in the heatmap indicate regions with low anomaly scores, whereas cool colors indicate regions with high anomaly scores.

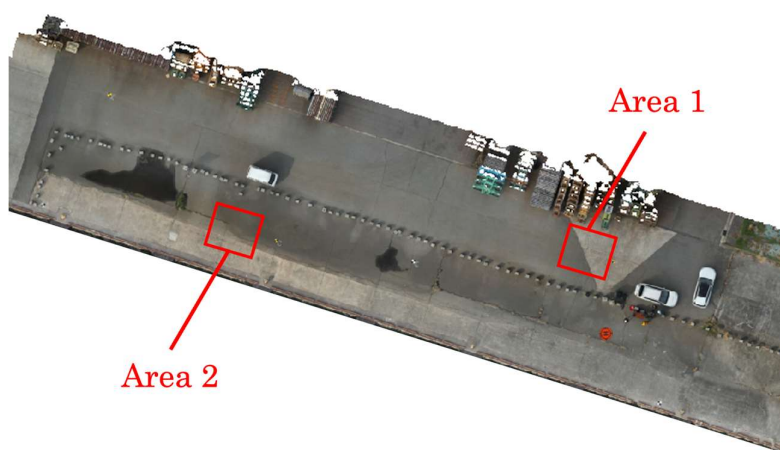
This anomaly-based formulation offers an important advantage over conventional closed-set classification. Because the method evaluates deviation from the crack feature distribution rather than explicitly classifying all non-crack categories, it does not require exhaustive prior learning of every possible false-positive object. This property is particularly beneficial in port inspection environments, where non-crack patterns are highly diverse and often unpredictable. Accordingly, the anomaly detection module functions as a second-stage verifier that systematically removes visually confusing non-crack regions while preserving broad crack candidate extraction in the preceding object detection stage.

#### 4. Experimental Setup

This section describes the study areas, image acquisition conditions, ground-truth annotation procedure, compared methods, and evaluation metrics used to validate the proposed false-positive reduction framework.

##### 4.1. Study Areas and Image Acquisition

The experimental validation was conducted in five verification areas distributed across three different port-related environments. Figures 6–8 show representative orthophotos of the verification areas for site illustration, whereas the crack-detection experiments were conducted at the single-image level. These areas were selected to evaluate the proposed method across a variety of concrete-surface conditions and false-positive patterns, while ensuring the presence of both crack regions and drift debris, which is a typical source of false positives in port quay wall inspection. For each verification area, the same spatial extent was then examined across different flight altitudes (5–20 m) and across the compared methods. For quantitative evaluation, one aerial image covering this predefined spatial extent was used for each verification area and flight-altitude condition. Accordingly, the reported metric values correspond to the measured results for the selected image under each condition, rather than averages over multiple images.



**Figure 6.** Representative orthophoto showing Verification Areas 1 and 2.

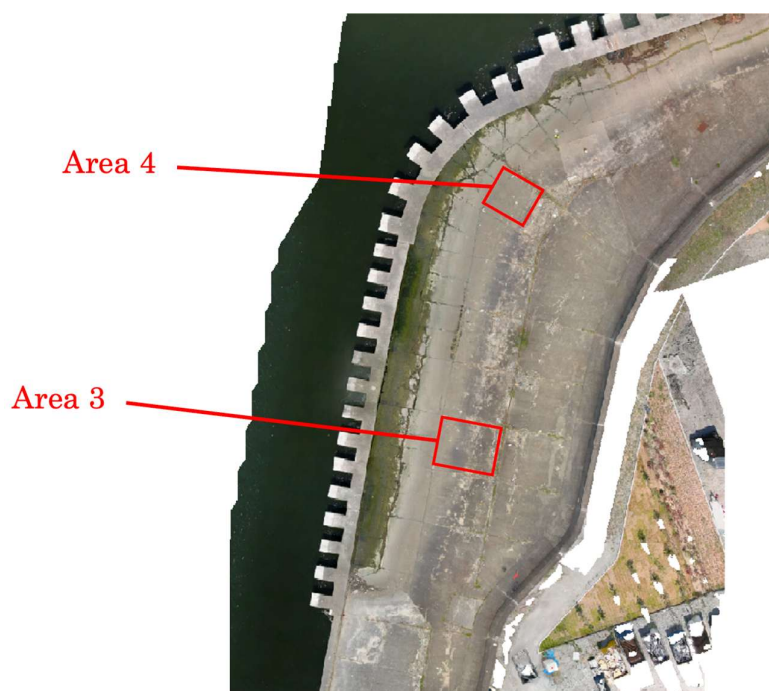


Figure 7. Representative orthophoto showing Verification Areas 3 and 4.



Figure 8. Representative orthophoto showing Verification Area 5.

A small, general-purpose UAV, the Autel Robotics EVO II Pro, was used to acquire aerial images across all verification areas. To examine the influence of flight altitude on crack detection performance, aerial images were captured at four flight altitudes: 5, 10, 15, and 20 m. To quantify the corresponding image resolution, the ground sample distance (GSD) was calculated for each altitude. The GSD values were approximately 1.14, 2.28, 3.42, and 4.56 mm/pixel at 5, 10, 15, and 20 m, respectively (Table 1). Because the target crack width in this study was 1.0 mm, the 5 m condition was treated as the baseline high-resolution setting. This relationship between target crack width and GSD is important for interpreting the subsequent results. As the GSD increases relative to the target crack width, fine cracks are represented by fewer effective image details, and their geometric continuity becomes more difficult to distinguish from small surface irregularities and background noise. At the same time, the higher-altitude datasets were used to evaluate the effect of reduced image resolution on crack detectability and false-positive suppression. The resulting aerial images and the corresponding annotation images were prepared for each verification area, from which one representative image was selected for each area-altitude condition for quantitative evaluation. In the present paper, the

main comparative discussion focuses on the image sets acquired at 5 and 10 m. The detailed results for the 15 m and 20 m cases of Verification Area 1 are additionally provided in Appendix B as supplementary evidence for the altitude-related analysis.

**Table 1.** Relationship between the EVO II Pro's flight altitude and ground sample distance (GSD).

Flight altitude (m)	GSD (mm/pixel)
5	1.14
10	2.28
15	3.42
20	4.56

To increase the diversity of imaging conditions, aerial surveys were conducted on multiple dates for each verification area. This design introduced variation in season, weather, and solar altitude, thereby increasing the diversity of brightness characteristics of the concrete surfaces.

#### 4.2. Ground Truth Annotation

For quantitative evaluation, ground-truth crack annotations were prepared for all verification areas. The annotation images were created using Labelme [25], and crack regions were manually delineated to serve as reference data for performance evaluation. In the annotation images, the crack regions were represented as the target class for comparison with the detection results.

In this study, construction joints on the concrete surface were excluded from the detection target. Although such joints may appear as line-like structures in aerial images, they were treated as known structural elements rather than damage. Accordingly, these regions were also excluded from the ground-truth annotation images. This setting was introduced to ensure that the evaluation focused specifically on actual crack detection performance rather than on the detection of structurally known non-damage features.

#### 4.3. Compared Methods

To verify the effectiveness of the proposed framework, five different methods were compared. Table 2 summarizes the configurations of the five compared methods for object detection threshold, preprocessing, and anomaly detection.

**Table 2.** Configurations of the five compared methods for object detection threshold and preprocessing. Methods 2, 4, and 5 additionally include anomaly-based post-processing.

Setting	Method 1	Method 2	Method 3	Method 4	Method 5
Object detection threshold	Standard	Standard	Low	Low	Low
Preprocessing for anomaly detection	None	Grayscale	None	Grayscale	Grayscale + Frangi
Anomaly detection	No	Yes	No	Yes	Yes

Method 1 corresponds to the conventional baseline configuration under the standard object detection threshold and without preprocessing. Method 2 adds grayscale preprocessing under the same threshold setting. Method 3 uses a lower object-detection threshold without preprocessing, whereas Method 4 combines the low-threshold setting with grayscale preprocessing. Method 5 further introduces Frangi filtering in addition to grayscale preprocessing under the low-threshold setting.

Method 1 serves as the conventional reference for comparison. Method 2 was designed to examine whether the proposed anomaly detection stage can improve the reliability of the baseline configuration under the standard-threshold setting. Method 3 was introduced to investigate whether

lowering the object detection threshold can reduce missed crack detections by collecting crack candidates more exhaustively. Because such low-threshold operation is expected to increase false positives caused by debris and irregular concrete textures, Methods 4 and 5 were evaluated to determine whether these newly induced false positives can be removed through different preprocessing strategies. Frangi filtering was introduced specifically to suppress the large number of false positives generated under the low-threshold setting. For this reason, a standard-threshold configuration with Frangi filtering was not included, as the present comparison primarily aimed to evaluate Frangi-based preprocessing as a countermeasure for low-threshold over-detection. In the current experimental design, the effectiveness of each method was primarily discussed in terms of false-positive suppression and comprehensive crack detection. In the experiments, the standard object detection threshold was set to 0.4, whereas the low-threshold setting was set to 0.03.

The anomaly detection thresholds were manually calibrated for each verification area, flight altitude, and preprocessing condition during the preliminary analysis to achieve a practical balance between false-positive suppression and crack retention. Accordingly, the threshold was treated as a condition-dependent calibration parameter in the present experimental framework, rather than as a universally fixed value. To improve transparency and reproducibility, the complete list of threshold values used in all experiments, including Method 5, is now provided in Table A1 in Appendix A.

