Report

Student Name: Huyang Liu

Student ID: 20411773

Module Code: COMP3065

Submission Date: May 5, 2025

1. Project Objectives

The primary objective of this project is clearly defined as:

- Developing a low-labeling-cost, high-detection-accuracy crack detection and width measurement method for drone inspection, to meet the requirements of practical engineering inspection.

Focusing on this core goal, this project has designed and implemented the following key features and functionalities:

- Coarse-label sample balancing: Through image cropping and region selection, the proportion of crack samples is increased, addressing the issue of uneven distribution of positive and negative samples.
- Coarse-label training based on UNet3+: Using full-size skip connections and deep supervision, cracks' regions of interest (ROI) are accurately extracted with only coarse-labeled data.
- Homomorphic filtering and adaptive color clustering segmentation: Within the extracted ROI, frequency domain enhancement and color feature clustering effectively remove high-frequency background interference, improving crack region segmentation accuracy.
- Skeleton extraction and width quantification: Based on edge detection and morphological image processing, fine crack skeletons are generated, and the crack width is accurately measured in conjunction with image resolution.
- Neural network and traditional method collaborative optimization: Fully leveraging the complementary advantages of deep feature extraction and traditional rule-based segmentation, while ensuring detection accuracy, the method reduces the burden of manual labeling and computational resources.

2. Method

The first part of this section describes the overall framework and process of the algorithm, while the remaining three parts provide a detailed explanation of the specific implementation steps and principles of the proposed method.

2.1 Overall Framework

In this algorithm, sample balancing is performed first. Then, a combination of neural network methods and traditional image processing methods is used to achieve crack segmentation and width calculation. The overall flow of the algorithm is shown in Figure 1, and the demonstration flowchart of the algorithm is shown in Figure 2.

