Article

Edge-Enhanced CrackNet for Underwater Crack Detection in 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; xbwu@nhri.cn (X.W.); wbzhang@nhri.cn (W.Z.); gzshen@nhri.cn (G.S.)
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: jbshen@nhri.cn; Tel.:+86-025-85828600
† This paper is an extension of a conference paper. Edge Enhanced Underwater CrackNet for Dam Underwater Crack Detection. In Proceedings of the International Symposium "Common Challenges, Shared Future, Better Dams" ICOLD-CIGB 2025, Chengdu, China, 16–23 May 2025.

Featured Application

This study provides an underwater image enhancement method that effectively improves the detection performance of crack recognition algorithms under complex flow field and low-illumination underwater conditions and provides a new solution for concrete crack detection algorithms applied to dam surfaces under deep-water conditions.

Abstract

Underwater crack detection in dam structures is of significant importance to ensure structural safety, assess operational conditions, and prevent potential disasters. Traditional crack detection methods face various limitations when applied to 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 study proposes a dual-stage underwater crack detection method based on Cycle-GAN and YOLOv11 called Edge-Enhanced Underwater CrackNet (E^2UCN) 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), and various complex underwater environments were simulated to construct a test dataset. Then, an improved Cycle-GAN image style transfer method was used to enhance the underwater images. Unlike conventional GAN-based underwater image enhancement methods that focus on global visual quality, our model specifically constrains edge preservation and high-frequency crack textures, providing a novel solution tailored for crack detection tasks. 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. The 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 (precision = 0.995, F1 = 0.99762, mAP@0.5 = 0.995, and mAP@0.5:0.95 = 0.736) 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

Currently, China has the largest number of high dams, with 588 dams over 70 m, 233 dams over 100 m, and 23 dams over 200 m. Thus, high dam safety is a major challenge for national water safety and public safety. Cracks are one of the most common structural defects in such 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 safety. In some cases, the cracks exist underwater, which is even more challenging, as mature methods for above-water detection cannot be directly applied. Therefore, determining methods for early detection and precise evaluation of cracks both above and below water is of critical importance for ensuring the stability of these engineering structures.

Traditional crack detection methods mainly rely on draining reservoirs to clear interference factors combined with manual inspections to clearly observe underwater structural defects [1,2], an approach that is both time-consuming and economically costly and requires a considerable amount of human resources, potentially having a negative impact on the ecological environment in certain cases. It is also not suitable for high dams and large reservoirs. In contrast, modern detection methods primarily rely on divers or portable detection equipment for underwater visual inspections, with saturation diving required for high dams, which is both dangerous and costly. Moreover, divers often lack professional knowledge of hydraulic engineering, lowering the reliability of their detection results. Most recently, with the advancements in ROVs 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 low 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, but have 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 [9,10]. Recent unpaired enhancement approaches (e.g., WaterGAN [11], Shallow-uwnet [12], UColor [13], PUIE-Net [14]) and datasets show that paired data are not strictly required, but these methods focus mainly on perceptual quality rather than preserving crack features for detection. Meanwhile, modern detectors such as YOLOv8–v11 demonstrate strong robustness, though pipelines that explicitly constrain edge/high-frequency cues for underwater cracks remain scarce. For example, Wang Yue et al. [15] proposed a multi-scale attention and contrast learning-based underwater image enhancement algorithm, which effectively extracts multi-level image features by combining an encoder–decoder structure with multi-scale channel pixel attention modules. The improvements achieved in PSNR and SSIM metrics were 4.4% and 2.8%, respectively, significantly improving image clarity. Additionally, Du Feiyu et al. [16] proposed a domain-adaptive underwater image enhancement method combining convolu-

tional neural networks with a multi-head attention mechanism and adversarial learning to achieve image enhancement under unsupervised conditions. Although these methods 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. In other words, current underwater enhancement pipelines rarely optimize for crack-salient edges and high-frequency cues, leaving a gap between visually pleasing restoration and useful feature retention for detectors. Furthermore, existing underwater image enhancement models require paired images for training [17,18], but paired underwater crack images are difficult to obtain as they require 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 [19] and Sobel operators [20], rely on edge detection algorithms in their image processing techniques. Although these methods perform well in simple scenarios, they are highly sensitive to noise and 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 [21], based on the YOLOv5 model, which introduces a feature fusion module, a Res2C3 feature extraction module, and a BCAtt attention mechanism, significantly improving crack detection performance. It achieved 94.3% mAP and a detection speed of 151 FPS in underwater crack detection tasks. Moreover, Mao Yingchi et al. [22] proposed a multi-task enhanced crack image detection method (ME-Faster R-CNN) based on Faster R-CNN, which improves the regional proposal network (RPN) and introduces the multi-source adaptive balancing TrAdaBoost method, effectively improving the detection capability for multiple targets and small target cracks. In experiments, it achieved an 82.52% average intersection over union (IoU) and 80.02% average precision (mAP), an improvement of 1.06% and 1.56%, respectively, compared to traditional Faster R-CNN methods. Additionally, Huang et al. [23] 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 [24], thereby substantially elevating real-time detection accuracy. Concurrently, a refined YOLOv8-derived architecture was developed, which 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 [25]. These research outcomes show that deep learning-based object detection methods can effectively handle complex underwater crack detection tasks in various fields and achieve a good balance between detection speed and accuracy. However, due to the lack of underwater crack data, research on dam underwater crack detection based on object detection algorithms is relatively limited [26].