#### 4.4. Evaluation Metrics

The detection performance of the baseline and proposed methods was quantitatively evaluated using four metrics: Precision, Recall, F1-score, and F2-score. Precision was used to evaluate the degree of false-positive suppression, whereas Recall measured the extent to which actual cracks were successfully detected. Because these two metrics are generally in trade-off, the F1-score was used as a balanced indicator of overall performance. In addition, the F2-score was introduced to place greater emphasis on Recall, reflecting the practical requirement in infrastructure inspection that missed crack detections should be minimized.

Let TP denote correctly detected crack evidence, FP denote falsely detected non-crack evidence, and FN denote missed crack evidence. Based on these definitions, the evaluation metrics were computed as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

$$\text{F2-score} = \frac{5 \times \text{Precision} \times \text{Recall}}{4 \times \text{Precision} + \text{Recall}} \quad (8)$$

In the present evaluation, Precision and Recall were computed using different counting units in accordance with the role of the proposed framework as a post-processing verifier of detected bounding-box candidates. Precision was evaluated on a candidate-region basis: each detected bounding box was counted as correct if it contained at least one annotated crack pixel; otherwise, it was counted as a false positive. By contrast, Recall was evaluated on a pixel-coverage basis: each annotated crack pixel was counted as successfully detected if it was included in at least one detected bounding box. Consequently, multiple detections overlapping the same crack could contribute independently to Precision, whereas they did not increase Recall.

Using this evaluation design, the experiments were intended to assess not only whether the proposed framework can suppress false positives among detected candidates, but also whether it can maintain or improve practical crack coverage relative to the conventional object-detection-based baseline. Accordingly, the evaluation should be interpreted as a candidate-region-level assessment of practical false-positive suppression performance, rather than as an instance-segmentation-style evaluation of complete crack geometry. Although this evaluation rule differs from one-to-one object

matching, it was applied identically to all compared methods and was therefore intended to support internally consistent relative comparison under a common practical verification protocol.

## 5. Results and Discussion

### 5.1. Results for Verification Area 1

#### 5.1.1. Flight Altitude of 5 m

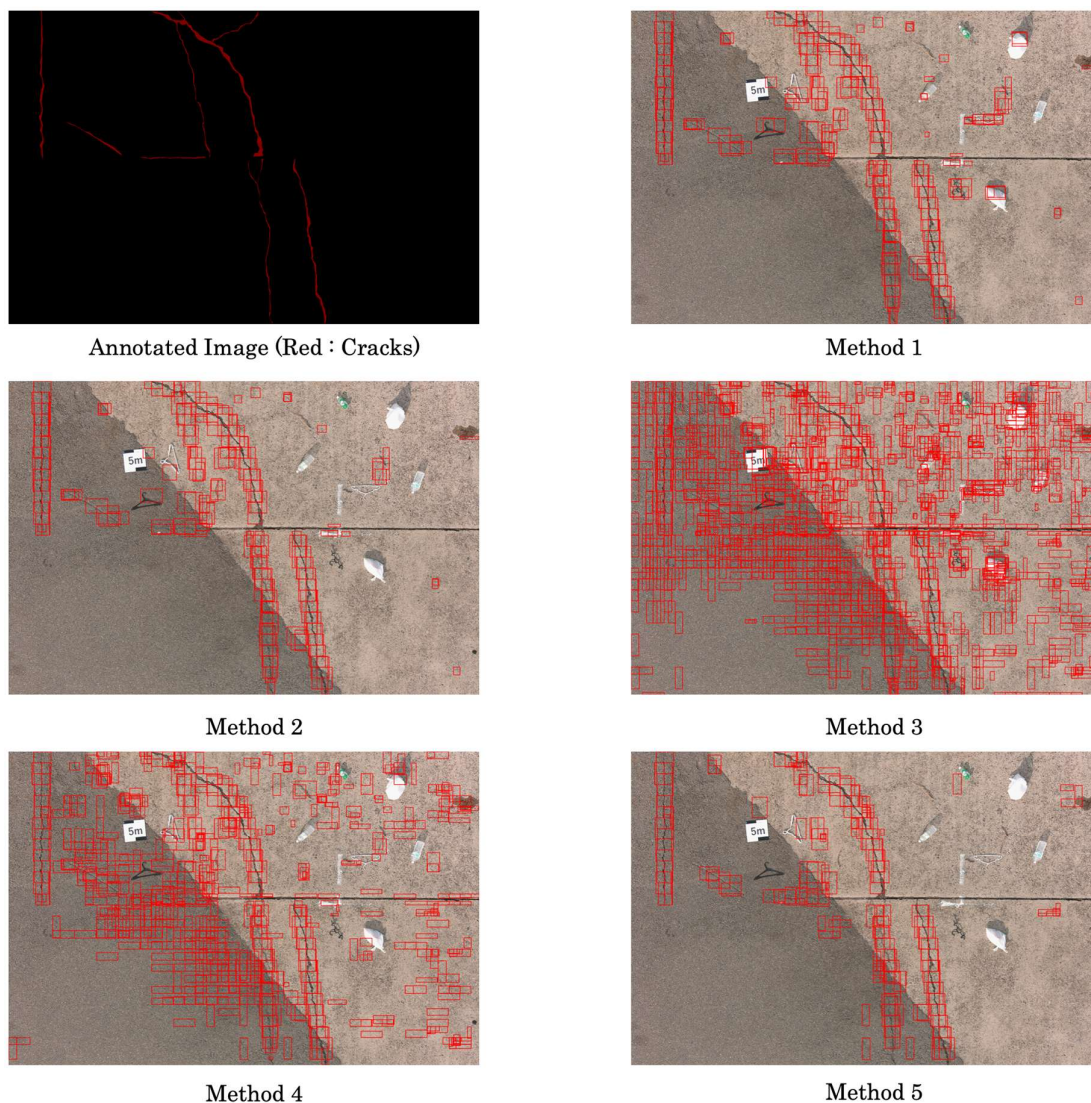
The quantitative results for Verification Area 1 at a flight altitude of 5 m are summarized as follows. Under the standard detection-threshold setting, Method 2 improved Precision while largely preserving the crack detections obtained by Method 1, resulting in higher overall performance. This indicates that introducing anomaly detection after object detection can effectively suppress false positives without substantially sacrificing crack detectability.

In contrast, Method 3, which employed a lower object-detection threshold, increased Recall by extracting crack candidates more comprehensively, but this was accompanied by a substantial decrease in Precision due to a large number of newly generated false positives. This result suggests that lowering the detection threshold improves candidate coverage, but it also leads to excessive responses to fine surface irregularities in concrete.

A clear difference was observed between Methods 4 and 5 in the low-threshold setting. In Method 4, many of the false positives generated by Method 3 persisted in the final results, indicating that anomaly detection based solely on grayscale information was insufficient to distinguish true cracks from small surface unevenness. By contrast, Method 5 removed most of these false positives, suggesting that preprocessing with the Frangi filter was more effective because it emphasized crack-like linear geometry rather than brightness alone. The results therefore demonstrate that, for suppressing false positives induced by low-threshold operation, geometric structure-based preprocessing is more suitable than grayscale-only preprocessing.

The numerical results are also consistent with this interpretation. Regarding Verification Area 1 at 5 m, Method 5 achieved a Precision of 0.7443, outperforming Method 4 (0.3475) and Method 3 (0.2380), while maintaining a Recall of 0.8674. Its F1-score and F2-score reached 0.8012 and 0.8396, respectively. These values indicate that Frangi-based preprocessing substantially improved the low-threshold results compared with Methods 3 and 4 by recovering Precision while retaining relatively high Recall. However, Method 5 did not surpass Method 2 in the balanced or recall-oriented metrics under this condition. Therefore, for Verification Area 1 at 5 m, the standard-threshold anomaly-based configuration (Method 2) remained the more reliable overall operating strategy, although Method 5 still demonstrated the practical benefit of Frangi-based preprocessing as a countermeasure for low-threshold over-detection.

Figure 9 visually supports the same conclusion as the quantitative results. Method 2 removed part of the false detections observed in Method 1 while preserving the main crack pattern. Method 3 produced numerous spurious detections over the concrete surface, whereas Method 5 substantially reduced them. Thus, at 5 m, the proposed framework improved not only the standard-threshold baseline but also the reliability of the low-threshold result through Frangi-based preprocessing.



**Figure 9.** Visual comparison of crack detection results obtained by the five methods for Verification Area 1 at a flight altitude of 5 m. Red boxes indicate detected crack regions, and the annotation image is shown in red.

#### 5.1.2. Flight Altitude of 10 m

The quantitative results for Verification Area 1 at a flight altitude of 10 m are summarized as follows. As in the 5 m case, Method 2 improved Precision relative to Method 1 while only slightly decreasing Recall, and consequently achieved higher F1 and F2 scores. This confirms that anomaly-based false-positive reduction remained effective even at the 10 m image resolution. In other words, under the standard threshold setting, the proposed second-stage verification still contributed to more reliable crack detection.

