

Article

Not peer-reviewed version

Edge Enhanced CrackNet for Underwater Crack Detection of Concrete Dams

[Xiaobian Wu](#), [Weibo Zhang](#), [Guangze Shen](#), [Jinbao Sheng](#)*

Posted Date: 19 August 2025

doi: 10.20944/preprints202508.1410.v1

Keywords: dam safety; small object detection; underwater crack detection



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.

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

Edge Enhanced CrackNet for Underwater Crack Detection of Concrete Dams

Xiaobian Wu ^{1,2}, Weibo Zhang ¹, Guangze Shen ^{1,3} and Jinbao Sheng ^{1,3,*}

¹ Nanjing Hydraulic Research Institute, Nanjing 210029, China

² NHRI R&D Tech Group Co., Ltd., Nanjing 210029, China

³ Dam Safety Management Center of the Ministry of Water Resources, Nanjing 210029, China

* Correspondence: jbshe@nhri.cn; Tel.: +8602585828600

Featured Application

This study provides an underwater image enhancement method, which effectively improves the detection performance of crack recognition algorithm under complex flow field and low illumination underwater conditions, and provides a new solution for concrete crack detection algorithm on dam surface under deep water conditions.

Abstract

Underwater crack detection in dam structures is of significant engineering importance and scientific value for ensuring the structural safety, assessing operational conditions, and preventing potential disasters. Traditional crack detection methods face various limitations when applied in underwater environments, particularly in high dam underwater environments where image quality is influenced by factors such as water flow disturbances, light diffraction effects, and low contrast, making it difficult for conventional methods to accurately extract crack features. This paper proposes a dual-stage underwater crack detection method based on Cycle-GAN and YOLOv11, called Edge Enhanced Underwater CrackNet (E²UCN), to overcome the limitations of existing image enhancement methods in retaining crack details and improving detection accuracy. First, underwater concrete crack images were collected using an underwater remotely operated vehicle (ROV), simulating various complex underwater environments to construct a test dataset. Then, an improved Cycle-GAN image style transfer method was used to enhance the underwater images, emphasizing crack edges and high-frequency details, avoiding the issues of blurred crack edges and detail loss in existing underwater image enhancement methods. Subsequently, the YOLOv11 model was employed to perform object detection on the enhanced underwater crack images, effectively extracting crack features and achieving high-precision crack detection. Experimental results show that the proposed method significantly outperforms traditional methods in terms of crack detection accuracy, edge clarity, and adaptability to complex backgrounds, effectively improving underwater crack detection accuracy and providing a feasible technological solution for intelligent inspection of high dam underwater cracks.

Keywords: dam safety; small object detection; underwater crack detection

1. Introduction

Cracks are one of the most common structural defects in underwater engineering facilities during long-term operation. They can lead to reduced structural strength, water leakage, and even catastrophic consequences, posing a serious threat to engineering safety. Therefore, early detection and precise evaluation of cracks are of critical importance for ensuring the stability of engineering structures. Currently, China has the largest number of high dams, with 588 dams over 70 meters, 233 dams over 100 meters, and 23 dams over 200 meters. High dam safety is a major challenge for national

water safety and public safety. Some cracks exist underwater, and mature methods for above-water detection cannot be directly applied. However, traditional crack detection methods mainly rely on draining reservoirs to clear interference factors for obtaining clear observations, combined with manual inspections for underwater structural defects[1,2]. This approach is not suitable for high dams and large reservoirs, and it consumes considerable time, economic costs, and human resources, potentially having a negative impact on the ecological environment in certain cases. Modern detection methods primarily rely on divers for underwater visual inspections or portable detection equipment, with saturation diving required for high dams, which is both dangerous and costly. Moreover, divers lack professional knowledge of hydraulic engineering, making the reliability of detection results lower. With the advancements in ROV and underwater imaging technologies, the possibility of intelligent inspection for underwater cracks has emerged. However, traditional image processing methods perform poorly in complex underwater environments, and their detection accuracy and adaptability are severely limited[3]. Therefore, research on crack defect detection models in complex underwater environments is of great significance.

One of the core challenges in underwater crack detection tasks is the quality of underwater images. Due to the influence of various disturbance fields in the underwater environment (such as flow fields and hydrodynamic fields), underwater images often exhibit significant motion blur and light diffraction effects. Additionally, the absorption and scattering of light in water result in reduced image contrast, color distortion, and detail blurring, which severely affects the performance of crack detection models[4]. Therefore, underwater image enhancement techniques have become a fundamental research direction for improving the accuracy of underwater crack detection. Traditional underwater image data enhancement algorithms, such as histogram equalization[5] and homomorphic filtering[6], are relatively simple, with poor robustness, making them unsuitable for image enhancement tasks in high-dam underwater multi-field disturbance environments. In recent years, with the rapid development of machine learning[7] and deep learning technologies[8], neural network-based image enhancement methods have made significant progress in the field of image processing. For example, Wang Yue et al.[9] proposed a multi-scale attention and contrast learning-based underwater image enhancement algorithm, which effectively extracted multi-level image features by combining an encoder-decoder structure with multi-scale channel-pixel attention modules. The improvement in PSNR and SSIM metrics was 4.4% and 2.8%, respectively, significantly improving image clarity. Additionally, Du Feiyu et al.[10] proposed a domain-adaptive underwater image enhancement method, combining convolutional neural networks with a multi-head attention mechanism and adversarial learning to achieve image enhancement under unsupervised conditions. Although these methods have achieved good results in restoring visual quality, they focus more on enhancing the overall visual effect of the image rather than preserving and enhancing crack details, such as crack boundaries, which may lead to blurred crack edges or feature loss, thereby affecting subsequent crack detection accuracy. Furthermore, existing underwater image enhancement models require paired images for training[11,12], but paired underwater crack images are difficult to obtain due to the use of ROVs for image collection, making it challenging to create paired image datasets for model training.