To address the deficiencies in the existing research on underwater image enhancement and crack detection, this paper proposes a dual-stage underwater crack detection method based on Cycle-GAN [27] and YOLOv11 [28] called Edge-Enhanced Underwater CrackNet (E^{2}UCN) in order to pay sufficient 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 a 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 Cycle-GAN-based underwater image style transfer method named the CycleGAN-Based Underwater Image Enhancement (CGBUIE) model to improve underwater image quality and highlight crack detail features. Although hybrid GAN-based approaches such as attention-guided CycleGAN [29,30] and ESRGAN variants [31,32] have achieved impressive results in natural image enhancement, they primarily optimize for perceptual quality rather than structural crack feature preservation. In contrast, our method explicitly constrains edge and frequency information, making it more suitable for downstream crack detection tasks. Furthermore, the CGBUIE model introduces Sobel operators [33] and high-frequency transformations [34,35] 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 [36–38]. Finally, in the crack detection stage, we first trained the YOLOv11 model on an above-water concrete crack dataset so it could learn key crack features and prominent edge features, then applied the trained model to the enhanced underwater crack images generated by the CGBUIE model 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, particularly in cases where the cracks are small and complex in shape [39]. Unlike existing underwater image enhancement approaches, which mostly emphasize overall image clarity, our method explicitly integrates the Sobel operator and high-frequency Fourier constraints into CycleGAN, ensuring the preservation of crack edges and the fine textures critical for subsequent detection. 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 showed that the proposed method performs well in terms of crack edge clarity, object localization accuracy, and adaptability to complex backgrounds.

The novelty of E^2UCN lies in its dual-stage architecture: (i) an enhanced CycleGAN that incorporates Sobel and Fourier constraints to explicitly preserve crack-specific features during the style transfer process, and (ii) its integration with YOLOv11 for robust crack detection. This was further substantiated by comprehensive ablation studies. The remainder of this paper is organized as follows: Section 2 describes the proposed E^2UCN and its edge-/texture-aware enhancement. Section 3 presents the datasets, annotation protocols, and ablation design, discusses the results, uncertainties, and analysis. Section 4 concludes with the study's contributions and limitations.

2. Proposed Method

The E^2UCN framework, the CGBUIE model, and the YOLOv11 model used in this study will be described in detail in this section. Specifically, the E^2UCN model is shown in Figure 1.
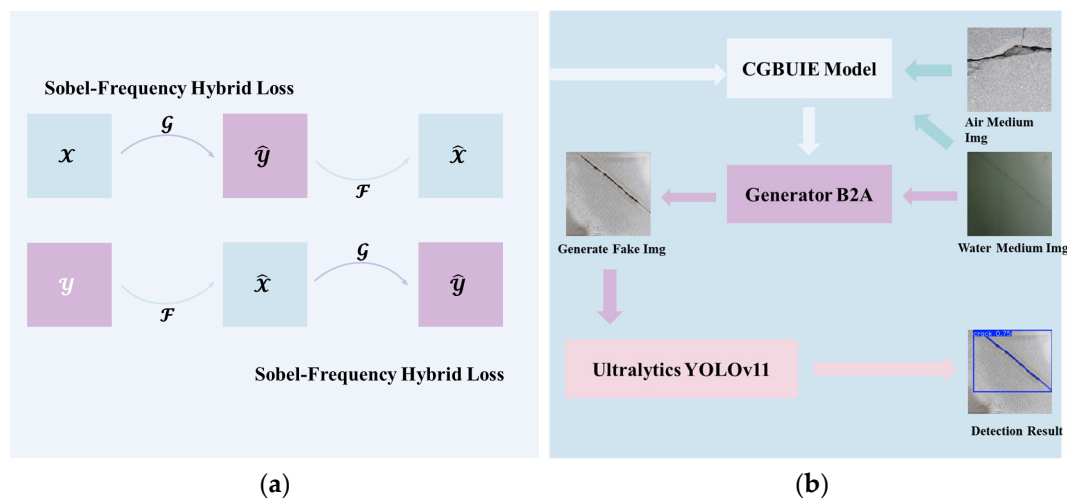
Figure 1. The architecture of the proposed model: (**a**) structural diagram of E^{2}UCN; (**b**) structural diagram of CGBUIE.