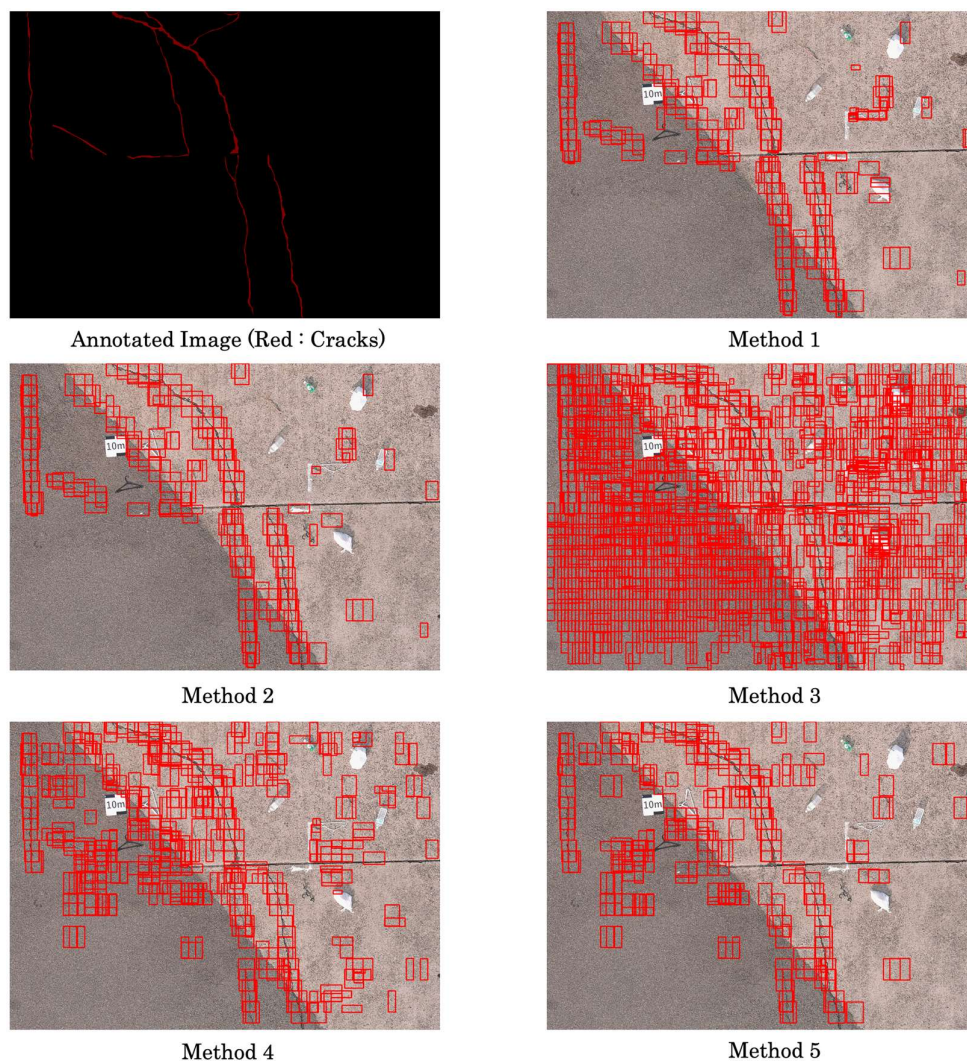
Method 3 again increased Recall relative to Method 1, but at the cost of a substantial drop in Precision, indicating that the lower detection threshold yielded more crack candidates while also generating many new false positives. This pattern is consistent with the 5 m case and suggests that low-threshold object detection remained highly sensitive to non-crack surface patterns.

However, unlike at 5 m, the effectiveness of anomaly detection in the low-threshold setting became markedly weaker at 10 m. Method 4 showed only a very limited improvement over Method 3, implying that grayscale-based preprocessing could no longer sufficiently distinguish cracks from concrete-surface irregularities under reduced resolution. Method 5 still outperformed Method 4,

which suggests that emphasizing line-like geometry remained beneficial. For example, in Verification Area 1, the Precision of Method 4 was 0.1604, whereas that of Method 5 increased to 0.5441. Nevertheless, this value was still far lower than the corresponding 5 m result for Method 5 (0.7443), showing that the benefit of Frangi-based preprocessing deteriorated considerably as image resolution decreased.

This degradation indicates that, at 10 m, the geometric differences between fine cracks and small concrete-surface irregularities became less distinguishable, thereby reducing the effectiveness of both the Frangi filter and the ViT-based anomaly detector. As a result, although Method 5 remained more effective than Method 4 at suppressing false positives generated by low-threshold detection, the overall balance between reliability and crack detectability remained inferior to that of Method 2. Therefore, for Verification Area 1 at 10 m, the most reliable operating strategy was not the low-threshold setting but rather the standard-threshold detection followed by anomaly detection.

Figure 10 visually supports the same interpretation as the quantitative results. Method 2 retained the principal crack detections while removing part of the false positives present in Method 1. In contrast, Method 3 produced dense spurious responses across the concrete surface. Although Method 5 reduced these responses more effectively than Method 4, residual false detections remained more noticeable than in the 5 m results, indicating a decline in discrimination performance due to reduced image resolution.



**Figure 10.** Visual comparison of crack detection results obtained by the five methods for Verification Area 1 at a flight altitude of 10 m. Red boxes indicate detected crack regions, and the annotation image is shown in red.

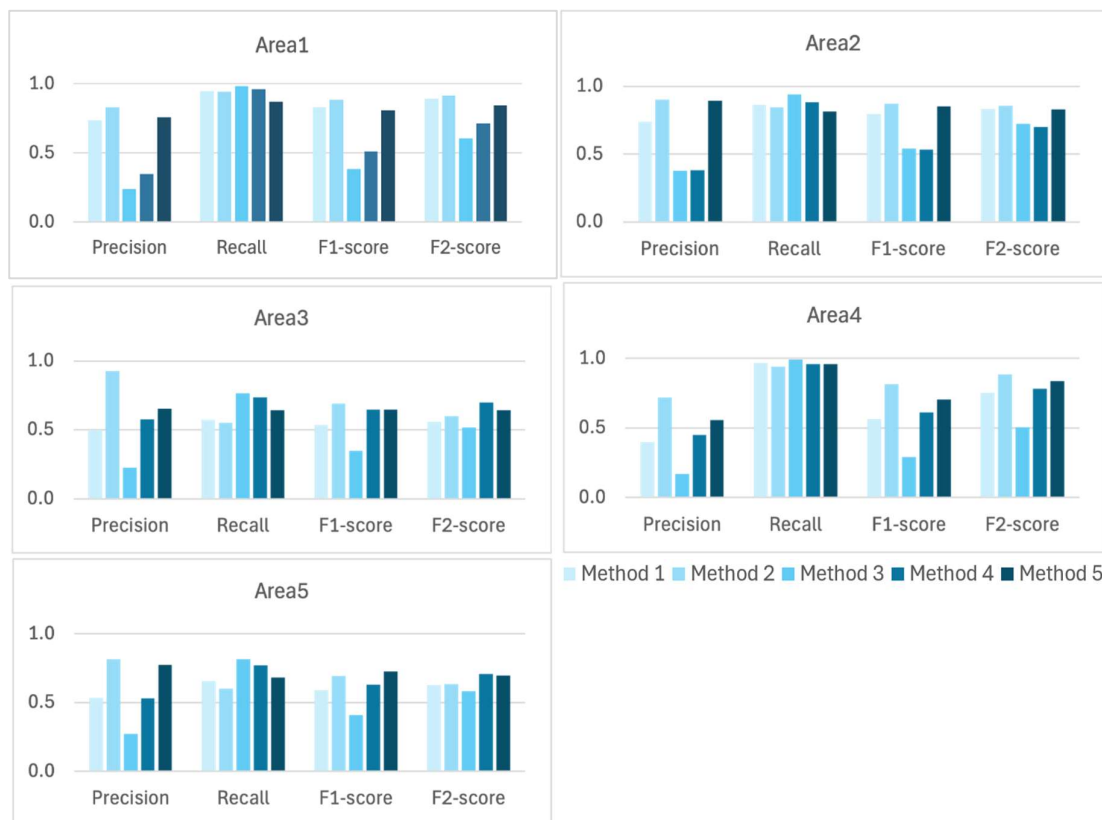
Overall, the comparison between 5 m and 10 m suggests that the proposed framework remained effective under the standard-threshold setting across both resolutions. In contrast, the benefit of low-threshold operation became increasingly limited as the image resolution decreased. This finding implies that the practical advantage of combining broad candidate extraction with anomaly-based filtering depends strongly on image quality and GSD. To improve the completeness of the altitude-dependent analysis, the detailed results for the 15 m and 20 m cases of Verification Area 1 are additionally provided in Appendix B. These supplementary results show the same overall tendency as observed at 10 m: Method 2 remained the most reliable configuration under the standard-threshold setting, whereas the low-threshold variants, particularly Methods 4 and 5, became less effective as image resolution deteriorated.

Figure 11 provides an overview of the detection performance of the five methods across Verification Areas 1–5 at a flight altitude of 10 m, and highlights the stable effectiveness of Method 2 under the standard-threshold setting despite the reduced image resolution.



**Figure 11.** Comparison of Precision, Recall, F1-score, and F2-score across Verification Areas 1–5 at a flight altitude of 10 m.

Because the subsequent discussion focuses on Verification Areas 2–5 under the 5 m condition, Figure 12 provides an overview of the detection performance of the five methods across Verification Areas 1–5 at a flight altitude of 5 m.



**Figure 12.** Comparison of Precision, Recall, F1-score, and F2-score across Verification Areas 1–5 at a flight altitude of 5 m.

### 5.2. Results for Verification Areas 2–5

Because the reliability of crack detection decreased with increasing flight altitude in Verification Area 1, the subsequent verification for Areas 2–5 was conducted using aerial images acquired at a flight altitude of 5 m. This setting was adopted in order to evaluate the proposed method under sufficiently high-resolution conditions while focusing on differences in environmental disturbances among sites.

As shown in Figure 12, the results at 5 m across Verification Areas 1–5 exhibited broadly consistent trends. Under the standard-threshold setting, Method 2 generally improved Precision relative to Method 1 while maintaining comparable Recall, indicating that anomaly detection remained effective as a post-processing step for suppressing false positives across different environments. By contrast, Method 3, which adopted a lower detection threshold, increased Recall across all areas but led to a substantial decline in Precision, confirming that broader candidate extraction inevitably induced many additional false positives. When anomaly detection was applied after low-threshold detection, both Method 4 and Method 5 recovered part of the lost Precision. However, their relative effectiveness varied depending on site conditions, suggesting that the benefit of preprocessing depended on the dominant surface patterns and disturbance types present in each area.