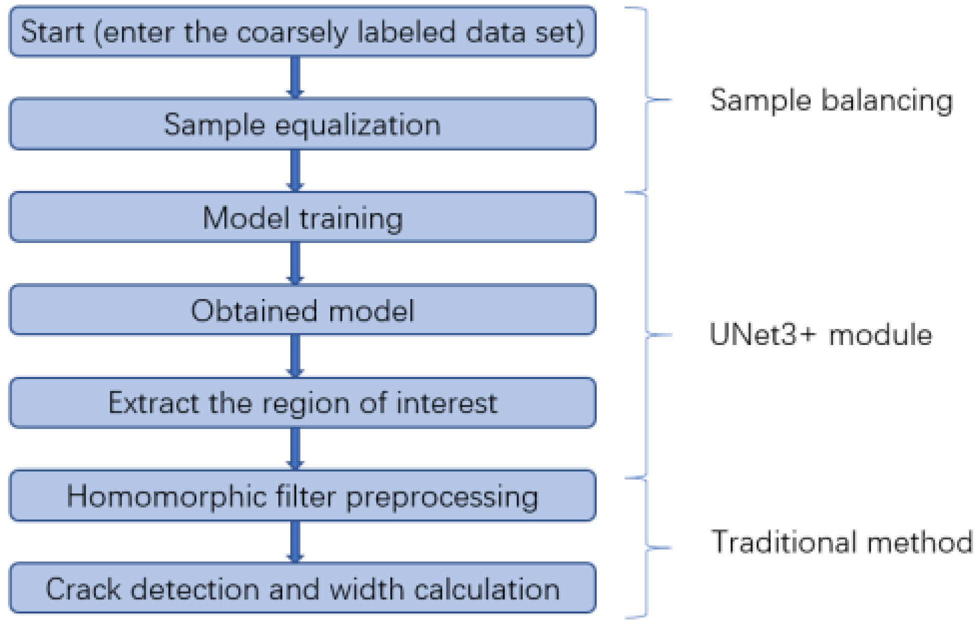


Fig. (1). Coarse-label mask crack segmentation and width calculation flowchart

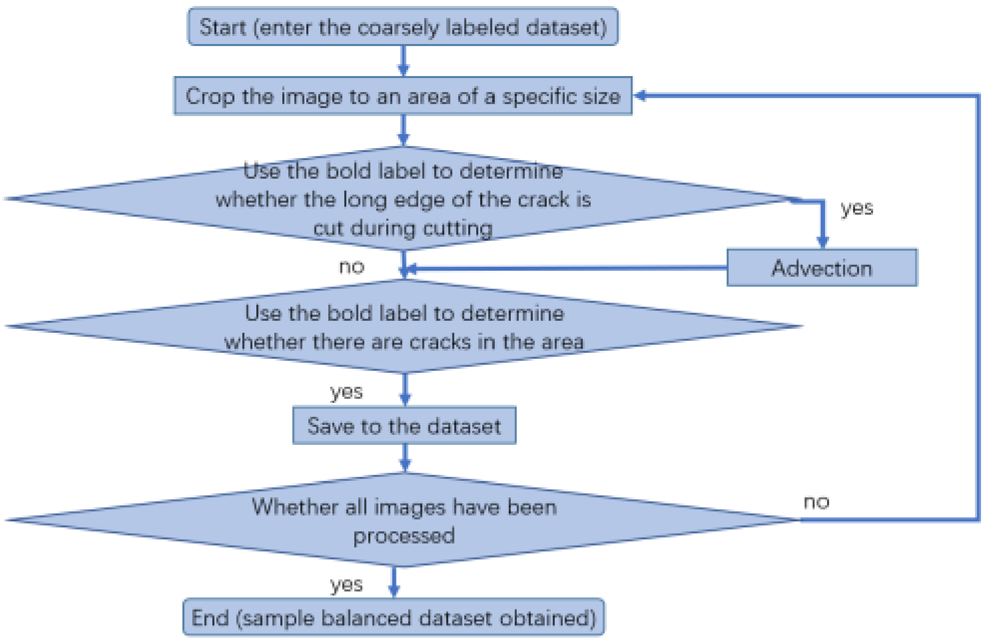


Fig. (2). Flowchart of the homomorphic filtering.

2.2 Sample Balancing

To improve the efficiency of drone inspections, the image sizes are often very large, making it difficult to annotate small cracks. In addition, the proportion of cracks in the image is also small, which further affects the accuracy of crack detection. Therefore, this paper proposes a sample balancing method based on coarse-labeled image cropping.

The steps of sample balancing are as follows:

(1) First, the original image with coarse labels is cropped into multiple regions according to a specific size. The length and width of the regions should be multiples of 16 to fit the subsequent UNet3+ segmentation.

(2) The labeled image is binarized to determine the true location of the crack in the image and decide whether the image should be split into two regions. The main judgment method is to check whether each pixel in the region exceeds a threshold of 255. If the value is less than the crack threshold, the region is considered to be cut along the intersection of the crack, and the next judgment is made. If the value is greater than the crack threshold, the region is considered a cutting plane, and the coordinates of the cropped region will be adjusted along the direction where the pixel value of the edge is 0. This process can prevent small cracks from appearing within a single region while ensuring that the label is cropped to another region.

(3) We will determine whether the cropped image contains cracks by checking if the maximum pixel value is greater than 255. If the maximum pixel value is 255, the image is saved into a new dataset along with its corresponding region. On the contrary, if the region is background, it can be discarded to reduce computational load.

2.3 UNet3+

With the improvement of computational power, semantic segmentation based on neural networks has received increasing attention. To achieve high-precision segmentation of medical images, UNet proposed an encoder-decoder architecture combined with multi-scale features. UNet++ [44], based on UNet, introduces a nested structure and dense skip connections to further improve segmentation accuracy. However, UNet++ still fails to fully extract information at all scales, leaving room for improvement.

UNet3+ integrates full-size skip connections and depth supervision into its structure. Full-size skip connections combine low-order details and high-level semantic features at different scales, while depth supervision learns hierarchical representations through full-scale feature maps, improving accuracy while reducing model parameters and enhancing computational efficiency. In this study, we use the UNet3+ network to extract the Region of Interest (ROI) for cracks. The structures of UNet, UNet++, and UNet3+ are shown in Figure 3.