2.1. CycleGAN-Based Underwater Image Enhancement Model

Based on the style transfer functionality of CycleGAN, this study 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, particularly when large annotated datasets are unavailable. CycleGAN achieves this goal by introducing two generators and two discriminators. Generators are used to generate images similar to the target style, while discriminators judge the difference between the generated image and the real image, thereby guiding the generator to optimize its generation effect.

In CycleGAN (https://github.com/junyanz/CycleGAN, accessed on 9 September 2025), 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 aim is to minimize the difference between generated 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.

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:

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

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

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 follows:

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

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 as follows:

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

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

However, experiments have shown 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 introduces the E^2UCN model.

The Sobel operator extracts edge information from an image by computing the gradients and is effective in retaining the image's edge features. Here, based on the original CycleGAN, the basic principle of the Sobel operator is applied 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 as follows:

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

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 follows:

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

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, by using the 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 such as textures and edges. 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, preserving and enhancing the texture information.

In detail, 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:

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

where N is the signal length, n is the time index, and k is the frequency index. After applying the 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:

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

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:

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

Therefore, the overall loss function of the E^{2}UCN model proposed in this paper, called Sobel-Frequency Hybrid Loss (SFHLoss), can be expressed as follows:

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

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 are obtained. Subsequently, this study uses the YOLOv11 (https://github.com/ultralytics/ultralytics, accessed on 9 September 2025), 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, particularly in the detection of fine targets such as cracks. The specific model architecture is shown in Figure 2.

The core structure of YOLOv11 includes the C3k2 module, the SPPF module, the 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.
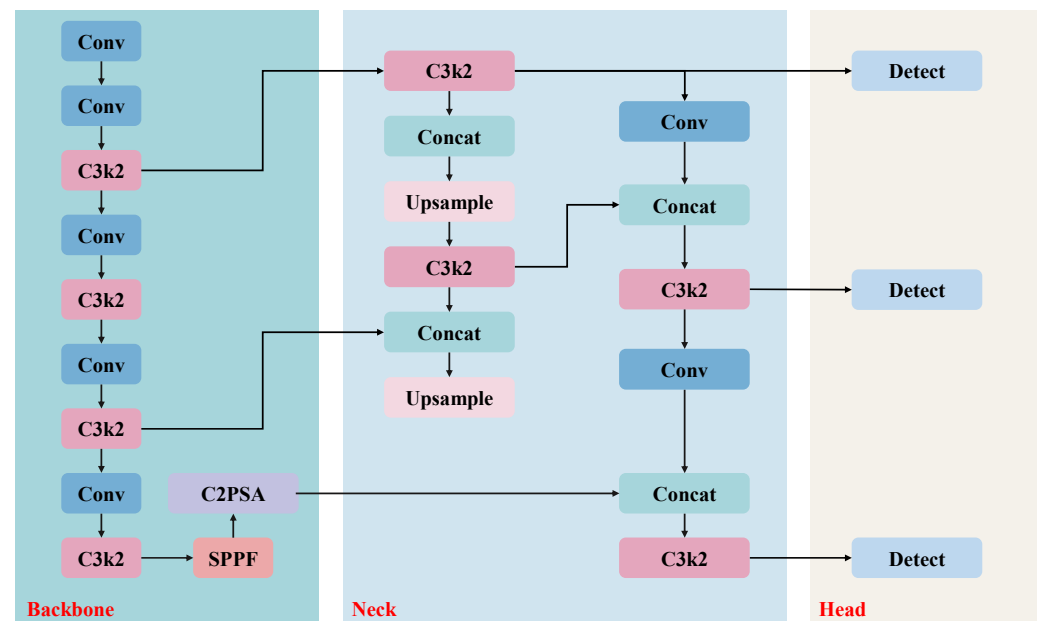
Figure 2. The architecture of YOLOv11.

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, another key component of YOLOv11, 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, which is particularly relevant as scale differences are common in crack images. By aggregating global features, the SPPF module improves the model's detection accuracy. It generates multi-level feature maps through pooling operations at different scales and ultimately merges 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 the 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 against 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.

To enhance the model's 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 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}). L_{cls} mainly

optimizes the prediction of object categories, L_{box} is used to optimize the prediction of object locations, and L_{dfl} 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 introduced. A series of experiments, including quantitative comparisons of image enhancement and crack detection metrics, detection result images, and ablation studies, is used to validate the effectiveness of the proposed E^{2}UCN.