#### 5.2.1. Verification Area 2

For Verification Area 2, Method 2 provided the most stable overall improvement over the baseline. Compared with Method 1, Precision increased from 0.7391 to 0.9042, while Recall decreased only slightly from 0.8635 to 0.8440. As a result, both F1-score and F2-score improved, from 0.7965 and 0.8354 to 0.8731 and 0.8554, respectively. These results indicate that, in this area, anomaly detection

effectively removed false positives generated by the baseline detector without causing a substantial loss of crack detectability.

Under the low-threshold setting, Method 3 increased Recall to 0.9422 but reduced Precision to 0.3778, demonstrating the expected trade-off between broader candidate extraction and severe over-detection. Method 4 showed only limited recovery, whereas Method 5 substantially improved Precision to 0.8941 while maintaining Recall at 0.8159. Although Method 5 did not surpass Method 2 in balanced metrics, it clearly outperformed Method 4, suggesting that Frangi-based preprocessing was more effective than grayscale-only preprocessing for suppressing false positives induced by the low-threshold operation in this area.

### 5.2.2. Verification Area 3

Verification Area 3 exhibited a slightly different tendency. As in the other areas, Method 2 markedly improved Precision relative to Method 1, increasing it from 0.5000 to 0.9254, while Recall changed only marginally from 0.5735 to 0.5507. Consequently, the F1-score increased from 0.5343 to 0.6905, confirming that the proposed anomaly detection step again contributed to more reliable crack detection under the standard-threshold setting.

However, under the low-threshold setting, the performance balance differed from that in Verification Area 2. Method 4 achieved Precision and Recall values of 0.5758 and 0.7350, respectively, yielding the highest F2-score among the low-threshold variants (0.6964), whereas Method 5 slightly improved the F1-score to 0.6474 but reduced the F2-score to 0.6438. This suggests that, in Verification Area 3, both preprocessing strategies were effective to some extent, but grayscale-based preprocessing retained a slight advantage in the recall-oriented evaluation, whereas Frangi-based preprocessing provided a marginally better balance between precision and recall.

### 5.2.3. Verification Area 4

In Verification Area 4, the superiority of Method 2 became particularly clear. Relative to Method 1, Method 2 increased Precision from 0.3981 to 0.7185 while maintaining a high Recall of 0.9408, resulting in substantial improvements in F1-score and F2-score, from 0.5641 and 0.7521 to 0.8148 and 0.8860, respectively. This confirms that anomaly detection introduced after standard-threshold object detection provided the most reliable crack detection for this area.

Method 3 again yielded the highest Recall (0.9915) but at the cost of extremely low Precision (0.1697), indicating the generation of a large number of false positives. Although both Method 4 and Method 5 improved Precision relative to Method 3 and even exceeded Method 1 in balanced metrics, neither method outperformed Method 2. The results likewise indicate that, for this area, introducing anomaly detection into the existing standard-threshold framework was more reliable than lowering the detection threshold and then attempting to remove the newly generated false positives.

### 5.2.4. Verification Area 5

Verification Area 5 showed another practically important pattern. Method 2 again improved Precision over the baseline, from 0.5352 to 0.8154, while Recall decreased only moderately from 0.6546 to 0.6018. This led to a higher F1-score than Method 1 (0.6925 vs. 0.5889), indicating that the standard-threshold anomaly-detection framework remained effective in this area as well.

At the same time, the low-threshold variants achieved competitive results in Verification Area 5. Method 5 increased Precision to 0.7736 while maintaining Recall at 0.6813, resulting in the highest F1-score among all five methods in this area (0.7245). By contrast, Method 4 achieved the highest F2-score (0.7074) because it retained a somewhat higher Recall of 0.7708. These results further suggest that Method 5 suppressed concrete-surface false positives more effectively than Method 4, but it also removed part of the fine crack regions as false positives. Thus, in Verification Area 5, Frangi-based preprocessing improved precision recovery under the low-threshold setting, whereas grayscale-based preprocessing preserved slightly better crack coverage.

### 5.2.5. Discussion of Verification Areas 2–5

Taken together, the results for Verification Areas 2–5 reinforce two main findings. First, Method 2 consistently provided a robust and reliable improvement over the conventional baseline across different site conditions. This supports the interpretation that anomaly detection can function as a stable false-positive suppression layer when applied after standard-threshold object detection. Second, the effectiveness of low-threshold detection followed by anomaly filtering was strongly site-dependent. In some areas, especially Verification Area 5, Method 5 achieved competitive or even superior F1 scores, whereas in others, Method 2 remained clearly preferable. These results suggest that the practical value of low-threshold operation depends not only on image resolution but also on the local characteristics of concrete-surface texture and environmental noise.

### 5.3. General Discussion

The experimental results obtained in this study reveal two important findings regarding the practical UAV-based crack inspection of port quay walls. First, integrating anomaly detection into the conventional object detection framework under the standard detection-threshold setting yielded a stable improvement in reliability across different flight conditions and site environments. Overall, Method 2 provided the most stable overall improvement over the conventional baseline across the tested conditions, indicating that anomaly detection can be incorporated into the existing system as a robust false-positive suppression layer without substantially sacrificing crack detectability. This finding is particularly important for practical deployment, because it suggests that the proposed framework can improve inspection reliability even in visually complex inspection environments on port quay walls, where debris and other disturbances frequently interfere with crack identification. Accordingly, the main contribution of this study is not the proposal of a fundamentally new anomaly-detection architecture, but the operational validation of a reliability-oriented inspection framework under practical conditions relevant to UAV-based geomatics and infrastructure monitoring.

Second, the experimental results clarify both the potential and the limitations of the low-threshold detection strategy. Lowering the object detection threshold increased crack candidate coverage and enabled capturing fine crack regions that were likely to be missed in the conventional setting. However, this broader candidate extraction also yielded a large number of false positives due to concrete-surface irregularities and environmental noise. In this context, Method 5, which combined low-threshold object detection with Frangi-based preprocessing and anomaly detection, demonstrated greater versatility than Method 4, which relied primarily on grayscale-based preprocessing. In particular, under the 5 m condition, Method 5 achieved a favorable balance between broader detection and false-positive suppression in some verification areas. These results suggest that emphasizing geometric line-like structures is more effective than relying only on brightness information when distinguishing cracks from visually similar non-crack patterns. In this sense, the study should be understood as a practically oriented comparative evaluation of false-positive suppression strategies across object-detection thresholds, preprocessing conditions, and image resolutions represented by different flight altitudes and GSDs, rather than as a benchmark for proposing a new deep learning backbone.

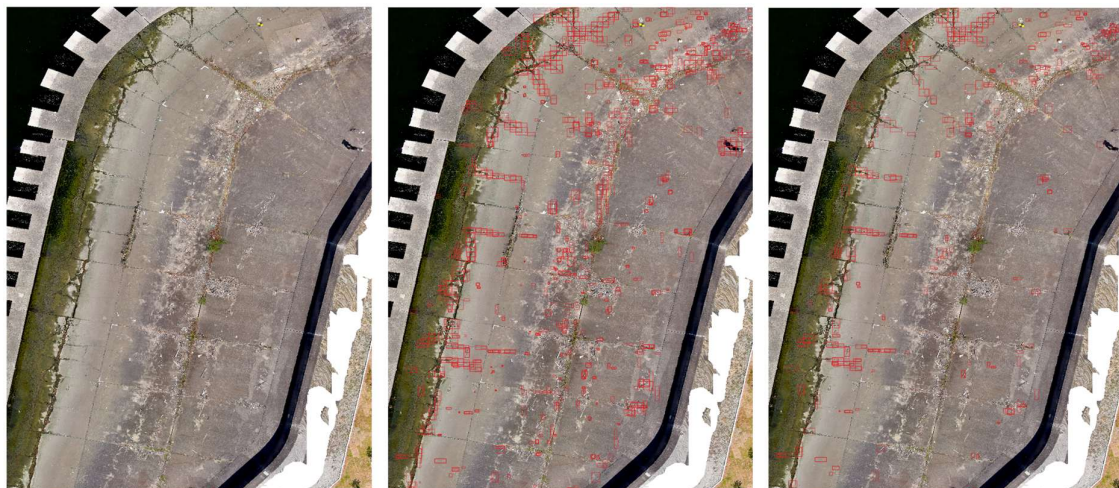
At the same time, the results also demonstrate that the effectiveness of this low-threshold strategy is strongly dependent on image resolution. As flight altitude increased, the geometric difference between fine cracks and small concrete-surface unevenness became less distinguishable, and the anomaly detection stage was no longer able to sufficiently eliminate the newly generated false positives. These findings indicate that, when the flight altitude exceeded 10 m, the current combination of Frangi filtering and the ViT-based anomaly detector approached its discrimination limit, as reduced image resolution weakened the geometric cues required for reliable separation. Accordingly, the present results imply that the practical benefit of the low-threshold strategy is conditional on sufficiently high-resolution imagery. In contrast, the standard-threshold anomaly-detection framework is more robust across operating conditions. The need for condition-specific threshold calibration is a current limitation of the proposed framework, and future work should

investigate more principled and automated threshold selection strategies. A supplementary sensitivity analysis for representative conditions is additionally provided in Appendix C, showing that the adopted thresholds corresponded to practical operating points near the balance between false-positive suppression and crack retention.

