electronics

MDPI

Article

Adaptive Visual Enhancement and Semi-Supervised Learning for Robust Underwater Object Detection

Wangpeng He, Xiaofei Zhang, Yue Zhou, Ye Tao and Juanjuan Zhu *

School of Aerospace Science and Technology, Xidian University, Xi'an 710071, China; hewp@xidian.edu.cn (W.H.); zhangxf@stu.xidian.edu.cn (X.Z.); yz@stu.xidian.edu.cn (Y.Z.); ty2026@stu.xidian.edu.cn (Y.T.)
* Correspondence: zhujj@xidian.edu.cn

Abstract

Underwater object detection remains a challenging task due to the degradation of image quality caused by light absorption and scattering. To address this, we propose an innovative method combining adaptive visual enhancement and semi-supervised learning for robust underwater object detection. Our approach incorporates a visual enhancement and joint training-based image preprocessing method (VEJT-IP), which utilizes existing visual enhancement algorithms guided by object detection loss to preprocess underwater images. Additionally, we introduce a semi-supervised object detection framework utilizing high-quality pseudo-labels (HQPL-SSOD), which includes a localization quality-based pseudo-label filter to ensure the use of only high-quality pseudo-labels during training. Extensive experiments on the URPC2022 dataset demonstrate that our method significantly improves detection accuracy compared to baseline models, achieving a good balance between accuracy and annotation cost. The proposed approach improves underwater object detection performance under the evaluated URPC2022 conditions and provides a basis for further validation in broader underwater environments.

Keywords: underwateroptical scene; object detection; visual enhancement; semi-supervised learning

1. Introduction

Underwater object detection is a challenging research topic in computer vision. Traditional ocean exploration relies heavily on divers, which is inefficient and demanding in terms of skills and physical endurance. As a result, underwater exploration using Remotely Operated Vehicles (ROV) and Autonomous Underwater Vehicles (AUVs) [1–3] has become a preferred approach. These vehicles can collect valuable data and images from deep-sea areas that humans cannot easily access, providing scientists and engineers with crucial information. Underwater object detection is a key technology for processing, analyzing, and utilizing these images, forming the core of automated exploration tasks. Algorithms can automatically recognize objects from sensor and camera data collected by ROVs and AUVs, significantly improving data processing efficiency while reducing manual intervention. The application of deep learning further enhances these algorithms, enabling accurate recognition under varying lighting and water quality conditions—critical for managing the uncertainty of underwater environments. Detection algorithms fall into two main categories: sonar-based and optical-based methods. Sonar images are limited by low resolution and slow acquisition speeds, restricting their real-time high-resolution imaging capabilities. In contrast, optical images offer higher resolution and contain rich

Check for updates

Academic Editor: Danilo Pau

Received: 21 July 2026
Revised: 20 August 2026
Accepted: 21 August 2026
Published: 25 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

color and texture information, showcasing greater potential. Ideal deep learning-based underwater optical detection algorithms must be accurate and robust in complex underwater environments. However, these algorithms face several challenges compared to general detection algorithms. The primary challenges include the poor visual quality of underwater images due to light absorption, scattering, artificial lighting, and water impurities, resulting in low brightness, contrast, and color distortion. These issues impact the quality of color and texture information, limiting the performance of deep learning-based detectors. Additionally, the high cost of annotation due to poor image quality makes manual labeling difficult and raises costs. The scarcity of labeled data in certain underwater environments further limits model performance, highlighting the importance of training methods that can utilize unlabeled data. Additionally, pseudo-labeled data can be augmented by generating underwater images [4].

This paper addresses these challenges by introducing recent technological advancements in image preprocessing and training frameworks. It incorporates a visual enhancement and joint training-based image preprocessing method, and a semi-supervised object detection framework with high-quality pseudo-labels. Together, these methods improve detection accuracy under the evaluated URPC2022 conditions. In summary, the main contributions of this paper are summarized as follows:

1. A visual enhancement and joint training-based image preprocessing method (VEJT-IP) is proposed, which utilizes existing image visual enhancement algorithms, guided by the optimization of object detection loss, to preprocess underwater images with low brightness, low contrast, and color distortion.
2. A semi-supervised object detection framework with high-quality pseudo-labels (HQPL-SSOD) is proposed. The framework includes a pseudo-label filter based on (LQF), allowing the model to learn not only from labeled data but also from high-quality pseudo-labels selected through the filter, thereby improving model performance.
3. Comprehensive experiments on the URPC2022 dataset indicate that, under the evaluated conditions, the integrated application of the proposed methods achieves a good balance between detection accuracy and annotation cost compared with the evaluated baseline models.

The rest of the paper is organized as follows: Section 2 provides a brief overview of the basic theory behind underwater optical object detection. Section 3 introduces a visual enhancement and joint training-based image preprocessing method, along with a semi-supervised object detection framework utilizing high-quality pseudo-labels. Section 4 presents the experimental evaluation of the proposed methods on the URPC2022 dataset. Finally, the conclusions are presented in Section 5.

2. Related Work

Underwater optical object detection aims to extract features from underwater optical images and classify and locate objects. Existing underwater optical object detection methods are primarily derived from General Object Detection (GOD) [5] algorithms and focus on improving detection performance in underwater environments through three main approaches: data augmentation, image preprocessing, and model structure innovation. At the transmission stage, Im and Pearmain [6] proposed an H.264 error-resilient coding method that prioritizes vulnerable macroblocks to improve video quality under transmission errors. Although their work concerns video transmission rather than object detection, it highlights the importance of preserving visual information before subsequent analysis. In deep learning-based object detection, data augmentation (DA) [7] is a common technique. It increases the diversity and quantity of training data by applying a series

of predefined transformations to the original data. Given the complex and varied nature of underwater optical imaging environments, limited training data often cannot cover all possible scenarios encountered in practical applications, making data augmentation especially crucial. Lin et al. proposed a data augmentation method called RoIMix [8] to simulate the characteristics of overlapping, occlusion, and blurring in underwater organisms. This method applies to the Faster R-CNN's RPN and RoI classifier, mixing RoIs generated by the RPN with random weights and using these mixed samples to train the model. Peng et al. suggested using underwater image restoration as data augmentation to improve object detection accuracy [9], testing the effects of four underwater image restoration methods—GDCP [10], IBLA [11], Water-Net [12], and UWCNN [13]—when applied to YOLOv5. To mitigate the effects of low contrast, blurriness, low brightness, and color deviation in underwater optical images on object detection models, many studies focus on image preprocessing steps to enhance images before inputting them into the object detector. For example, Yeh et al. [14] proposed a method that combines a CNN-based underwater image color conversion module with an object detection model to reduce the impact of color absorption on detection performance by converting color images to grayscale. Han et al. [15] introduced a method combining max-RGB and Gray World techniques for visual enhancement of underwater images and a CNN-based approach to address low-light issues, thereby improving detection performance. Chen et al. [16] highlighted that "image visual restoration can reduce domain shift between training data and real scenes, thereby enhancing online detection performance." Wang et al. [17] proposed a reinforcement learning-based method to configure appropriate visual enhancement algorithms in underwater scenarios to improve detection accuracy. This method models the problem as a Markov decision process, selecting the most suitable visual enhancement algorithm based on the current state to gradually transform raw underwater images into a form suitable for object detection algorithms. More recently, Wen et al. [18] proposed SSD-UIE, a semi-supervised domain-adaptive framework that addresses the gap between synthetic and real underwater images and the variability of real-world conditions.