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 $\times$ width $\times$ depth*), with a depth of 3.4 m below the ground surface and a surrounding wall height of 0.8 m, and the pool's sidewalls are reinforced with carbon-fiber fabric. 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 $\times$ width $\times$ height*), with typical defects set inside. The tunnel is 3.0 m wide and 2.08 m high, providing sufficient space for robotic operations, as shown in Figure 3.
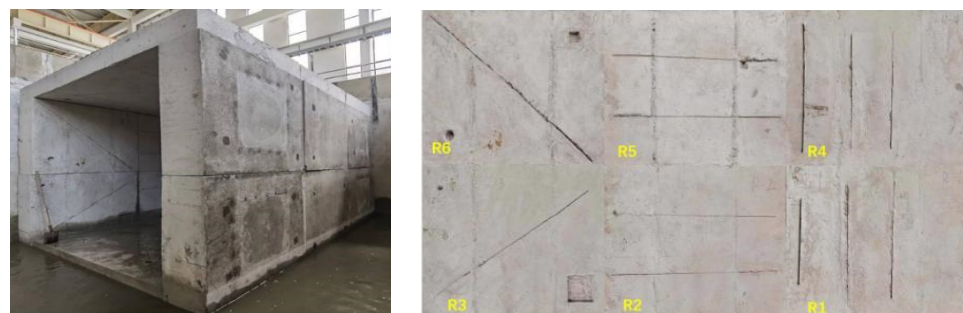
Figure 3. Schematic diagram of the construction layout of the test platform and the cracked wall. Specially, R1–R6 represent six different types of cracks with varying orientations.

During the data collection process, the mini underwater P200 robot "Qianjiao" (Manufacturer: Qianxin Innovation Technology Co., Ltd.; City: Shenzhen; Country: China) 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 4a.

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. For the CycleGAN training, 10 underwater crack images and 26 above-water crack images were collected, among which 8 typical underwater images (with scattering, insufficient illumination, and blur) were selected for testing. In addition, the dataset for validation contained 11 images, covering cracks with diverse widths, depths, and orientations and several real underwater crack images. Although the number of self-collected underwater images was limited, representative cases with scattering, blurring, and color distortion were selected to ensure diversity in crack characteristics. 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 obtained using high-frequency loss, and (**e**) is the image obtained 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 [40–42]. First, the original CycleGAN was used for the style transfer of underwater images. Then, Sobel operators and high-frequency loss were added, and subjective evaluation of the resulting images compared to the designed CGBUIE model was performed. 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 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.

Table 1 shows the models used for target detection, where the checked boxes represent the models used in this experiment, and the first row, with no checked boxes, represents the use of the original underwater crack images. 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. However, 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 the 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
√	√	√	0.995	1	0.997	0.995	0.732

Secondly, high-frequency loss focuses on enhancing the details and high-frequency information in the image, particularly 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, particularly 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.

Overall, the goal of image enhancement is to improve detection performance by enhancing edge clarity and restoring details, and, when Sobel operators and high-frequency loss are combined, their advantages complement each other toward achieving that goal. In other words, 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, effectively improving YOLOv11's overall performance in crack detection, particularly for small cracks and complex backgrounds, with precision, recall, and F1-scores reaching near-perfect levels.