Another limitation of the present study is the potential domain gap between the SDNET2018 crack images used to construct the crack feature center and the UAV-acquired images of port quay walls used for verification. These two image domains differ in acquisition geometry, image resolution, lighting conditions, and background texture, and such differences may affect the transferability of the learned crack feature distribution. Therefore, the present results should be interpreted as evidence that the SDNET2018-based crack feature center retained practical utility under the investigated conditions, rather than as proof that the same transferability will hold universally across different UAV inspection settings. Future work should examine feature-center construction using UAV-specific crack datasets and more systematic strategies for reducing domain mismatch.

From a geomatics perspective, these findings have two implications. One is methodological: reliability enhancement should not be discussed only in terms of georeferenced output, but also in terms of the fundamental discrimination capability of the image-analysis pipeline itself. The other is operational: for low-cost UAV-based infrastructure inspection, maintaining stable performance under practical field conditions requires not only efficient image acquisition but also an appropriate balance between crack candidate extraction and post-detection verification. In this sense, the proposed anomaly-based framework can be interpreted as a reliability-oriented extension to a low-cost UAV inspection workflow. In other words, the present study focuses on reliability validation at the image-analysis level, while the extension to a full orthophoto-based workflow should be understood as the next implementation step rather than a prerequisite for interpreting the core discrimination results.

To illustrate the practical integration of the proposed framework into a georeferenced inspection workflow, Figure 13 presents an example in which crack detection results were projected onto an orthophoto of the same inspection area. The comparison visually indicates that the anomaly-based configuration (Method 2) reduced many spatially scattered false positives in the orthophoto-based output while preserving the main crack-related detections. In particular, the anomaly-based result shows fewer false detections around the drift-debris-rich central portion of the orthophoto. Because the orthophoto used in Figure 13 was generated from aerial images acquired at 30 m due to site constraints, its spatial resolution is lower than that of the images used in the main quantitative evaluation. Accordingly, some cracks are not clearly represented in the orthophoto-based output, and this figure should be interpreted as an illustrative demonstration of georeferenced integration rather than as a complete crack-detection result. This figure is therefore intended only as an end-to-end illustration of georeferenced output, whereas the main quantitative evaluation in this study remains based on single-image analysis.



**Figure 13.** Illustrative end-to-end demonstration of orthophoto-based georeferenced output. Left: original orthophoto of the inspection area (approximately 50 m × 30 m). Center: crack detection results from the conventional pipeline projected onto the orthophoto. Right: crack detection results from the anomaly-based configuration (Method 2) projected onto the same orthophoto.

Nevertheless, several issues remain. The current anomaly detection framework may still remove parts of true crack regions, especially under aggressive filtering conditions, and its effectiveness decreases with lower image resolution. These results also suggest several future directions, including dedicated learning strategies tailored to port quay wall inspection, geometric or morphological completion of crack regions partially removed during anomaly filtering, threshold optimization strategies, and the use of super-resolution techniques. Additional improvements may also be achieved by updating the baseline object detection model itself and by developing hybrid discrimination frameworks that combine geometric, color, and depth-related features. These directions are expected to further improve robustness against changes in altitude and imaging environment.

It should also be noted that the scope of the present study was intentionally limited to port quay walls, which represent a practically important but visually challenging inspection target. Accordingly, the current findings should be interpreted as evidence of effectiveness within this domain, based on comparisons across flight altitudes, preprocessing conditions, and threshold settings, rather than as proof of immediate generalization to other infrastructure types or external crack datasets. Validation across different infrastructure assets and cross-dataset settings remains an important subject for future work. It should also be noted that the present study did not include formal statistical significance testing, because the evaluation was designed as a practical comparative analysis across a limited number of site-specific verification cases rather than as a large-scale benchmark experiment with repeated randomized trials. Therefore, the present findings should be interpreted primarily in terms of the consistency of performance trends observed across the investigated areas, flight altitudes, and preprocessing conditions. Because all compared methods were evaluated on the same predefined spatial extent for each verification area and flight-altitude condition, the observed differences should be interpreted as method-dependent changes under matched visual conditions rather than as differences caused by scene selection. Accordingly, the present results should be interpreted as evidence of the comparative benefit of anomaly-based post-processing under condition-dependent calibration, rather than as proof of a universally threshold-free inspection workflow.

To examine the practical applicability of the proposed framework, the computational environment and runtime were also recorded. Table 3 summarizes the hardware environment used for runtime measurement. Because YOLOR-based crack detection and anomaly detection (for both

grayscale- and Frangi-based settings) were executed in different software environments, the corresponding Python, PyTorch, and CUDA versions are listed separately.

**Table 3.** Computational environment used for runtime measurement.

Component	GPU	CPU	Memory	Python	PyTorch	CUDA
YOLOR-based crack detection	NVIDIA GeForce RTX 3090	12th Gen Intel Core i9-12900KF	64 GB	3.8.16	1.7.1+cu110	11.0
Anomaly detection (grayscale / Frangi)	NVIDIA GeForce RTX 3090	12th Gen Intel Core i9-12900KF	64 GB	3.11.13	2.5.1+cu121	12.1

Table 4 reports the average number of candidate regions and the average total processing time under representative experimental conditions. The runtime values were measured in seconds from two representative site images in Verification Areas 1 and 2, using a single execution for each condition and including image loading and saving. In contrast, GeoJSON output was not included because the present study evaluated the framework at the single-image level. Table 4 focuses on representative configurations that include anomaly-based post-processing, because the purpose of the runtime comparison was to examine the additional computational cost introduced by preprocessing and anomaly detection. The runtime comparison was intended solely as a practical reference, not a strict benchmark analysis.

**Table 4.** Average number of candidate regions and average total processing time under representative experimental conditions.

Method	Object detection threshold	Average candidate regions per image	Average total processing time (s)
Method 2	0.40	555.6	23.34
Method 4	0.03	2333.9	27.89
Method 5	0.03	2333.9	41.07

Note: The average number of candidate regions was calculated over all verification areas, whereas the total processing time was measured as the average of two representative site images from Verification Areas 1 and 2. The representative original aerial images used for runtime measurement had a resolution of  $5472 \times 3648$  pixels. The runtime includes image loading and saving, but does not include GeoJSON output.

As expected, the low-threshold setting substantially increased the number of candidate regions and, consequently, the total processing time of the overall workflow.

## 6. Conclusions

This study investigated a framework for reducing false positives in UAV-based crack inspection of port quay walls using aerial images acquired by a small general-purpose UAV. The proposed approach introduced anomaly detection after object detection in order to suppress false detections caused by debris and other environmental disturbances, which have been a major obstacle to practical implementation. In addition, this study examined a strategy in which the object detector's detection threshold was intentionally lowered to extract crack candidates more comprehensively, which were then filtered using anomaly detection.

The experimental results demonstrated that integrating anomaly detection into the conventional object detection framework, with the standard threshold setting, improved the reliability of crack detection across flight conditions and site environments. In particular, Method 2 consistently achieved better overall performance than the baseline method, indicating that the proposed post-processing strategy can stably suppress false positives while preserving practical crack detectability. This result suggests that anomaly detection is an effective means of enhancing the reliability of low-cost UAV-based crack inspection systems for port quay walls.

The results also showed that lowering the object detection threshold can improve the comprehensiveness of crack candidate extraction, especially for fine cracks that may otherwise be

overlooked. When this low-threshold strategy was combined with anomaly detection and Frangi-based preprocessing, a favorable balance between broader detection coverage and false-positive suppression was achieved in some cases under the 5 m condition. However, this benefit was not maintained at higher altitudes, where reduced image resolution made it difficult to distinguish fine cracks from concrete surface irregularities. Thus, although the low-threshold strategy has practical potential, its effectiveness is strongly constrained by image quality and GSD.

Overall, the findings indicate that the most robust and practically reliable configuration in the current framework is the combination of standard-threshold object detection and anomaly-based false-positive suppression. Therefore, the significance of the present study lies in demonstrating how a practical anomaly-based post-processing framework can improve inspection reliability under realistic UAV imaging conditions, rather than in proposing a new deep learning backbone itself. At the same time, the results suggest that broader crack extraction using a lower threshold remains a promising direction when sufficiently high-resolution imagery is available. Therefore, the proposed framework provides a useful basis for improving the reliability of practical UAV-based crack inspection of port quay walls while clarifying the present limitations associated with image resolution and anomaly discrimination.

A practical limitation of the proposed framework is that the anomaly-detection threshold must be adjusted based on site conditions and GSD. The current framework may also remove part of the true crack region in some cases, indicating the need for improved robustness against over-filtering.

Future work will focus on improving discrimination capability under more challenging conditions, including more principled threshold selection, enhanced robustness against over-filtering, super-resolution for higher-altitude imagery, and multimodal feature integration. Although the present study intentionally evaluated the framework at the single-image level, the workflow can also be extended to orthophoto-based inspection and linked with GeoJSON and real-world coordinates for geomatics-oriented asset management and infrastructure DX applications.

**Author Contributions:** Conceptualization, M.A. and D.Y.; methodology, M.A. and D.Y.; software, M.A. and D.Y.; validation, M.A., D.Y., and W.F.; investigation, M.A., D.Y., and W.F.; data curation, M.A., D.Y., and W.F.; writing—original draft preparation, M.A. and D.Y.; writing—review and editing, D.Y. and M.A.; visualization, M.A., D.Y., and W.F.; supervision, D.Y.; project administration, D.Y.; funding acquisition, D.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Japan Science and Technology Agency (JST), Co-Creation Field Formation Support Program, grant number JPMJPF2115.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The scripts and datasets used in this study are publicly available in the GitHub repository “Scripts-for-Anomaly-Detection” (<https://github.com/omu-geolab/Scripts-for-Anomaly-Detection>, accessed on 23 March 2026).