In the field of underwater crack detection, traditional crack detection methods, such as Markov random fields[13] and Sobel operators[14], rely on edge detection algorithms in image processing techniques. Although these methods perform well in simple scenarios, they are highly sensitive to noise and are easily affected by background interference and optical artifacts in complex underwater environments, making it difficult to effectively extract crack features. In recent years, with the continuous development of object detection technologies, deep learning-based models have achieved significant progress in underwater crack detection. For instance, Shi et al. proposed a method called CrackYOLO[15], based on the YOLOv5 model, which introduces a feature fusion module, Res2C3 feature extraction module, and BCAttn attention mechanism, significantly improving crack detection performance. It achieved 94.3% mAP and 151 FPS detection speed in underwater crack detection tasks. Moreover, Mao Yingchi et al.[16]. proposed a multi-task enhanced crack image detection

method (ME-Faster R-CNN) based on Faster R-CNN, which improved the regional proposal network (RPN) and introduced a multi-source adaptive balancing TrAdaBoost method, effectively improving the detection capability for multi-target and small-target cracks. In experiments, it achieved 82.52% average intersection over union (IoU) and 80.02% average precision (mAP), improving by 1.06% and 1.56% respectively compared to traditional Faster R-CNN methods. Additionally, Huang et al[17]. tackled the inherent limitations of redundant architectural components and deficient multi-scale feature extraction in the canonical YOLOv5 framework by introducing an enhanced model that synergistically integrates attention mechanisms with the Complete-IoU (CIoU) loss[19], thereby substantially elevating real-time detection accuracy. Concurrently, the refined YOLOv8-derived architecture exhibits markedly superior robustness and detection fidelity when confronted with the severe visual degradations characteristic of the underwater domain, thereby advancing the state-of-the-art in marine object detection[20]. These research outcomes show that deep learning-based object detection methods can effectively handle complex underwater crack detection tasks and achieve a good balance between detection speed and accuracy. In summary, deep learning-based object detection models have performed well in crack detection tasks in various fields. However, due to the lack of underwater crack data, research on dam underwater crack detection based on object detection algorithms is relatively limited[18].

To address the deficiencies in existing research on underwater image enhancement and crack detection, this paper proposes a dual-stage underwater crack detection method based on CycleGAN[21] and YOLOv11[22], called Edge Enhanced Underwater CrackNet (E²UCN), to solve the issue of insufficient attention to crack feature details in complex underwater environments and achieve high-precision crack detection. First, during the underwater crack image collection process, we used the P200 underwater remotely operated vehicle (ROV) to capture artificial crack images in an underwater concrete tank, simulating various complex underwater scenarios, including flow field disturbances, optical diffraction, and low contrast environments to simulate the real environment of high dams. Next, in the image enhancement stage, we designed a CycleGAN-based underwater image style transfer method, named the CycleGAN-Based Underwater Image Enhancement Model (CGBUIE), to improve underwater image quality and highlight crack detail features. Unlike traditional CycleGAN, CGBUIE introduces Sobel operators[23] and high-frequency transformations[24,25] in the loss function to constrain the edge information and high-frequency detail retention in the generated image, preventing crack edges from becoming blurred or details from being lost. The Sobel operator extracts prominent edge information from the image, while high-frequency transformations enhance crack texture features, enabling the enhanced image to achieve both visual style transfer to an above-water environment and crack boundary and detail retention at the feature level[26–28]. Finally, in the crack detection stage, we first trained the YOLOv11 model on the above-water concrete crack dataset to learn key crack features and prominent edge features, then applied the trained model to the enhanced underwater crack images generated by CGBUIE for accurate underwater crack detection. During the detection process, YOLOv11, with its optimized network architecture and multi-scale feature extraction capability, is able to better capture the subtle features and irregular edges of cracks, especially in cases where cracks are small and complex in shape[29]. The enhanced underwater crack images not only significantly improved the visibility of cracks but also provided high-quality input for the model, enabling fast and accurate crack localization and classification in complex underwater environments. Experimental results show that the proposed method performs well in terms of crack edge clarity, object localization accuracy, and adaptability to complex backgrounds.

2. Proposed Method

The E²UCN framework, the CGBUIE model, and the YOLOv11 model used in this study will be described in detail in this section. Specifically, the E²UCN model is shown in Figure 1.

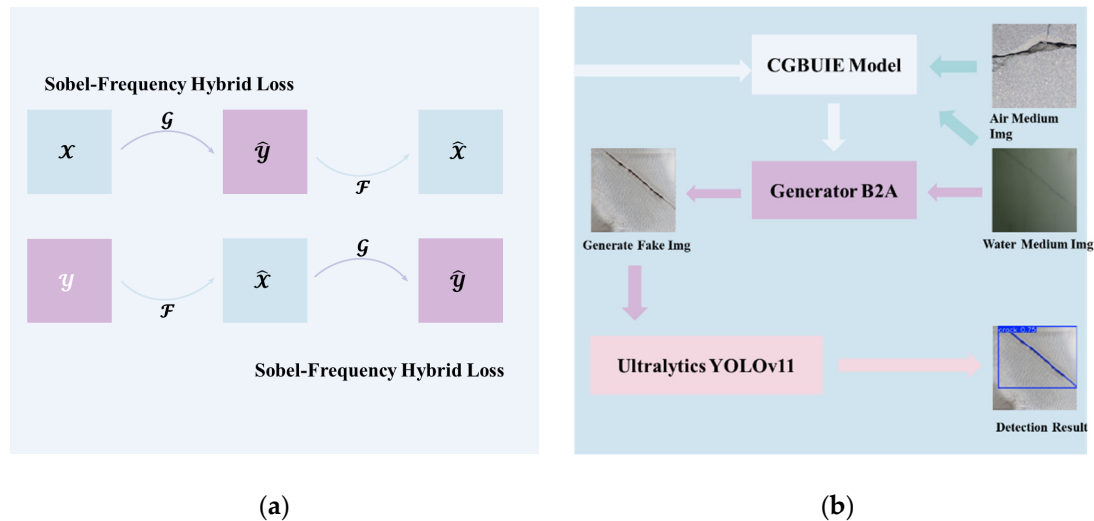


Figure 1. The architecture of the proposed model, (a) structural diagram of E^2UCN ; (b) structural diagram of CGBUIE.