The changes in the loss function and performance metrics during YOLOv11 training are shown in Figure 5, with the ground truth illustrated in line (a). The detection results after applying different enhancement models are shown in Figure 6.
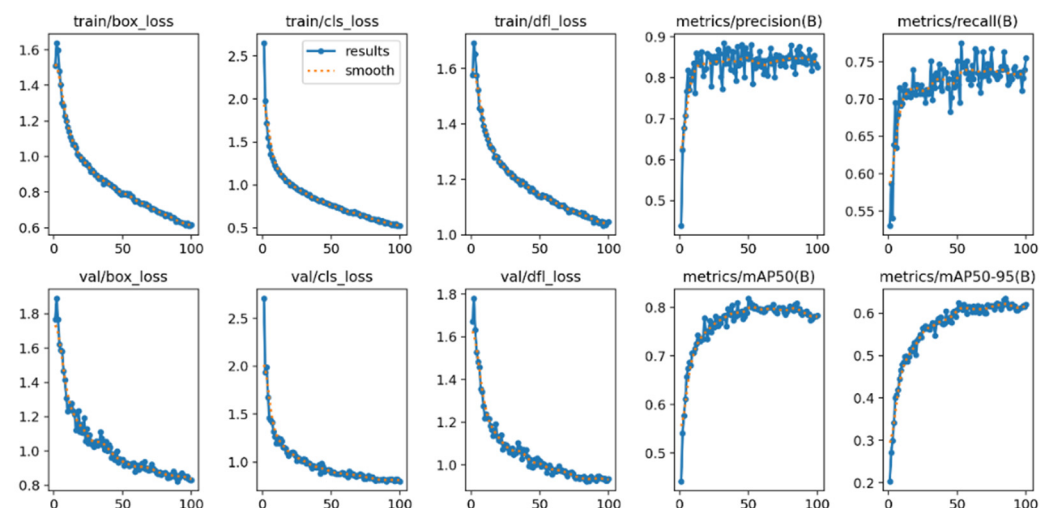
Figure 5. Loss function curve and performance index variation curve after training YOLOv11 on the Crack-Seg dataset.
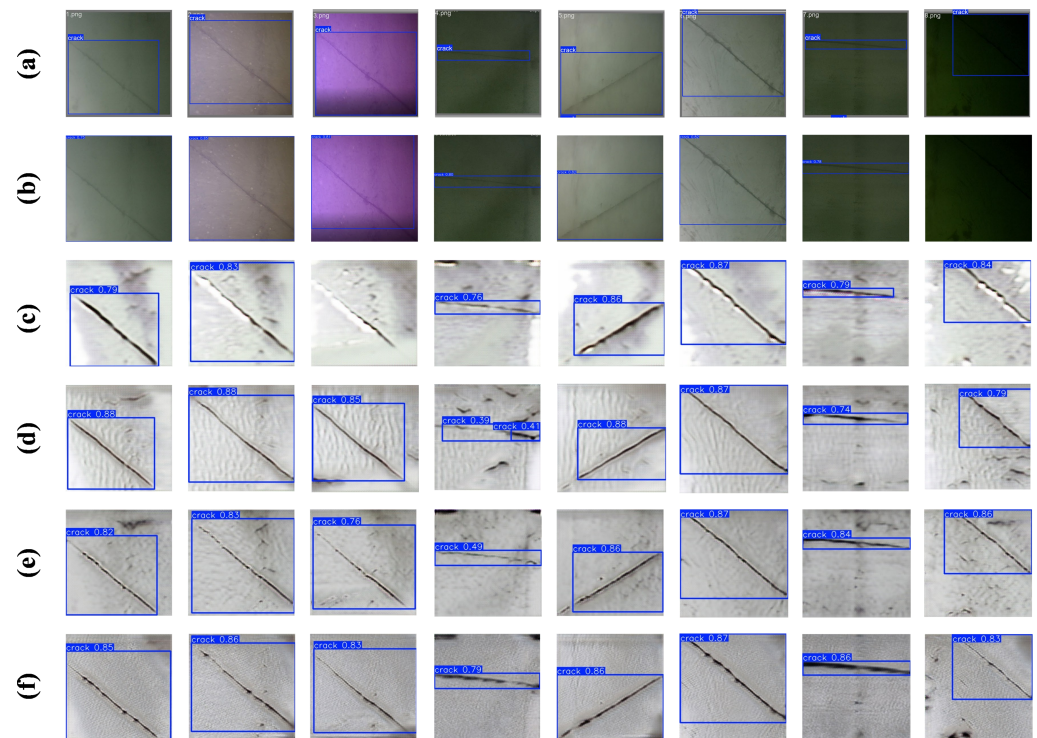
Figure 6. Target detection results, where (**a**) is the ground truth, (**b**) is the original image detection results, (**c**) is the images generated by the original CycleGAN model detection results, (**d**) is the images generated by SobelLoss detection results, (**e**) is the images generated by high-frequency loss detection results, and (**f**) is the images generated by SobelLoss and high-frequency loss detection results.

It can be observed that YOLOv11 exhibits some deficiencies in the original images (Figure 6b), particularly 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 6c), 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 improved 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 6d), 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 6e), focusing on the high-frequency components of the image, the fine details and subtle features of the cracks are more effectively restored, improving YOLOv11's ability to recognize small cracks. The enhanced image made crack detection more precise, and confidence increased, particularly for 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 6f) 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 its overall detection precision and recall being 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 for enhancing image detail restoration and model robustness.