**Acknowledgments:** The authors would like to express their sincere gratitude to the Osaka Ports and Harbors Bureau for providing access to experimental sites and inspection data. Additionally, we are grateful to Kazuki Ata for his dedicated help in preparing the experiment data.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

UAV	Unmanned Aerial Vehicle
ViT	Vision Transformer
GSD	Ground Sample Distance

GeoJSON	Geographic JavaScript Object Notation
TP	True Positive
FP	False Positive
FN	False Negative
F1	F1-score
F2	F2-score
ITU-R	Radiocommunication Sector of the International Telecommunication Union
BT.601	Recommendation BT.601
CNN	Convolutional Neural Network
YOLOv3	You Only Learn One Representation

## Appendix A

**Table A1.** Anomaly detection threshold values used for each verification area, flight altitude, and experimental method.

Verification Area	Flight Altitude	Method 2	Method 4	Method 5
1	5 m	20	18	4.3
1	10 m	21	19	4.8
1	15 m	21	21	4.8
1	20 m	22	21	5.5
2	5 m	20	21	4.8
2	10 m	22	21	4.8
2	15 m	23	23	5.5
2	20 m	23	24	5.5
3	5 m	18	18	5
3	10 m	19	17	5.2
3	15 m	20	20	5.4
3	20 m	—	20	5.4
4	5 m	17	17	5
4	10 m	20	19	5.4
4	15 m	20	20	5.4
4	20 m	—	19	5.4
5	5 m	19	18	5
5	10 m	20	19	5.2
5	15 m	21	19	5.2
5	20 m	21	21	5.2

Note: The anomaly detection thresholds shown in this table were manually calibrated for each verification area, flight altitude, and experimental method. The symbol “—” indicates that the corresponding condition was not evaluated.

## Appendix B

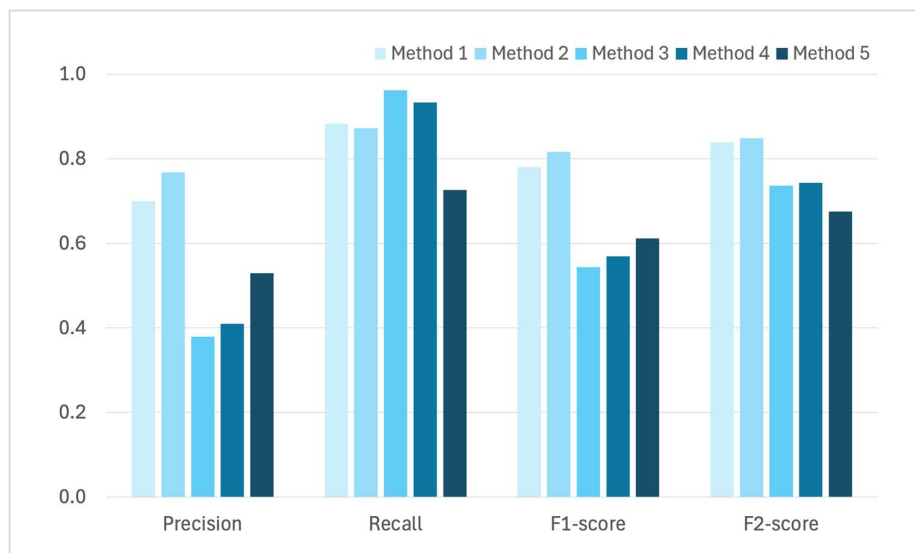
This appendix provides the detailed results for the 15 m and 20 m cases of Verification Area 1, which were acquired for the altitude-dependent analysis. These results are presented here to improve the completeness of the study and to clarify the deterioration trend observed under reduced image resolution. Table A2 summarizes the quantitative results for the 15 m and 20 m cases, while Figures A1–A4 provide corresponding metric comparisons and visual examples of the detection results.

**Table A2.** Quantitative results for Verification Area 1 at flight altitudes of 15 m and 20 m.

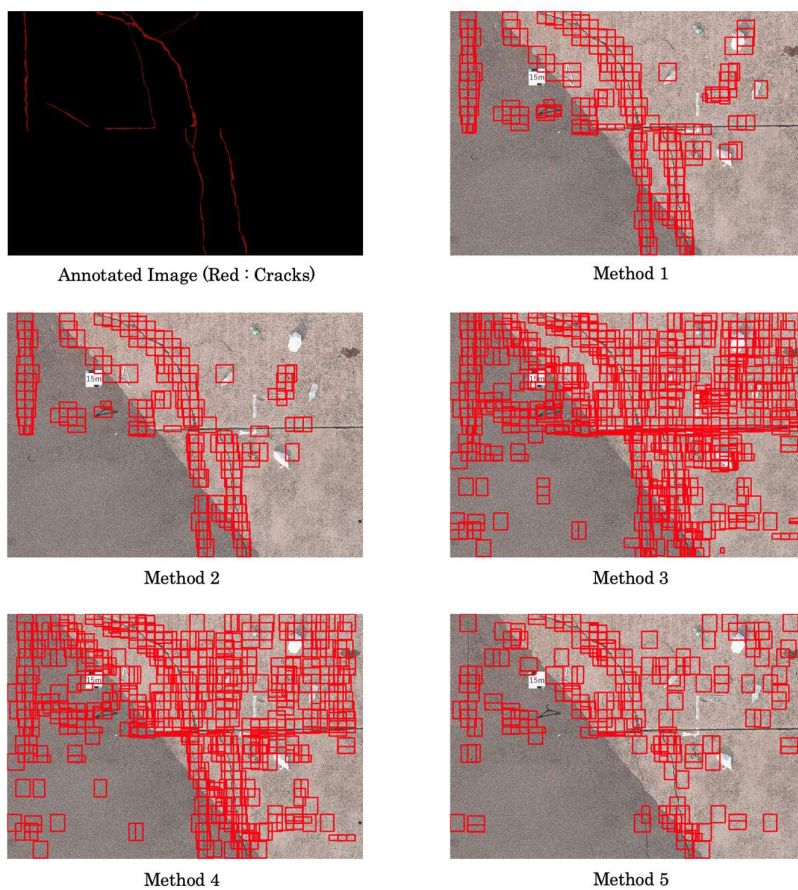
Altitude	Method	Precision	Recall	F1-score	F2-score
15 m	1	0.7000	0.8829	0.7809	0.8390
15 m	2	0.7677	0.8717	0.8164	0.8487
15 m	3	0.3793	0.9624	0.5441	0.7361

15 m	4	0.4093	0.9335	0.5691	0.7432
15 m	5	0.5291	0.7258	0.6120	0.6756
20 m	1	0.7353	0.7676	0.7511	0.7609
20 m	2	0.7983	0.7519	0.7744	0.7607
20 m	3	0.2661	0.9761	0.4182	0.6365
20 m	4	0.2843	0.9565	0.4383	0.6494
20 m	5	0.2909	0.9391	0.4443	0.6497

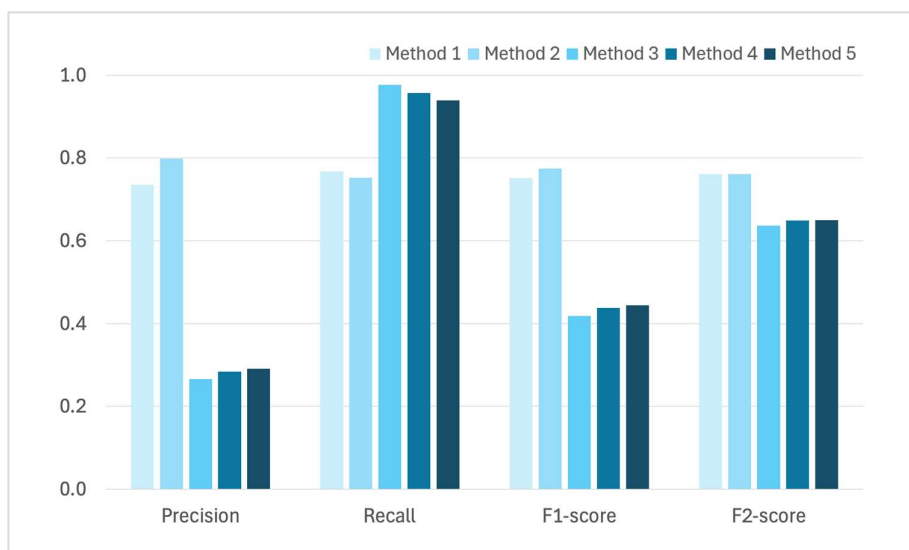
Note: The values correspond to the representative image selected for each altitude condition, following the same evaluation protocol as in the main text.