Recent studies have expanded underwater enhancement from perceptual restoration toward detector-oriented joint optimization. Zhang et al. [19] proposed PU-Net, which uses multi-angle polarization and intensity inputs within dense U-Net variants to recover images in turbid water. Rehman et al. [20] introduced DeepAquaEnhance, combining hierarchical multi-head attention, adaptive feature integration, Transformers, and adversarial learning to improve color, contrast, and structural detail. Sarkar et al. [21] developed UICE-MIRNet to enhance colorfulness and constrain excessive brightness before YOLOv4 detection. More closely related to task-driven processing, Liu et al. [22] proposed UnitModule, a lightweight enhancement module jointly trained with detectors using an unsupervised enhancement loss. These approaches show a progression from perceptual and polarization-based recovery to enhancement designed for downstream detection. Unlike restoration-oriented DeepAquaEnhance and PU-Net, VEJT-IP predicts parameters for three interpretable enhancement functions directly from each input. Its parameters are optimized jointly with the detector using the detection loss, without requiring polarization hardware or a separate perceptual enhancement objective.

Additionally, various works have focused on improving model structures to enhance the performance of general object detection models in underwater scenarios. For instance, to address the issue of varying object scales in underwater detection, Pan et al. proposed M-ResNet [23], which uses multi-scale operations to accurately detect objects of various sizes, especially small ones. To tackle the problem of many small and blurred objects in underwater scenes, Long Chen introduced SwipeNet [24], a network composed of high-resolution and semantically rich super feature maps, significantly improving the detection accuracy of small

objects. Liu et al. developed DP-FishNet [25], a dual-path pyramid visual transformer underwater fish detection network. This method effectively extracts global and local features from underwater images by integrating visual transformers and CNNs and addresses sample imbalance issues with balanced sample loss and multi-scale training. Zhang et al. [26] proposed an improved YOLOv5-based underwater object detection framework, combining techniques such as multi-head self-attention mechanisms, minimal color loss, and local adaptive contrast enhancement to enhance the accuracy of underwater biological identification.

Semi-supervised learning reduces annotation requirements by exploiting unlabeled data. FixMatch generates pseudo-labels from weakly augmented images, retains predictions above a confidence threshold, and enforces prediction consistency on strongly augmented views [27]. For object detection, STAC applies the same core principle to localized objects by combining confidence-based pseudo-label selection with strong data augmentation [28]. More recent teacher–student methods, including Soft Teacher and Unbiased Teacher v2, further improve pseudo-label reliability through online teacher updates and localization-aware selection strategies [29,30]. Nevertheless, classification confidence does not necessarily reflect bounding-box localization quality, particularly for degraded underwater images. In contrast to confidence-only filtering, the proposed HQPL-SSOD framework introduces a localization quality-based filter (LQF) that estimates localization reliability from the agreement among five independently trained detectors with identical architectures.

3. The Proposed Method

In this section, the detailed procedure of the proposed adaptive visual enhancement ensembled semi-supervised learning method is presented, as shown in Figure 1.

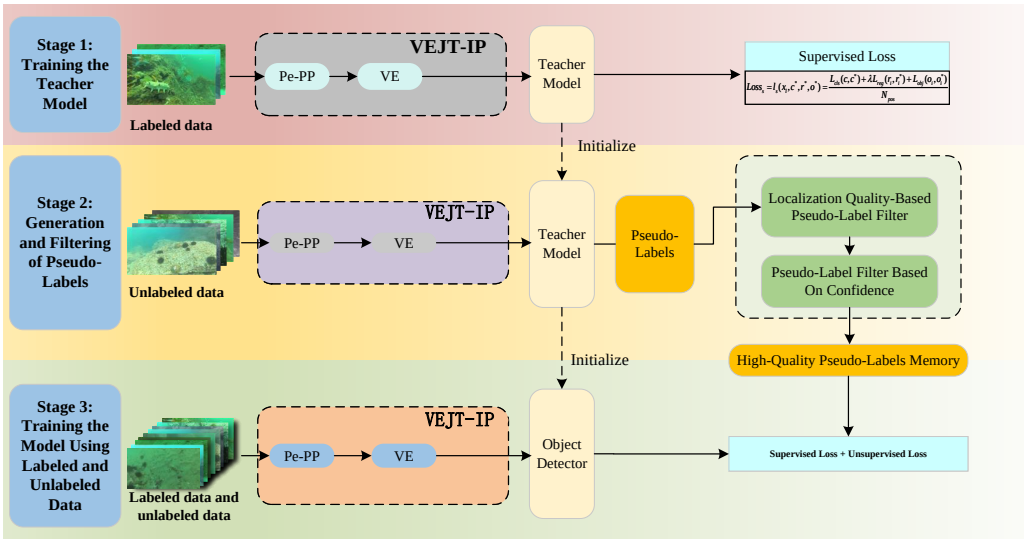


Figure 1. The overview of our proposed method. First, the proposed VEJT-IP method is applied for image preprocessing during model training. Then, a semi-supervised object detection framework based on high-quality pseudo-labels is utilized for self-training, enabling the object detection model to be trained on both labeled and unlabeled data.

To address problems encountered in the research area of underwater object detection, several techniques are adopted, which include data augmentation, image preprocessing, and training frameworks. The method proposed in this paper includes the following two parts. The first part focuses on developing an image preprocessing method based on visual enhancement and joint training (VEJT-IP) (see Section 3.1). This approach simultaneously addresses issues of low image brightness, low contrast, and color distortion through adaptive param-

eter adjustment and aligned visual enhancement. After image preprocessing, the second part focuses on formulating a semi-supervised object detection framework based on high-quality pseudo-labels (HQPL-SSOD) (see Section 3.2). This method effectively eliminates noisy pseudo-labels, thereby providing high-quality data for model training.

3.1. Image Preprocessing Method Based on Visual Enhancement and Joint Training

As shown in Figure 2, the VEJT-IP framework consists of three parts: the Parameter Prediction Module based on PeleeNet [31] (Pe-PP), the Visual Enhancement Module (VE), and the Deep Learning-based Object Detector. The Pe-PP module is a convolutional neural network structure that can adaptively generate parameters for configuring the VE module based on different images. The VE module is responsible for the visual enhancement preprocessing of underwater images, and the processed images serve as input for the object detector. The object detector is responsible for extracting and analyzing features from the preprocessed images, predicting the object's location coordinates and category. The learnable component in VEJT-IP, Pe-PP, is jointly trained with the object detector, using the object detection loss as the total loss function to update the weights of the Pe-PP module through backpropagation. The inference process of the model follows the path indicated by the red arrows in Figure 2. The Pe-PP module, VE module, and object detector work in series. The input image is divided into two paths: one resized to 416×416 or 640×640 for the VE module, and the other resized to 256×256 for the Pe-PP module. The Pe-PP module infers and outputs the configuration parameters for the VE module. Upon receiving these configuration parameters, the VE module performs visual enhancement on the input image of 416×416 or 640×640, and the enhanced image is fed into the object detector. The object detector then infers the VE module's output image and produces the object detection results, completing the task. The training process follows the path indicated by the blue arrows in Figure 2. The main difference between the training and inference stages is the need to update the model parameters. Specifically, both the Pe-PP module and the object detector require parameter updates. During training, the Pe-PP module and the object detection network share the same loss function, which is the object detection loss. The object detector's predictions are used to calculate this loss, and both the Pe-PP module and the object detector are updated to minimize this loss. This design allows the Pe-PP module to learn how to adjust the visual enhancement parameters to maximize object detection performance during training. The joint training mechanism permits the Pe-PP module and the object detector to be optimized together using gradient descent and backpropagation, achieving synchronous parameter updates. By using the object detection loss as the training guide, the optimization goal of the Pe-PP module aligns directly with the final task performance (i.e., object detection accuracy). This method is superior to traditional staged optimization approaches (i.e., performing visual enhancement before object detection) because it takes into account the actual impact of image enhancement on object detection results.