2.1. CycleGAN Based Underwater Image Enhancement Model

Based on the style transfer functionality of CycleGAN, this paper combines Sobel operators and high-frequency filtering to perform style transfer between underwater crack images and above-water crack images, aiming to enhance the images. The core idea of CycleGAN is to map between different domains through unsupervised learning without paired training data. This enables its widespread application in underwater crack detection, especially when large annotated datasets are unavailable. CycleGAN achieves this goal by introducing two generators and two discriminators. The generator is used to generate images similar to the target style, while the discriminator judges the difference between the generated image and the real image, thereby guiding the generator to optimize its generation effect.

The standard CycleGAN model consists of two generators and two discriminators. Generator G_x maps source domain images to target domain images. Meanwhile, generator G_y maps target domain images back to the source domain. Discriminators D_x and D_y are used to distinguish generated images from real images, thus guiding the generator to optimize its style transfer effect. The goal of CycleGAN is to minimize the difference between generated images and real images while ensuring that the generated image can restore the original image after being mapped back. This process is achieved by introducing cyclic consistency loss based on the adversarial loss of the original GAN network.

Adversarial loss is used to train the generator to produce realistic images, forcing the generated images to "fool" the discriminator. For generator G_x , the goal is to minimize the following loss function:

$$L_{GAN}(G_x, D_y, x, y) = \mathbb{E}_{\{y \sim p_{data}(y)\}} [\log D_y(y)] + \mathbb{E}_{\{x \sim p_{data}(x)\}} [\log(1 - D_y(G_x(x)))] \quad (1)$$

Where x and y are real images from the source and target domains, and $G_x(x)$ is the image generated by generator G_x , and D_y is the discriminator.

Moreover, to ensure that the generated image can still restore the original image after being mapped back, CycleGAN introduces cyclic consistency loss. For generators G_x and G_y , their goals are defined as:

$$L_{Cycle}(G_x, D_y, x, y) = \mathbb{E}_{\{y \sim p_{data}(x)\}} [G_y(G_x(x)) - x_1] + \mathbb{E}_{\{x \sim p_{data}(y)\}} [G_x(G_y(y)) - y_1] \quad (2)$$

This loss uses the L1 norm to measure the difference between the generated image and the original image, forcing the generated image to maintain the structural features of the source image. Therefore, the total loss function of the original CycleGAN is:

$$L_{CycleGAN} = L_{GAN}(G_x, D_y, x, y) + L_{GAN}(G_y, D_x, y, x) + \lambda_{cycle} L_{Cycle}(G_x, D_y, x, y), \quad (3)$$

Where λ_{cycle} is the weight balancing adversarial loss and cyclic consistency loss.

However, experiments showed that the style transfer images generated by the original CycleGAN, focusing mainly on style transfer (e.g., color, contrast of the above-water images), fail to retain the edge information and texture details of the image. To enhance the edge details of the style-transferred images, this paper designs the E²UCN model.

Specifically, based on the original CycleGAN, Sobel operators are introduced and constructed into Sobel loss. The Sobel operator extracts edge information from the image by computing the gradients and is effective in retaining the image's edge features. The basic principle of the Sobel operator is applied here and further improved to construct Sobel loss. First, the Sobel operator is applied to compute the gradients of the input image, obtaining the gradients in the horizontal x and vertical y directions. Specifically, the Sobel operator used in this paper is:

$$Sobel_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, Sobel_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \quad (4)$$

Where $Sobel_x$ and $Sobel_y$ are the convolution kernels of the Sobel filter in the x and y directions, respectively. The gradients calculated by the two operators can be represented as:

$$\nabla_{x,y}(i, j) = \sum_{m=0}^2 \sum_{n=0}^2 I(i+m, j+n) \cdot sobel_{x,y}(m, n), \quad (5)$$

Where i and j represent the pixel positions in the image, and $\nabla_x(\cdot)$ and $\nabla_y(\cdot)$ represent the gradients in the x and y directions, respectively. Then, the gradient magnitude of the image is calculated to represent the edge information of the restored image. The specific formula is as follows:

$$\nabla(i, j) = \sqrt{\nabla_x(i, j)^2 + \nabla_y(i, j)^2 + \varepsilon}, \quad (6)$$

Where ε is a small constant to avoid numerical instability when the gradient is zero.

The Sobel loss proposed in this paper measures the gradient magnitude difference between the generated image and the real image using the L1 norm, comparing the edge differences in both the x and y directions between the generated image and the original image to simulate the edge enhancement effect of the Sobel operator. The specific formula is as follows:

$$L_{sobel} = \mathbb{E}_{x \sim P_{data}(x)} [\nabla G_x(x) - \nabla x_1], \quad (7)$$

Where $G_x(x)$ is the generated image, x is the real image, and $\nabla G_x(x)$ and ∇x are the gradients of the generated image and the real image, respectively.

By constructing Sobel loss, CycleGAN is forced to preserve the edge information of the real image during the image generation process, better retaining the crack edge features in the image and improving the accuracy of subsequent object detection tasks. However, while the generated image with Sobel loss retains the edge features of the cracks, there is still a significant amount of blurring and loss of the internal texture information of the cracks. Therefore, to further enhance the texture information in the image, and enable the subsequent object detection model to recognize and extract cracks from the image, this paper constructs a high-frequency loss function. Specifically, high-frequency information corresponds to the detailed parts of the image, such as textures and edges. By

using Fourier transform and other high-frequency transformations, the image can be converted from the spatial domain to the frequency domain, allowing for better extraction and retention of high-frequency information. Based on this, in the CGBUIE model, a high-frequency loss function is used to compare the high-frequency information between the real image and the generated image, achieving the goal of preserving and enhancing the texture information.