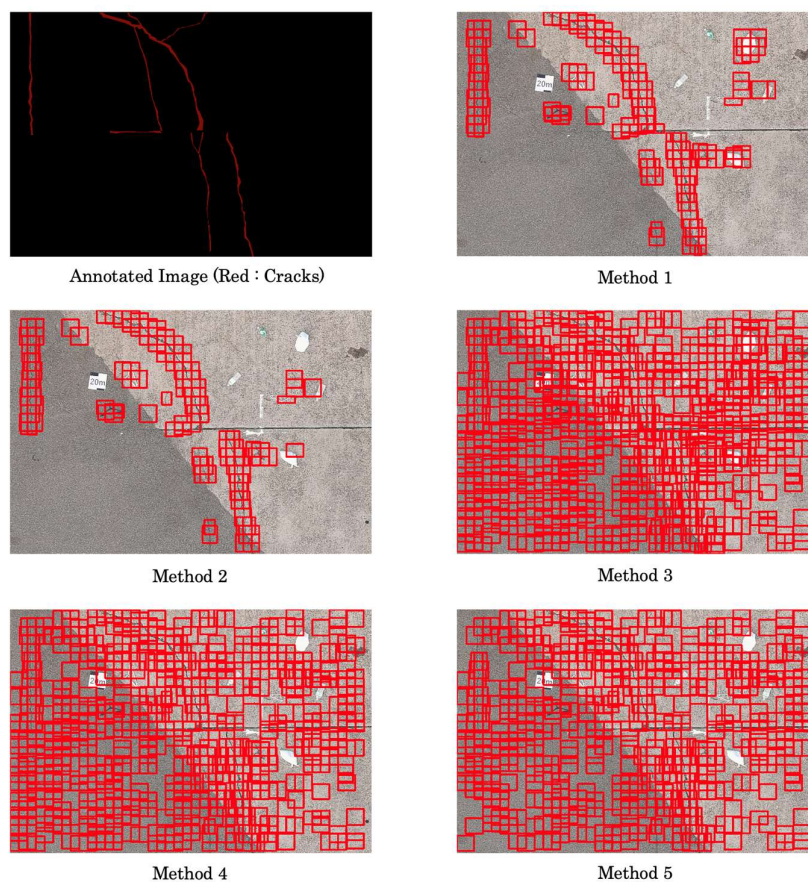
**Figure A1.** Comparison of Precision, Recall, F1-score, and F2-score for the five methods in Verification Area 1 at a flight altitude of 15 m.



**Figure A2.** Visual comparison of crack detection results obtained by the five methods for Verification Area 1 at a flight altitude of 15 m. Red boxes indicate detected crack regions, and the annotation image is shown in red.



**Figure A3.** Comparison of Precision, Recall, F1-score, and F2-score for the five methods in Verification Area 1 at a flight altitude of 20 m.



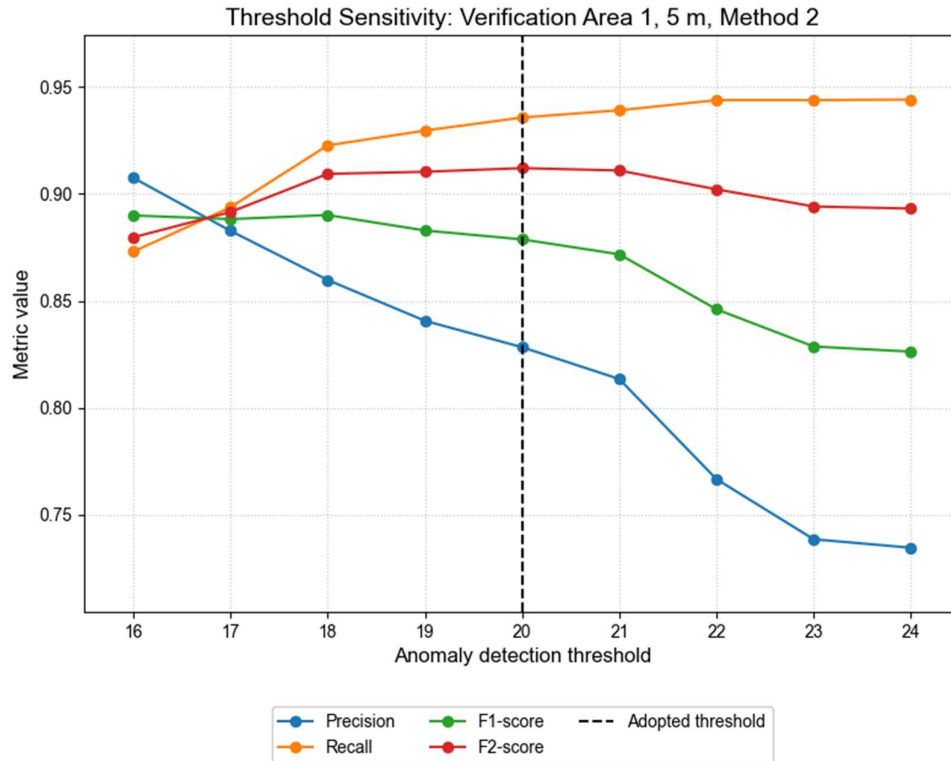
**Figure A4.** Visual comparison of crack detection results obtained by the five methods for Verification Area 1 at a flight altitude of 20 m. Red boxes indicate detected crack regions, and the annotation image is shown in red.

## Appendix C

This appendix presents a supplementary sensitivity analysis of the anomaly detection threshold for representative experimental conditions. For each selected condition, the threshold was varied around the adopted value listed in Appendix A, while keeping the image, method configuration, and evaluation protocol unchanged. The purpose of this analysis was to examine whether the adopted thresholds corresponded to a practical balance between false-positive suppression and crack retention. Because the main evaluation in this study places particular emphasis on Recall in order to minimize missed crack detections in infrastructure inspection, the results were interpreted not only in terms of F1-score but also in terms of F2-score.

### *Area 1, 5 m, Method 2*

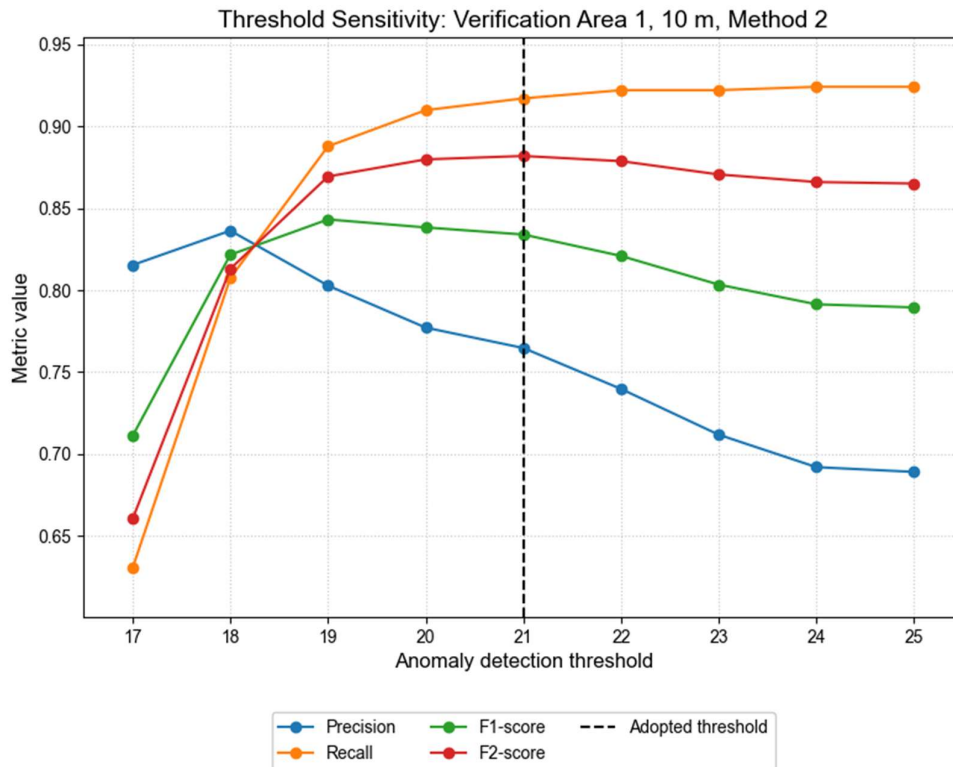
For Verification Area 1 at a flight altitude of 5 m under Method 2, the F1-score reached its maximum at a threshold of 18 (0.890), whereas the F2-score reached its maximum at the adopted threshold of 20 (0.912). As the threshold increased beyond this range, Precision gradually decreased while Recall increased slightly, reflecting the expected trade-off between false-positive suppression and crack retention. These results indicate that the adopted threshold of 20 did not correspond to the single best F1-score, but rather to a practically balanced operating point that favored Recall, consistent with the inspection-oriented role of the framework. The corresponding sensitivity curves are shown in Figure A5.



**Figure A5.** Sensitivity of Precision, Recall, F1-score, and F2-score to the anomaly detection threshold for Verification Area 1 at a flight altitude of 5 m under Method 2. The dashed vertical line indicates the threshold adopted in the main experiment.