3.1.1. Visual Enhancement Algorithms

Underwater image enhancement is an important branch of image processing, aimed at improving the visual quality of images captured in underwater environments. Due to the unique characteristics of underwater settings, such as light absorption and scattering, underwater images often suffer from color distortion, reduced contrast, and diminished brightness. Underwater visual enhancement algorithms are a crucial component of the method proposed in this chapter. To address the issues of low contrast, low brightness, and color distortion in underwater images, this paper employs three existing basic visual enhancement algorithms. Below are their definitions and functions:

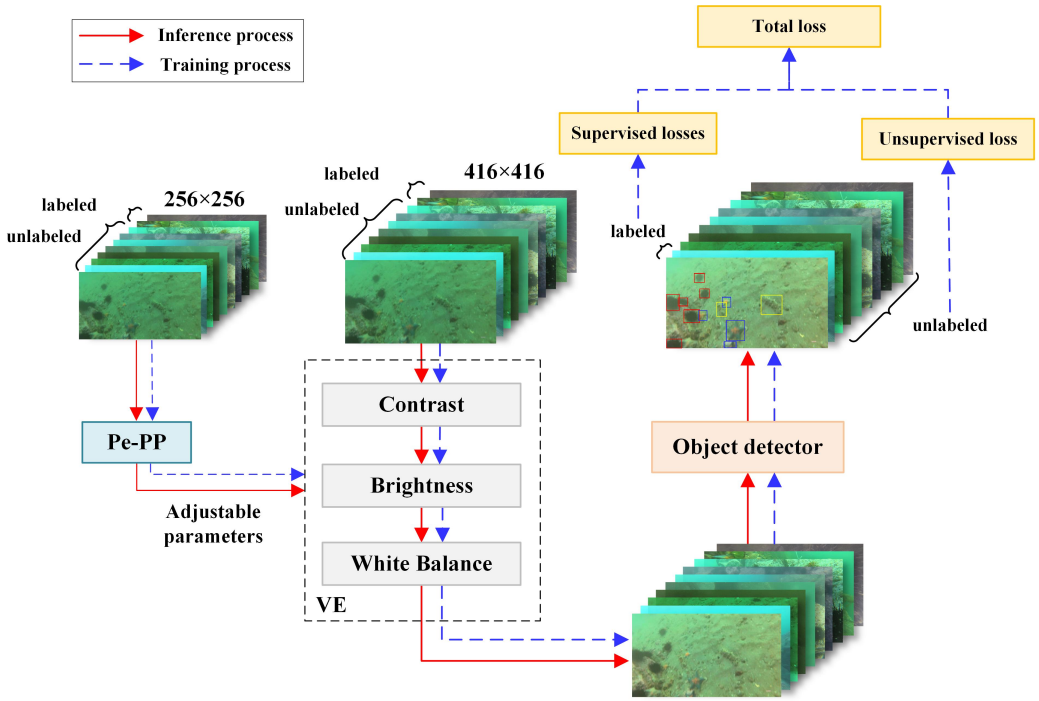


Figure 2. The structure of VEJT-IP.

1. Contrast Adjustment: the goal of the contrast adjustment algorithm is to modify the differences between the light and dark areas in an image. By enhancing the contrast, the image becomes clearer and more detailed, which is especially beneficial in underwater image processing. Contrast adjustment algorithms not only significantly improve the image's clarity and detail but also effectively optimize the adverse visual effects caused by uneven lighting conditions. For underwater images that inherently have low contrast, the contrast-enhanced images can better reveal object contours and texture details. The contrast adjustment algorithm used in this chapter is defined as follows:

$$I_o = \alpha \times I_i + (1 - \alpha) \times \text{mean} \quad (1)$$

Let I_o be the output image, I_i the input image, α the contrast adjustment coefficient, and mean the uniform brightness image. mean has the same shape as the input image, with all pixels having the same brightness, which is the average brightness of the entire image, i.e., the average value of all pixel values. The contrast adjustment coefficient α controls the weighting of the original image and the uniform brightness image in the final result. When $\alpha > 1$, the weight of the original image in the blend increases, resulting in higher contrast; when $\alpha < 1$, the weight of the uniform brightness image increases, resulting in lower contrast. When $\alpha = 1$, the image remains unchanged.

2. Brightness Adjustment: brightness adjustment modifies the brightness value of each pixel in an image to adjust the overall brightness of the image. This technique can address issues in underwater images caused by insufficient or excessive lighting. Adjusting the brightness of underwater images can increase the brightness of low-light areas, helping to reveal details and features that are difficult to observe under natural lighting conditions. This is particularly crucial for accurately locating and identifying underwater objects, as these details are often essential for distinguishing between different objects or identifying specific ones. The brightness adjustment algorithm used in this paper is defined as follows:

$$I_o = \beta \times I_i + (1 - \beta) \times \text{dark} \quad (2)$$

Let I_o be the output image, I_i the input image, β the brightness adjustment coefficient, and dark the dark image. dark has the same shape as the input image, with all pixels having a brightness of 0, making it a "black" image. The original image is blended with the black image, with the brightness adjustment coefficient β controlling the blend ratio. When $\beta > 1$, the image brightness increases; when $\beta = 1$, the image remains unchanged; when $\beta < 1$, the image brightness decreases. By adjusting the value of β, precise control over the image brightness can be achieved. It is important to note that this blending process is actually implemented by multiplying each pixel value of the original image by β, and then adding a zero matrix scaled by $1 - \beta$, effectively adjusting the image brightness proportionally.

3. White Balance: underwater images often exhibit a blue-green tint due to the absorption of red light by water. White balance adjustment aims to correct this color imbalance by adjusting the intensities of the red, green, and blue color components in the image, ensuring that white objects appear white and thereby restoring the natural colors of the image. In this chapter, the image adjustment is tailored to the practical needs of the object detection model; hence, the white balance processing in this paper does not necessarily aim to restore the natural colors of the image. The white balance algorithm used in this paper is defined as follows:

$$\begin{aligned} I_o[:,:,0] &= b \times I_i[:,:,0], \\ I_o[:,:,1] &= g \times I_i[:,:,1], \\ I_o[:,:,2] &= r \times I_i[:,:,2]. \end{aligned} \tag{3}$$

where I_o is the output image and I_i is the input image. r, g, and b represent the gain coefficients for the red, green, and blue channels, respectively. These gain coefficients can be adjusted according to the actual conditions. The brightness values of the red, green, and blue channels are multiplied by the corresponding gain coefficients r, g, and b, respectively, to adjust the brightness proportion of each channel and thereby achieve color adjustment.

In the implementation, input images are normalized to the range $[0, 1]$. For each transformation in Equations (1)–(3), the algebraic operation first produces a pre-clipped output $\widetilde{I}_o$, after which the pixel values are restricted to the valid range as follows:

$$I_o = \mathrm{clip}(\widetilde{I}_o, 0, 1).$$

The clipping operation is piecewise differentiable. Its derivative with respect to the pre-clipped value is

$$\frac{\partial I_o}{\partial \widetilde{I}_o} = \begin{cases} 1, & 0 < \widetilde{I}_o < 1, \\ 0, & \widetilde{I}_o < 0 \text{ or } \widetilde{I}_o > 1. \end{cases}$$

At the two boundary points, the automatic-differentiation framework uses its defined subgradient. Therefore, gradients propagate normally through unsaturated pixels, whereas saturated pixels contribute zero gradients. Accordingly, the complete enhancement pipeline with clipping is differentiable almost everywhere rather than fully differentiable.

3.1.2. Parameter Prediction Module (Pe-PP)

The Parameter Prediction Module (Pe-PP) is the core component of the VEJT-IP architecture. Its function is to predict the hyperparameters needed by the Visual Enhancement Module (VE), adaptively generating different hyperparameters for different underwater images to achieve adaptive image enhancement. Essentially, Pe-PP is a lightweight convolutional neural network called PeleeNet. PeleeNet is an efficient network built using

traditional convolutions, introduced by Robert J. Wang et al. in 2018 as a variant of the DenseNet [32] architecture.