In addition to YOLOv11, we conducted comparative experiments with YOLOv5 and YOLOv8 to provide a broader reference for detection performance. As summarized in Table 2 and Figure 7, all three models benefited from the proposed image enhancement strategy. YOLOv5 achieved a precision of 0.91 and a recall of 1.0, resulting in an F1-score of 0.953, with mAP@0.5 = 0.876 and mAP@0.5:0.95 = 0.752. YOLOv8 exhibited higher overall precision and recall (1.0 and 0.995, respectively), yielding an F1-score of 0.997, and the highest mAP@0.5 value of 0.995, although its performance at stricter IoU thresholds (mAP@0.5:0.95 = 0.685) was comparatively lower. YOLOv11 attained balanced and robust performance, with precision = 0.995, recall = 1.0, F1-score = 0.998, and mAP@0.5 = 0.995, while maintaining a competitive mAP@0.5:0.95 of 0.732.

Table 2. Comparative experiments and corresponding evaluation metrics. The highest score of each evaluation metric is highlighted in bold.

Model Name	Detection Evaluation Indicators				
	Precision	Recall	F1-Score	mAP50	mAP50-95
YOLOv5	0.91	**1**	0.953	0.876	**0.752**
YOLOv8	**1**	0.995	0.997	0.995	0.685
YOLOv11	0.995	**1**	**0.998**	**0.995**	0.732
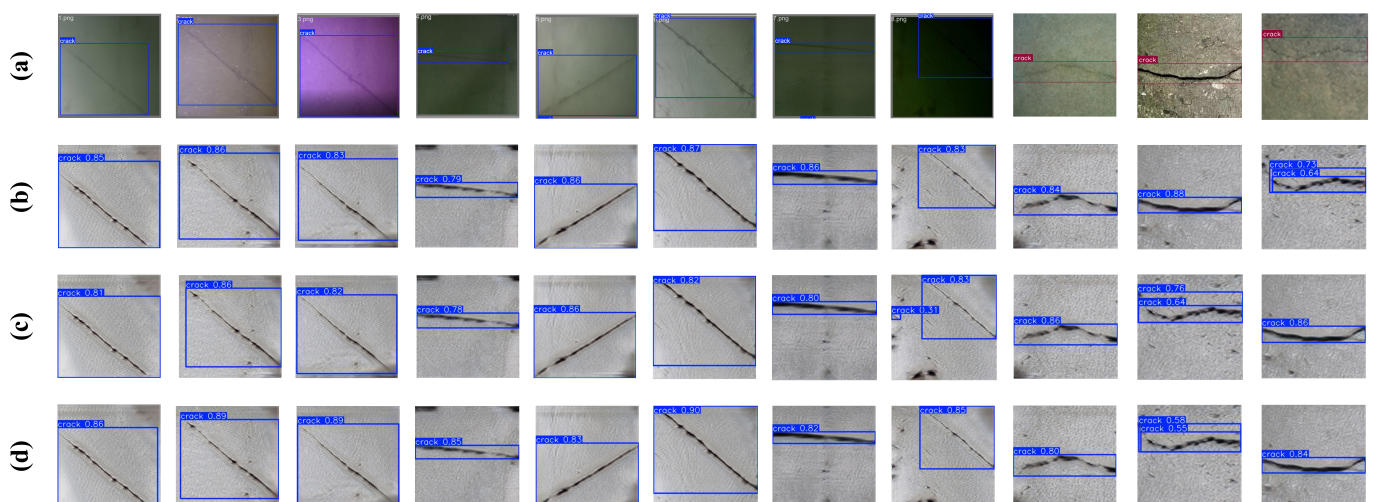
Figure 7. Target detection results, where (**a**) is the ground truth, (**b**) is the YOLOv11 results, (**c**) is the YOLOv5 results, and (**d**) is the YOLOv8 results.

These results show that, although YOLOv8 achieved the highest detection accuracy under lenient IoU criteria, YOLOv11 demonstrated superior robustness in handling small and irregular crack patterns, reflected in its higher F1-score and improved performance at

more stringent IoU thresholds compared to YOLOv8. Therefore, YOLOv11 was selected as the primary detection model in this study. More comprehensive comparisons with other detectors, such as Faster R-CNN and Transformer-based architectures, will be explored in future work to further validate the generality of the proposed approach.

4. Conclusions

This study 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 (the CGBUIE model) 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 on 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 the model on the Crack-Seg dataset, constructing a detection model capable of effectively recognizing cracks. The experimental results show that the enhanced underwater crack images significantly improved YOLOv11's detection performance, detection confidence, and accuracy with complex backgrounds and in low-contrast conditions. Specifically, on a real underwater validation set, E^2UCN achieved precision = 0.995, F1 = 0.99762, mAP@0.5 = 0.995, and mAP@0.5:0.95 = 0.736.

In summary, the primary contribution of this work lies in the integration of Sobel operators and high-frequency Fourier constraints into the CycleGAN framework, which ensures the preservation of critical crack-related details during the enhancement process. Together with YOLOv11, this creates a powerful and reliable system for underwater crack detection in challenging environments.