Specifically, the original image x and the generated image G_x are first transformed from the spatial domain to the frequency domain using the fast Fourier transform. Assuming the discrete image signal is Ω , the specific formula is as follows:

$$\mathcal{F}(\Omega(h, w)) = \sum_{n=0}^{N-1} \Omega[n] e^{-i2\pi \frac{hwn}{N}}, \quad (8)$$

Where N is the signal length, n is the time index, and k is the frequency index. After applying fast Fourier transform, the original image and the generated image are represented as $\mathcal{F}_{real}(h, w)$ and $\mathcal{F}_{fake}(h, w)$, respectively. High-frequency components are extracted by applying a high-frequency filter to remove low-frequency parts of the frequency domain signal, and the absolute value is taken to retain the magnitude of the high-frequency components, i.e., the texture details in the original image. The specific rule is as follows:

$$H(h, w) = \begin{cases} |\mathcal{F}(\Omega(h, w))|, & \text{if } h < \text{cutoff or } w < \text{cutoff} \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

Where h and w represent the height and width in the frequency domain, i.e., the dimensions of the frequency domain, and cutoff is a manually set threshold for determining the high-frequency filtering threshold. Finally, the L1 norm loss is calculated between the original image and the generated image based on the magnitude of the high-frequency components to further enhance the texture details of the generated image. The formula is as follows:

$$L_{HF} = \mathbb{E}_{x \sim P_{data}(x)} [H(G_x(x)) - H(x)_1]. \quad (10)$$

Therefore, the overall loss function of the E²UCN model proposed in this paper, called Sobel-Frequency Hybrid Loss (SFHLoss), can be expressed as:

$$SFHLoss = L_{GAN}(G_x, D_y, x, y) + L_{GAN}(G_y, D_x, y, x) + L_{cycle}(G_x, D_y, x, y) + L_{sobel} + L_{HF} \quad (11)$$

By adding Sobel loss and high-frequency loss, the edge and texture details of the image are effectively preserved, which helps the subsequent object detection model accurately extract the crack locations.

2.2. YOLOv11 Model

After image enhancement, images with enhanced texture details were obtained. Subsequently, this study uses the YOLOv11 model for automatic underwater crack detection. YOLOv11, based on the previous generations of the YOLO model, significantly improves detection accuracy and speed through various innovations and optimizations, demonstrating excellent performance, especially in the detection of fine targets such as cracks. The specific model architecture is shown in Figure 2.

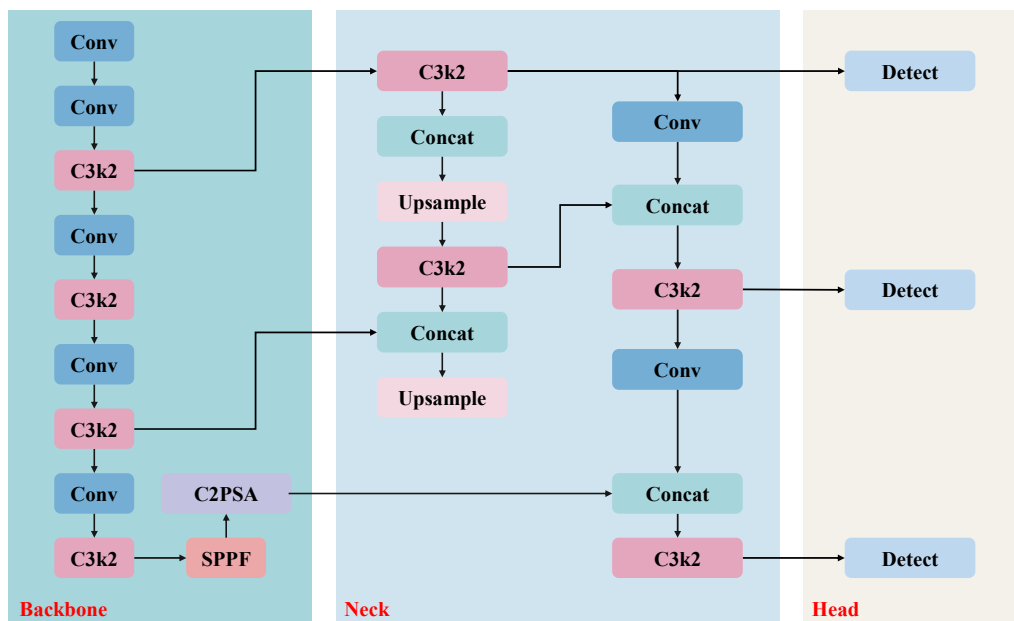


Figure 2. The architecture of YOLOv11.

The core structure of YOLOv11 includes the C3k2 module, SPPF module, C2PSA module, and a lightweight design. These innovations enable YOLOv11 to efficiently process the enhanced underwater crack images and achieve accurate crack localization and classification.

First, the C3k2 module in YOLOv11 adopts an improved CSP (Cross Stage Partial) structure, which optimizes the feature extraction process by using smaller convolution kernels (e.g., 3×3 kernels). The C3k2 module splits the input feature map into two parts, performs convolution on each part, and then merges them. This design effectively reduces the number of parameters while maintaining feature extraction capabilities. Next, the SPPF (Spatial Pyramid Pooling - Fast) module is another key component of YOLOv11. The SPPF module quickly merges feature maps of different scales through multi-scale pooling. This module significantly enhances the network's ability to detect targets of different sizes, especially when scale differences are common in crack images. By aggregating global features, the SPPF module improves the model's detection accuracy. The SPPF generates multi-level feature maps through pooling operations at different scales, ultimately merging these feature maps into a global feature representation, thus enhancing the network's sensitivity to multi-scale information in crack images.