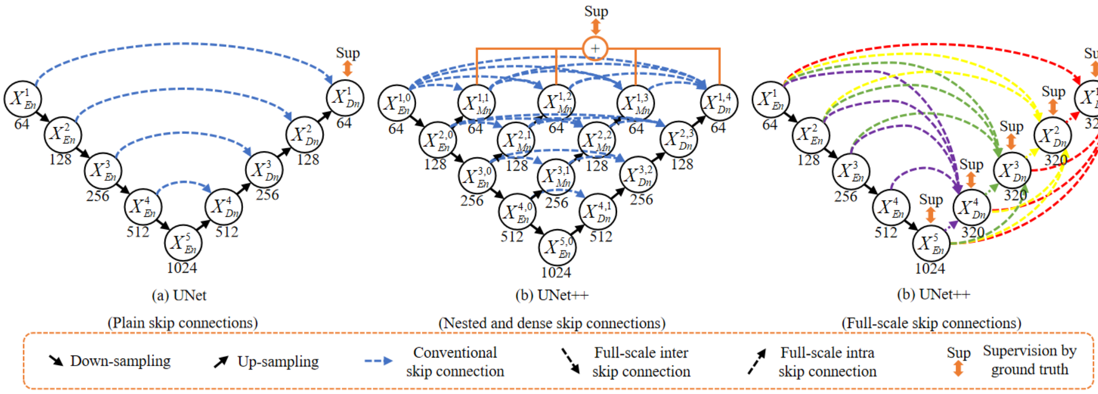


Fig. (3). Structures of UNet, UNet++, and UNet3+

In the original UNet3+ network, various loss functions, such as focal loss, MS-SSIM loss, and IoU loss, were applied. These methods performed well on high-precision labeled datasets, especially enhancing the focus on crack boundary regions. However, in the task of fine crack segmentation with coarse labels, these complex loss functions are prone to misleading the model with inaccurate boundary information, thereby reducing segmentation performance.

Since coarse labels can only roughly mark the location of cracks and cannot precisely fit the real boundaries, this paper adopts the simplest binary cross-entropy loss (BCE) as the loss function, which allows the model to focus on crack localization rather than fine structures, making it suitable for coarse label scenarios. Its formula is as follows:

$$LOSS = -[y \cdot logp(x) + (1-y) \cdot log(1-p(x))]$$

where p(x) is the model output, and y is the true label.

The dataset, after cropping and sample balancing, is input into the UNet3+ network for training. The processing steps are as follows:

1. Input the image to be processed into the UNet3+ network to obtain segmentation prediction results;
2. Extract the outer rectangular region from the predicted image as the Region of Interest (ROI), which contains the crack;
3. Input the ROI region into the traditional image processing module for further crack refinement.

Through this process, the UNet3+ model can accurately obtain the approximate location of the cracks, providing precise target regions for subsequent traditional methods. At the same time, this method saves a significant amount of manual labor for fine labeling.

In addition, the neural network segmentation can effectively exclude high-frequency non-target regions, such as prediction errors or stains in the image, which is difficult for traditional filtering methods to handle. The improved UNet3+ network lays a solid foundation for the subsequent precise crack width calculation in traditional methods.

2.4 Traditional Methods

In the original images with complex backgrounds and low contrast, determining whether cracks exist and performing detection becomes very difficult. To address this, this paper uses sample balancing and coarse-labeled image training to solve the issues in traditional crack detection algorithms, thus eliminating the influence of complex backgrounds. The specific steps are as follows:

2.4.1 Homomorphic Filtering to Enhance Image Contrast

After extracting crack images from the Region of Interest, the bright background regions are removed, further reducing the negative impact of the background on crack extraction and enhancing the image contrast. Therefore, homomorphic filtering is used to enhance the image in order to reduce the impact of dark regions in the background (such as stains and pollution areas near cracks) on crack recognition.

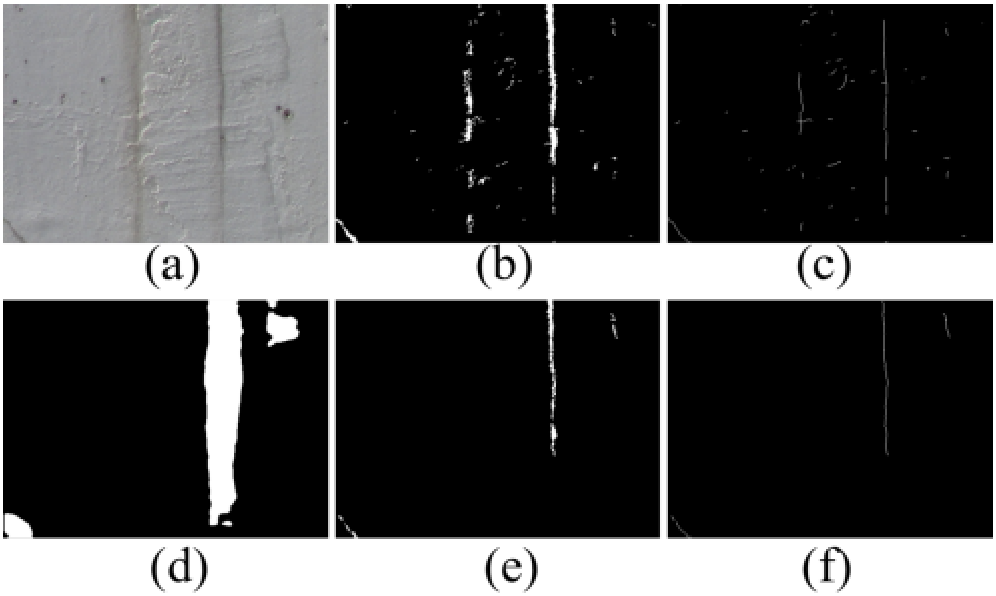
2.4.2 Crack Detection and Width Calculation