PeleeNet, in addition to having a lightweight network structure, inherits several key characteristics of DenseNet, such as feature reuse and deep supervision. Feature reuse enables PeleeNet to comprehensively and efficiently understand and capture the details and characteristics of images, which is crucial for accurately generating hyperparameters for specific image enhancement tasks. Deep supervision allows PeleeNet to effectively learn and extract deep and high-level features that are critical for image enhancement tasks, thereby generating more precise hyperparameters for different types of underwater environment images. These advantages enable PeleeNet to perform the parameter prediction task with minimal additional training and inference time.

3.1.3. Visual Enhancement Module (VE)

The Visual Enhancement Module (VE) is also a core component of the VEJT-IP architecture, responsible for performing visual enhancement processing on the input images. This paper identifies a set of effective visual enhancement algorithms, including Contrast Adjustment, Brightness Adjustment, and White Balance. These three visual enhancement algorithms are used to construct the VE module.

The algebraic forms of the three enhancement operations are differentiable, and the complete pipeline with pixel clipping is differentiable almost everywhere, as described in Section 3.1.1. Consequently, the object detection loss can be backpropagated through the VE module to train the Pe-PP module. To prevent uncontrolled growth of the predicted enhancement parameters, each raw network output z_θ is mapped to a predefined bounded interval using a sigmoid function:

$$\theta = \theta_{\min} + (\theta_{\max} - \theta_{\min})\sigma(z_\theta), \qquad \theta \in \{\alpha, \beta, r, g, b\}.$$

Here, $\sigma(\cdot)$ denotes the sigmoid function, and $\theta_{\min}$ and $\theta_{\max}$ are the lower and upper bounds specified for the corresponding parameter. This parameterization guarantees that the contrast coefficient α, brightness coefficient β, and channel gains r, g, and b cannot grow without bound during joint training. In addition, the training objective includes the parameter-regularization term $\mathcal{L}_{\text{reg}}$ with weight λ_{reg}:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{det}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}}.$$

The regularization term discourages the predicted enhancement parameters from remaining near their interval boundaries, while the sigmoid mapping provides the hard bounds. Together, these mechanisms stabilize joint optimization. The parameters of the visual enhancement operations are independent of image resolution, allowing the input to the Pe-PP module to be resized to 256×256 to reduce computational cost.

As shown in Figure 2, within the VE module, the three visual enhancement algorithms are applied in sequence: Contrast, Brightness, and White Balance. The input image undergoes contrast adjustment, brightness adjustment, and white balance operations in this order. The adjustable parameters for each visual enhancement algorithm are listed in Table 1: Contrast and Brightness each have one adjustable parameter, while White Balance has three adjustable parameters. The optimal parameter values for different input images are predicted by the Pe-PP module, which can automatically select the most suitable parameter configuration based on the specific characteristics of the image, thereby achieving adaptive image enhancement.

Table 1. Adjustable Parameters for Different Visual Enhancement Algorithms.

Visual Enhancement Algorithms	Parameters
Contrast	α
Brightness	β
White Balance	b, g, r

3.2. Semi-Supervised Object Detection Framework Based on High-Quality Pseudo-Labels

HQPL-SSOD follows the three-stage self-training paradigm of STAC [28]. Like STAC, it uses classification confidence to remove unreliable pseudo-labels and applies strong augmentation to pseudo-labeled images. However, confidence-only filtering primarily reflects classification certainty and cannot directly assess bounding-box localization reliability. HQPL-SSOD therefore introduces LQF as a complementary filtering criterion. LQF measures the agreement among bounding boxes predicted by five independently trained detectors with identical architectures, and only pseudo-labels satisfying both the confidence and localization quality thresholds are retained.

As shown in Figure 3, the framework contains three stages. First, all available labeled data are used to train the teacher model with supervised loss. Second, the trained teacher model predicts rough pseudo-labels for the unlabeled data, which are filtered according to confidence and localization quality. The retained pseudo-labels contain bounding-box coordinates and object categories in the same format as the manual annotations. Third, the labeled data and the unlabeled data with retained pseudo-labels are used to train the detector from scratch. Strong augmentation, including global color transformation, CoarseDropout, Mosaic, and random affine transformation, is applied to the pseudo-labeled data. The detector parameters are then updated by minimizing the weighted sum of the supervised and unsupervised losses.

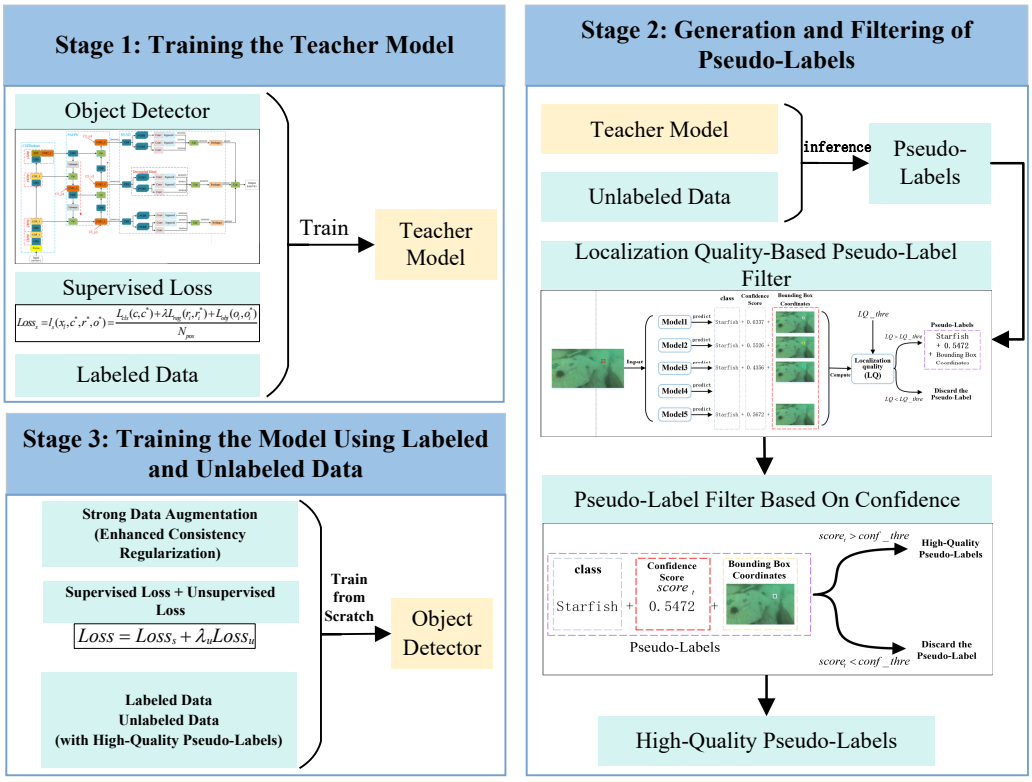


Figure 3. The structure of HQPL-SSOD.

HQPL-SSOD inherits two hyperparameters from STAC: $\text{conf}_{\text{thre}}$, the confidence threshold, and λ_u, the unsupervised loss weight. It additionally introduces LQ_{thre}, the localization quality threshold. These three hyperparameters control the classification confidence of retained pseudo-labels, the contribution of the unsupervised loss, and the required localization reliability, respectively.

3.2.1. Localization Quality Estimation and Filtering Procedure

As shown in Figure 4, the localization quality-based pseudo-label filter includes the following key steps:

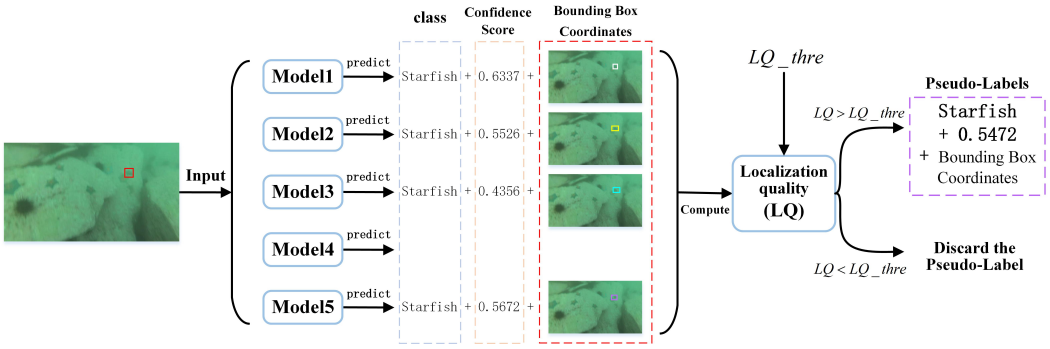


Figure 4. Localization quality-based pseudo-label filter.