Furthermore, the Convolutional block with Parallel Spatial Attention (C2PSA) module introduced by YOLOv11 further optimizes the extraction of spatial features. The C2PSA module uses parallel spatial attention mechanisms to focus on key areas in the image (such as the edges of cracks or the cracks themselves), effectively improving the model's recognition ability in complex backgrounds. The C2PSA module combines both channel attention and spatial attention mechanisms, and through multi-head attention, it further enhances the feature expression capabilities, allowing YOLOv11 to more accurately localize cracks.

In terms of model lightweight design, YOLOv11 introduces the MobileViT backbone network and depthwise separable convolutions (DWConv), significantly reducing the computational load and the number of parameters while maintaining high accuracy. MobileViT combines the advantages of convolutional neural networks (CNN) and Transformers, enabling efficient information encoding and fusion, allowing it to capture complex features in crack images while maintaining a low computational overhead. This design makes YOLOv11 suitable for resource-constrained devices such as embedded systems and drones, meeting the demand for crack detection on edge devices.

The loss function of YOLOv11 consists of three main components: classification loss (L_{cls}), bounding box regression loss (L_{box}), and distribution focal loss (L_{dfl}). Among them, L_{cls} mainly

optimizes the prediction of object categories, L_{box} is used to optimize the prediction of object locations, and L_{cfl} optimizes the confidence of the bounding boxes.

Through this multi-task loss function optimization strategy, YOLOv11 can effectively balance accuracy and speed in crack detection tasks, particularly showing high accuracy and robustness in detecting small cracks in complex backgrounds.

3. Experimental Analysis

In this section, the data collection process and experimental setup are first introduced. A series of experiments, including quantitative comparisons of image enhancement and crack detection metrics, detection result images, and ablation studies, are used to validate the effectiveness of the proposed E²UCN.

3.1. Data Collection and Experimental Setup

The underwater image dataset used in this study was collected from the physical model pool at the Tangtu Experimental Base of the Nanjing Hydraulic Research Institute. The test pool dimensions are $11.0\text{m} \times 5.9\text{m} \times 4.2\text{m}$ (*length*×*width*×*depth*), with a depth of 3.4m below the ground surface and a surrounding wall height of 0.8m, reinforced with carbon fiber fabric on the pool's sidewalls. The pool contains an underwater tunnel and a dam test scenario, with concrete of grade C30. The underwater tunnel dimensions are $6\text{m} \times 4\text{m} \times 3.3\text{m}$ (*length*×*width*×*height*), with typical defects set inside. The tunnel is 3.0m wide and 2.08m high, providing sufficient space for robotic operations, as shown in Figure 3.

During the data collection process, the mini underwater robot P200 "Qianjiao" was used to capture underwater optical images. The original video data were processed into eight typical images with a size of 256×256 pixels, simulating various lighting and visual conditions, including normal light, low contrast, light scattering, and non-uniform lighting, as shown in Figure 4.(a)



Figure 3. Schematic diagram of the construction layout of the test platform and the cracked wall.

The dataset used for training YOLOv11 was the Roboflow crack dataset, which was collected by researchers working on transportation and public safety. The dataset contains 4029 different static images of cracks, divided into training, testing, and validation sets, each with corresponding labels. The experiments were conducted on a high-performance personal computer equipped with an NVIDIA RTX 3080ti 12GB GPU and an AMD5800X CPU. All models were constructed and tested using PyTorch version 1.10.0. During network training, the high-frequency filtering parameter *cutoff* was set to 10.

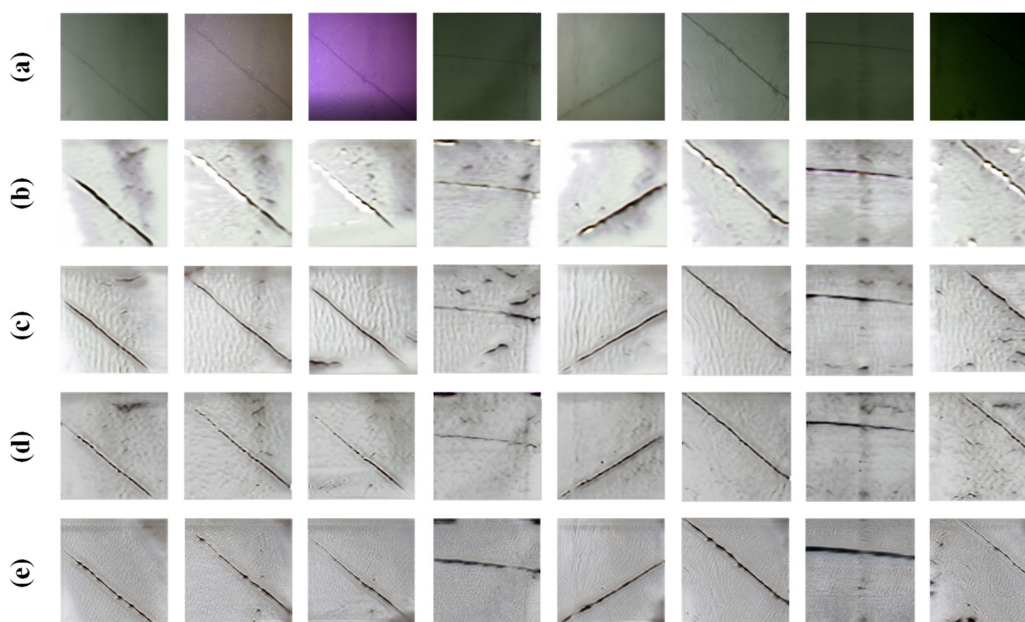


Figure 4. Image enhancement results, where (a) is the original image, (b) is the image generated using the original CycleGAN model, (c) is the image generated after using SobelLoss, (d) is the image using high frequency loss, and (e) is the image using SobelLoss and high frequency loss.