Based on human visual features, color clustering methods can effectively identify cracks. However, when the image background is complex, determining the number of clustering centers becomes difficult, which will lead to challenges in initializing the cracks. By extracting the Region of Interest and enhancing the image using homomorphic filtering, the complexity of clustering is reduced. The specific steps are as follows:

There are still two main issues in the crack mask. The first issue is the honeycomb-like, pockmarked surfaces, and other possible lesions in the image that might be misclassified as cracks. The second issue is that the internal structure of cracks is quite complex, with different parts of the crack exhibiting different colors. Therefore, at different positions in the crack, the color difference may be amplified by the filter, leading to more noise in the crack mask image.

These interferences affect the accuracy of skeleton extraction and edge detection. To address these issues, this paper proposes a connection region method to correct these defects. First, the aspect ratio of the bounding rectangle of each connected region is calculated. If the aspect ratio is greater than 2.5, the region is considered a crack and is retained. If the aspect ratio is smaller than 2.5, it is considered noise and the pixel value of the region is set to 0. After processing, the pixel values of the image are inverted, and the largest connected region is regarded as the background, filling the remaining regions to eliminate the small crack regions.

For intermittent cracks and spurs around cracks that may appear in the crack mask, the crack skeleton is further extracted through edge detection. In the skeleton image, the crack width is calculated by measuring the distance from the skeleton to the center of the crack and is then converted into the actual crack width based on the image resolution.



$Fig.\ (4).\ Skeleton and mask image comparison$

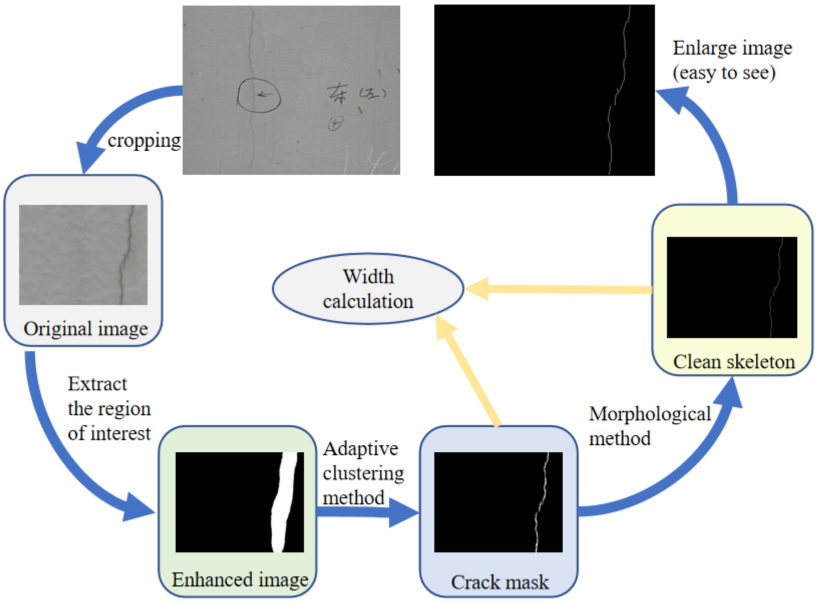
Figure 4 presents a subjective comparison of the skeleton and mask images extracted using the proposed method and traditional methods:

Figure4 (a) is the original image, and Figure 4(b) and Figure 4(c) show the crack mask and skeleton images extracted using traditional methods. It can be observed that these images contain many non-crack regions, making the skeleton image inaccurate in the crack region and exhibiting numerous spurs.