Step 1: Place the N ($N \leq 5$) sets of bounding box coordinates predicted by five models into set B. The corresponding N class predictions and N confidence scores are placed in sets C and S, respectively.

Step 2: For the bounding boxes $B = \{b_1, b_2, \ldots, b_N\}$, calculate the intersection over union (IoU) for each pair of bounding boxes and compute the mean value to obtain the localization quality LQ.

Step 3: Determine whether the value of LQ is greater than the predefined localization quality threshold LQ_{thre}. If $LQ > LQ_{\text{thre}}$, retain the pseudo-label for the object; otherwise, discard it.

Step 4: If the pseudo-label for the object is retained, take the mean of the bounding box coordinates in set B as the bounding box coordinates in the pseudo-label.

Step 5: Use majority voting to select the class prediction that appears most frequently in set C as the class in the pseudo-label.

Step 6: Take the highest confidence score in set S as the confidence score in the pseudo-label. This value is carried forward as the confidence assigned to the retained pseudo-label.

At this point, whether to retain the pseudo-label for the object and the position coordinates, class, and confidence score of the retained pseudo-label are determined.

The maximum score is used as a simple heuristic to avoid attenuating the strongest classification response through averaging, rather than as evidence that this rule is optimal. A pseudo-label is not accepted on this score alone: its class is determined by majority voting, and its predicted boxes must also satisfy the localization quality threshold. These criteria can reduce the influence of an isolated high-confidence prediction. Nevertheless, an incorrect prediction may still be overconfident, and maximum-score selection can preserve this error. The present study does not compare maximum selection with mean or weighted confidence aggregation and post-hoc calibration, so no superiority over these alternatives is claimed. A controlled comparison of these confidence aggregation strategies will be considered in future work.

Localization quality is defined based on the prediction results of five homomorphic models with different parameters for a given object. First, the 9000 labeled data from the URPC2022 dataset are randomly divided into training, validation, and test sets in a ratio of

7:1:2, and this random division is repeated five times to obtain five different training sets. The same object detector is trained on these five different training sets, resulting in five object detectors with the same structure but different parameters. Except for the different training datasets, all other training configurations are identical.

After obtaining the five homomorphic models, the N ($N \leq 5$) sets of bounding box coordinates predicted by the five models for a given object are placed in set B. Let the bounding boxes in B be $b_1, b_2, \ldots, b_N$. The intersection over union (IoU) is calculated for each pair of bounding boxes in B, and the mean value is computed, yielding the localization quality LQ for the pseudo-label of the given object, as given in Equation (4):

$$LQ = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathrm{IoU}(b_i, b_j) \tag{4}$$

where the pairwise IoU is defined in Equation (5) as

$$\mathrm{IoU}(b_i, b_j) = \frac{\mathrm{Area}(b_i \cap b_j)}{\mathrm{Area}(b_i \cup b_j)} \tag{5}$$

The coefficient $\frac{2}{N(N-1)}$ is used because, for each pair of bounding boxes in set B, the IoU between them needs to be calculated. In a set containing N elements, there are $\frac{N(N-1)}{2}$ different pairing combinations. This coefficient ensures that, when calculating the average, all possible pairs of bounding boxes in the set are considered.

3.2.2. Generation and Filtering of Pseudo-Labels

In the second stage, the teacher model generates rough pseudo-labels for the unlabeled images. As shown in Figure 5, these candidates are first screened by classification confidence and are then evaluated by the localization quality-based filter described in Section 3.2.1.

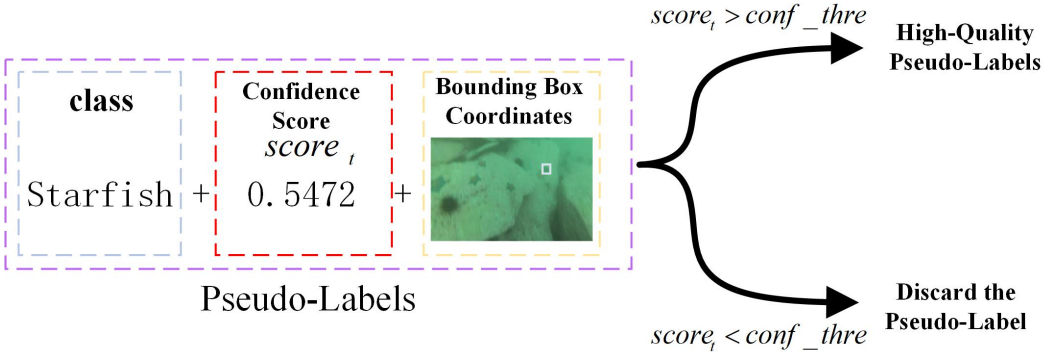


Figure 5. Confidence-based pseudo-label filter.

For each predicted bounding box b and its predicted class c, the confidence score is calculated using Equation (6):

$$s_{\mathrm{conf}}(b, c) = p_{\mathrm{obj}}(b) \times p_{\mathrm{cls}}(c \mid b) \tag{6}$$

where $p_{\mathrm{obj}}(b)$ denotes the predicted probability that bounding box b contains an object, and $p_{\mathrm{cls}}(c \mid b)$ denotes the conditional probability that the object belongs to class c. A candidate pseudo-label is retained by the confidence-based filter when the condition in Equation (7) is satisfied:

$$s_{\mathrm{conf}}(b, c) \geq \mathrm{conf}_{\mathrm{thre}} \tag{7}$$

where $\text{conf}_{\text{thre}}$ is the predefined confidence threshold. Candidate pseudo-labels that do not satisfy this criterion are discarded. The retained candidates are subsequently evaluated using the localization quality-based filter, and only pseudo-labels satisfying both the confidence and localization quality thresholds are used for semi-supervised training.

4. Experiments

4.1. Experimental Setup

The GPU is GeForce RTX 3090 with a single card memory of 24 GB, and the system has two cards. The CPU is Intel(R) Xeon(R) Silver 4214R CPU @ 2.40 GHz with 12 cores and 90 GB of memory. The operating system is Ubuntu 20.04, and the programming language is Python 3.8. The deep learning framework used is PyTorch 2.0.0, and the underlying library is CUDA 11.8. The experimental dataset is URPC2022, which contains 9000 labeled data samples. These 9000 labeled samples are divided into training set (train), validation set (val), and test set (test) in a 7:2:1 ratio, respectively. The training set, validation set, and test set contain 6300, 1800, and 900 samples, respectively. Conduct experiments using YOLOX [33] as the object detector. The backbone network CSPDarknet of YOLOX is initialized using the pretrained weights on the MSCOCO dataset. A total of 200 epochs are trained with a batch size of 8, using the cosine learning rate schedule. The following ablation experiments are designed to validate the effectiveness of the VEJT-IP method proposed in this chapter for improving object detection performance and to assess the necessity of its internal components:

a. YOLOX baseline model.
b. YOLOX + Pe-PP + Contrast, abbreviated as P + C.
c. YOLOX + Pe-PP + Brightness, abbreviated as P + B.
d. YOLOX + Pe-PP + White Balance, abbreviated as P + W.
e. YOLOX + VEJT-IP, abbreviated as VEJT-IP.