However, this study has certain limitations. The underwater dataset used is relatively small, with only three real underwater images supplementing the dataset. This limited dataset size restricts the diversity and robustness of the model, which may affect its performance in more varied scenarios. Future research should focus on expanding this dataset to increase its diversity and improve the model's generalization capabilities.

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.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Conflicts of Interest: Author X.W. was employed by the company NHRI R&D Tech Group Co., Ltd. The remaining authors declare that the re-search was con-ducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Gong, X.N.; Zhang, C.S.; Jia, J.S. *Dam Danger Assessment and Risk Reinforcement Technology*; China Construction Industry Press: Beijing, China, 2021. (In Chinese)

2. Xiang, Y.; Wang, Y.; Chen, Z.; Dai, B.; Chen, S.; Shen, G. Underwater defect detection and diagnosis assessment for high dams: Current status and challenges. *Adv. Water Sci.* **2024**, *35*, 153–164.

3. Raveendran, S.; Patil, M.D.; Birajdar, G.K. Underwater image enhancement: A comprehensive review, recent trends, challenges and applications. *Artif. Intell. Rev.* **2021**, *54*, 5413–5467. [CrossRef]

4. Akkaynak, D.; Treibitz, T. Sea-thru: A Method for Removing Water from Underwater Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 1682–1691.

5. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J.B.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.* **1987**, *39*, 355–368. [CrossRef]

6. Chen, L.; Liang, X.; Peng, Q.; Liu, W. Adaptive homomorphic filtering algorithm for low illumination image enhancement. *Comput. Sci. Appl.* **2023**, *13*, 450–457. (In Chinese) [CrossRef]

7. Li, C.; Zhang, W.; Zhang, Y.; Chen, Z. Adaptively Dictionary Construction for Hyperspectral Target Detection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5502005. [CrossRef]

8. Wang, Y.; Xiang, Y.; Dai, B.; Li, J. Dam early warning model based on structural anomaly identification and dynamic effect variables selection. *Structures* **2025**, *74*, 108507. [CrossRef]

9. Habib, A.; Yildirim, U. Distribution of strong input energy in base-isolated structures with complex nonlinearity: A parametric assessment. *Multidiscip. Model. Mater. Struct.* **2023**, *19*, 324–340. [CrossRef]

10. Habib, A.; Barakat, S.; Al-Toubat, S.; Junaid, M.T.; Maalej, M. Developing machine learning models for identifying the failure potential of fire-exposed FRP-strengthened concrete beams. *Arab. J. Sci. Eng.* **2025**, *50*, 8475–8490. [CrossRef]

11. Li, J.; Skinner, K.A.; Eustice, R.M.; Johnson-Roberson, M. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot. Autom. Lett.* **2017**, *3*, 387–394. [CrossRef]

12. Naik, A.; Swarnakar, A.; Mittal, K. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; Volume 35, pp. 15853–15854.

13. Lin, P.; Wang, Y.; Li, Y.; Fan, Z.; Fu, X. Underwater color correction network with knowledge transfer. *IEEE Trans. Multimed.* **2024**, *26*, 8088–8103. [CrossRef]

14. Fu, Z.; Wang, W.; Huang, Y.; Ding, X.; Ma, K. Uncertainty inspired underwater image enhancement. In *Computer Vision—ECCV 2022, 17th European Conference, Tel Aviv, Israel, 23–27 October 2022*; Springer Nature: Cham, Switzerland, 2022; pp. 465–482.

15. Wang, Y.; Fan, H.; Liu, S.; Tang, Y. Underwater Image Enhancement Based on Multi-scale Attention and Contrastive Learning. *Laser Optoelectron. Prog.* **2024**, *61*, 0437008. (In Chinese)

16. Du, F.; Wang, H.; Yao, H.; Chen, X. Domain-Adaptive Underwater Image Enhancement Algorithm. *Comput. Mod.* **2024**, 55–60. (In Chinese) [CrossRef]

17. Li, J.; Skinner, K.; Eustice, R.M.; Johnson-Roberson, M. WaterGAN: Unsupervised Generative Network to Enable Real-time Color Correction of Monocular Underwater Images. *arXiv* **2017**, arXiv:1702.07392. [CrossRef]

18. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [CrossRef]

19. Hambon, J. Electronical imaging of structural concrete: Investigation by Bayesian stationary wavelet field. *Electron. Imaging* **2009**, *9*, 764–773.

20. Gao, Y.; Wang, Y.; Zhou, D. The Image Recognition, Automatic Measurement and Seam Tracking Technology in Arc Welding Process. In Proceedings of the 2010 8th World Congress on Intelligent Control and Automation, Jinan, China, 7–9 July 2010; pp. 2327–2332.