3.2. Experimental Results and Analysis

To validate the effectiveness of the CGBUIE model, several ablation experiments were designed in this section[30–33]. First, the original CycleGAN was used for style transfer of underwater images. Then, Sobel operators and high-frequency loss were added, followed by subjective evaluation of the result images compared to the designed CGBUIE. The enhanced results from different models are shown in Figure 4. The images were then input into the trained YOLOv11 model for detection. During the evaluation process, precision, recall, mAP (mean average precision), F1-score, and other metrics were recorded for each case to comprehensively evaluate the performance differences of YOLOv11 with different input images. Visual comparisons were also made between the original and enhanced images in terms of crack detection, analyzing the enhancement method's effect on detection accuracy and feature recognition. The specific methods and corresponding experimental results are shown in the table below.

The Table 1 shows the models used for target detection, where the checked boxes represent the models used in this experiment. Specifically, the first row, with no checked boxes, represents the use of the original underwater crack images in this case. From the experimental results, it can be seen that the image enhancement methods significantly affect the performance of YOLOv11 in underwater crack detection. Firstly, Sobel operators primarily enhance the edge information of the image, which helps YOLOv11 achieve better detection results in crack edge localization. Although Sobel operators improve edge clarity, they may lose some texture and detail information when enhancing the edges, leading to relatively poor performance in complex textured regions. Therefore, although recall is very high, the mAP50-95 shows a certain decline, reflecting the model's performance in broader detection areas, which may be affected, thus reducing overall detection performance.

Table 1. Ablation Experiment Module Design and Evaluation Metrics.

CycleGAN	SobelLoss	HFLoss	Detection Evaluation Indicators				
			Precision	Recall	F1-Score	mAP50	mAP50-95
			0.93	0.875	0.902	0.876	0.565
√			0.994	1	0.930	0.982	0.624
√	√		0.869	1	0.996	0.991	0.548

√		√	0.993	1	0.997	0.993	0.72
√	√	√	1	1	0.999	0.995	0.732

High-frequency loss, on the other hand, focuses on enhancing details and high-frequency information in the image, especially the finer parts of cracks. Compared with Sobel operators, the high-frequency loss image enhancement method is better at preserving the fine texture information of cracks, thus improving detection precision and recall. However, high-frequency enhanced images may also introduce some noise, especially in more complex backgrounds, leading to minor errors in the detection boxes. Nevertheless, the improvements in the F1 score and mAP50 indicate that high-frequency loss has a significant effect on detail restoration.

When Sobel operators and high-frequency loss are combined, their advantages complement each other. The image not only enhances edge information but also retains details and textures. This combination effectively improves YOLOv11's overall performance in crack detection, with precision, recall, and F1-scores [16,17] reaching near-perfect levels. By combining both enhancement methods, the image's details and edges are maximized, allowing YOLOv11 to achieve the best results in crack detection tasks, especially in small cracks and complex backgrounds, where detection accuracy is significantly improved.

Overall, the goal of image enhancement is to improve detection performance by enhancing edge clarity and restoring details, but the effects of different enhancement methods complement each other. Sobel operators effectively enhance edges but may lose details, while high-frequency loss restores details but may introduce noise. Therefore, using both methods together maximizes the retention of crack feature information in the enhanced underwater images.

In addition, the changes in the loss function and performance metrics during YOLOv11 training are shown in Figure 5. The detection results after applying different enhancement models are shown in Figure 6.

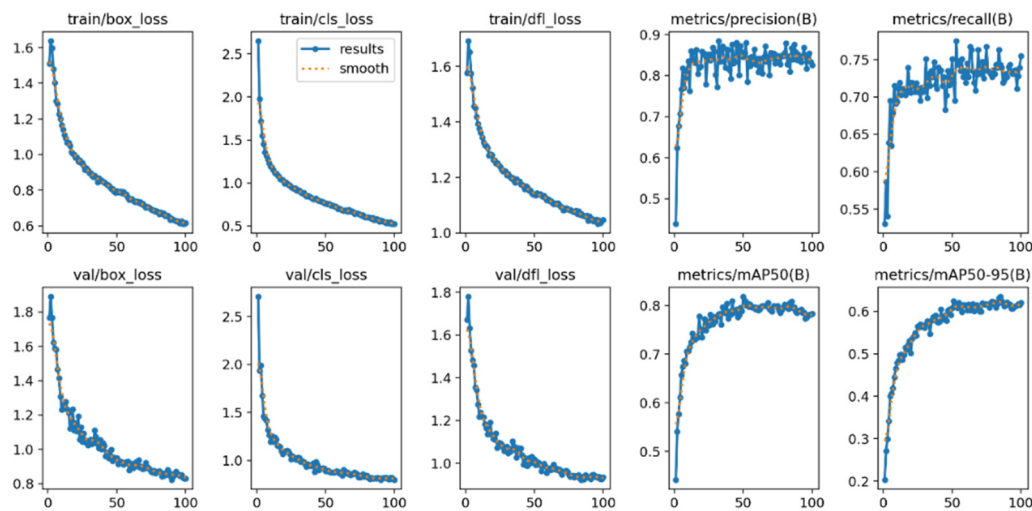


Figure 5. Loss function curve and performance index variation curve after training yolov11 on crack-seg dataset.