In these five groups of experiments, we used the medium-sized YOLOX-s as the object detector. Experiment a uses the original YOLOX-s as the object detection model for training and testing. Experiments b, c, and d use the VEJT-IP method proposed in this paper for image preprocessing, but each retains only one visual enhancement algorithm in the VE module. Experiment e uses the VEJT-IP method, which integrates all visual enhancement algorithms, for image preprocessing.

To validate the effectiveness of the VEJT-IP image processing method proposed in this paper on object detection models of different scales, the following comparative experiments were designed:

a. YOLOX-Nano.
b. YOLOX-Nano + VEJT-IP.
c. YOLOX-s.
d. YOLOX-s + VEJT-IP.
e. YOLOX-l.
f. YOLOX-l + VEJT-IP.

These six groups of experiments are organized as follows: Experiments a and b form one group, c and d form another group, and e and f form the third group. Each group verifies the effectiveness of VEJT-IP on the three model sizes, YOLOX-Nano, YOLOX-s, and YOLOX-l, respectively.

The hyperparameter settings for HQPL-SSOD are shown in Table 2. We used $\text{conf}_{\text{thre}} = 0.5$, $\text{LQ}_{\text{thre}} = 0.7$, and $\lambda_u = 2.0$ as a single operating point. Increasing either threshold makes pseudo-label selection more conservative, whereas decreasing it retains more pseudo-labels but may introduce additional noise. Similarly, increasing λ_u

strengthens the contribution of pseudo-labeled data but can also amplify pseudo-label errors, while a smaller value reduces the benefit of unlabeled data. The three values were kept unchanged in all reported HQPL-SSOD experiments and were not tuned separately for the evaluated YOLOX model scales. To compare different pseudo-label filtering strategies, we conducted three experiments using YOLOX-s, 800 labeled images, and 7000 unlabeled images:

a. Supervised YOLOX-s.
b. STAC-style confidence-only filtering (HQPL-SSOD without LQF).
c. Confidence-plus-localization quality filtering (HQPL-SSOD with LQF).

Table 2. Hyperparameter Settings of the HQPL-SSOD Framework.

Hyperparameter	Value
$\text{conf}_{\text{thre}}$	0.5
LQ_{thre}	0.7
λ_u	2.0

All three experiments used the same data partition, detector, training schedule, and evaluation protocol. Experiment b retained pseudo-labels using only the confidence threshold, whereas Experiment c additionally applied the proposed localization quality threshold. Therefore, the comparison between Experiments b and c isolates the contribution of LQF.

To validate the effectiveness of the HQPL-SSOD framework for object detection models of different scales, the following comparative experiments are designed:

a. YOLOX-Nano.
b. YOLOX-Nano + HQPL-SSOD.
c. YOLOX-s.
d. YOLOX-s + HQPL-SSOD.
e. YOLOX-l.
f. YOLOX-l + HQPL-SSOD.

With 800 labeled samples and 7000 unlabeled samples, experiments a–f are compared to validate the effectiveness of the HQPL-SSOD framework for object detection models of different scales.

4.2. Quantitative Evaluations

Table 3 records the results of the ablation experiments. In Experiment a, YOLOX-s is used as the baseline for the entire ablation study. Observing the results from Experiments c, d, and e, we can see that each visual enhancement algorithm, when used individually within the VEJT-IP framework, improves the mAP of the object detection model. This proves the necessity of each visual enhancement algorithm in VEJT-IP. However, the improvements in both mAP@0.5 and mAP@0.5:0.95 are not significant when using a single visual enhancement algorithm. In contrast, in Experiment f, where the complete VEJT-IP framework incorporating all three visual enhancement algorithms is used, the results show a much more significant improvement in mAP compared to using a single visual enhancement algorithm. Specifically, mAP@0.5 increased by 1.03%, and mAP@0.5:0.95 increased by 0.81% compared to the baseline experiment. This result demonstrates the VEJT-IP framework's capability in enhancing object detection accuracy. The improvement in mAP@0.5:0.95 indicates that VEJT-IP can improve mAP under different localization accuracy requirements (corresponding to different IoU thresholds).

Table 3. Results of Ablation Experiments on Different Visual Enhancement Algorithms.

Exp.	P+C	P+B	P+W	mAP@0.5 (%)	mAP@0.5:0.95 (%)
a				80.95	47.79
b	✓			81.32	47.98
c		✓		81.45	48.02
d			✓	81.43	47.92
e	✓	✓	✓	**81.98**	**48.60**

Note: P, Pe-PP; C, contrast adjustment; B, brightness adjustment; W, white-balance adjustment. A checkmark indicates that the corresponding component is included. Bold values indicate the best result for each metric.

Table 4 summarizes the results of the three paired comparisons in Experiment 2. Experiments a and b evaluated YOLOX-Nano, experiments c and d evaluated YOLOX-s, and experiments e and f evaluated YOLOX-l. Within each pair, VEJT-IP was the only added component, allowing its contribution to be assessed under the same detector scale.

Table 4. Experimental Results of Models at Different Scales Before and After Applying VEJT-IP.

Exp.	Method	mAP@0.5 (%)	mAP@0.5:0.95 (%)
a	YOLOX-Nano	70.38	37.74
b	YOLOX-Nano + VEJT-IP	**71.35**	**37.96**
c	YOLOX-s	80.95	47.79
d	YOLOX-s + VEJT-IP	**81.98**	**48.60**
e	YOLOX-l	81.80	50.03
f	YOLOX-l + VEJT-IP	**82.36**	**50.42**

Note: Bold values indicate the higher result within each baseline/VEJT-IP pair.

After introducing VEJT-IP, mAP@0.5 increased by 0.97%, 1.03%, and 0.56% for YOLOX-Nano, YOLOX-s, and YOLOX-l, respectively. The corresponding mAP@0.5:0.95 gains were 0.22%, 0.81%, and 0.39%, respectively. The largest gains were observed for YOLOX-s, which improved by 1.03% in mAP@0.5 and 0.81% in mAP@0.5:0.95.

All three model scales achieved improvements on both metrics after VEJT-IP was introduced. Thus, the benefit was not confined to a single evaluated YOLOX configuration. The results indicate that adaptive adjustment of brightness, contrast, and color can improve detection under the evaluated underwater imaging conditions. Overall, VEJT-IP served as an effective preprocessing component across YOLOX-Nano, YOLOX-s, and YOLOX-l.

The improvements for YOLOX-Nano and YOLOX-l are modest and should be considered together with the additional preprocessing cost. VEJT-IP adds Pe-PP inference and three pixel-level enhancement operations, although the lightweight PeleeNet design and 256 × 256 Pe-PP input are intended to limit this overhead. Because platform-specific latency and energy consumption were not measured, the present results do not establish that the accuracy gains justify the added cost in every deployment. The trade-off may be acceptable in accuracy-oriented applications with strongly degraded images, whereas the original detector may be preferable under strict latency or energy constraints.

Table 5 compares the supervised baseline, STAC-style confidence-only pseudo-label filtering, and the proposed confidence-plus-localization quality filtering strategy. Compared with supervised YOLOX-s, confidence-only filtering improved mAP@0.5 from 68.80% to 72.01% and mAP@0.5:0.95 from 36.17% to 39.59%. After incorporating LQF, the corresponding results further increased to 74.73% and 42.13%, representing improvements of 2.72 and 2.54 percentage points over confidence-only filtering, respectively. Because the two semi-supervised variants used the same experimental settings and differed only in whether LQF was applied, these results indicate that localization-quality information provides a useful complement to classification confidence when filtering pseudo-labels.

Table 5. Comparison of Pseudo-Label Filtering Strategies on URPC2022 Using YOLOX-s with 800 Labeled and 7000 Unlabeled Images.

Method	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Supervised YOLOX-s	68.80	36.17
STAC-style confidence-only filtering	72.01	39.59
HQPL-SSOD (confidence + LQF)	**74.73**	**42.13**

Note: Bold values indicate the best result among the compared pseudo-label filtering strategies.