21. Shi, P.; Shao, S.; Fan, X.; Xin, Y.; Zhou, Z.; Cao, P.; Li, X.; Zhu, S. CrackYOLO: Towards efficient dam crack detection for underwater scenes. *Pattern Anal. Appl.* **2024**, *27*, 105. (In Chinese) [CrossRef]

22. Mao, Y.; Ping, P.; Chen, J.; Chen, H. Crack Detection with Multi-task Enhanced Faster R-CNN Model. In Proceedings of the IEEE International Conference on Image Processing, Online, 25–28 October 2020; pp. 1234–1238.

23. Huang, B.; Kang, F.; Tang, Y. Real-Time Crack Detection Method for Concrete Dams Based on Object Detection. *J. Tsinghua Univ. (Sci. Ed.)* **2023**, *63*, 1078–1086. [CrossRef]

24. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. In Proceedings of the the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 12993–13000.

25. Huang, J.; Fang, C.; Zheng, X.; Liu, J. YOLOv8-UC: An Improved YOLOv8-Based Underwater Object Detection Algorithm. *IEEE Access* **2024**, *12*, 172186–172195. [CrossRef]

26. Liu, C.; Li, H.; Wang, S.; Zhu, M.; Wang, D.; Fan, X.; Wang, Z. A dataset and benchmark of underwater object detection for robot picking. In Proceedings of the 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shenzhen, China, 5–9 July 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.

27. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2242–2251.

28. Zhang, Y.; Wu, K.; Fang, K. Comprehensive Performance Evaluation of YOLOv11: Advancements, Benchmarks, and Real-World Applications. *arXiv* **2024**, arXiv:2411.18871.

29. Tang, H.; Liu, H.; Xu, D.; Torr, P.H.S.; Sebe, N. AttentionGAN: Unpaired Image-to-Image Translation Using Attention-Guided Generative Adversarial Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 1972–1987. [CrossRef]

30. Hou, M.; He, X.; Dou, F.; Zhang, X.; Guo, Z.; Feng, Z. Semi—Supervised image super-resolution with attention CycleGAN. *IET Image Process.* **2022**, *16*, 1181–1193. [CrossRef]

31. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.

32. Wei, Z.; Huang, Y.; Chen, Y.; Zheng, C.; Gao, J. A-ESRGAN: Training real-world blind super-resolution with attention U-Net Discriminators. In *PRICAI 2023: Trends in Artificial Intelligence, 20th Pacific Rim International Conference on Artificial Intelligence, PRICAI 2023, Jakarta, Indonesia, 15–19 November 2023*; Springer Nature: Singapore, 2023; pp. 16–27.

33. Kanopoulos, N.; Vasanthavada, N.; Baker, R.L. Design of an image edge detection filter using the Sobel operator. *IEEE J. Solid-State Circuits* **1988**, *23*, 358–367. [CrossRef]

34. Cooley, J.W.; Tukey, J.W. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.* **1965**, *19*, 297–301. [CrossRef]

35. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 3rd ed.; Pearson: London, UK, 2008.

36. Jiang, L.; Dai, B.; Wu, W.; Loy, C.C. Focal Frequency Loss for Image Reconstruction and Synthesis. In Proceedings of the 2021 Conference on Neural Information Processing Systems, Online, 6–14 December 2021.

37. Kim, M.W.; Cho, N.I. Wavelet-Domain High-Frequency Loss for Perceptual Quality. In Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2–7 January 2023.

38. Wu, J.; Zhang, G.; Fan, Y. LM-CycleGAN: Improving Underwater Image Quality Through Learned Perceptual Image Patch Similarity and Multi-Scale Adaptive Fusion Attention. *Sensors* **2024**, *24*, 7425. [CrossRef] [PubMed]

39. Chen, D.; Kang, F.; Li, J.; Zhu, S.; Liang, X. Enhancement of underwater dam crack images using a multi-feature CycleGAN. *Autom. Constr.* **2024**, *167*, 105727. [CrossRef]

40. Powers, D.M.W. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.

41. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The PASCAL Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]

42. Yacouby, R.; Axman, D. Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, Online, November 2020; pp. 79–91. Available online: https://aclanthology.org/2020.eval4nlp-1.9/ (accessed on 9 September 2025).

43. Zhang, W.; Wu, X.; Xiang, Y. Edge enhanced underwater CrackNet for dam underwater crack detection. In Proceedings of the International Symposium "Common Challenges, Shared Future, Better Dams," ICOLD-CIGB 2025, Chengdu, China, 16–23 May 2025; International Commission on Large Dams: Paris, France, 2025.