#### *Area 1, 10 m, Method 2*

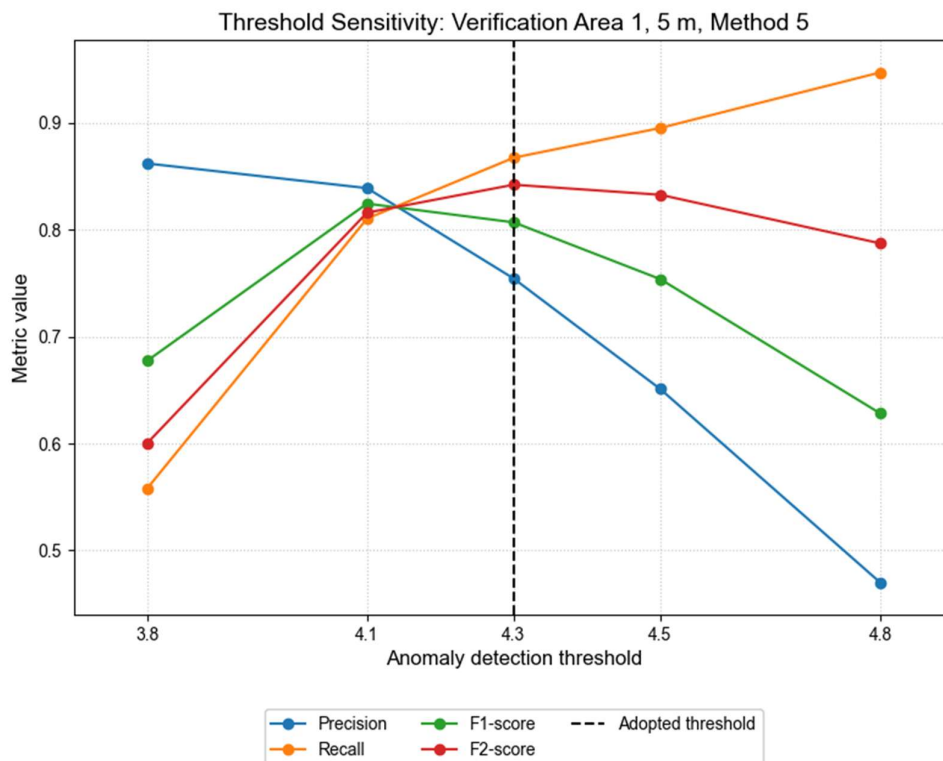
For Verification Area 1 at a flight altitude of 10 m under Method 2, the F1-score reached its maximum at a threshold of 19 (0.843), whereas the F2-score reached its maximum at the adopted threshold of 21 (0.882). In the tested range, lower thresholds tended to improve Precision but reduce Recall, whereas higher thresholds slightly increased Recall at the expense of Precision. Accordingly, the adopted threshold of 21 can be interpreted as a practical operating point selected under a Recall-oriented criterion rather than as an arbitrary threshold chosen without sensitivity confirmation. The corresponding sensitivity curves are shown in Figure A6.



**Figure A6.** Sensitivity of Precision, Recall, F1-score, and F2-score to the anomaly detection threshold for Verification Area 1 at a flight altitude of 10 m under Method 2. The dashed vertical line indicates the threshold adopted in the main experiment.

#### Area 1, 5 m, Method 5

For Verification Area 1 at a flight altitude of 5 m under Method 5, the adopted threshold of 4.3 yielded an F1-score of 0.8012 and an F2-score of 0.8396, consistent with the main experiment. Within the tested range, the balanced performance peaked around a threshold of 4.1, whereas the recall-oriented performance was highest at the adopted threshold of 4.3. As the threshold increased from 4.1 to 4.8, Recall increased while Precision decreased markedly, indicating a clear trade-off between broader crack retention and false-positive suppression. These results suggest that the adopted threshold of 4.3 did not maximize the balanced F1-score, but it provided the most favorable F2-score and therefore represented a practical Recall-oriented operating point under the high-resolution low-threshold setting with Frangi-based preprocessing. The corresponding sensitivity curves are shown in Figure A7.



**Figure A7.** Sensitivity of Precision, Recall, F1-score, and F2-score to the anomaly detection threshold for Verification Area 1 at a flight altitude of 5 m under Method 5. The dashed vertical line indicates the threshold adopted in the main experiment.

Taken together, these sensitivity analyses support the interpretation that the adopted thresholds were not isolated arbitrary values. Rather, they were located in practically stable operating regions near the balance point between false-positive suppression and crack retention across both the standard-threshold configuration and the representative low-threshold Frangi-based configuration, with the final selection guided by the Recall-oriented requirement of infrastructure crack inspection.

## References

1. Global Infrastructure Hub. Global Infrastructure Outlook. Available online: <https://outlook.gihub.org/> (accessed on 15 March 2026).
2. Ministry of Land, Infrastructure, Transport and Tourism, Port and Harbor Bureau. Guidelines for Inspection and Diagnosis of Port Facilities [Part 1: General Provisions]. 2021. Available online: <https://www.mlit.go.jp/kowan/content/001734485.pdf> (accessed on 18 March 2026).
3. Zhang, L.; Gong, L.; Wang, L.; Wang, Z.; Yan, S. A Building Crack Detection UAV System Based on Deep Learning and Linear Active Disturbance Rejection Control Algorithm. *Electronics* **2025**, *14*, 2975, doi:10.3390/electronics14152975.
4. Komi, D.; Yoshida, D.; Kameyama, T. Development of an Automated Crack Detection System for Port Quay Walls Using a Small General-Purpose Drone and Orthophotos. *Sensors* **2025**, *25*, 4325, doi:10.3390/s25144325.
5. Chen, Y.; Li, H.; Zhu, H.; Ren, T.; Cao, Z. Concrete Bridge Crack Detection Using Unmanned Aerial Vehicles and Image Segmentation. *Infrastructures* **2025**, *10*, 161, doi:10.3390/infrastructures10070161.
6. Gou, L.; Liang, Y.; Zhang, X.; Yang, J. RoadNet: A High-Precision Transformer-CNN Framework for Road Defect Detection via UAV-Based Visual Perception. *Drones* **2025**, *9*, 691, doi:10.3390/drones9100691.

7. Ma, Y.; Bao, T.; Li, Y.; Zhao, M.; Wu, Z.; Fan, C. GANFormerNet: A UAV-Based Concrete Crack Segmentation Model for Water-Related Structures Using Vision Transformer and Graph Attention Network. *Adv. Eng. Inform.* **2025**, *68*, 103725, doi:10.1016/j.aei.2025.103725.
8. Alamouri, A.; De Arriba López, V.; Achanccaray Diaz, P.; Backhaus, J.; Gerke, M. High-Resolution Data Capture and Interpretation in Support of Port Infrastructure Maintenance. In Proceedings of the 44. Wissenschaftlich-Technische Jahrestagung der DGPF, Publikationen der DGPF, Band 32, 2024; pp. 269–278.
9. Yu, M.; Chen, W.; Hou, J. A Lightweight Network for Concrete Crack Detection in Complex Environments Based on Cloud-Edge Collaboration. *Expert Syst. Appl.* **2026**, *299*, 130204, doi:10.1016/j.eswa.2025.130204.
10. Xu, J.; Wang, S.; Han, R.; Wu, X.; Zhao, D.; Zeng, X.; Yin, R.; Han, Z.; Liu, Y.; Shu, S. Crack Segmentation and Quantification in Concrete Structures Using a Lightweight YOLO Model Based on Pruning and Knowledge Distillation. *Expert Syst. Appl.* **2025**, *283*, 127834, doi:10.1016/j.eswa.2025.127834.
11. Nagahama, A. Learning and Predicting the Unknown Class Using Evidential Deep Learning. *Sci. Rep.* **2023**, *13*, 14904, doi:10.1038/s41598-023-40649-w.
12. Yang, Q.; Guo, R. An Unsupervised Method for Industrial Image Anomaly Detection with Vision Transformer-Based Autoencoder. *Sensors* **2024**, *24*, 2440, doi:10.3390/s24082440.
13. Cui, Y.; Liu, Z.; Lian, S. A Survey on Unsupervised Anomaly Detection Algorithms for Industrial Images. *IEEE Access* **2023**, *11*, 55297–55315, doi:10.1109/ACCESS.2023.3282993.
14. Wu, J.; Liu, C. VQ-VAE-2 Based Unsupervised Algorithm for Detecting Concrete Structural Apparent Cracks. *Mater. Today Commun.* **2025**, *44*, 112075, doi:10.1016/j.mtcomm.2025.112075.
15. Lee, Y.; Kang, P. AnoViT: Unsupervised Anomaly Detection and Localization With Vision Transformer-Based Encoder-Decoder. *IEEE Access* **2022**, *10*, 46717–46724, doi:10.1109/ACCESS.2022.3171559.
16. Li, H.; Zhang, H.; Zhu, H.; Gao, K.; Liang, H.; Yang, J. Automatic Crack Detection on Concrete and Asphalt Surfaces Using Semantic Segmentation Network with Hierarchical Transformer. *Eng. Struct.* **2024**, *307*, 117903, doi:10.1016/j.engstruct.2024.117903.
17. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, doi:10.48550/arXiv.2010.11929.
18. Wang, C.-Y.; Yeh, I.-H.; Liao, H.-Y.M. You Only Learn One Representation: Unified Network for Multiple Tasks. *arXiv* **2021**, doi:10.48550/arXiv.2105.04206.
19. International Telecommunication Union Radiocommunication Sector (ITU-R). BT.601-7: Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios. 2011. Available online: <https://www.itu.int/rec/R-REC-BT.601> (accessed on 16 March 2026).
20. Frangi, A.F.; Niessen, W.J.; Vincken, K.L.; Viergever, M.A. Multiscale Vessel Enhancement Filtering. In Medical Image Computing and Computer-Assisted Intervention — MICCAI'98; Wells, W.M.; Colchester, A.; Delp, S., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Heidelberg; 1998; Vol. 1496; pp. 130–137. doi:10.1007/BFb0056195.
21. Wang, W.; Liang, Y. Rock Fracture Centerline Extraction Based on Hessian Matrix and Steger Algorithm. *KSII Trans. Internet Inf. Syst.* **2015**, *9*, 5073–5086, doi:10.3837/tiis.2015.12.018.
22. Van Der Walt, S.; Schönberger, J.L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J.D.; Yager, N.; Gouillart, E.; Yu, T. Scikit-Image: Image Processing in Python. *PeerJ* **2014**, *2*, e453, doi:10.7717/peerj.453.
23. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252, doi:10.1007/s11263-015-0816-y.
24. Dorafshan, S.; Thomas, R.J.; Maguire, M. SDNET2018: An Annotated Image Dataset for Non-Contact Concrete Crack Detection Using Deep Convolutional Neural Networks. *Data Brief* **2018**, *21*, 1664–1668, doi:10.1016/j.dib.2018.11.015.
25. Wada, K. Labelme: Image Polygonal Annotation with Python, version 4.6.0; Zenodo, 2021.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s)

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.