Table 5 is a component-level ablation rather than a comprehensive benchmark against recent end-to-end SSOD frameworks. Soft Teacher and Unbiased Teacher v2 were discussed in Section 2 but were not reimplemented under the same split of 800 labeled and 7000 unlabeled images. Directly importing their published results would be inappropriate because the backbones, data partitions, and training protocols differ. Therefore, the current comparison supports the incremental value of LQF over matched confidence-only self-training but does not establish superiority over recent SSOD methods. A unified comparison under identical data splits and training settings remains an important direction for future work.

LQ should be interpreted as a consistency-based proxy for localization reliability rather than direct proof of bounding-box correctness. High agreement can still occur when the five detectors share a systematic localization bias. Training the detectors on different random data partitions introduces model diversity, which can reduce but cannot eliminate correlated errors. The matched comparison in Table 5 provides indirect task-level evidence: adding LQF to confidence-only filtering increased mAP@0.5 by 2.72 percentage points and mAP@0.5:0.95 by 2.54 percentage points. These gains suggest that agreement-based selection tends to retain more useful pseudo-labels under the evaluated setting. However, this ablation does not establish a direct sample-level correlation between LQ and ground-truth IoU. A dedicated correlation analysis and ensembles with more diverse architectures will be considered in future work.

Table 6 records the experimental results of applying HQPL-SSOD to different scales of YOLOX models using the data division scheme c (800 labeled data and 7000 unlabeled data). The results show that HQPL-SSOD can significantly improve the mAP@0.5 and mAP@0.5:0.95 of YOLOX-Nano, YOLOX-s, and YOLOX-l when using 800 labeled data. Specifically, the mAP@0.5 of YOLOX-Nano, YOLOX-s, and YOLOX-l increased by 4.11%, 5.93%, and 3.53%, respectively, while mAP@0.5:0.95 increased by 3.38%, 5.96%, and 3.79%, respectively. These consistent improvements provide limited indirect evidence that the selected operating point transfers across the evaluated detector scales. However, this comparison is not a controlled sensitivity analysis, and performance stability around 0.5, 0.7, and 2.0 has not been quantified. We therefore make no claim of hyperparameter insensitivity; a one-factor-at-a-time parameter sweep will be included in future work.

Table 6. Experimental Results of Models at Different Scales Before and After Applying HQPL-SSOD

Method	mAP@0.5 (%)	mAP@0.5:0.95 (%)
YOLOX-Nano	55.77	27.55
YOLOX-Nano + HQPL-SSOD	**59.88**	**30.93**
YOLOX-s	68.80	36.17
YOLOX-s + HQPL-SSOD	**74.73**	**42.13**
YOLOX-l	71.80	38.75
YOLOX-l + HQPL-SSOD	**75.33**	**42.54**

Note: Bold values indicate the higher result within each baseline/HQPL-SSOD pair.

4.3. Qualitative Evaluations

Figure 6 shows a visual comparison of the YOLOX-s detection results before and after applying VEJT-IP. The left and right columns show the results without and with VEJT-IP, respectively, and the red dashed ellipses highlight representative differences. Some enhanced images appear darker or greener than the raw inputs, although their detection results improve. This occurs because the enhancement parameters are optimized by the detection loss rather than by a perceptual image-quality objective. At the feature level, the resulting brightness, contrast, and channel adjustments may suppress background interference and strengthen task-relevant edges, textures, and object–background differences, thereby facilitating classification and localization. Consequently, VEJT-IP may favor detector-specific visual states that are not necessarily brighter, more natural, or more pleasing to human observers. This distinguishes task-driven enhancement from conventional perceptual enhancement, which primarily aims to improve visual naturalness, color fidelity, and human interpretability. The present evaluation therefore demonstrates task-level utility rather than perceptual image-quality improvement. PUIQA, UIQM, and UCIQE were not calculated, so we do not claim that VEJT-IP consistently improves these metrics or produces more natural-looking images. Future work will report both detection performance and underwater image-quality metrics to characterize the relationship between task-oriented and perceptual enhancement.

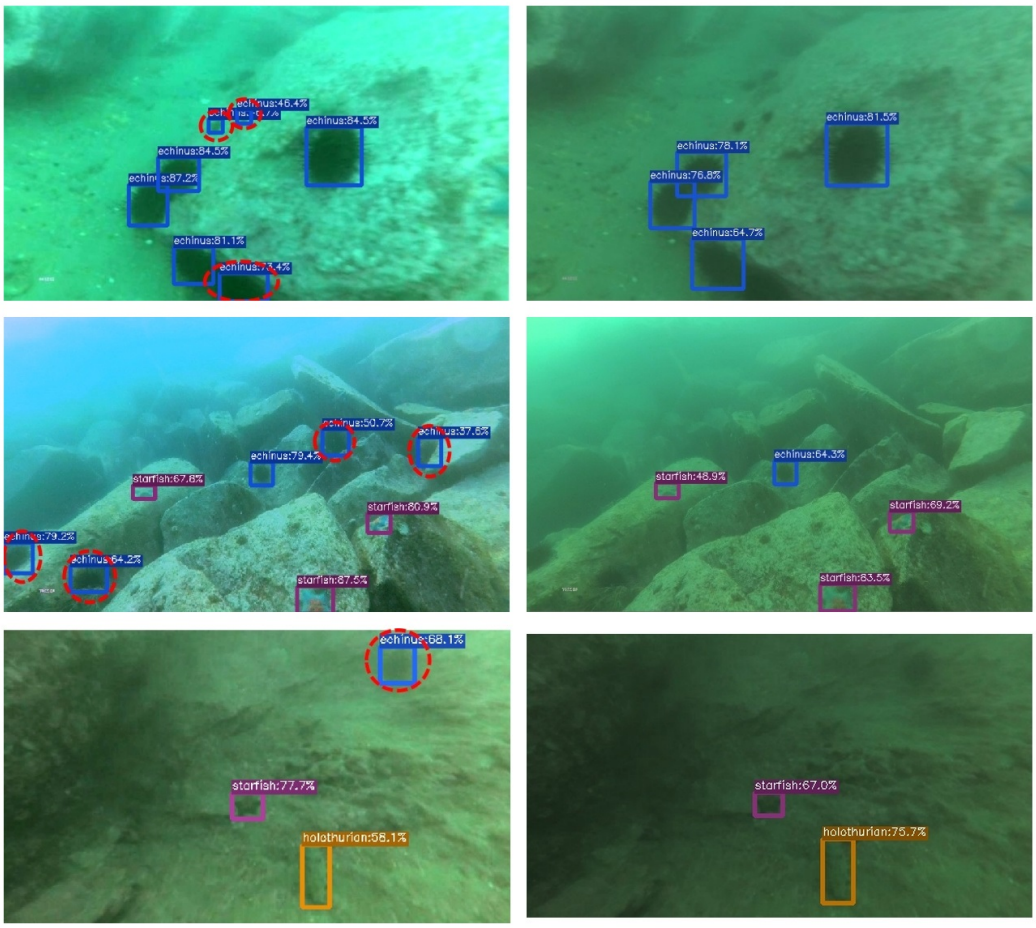


Figure 6. *Cont.*

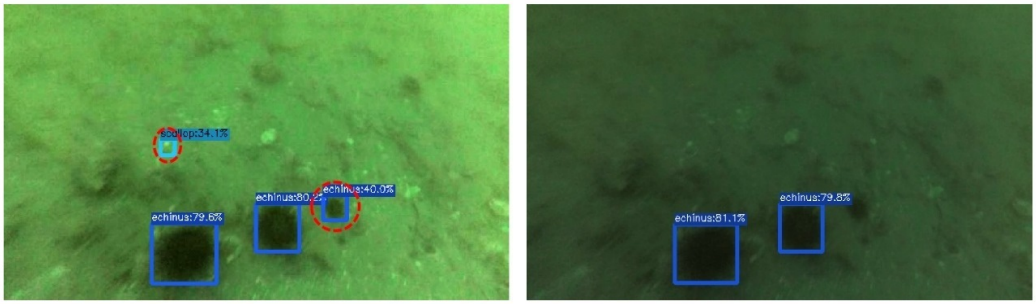


Figure 6. Visual comparison of YOLOX-s object detection results before and after applying VEJT-IP. The red dashed ellipses mark representative low-quality detections of the echinus class in the results obtained without VEJT-IP.