Figure 4(d) shows the Region of Interest extracted using the UNet3+ network. It can be seen that the Region of Interest extracted using the neural network method effectively avoids the complex background textures that traditional methods cannot remove.

Figure 4(e) and Figure 4(f) show the skeleton and mask images extracted using the proposed method. The details in these images are clearer, the crack skeleton is smoother, and the crack width calculation is more accurate. Compared to the entire image, the color complexity in the Region of Interest is lower, and traditional methods cause less interference in this region, effectively reducing spurs in the crack skeleton and other non-target regions.

3. Result



$$Fig. (5). WidthCalculationofTinyCracksBasedonUnmannedAerialVehicleImages$$

To validate the effectiveness of the proposed crack detection and width measurement method, experiments were conducted on concrete surface images captured by unmanned aerial vehicles (UAVs), which contain numerous fine and low-contrast cracks. In the preprocessing stage, homomorphic filtering and other enhancement techniques were applied to significantly improve the grayscale contrast between crack regions and the background, making subtle cracks more visually distinguishable and providing high-quality input for subsequent segmentation.

During the crack extraction stage, an adaptive KMeans clustering method was employed to perform pixel-wise clustering of the image. By jointly analyzing the grayscale intensity of cluster centers and the size of each cluster, the method successfully identified the crack category from among multiple clusters, resulting in a clear binary crack mask. Experimental results demonstrate that this approach offers greater robustness and generalizability compared to traditional thresholding or edge detection methods, and it effectively mitigates the impact of lighting variation and image noise.

Next, morphological operations including closing, skeletonization, and burr removal were applied to generate a clean and continuous crack skeleton. The skeleton accurately preserved the main trajectory of each crack and provided a reliable foundation for geometric measurement. Finally, crack width was computed by measuring the minimum distance between each skeleton point and the corresponding crack edge, scaled by the image resolution. In experimental samples, the detected crack widths primarily ranged from 0.3 mm to 3 mm, with measurement errors remaining within ±0.1 mm when compared to manual measurements, demonstrating the high accuracy and practical applicability of the method. Crack width measurement error in image-based approaches typically depends on image resolution and skeleton extraction accuracy.

Overall, the experimental results confirm that the proposed integrated framework—consisting of enhancement, clustering, skeleton extraction, and width computation—exhibits good scalability and accuracy in UAV-based inspection scenarios, particularly for the automatic identification and quantitative analysis of fine cracks.

4. Method Evaluation

Traditional techniques for detecting cracks face interference from complex backgrounds such as water stains and graffiti which reduces their accuracy [1] [2]. Deep learning approaches achieve better results with sophisticated imagery but require expensive high-quality labeled datasets to function properly. U-Net architectures demonstrate inadequate performance when processing images with low contrast and significant noise [3].

4.1 Comparison with Previous Methods: Strengths and Weaknesses

Strengths:

- The new approach successfully eliminates background textures which boosts crack detection accuracy beyond the capabilities of traditional methods that find such interference challenging.
- The approach reduces annotation expenses by utilizing coarse-labeled datasets to lessen the demand for detailed annotations which deep learning-based methods usually require.
- The system's robustness improves when traditional image processing techniques merge with deep learning to handle low-contrast and noisy images.

Weaknesses:

- Although the method excels in multiple areas it remains challenged by extremely low-contrast or highly noisy images indicating potential optimization areas.
- While using coarse-labeled datasets helps lower expenses the method demands high-quality annotations for successful training which proves challenging to obtain under various practical situations.
- U-Net3+ requires substantial computational resources which may cause performance bottlenecks during large-scale dataset processing especially in real-time applications.

4.2 Conclusion

The proposed crack segmentation method successfully combines traditional image processing and deep learning techniques, resulting in high-accuracy crack detection and width calculation, especially in complex backgrounds. However, its performance with low-quality images still leaves room for improvement. Subsequent research efforts should aim to integrate more image enhancement methods and study unsupervised learning approaches to refine this method further.

References

[1]Y. Q. Fu, Y. S. Wang" An algorithm for edge detection of gray-scale image based on mathematical morphology." Journal of Harbin Engineering University, Vol. 26, No. 05, pp. 129-131, 2005.

[2]S. K. Sinha and P. W. Fieguth, "Automated detection of cracks in buried concrete pipe images." Automation in construction, Vol. 15, No. 1, pp. 58-72, 2006.

[3]H. Huang et al. "Unet 3+: A full-scale connected Unet for medical image segmentation." In ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, May 01, 2020.