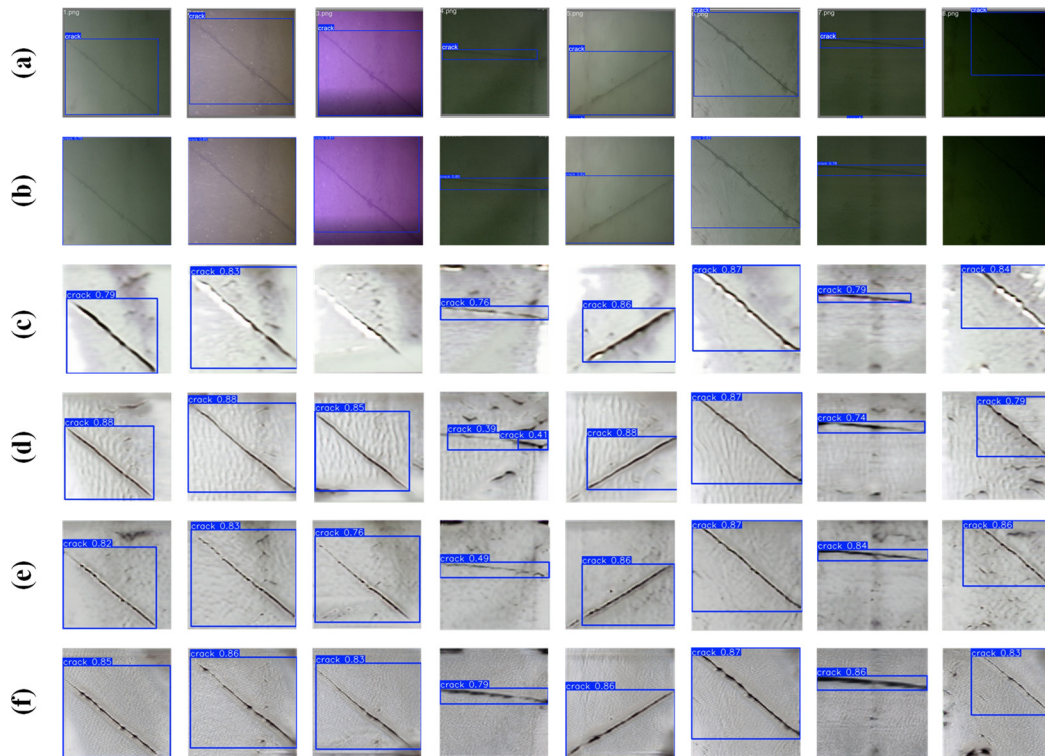


Figure 6. Target detection results, where (a) ground truth (b) the original image detection results, (c) images generated by the original CycleGAN model detection results, (d) images generated by SobelLoss detection results, (e) images generated by high frequency loss detection results, and (f) the images generated by SobelLoss and high frequency loss detection results.

From the detection result images, it can be observed that YOLOv11 exhibits some deficiencies in the original images (Figure 5 (b)), especially when the crack details are blurry or the background is complex. In these cases, the detection confidence is generally lower, and crack localization accuracy is affected. Specifically, the original image, due to its blurry details and low contrast, poses challenges for YOLOv11 in accurately detecting cracks.

With the application of the original CycleGAN model for style transfer (Figure 5 (c)), the detection results of YOLOv11 showed significant improvement. CycleGAN enhanced the overall image clarity, particularly improving the representation of edges and textures, which effectively increased the model's crack localization accuracy. The enhanced image's improvements in details and contrast allowed YOLOv11 to more accurately identify cracks, with a significant increase in detection confidence, reflecting the positive impact of image enhancement on detection results.

After further introducing SobelLoss (Figure 5 (d)), YOLOv11's crack detection performance improved further. The Sobel operator enhanced the image's edge details, significantly improving the clarity of the crack contours, and the model's precision in crack localization was improved. However, despite the positive effect of Sobel on edge enhancement, its ability to preserve image texture information is relatively weak, which may lead to the loss of details in small cracks, affecting detection accuracy. Some false positives and missed detections were still observed, suggesting that, while edge enhancement is beneficial, detail restoration remains a challenge for the model.

When high-frequency loss is applied for image enhancement (Figure 6 (e)), the fine details and subtle features of the cracks are more effectively restored. The introduction of high-frequency loss, focusing on the high-frequency components of the image, helps restore the fine features of the cracks, improving YOLOv11's ability to recognize small cracks. The enhanced image made crack detection more precise, and confidence increased, especially in low contrast and complex backgrounds, where the model demonstrated stronger robustness. However, excessive high-frequency enhancement could introduce background noise, causing instability in some areas of the detection results, highlighting the sensitivity of the enhancement method to background interference.

Finally, the combination of SobelLoss and high-frequency loss (Figure 6 (f)) demonstrated the best crack detection performance. This combination not only strengthened the edge details of the image but also effectively restored more texture information, making crack localization more precise, and the image details richer. YOLOv11 performed at its best with these enhanced images, with overall detection precision and recall significantly improved. By integrating both edge enhancement and detail restoration, the model's adaptability to complex environments was significantly improved, and detection confidence was generally higher, further validating the superiority of combining SobelLoss and high-frequency loss in enhancing image detail restoration and model robustness.

4. Conclusion

This paper combines the advantages of image style transfer, detail restoration, and edge enhancement, fully leveraging the complementary effects of Sobel operators and high-frequency filtering. A CycleGAN-based underwater image enhancement method (CGBUIE) was proposed, which effectively improves the edge and detail information of the generated images by introducing Sobel operators and high-frequency filtering as loss functions. Specifically, by training underwater crack images and above-water crack images, the style transfer of underwater images to above-water image styles was achieved, enhancing image visibility and detail expression while improving the robustness of the crack detection model. On this basis, YOLOv11 was used to train on the Crack-Seg dataset, constructing a detection model capable of effectively recognizing cracks. Experimental results show that the enhanced underwater crack images significantly improved YOLOv11's detection performance under complex backgrounds and low contrast conditions, with detection confidence and accuracy significantly increased. Future research can further optimize image enhancement strategies and combine other advanced deep learning techniques to improve the model's performance in more complex environments, achieving a more lightweight and accurate underwater crack detection model.

Author Contributions: methodology, X.W. and J.S.; software, W.Z. and X.W.; validation, W.Z.; investigation, X.W. and W.Z.; resources, X.W. and G.S.; writing—original draft preparation, W.Z.; writing—review and editing, X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Key Research and Development Program of China (No.2022YFC3005405), the National Natural Science Foundation of China - Joint Fund under Grant (No.U23B20150), the Special Fund of Chinese Central Government for Basic Scientific Research Operations in Commonweal Research Institutes under Grant(No.Y722003).

Data Availability Statement: The data presented in this study are available on request from the author.

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

References

1. Gong Xiaonan et al Dam Danger Assessment and Risk Reinforcement Technology [M], China Construction Industry Press 2021 (in Chinese).
2. XIANG Yan, WANG Yakun, CHEN Zhe, DAI Bo, CHEN Siyu, SHEN Guangze. Underwater defect detection and diagnosis assessment for high dams: current status and challenges[J]. Advances in Water Science.
3. Morlando G, Panei M, Porcino D, Pasquali M, Saponara S. Underwater Surveys of Hydraulic Structures: A Review Focused on the Use of Underwater Locomotion Systems[J]. Frontiers in Robotics and AI, 2021/2023.
4. Akkaynak D, Treibitz T. Sea-thru: A Method for Removing Water from Underwater Images[C]//Proc. CVPR. 2019: 1682–1691.
5. Pizer S M, Amburn E P, Austin J D, et al. Adaptive histogram equalization and its variations[J]. Computer Vision, Graphics, and Image Processing, 1987, 39(3): 355–368.

6. CHEN Lifeng, LIANG Xiaogang, PENG Qianqian, et al. Adaptive homomorphic filtering algorithm for low illumination image enhancement[J]. *Computer Science and Application*, 2023, 13: 450.(in Chinese)
7. C. Li, W. Zhang, Y. Zhang, Z. Chen and H. Gao, "Adaptively Dictionary Construction for Hyperspectral Target Detection," in *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1-5, 2023
8. Wang Y, Xiang Y, Dai B, et al. "Dam early warning model based on structural anomaly identification and dynamic effect variables selection," in *Structures*, Elsevier, vol-74, PP.108507, 2025.
9. Wang, Y., Fan, H., Liu, S., & Tang, Y. (2024). Underwater Image Enhancement Based on Multi-scale Attention and Contrastive Learning. *Laser & Optoelectronics Progress*. (in Chinese)
10. Du Feiyu, Wang Haiyan, Yao Haiyang, et al. Domain-Adaptive Underwater Image Enhancement Algorithm [J]. *Computers and Modernization*, 2024(10):55-60. (in Chinese)
11. Li J, Skinner K, Eustice R M, Johnson-Roberson M. WaterGAN: Unsupervised Generative Network to Enable Real-time Color Correction of Monocular Underwater Images[EB/OL]. arXiv:1702.07392, 2017.
12. OceanVision Lab. EUVP: A Large-Scale Paired and Unpaired Underwater Image Enhancement Dataset[EB/OL]. 2019.
13. Hambon J. Electronical imaging of structural concrete: Investigation by Bayesian stationary wavelet field[J]. *Electronic Imaging*, 2009, 9: 764-773.
14. Gao Y, Wang Y, Zhou D. The Image Recognition, Automatic Measurement and Seam Tracking Technology in Arc Welding Process[C]//2010 8th World Congress on Intelligent Control and Automation. 2010: 2327-2332.
15. Shi Z, Tao G, Cao Z, et al. CrackYOLO: A More Compatible YOLOv8 Model for Crack Detection[J]. *PAA (Pattern Analysis and Applications)*, 2024.
16. Mao, Y., et al. (2020). Crack Detection with Multi-task Enhanced Faster R-CNN Model. In *Proceedings of the IEEE International Conference on Image Processing* (pp. 1234-1238). IEEE.
17. Huang Y, Huang H, Zhang J, et al. Intelligent recognition of crack targets based on improved YOLOv5[J]. *Tsinghua Science and Technology*, 2023.
18. He K, Wang K, Liu Y, et al. URPC2019: Underwater Robot Picking Contest and Dataset[EB/OL]. LNCS (Springer) 2020/2019.
19. Zheng Z, Wang P, Liu W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[C]//Proc. AAAI. 2020: 12993-13000.
20. Hou X, Sheng H, You L, et al. An Improved YOLOv8 for Underwater Object Detection[J]. *Journal of Marine Science and Engineering*, 2024, 12(1): 123.
21. Zhu J-Y, Park T, Isola P, Efros A A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[C]//Proc. ICCV. 2017: 2242-2251.
22. Zhang Y, Wu K, Fang K. Comprehensive Performance Evaluation of YOLOv11: Advancements, Benchmarks, and Real-World Applications[EB/OL]. arXiv:2411.18871, 2024.
23. Chen H, Zhang H, Yin H, et al. Multi-scale reconstruction of edges in images using a morphological Sobel algorithm[J]. *Scientific Reports*, 2024, 14: 14202.
24. Cooley J W, Tukey J W. An algorithm for the machine calculation of complex Fourier series[J]. *Mathematics of Computation*, 1965, 19(90): 297-301.
25. Gonzalez R C, Woods R E. *Digital Image Processing (3rd ed.)*[M]. Pearson, 2008.
26. Jiang L, Dai B, Wu W, Loy C C. Focal Frequency Loss for Image Reconstruction and Synthesis[C]//NeurIPS. 2021.
27. Kim et al. Wavelet-Domain High-Frequency Loss for Perceptual Quality[C]//WACV. 2023.
28. El-askary A, et al. LM-CycleGAN: Improving Underwater Image Quality Through Loss Functions[EB/OL]. Scientific Foundation, 2024.
29. Enhancement of underwater dam crack images using a multi-feature CycleGAN[J]. *Automation in Construction*, 2024.
30. Powers D M W. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation[J]. *Journal of Machine Learning Technologies*, 2011, 2(1): 37-63.
31. Everingham M, Van Gool L, Williams C K I, Winn J, Zisserman A. The PASCAL Visual Object Classes (VOC) Challenge[J]. *IJCV*, 2010, 88: 303-338.

32. Future Machine Learning. Differences Between Precision, Recall, and F1 Score[EB/OL]. 2025.
33. Kintu J P, Kaini S, Shakya S, et al. Evaluation of a cough recognition model in real-world noisy environment[J]. BMC Medical Informatics and Decision Making, 2025, 25(1): 381.

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.