5. Conclusions

This study proposed VEJT-IP for adaptive visual enhancement and HQPL-SSOD for semi-supervised underwater object detection. Experiments on URPC2022 showed that the proposed components improved detection performance across the evaluated YOLOX model scales and labeled–unlabeled data configurations. These results support the effectiveness of the method under the evaluated URPC2022 conditions, but they do not establish cross-dataset generalizability.

A limitation of this study is that all experiments were conducted exclusively on URPC2022. Differences among underwater datasets in turbidity, depth, illumination, acquisition devices, object categories, and annotation protocols may affect detection performance. Fair comparison with DUO, UTDAC2020, and Brackish requires harmonized categories, annotations, and data partitions. Therefore, the conclusions are restricted to the evaluated URPC2022 setting. Future work will evaluate additional underwater datasets and cross-dataset transfer under diverse imaging conditions. In addition, the potential for deployment on mobile or embedded platforms remains insufficiently explored. Future work will also investigate lightweight algorithmic optimization to facilitate efficient inference on resource-constrained edge devices.

Author Contributions: Conceptualization, W.H., X.Z., Y.Z., Y.T. and J.Z.; Methodology, W.H., X.Z. and Y.Z.; Software, W.H., Y.Z., Y.T. and J.Z.; Validation, W.H., X.Z., Y.Z., Y.T. and J.Z.; Formal analysis, W.H., X.Z. and Y.T.; Investigation, W.H., X.Z., Y.T. and J.Z.; Resources, W.H., X.Z., Y.Z., Y.T. and J.Z.; Data curation, Y.Z., Y.T. and J.Z.; Writing—original draft, W.H., X.Z., Y.Z. and Y.T.; Writing—review & editing, W.H. and Y.Z.; Visualization, X.Z. and Y.T.; Supervision, W.H., X.Z., Y.Z. and Y.T.; Project administration, W.H., X.Z., Y.Z. and Y.T.; Funding acquisition, W.H., Y.T. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Key Research and Development Program of Shaanxi (Grant No. 2026CY-YBXM-687).

Data Availability Statement: The publicly available URPC2022 dataset was analyzed in this study and can be accessed through the 2022 China Underwater Robot Professional Contest platform at https://challenge.datacastle.cn/v3/cmptDetail.html?id=680. No new primary datasets were generated during the current study.

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

References

1. Lee, S.; Kim, M.; Seol, S.; Chung, J. AoI Minimization Scheduling Using Integrated Collection-Relay in Multi-AUV Multi-Hop Underwater Networks. *Electronics* **2026**, *15*, 2930 . [CrossRef]
2. Ma, M.; Zhou, Y.; Hu, Y.; Bu, L. Dual-Transmitter Wireless Power Transfer Based on Parity–Time Symmetry for Rapid and Reliable Deep-Sea AUV Recharging. *Electronics* **2026**, *15*, 1228. [CrossRef]

3. Arnaoutoglou, D.G.; Empliouk, T.M.; Papamoschou, D.-N.; Kyriacou, Y.; Papanastasiou, A.; Kaifas, T.N.F.; Kyriacou, G.A. Underwater Antenna Technologies with Emphasis on Submarine and Autonomous Underwater Vehicles (AUVs). *Electronics* **2026**, *15*, 219. [CrossRef]
4. Yan, H.; Zhang, H.; Shi, J.; Ma, J.; Xu, X. Inspiration Transfer for Intelligent Design: A Generative Adversarial Network With Fashion Attributes Disentanglement. *IEEE Trans. Consum. Electron.* **2023**, *69*, 1152–1163. [CrossRef]
5. Wang, G.; Sun, J.; Liu, Y.; Huang, M.; Zhang, P. Mamba-KGSC: Knowledge-Guided Semantic Communication for Robust V2V Cooperative Object Detection. *Electronics* **2026**, *15*, 2925. [CrossRef]
6. Im, S.K.; Pearmain, A.J. Error Resilient Video Coding with Priority Data Classification Using H.264 Flexible Macroblock Ordering. *IET Image Process.* **2007**, *1*, 197–204. [CrossRef]
7. Liu, P.; Wang, Z.; Wang, H.; Jiang, B.; Yu, W. Efficient Data Construction Method for Electromagnetic Fault Diagnosis of Radar Timing Control Boards. *Electronics* **2026**, *15*, 3166. [CrossRef]
8. Lin, W.H.; Zhong, J.X.; Liu, S.; Li, T.; Li, G. ROIMIX: Proposal-Fusion Among Multiple Images for Underwater Object Detection. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Online, 4–8 May 2020 ; pp. 2588–2592. [CrossRef]
9. Peng, W.; Peng, Y.T.; Lien, W.C.; Chen, C.S. Unveiling of How Image Restoration Contributes to Underwater Object Detection. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Penghu, Taiwan, 15–17 September 2021; pp. 1–2.
10. Peng, Y.T.; Cao, K.; Cosman, P.C. Generalization of the Dark Channel Prior for Single Image Restoration. *IEEE Trans. Image Process.* **2018**, *27*, 2856–2868. [CrossRef]
11. Peng, Y.T.; Cosman, P.C. Underwater Image Restoration Based on Image Blurriness and Light Absorption. *IEEE Trans. Image Process.* **2017**, *26*, 1579–1594. [CrossRef]
12. Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An Underwater Image Enhancement Benchmark Dataset and Beyond. *IEEE Trans. Image Process.* **2020**, *29*, 4376–4389. [CrossRef]
13. Li, C.; Anwar, S.; Porikli, F. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognit.* **2020**, *98*, 107038. [CrossRef]
14. Yeh, C.H.; Lin, C.H.; Kang, L.W.; Huang, C.H.; Lin, M.H.; Chang, C.Y.; Wang, C.C. Lightweight Deep Neural Network for Joint Learning of Underwater Object Detection and Color Conversion. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6129–6143. [CrossRef]
15. Han, F.; Yao, J.; Zhu, H.; Wang, C. Underwater Image Processing and Object Detection Based on Deep CNN Method. *J. Sens.* **2020**, *2020*, 6707328. [CrossRef]
16. Chen, X.; Lu, Y.; Wu, Z.; Yu, J.; Wen, L. Reveal of Domain Effect: How Visual Restoration Contributes to Object Detection in Aquatic Scenes. *arXiv* **2020**, arXiv:2003.01913. [CrossRef]
17. Wang, H.; Sun, S.; Bai, X.; Wang, J.; Ren, P. A Reinforcement Learning Paradigm of Configuring Visual Enhancement for Object Detection in Underwater Scenes. *IEEE J. Ocean. Eng.* **2023**, *48*, 443–461. [CrossRef]
18. Wen, J.; Yang, G.; Zhao, B.; Huang, D.; Lei, L.; Zhang, B.; Gao, Z.; Chen, X.; Chen, B.M. A Semi-Supervised Domain-Adaptive Framework for Real-World Underwater Image Enhancement. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4209715. [CrossRef]
19. Zhang, R.; Gui, X.; Cheng, H.; Chu, J. Underwater Image Recovery Utilizing Polarimetric Imaging Based on Neural Networks. *Appl. Opt.* **2021**, *60*, 8419–8425. [CrossRef]
20. Rehman, M.U.; Bakht, A.B.; Hussain, I. DeepAquaEnhance: Hierarchical Multi-Head Attention and Adaptive Feature Fusion for Underwater Image Enhancement. *Ecol. Inform.* **2026**, *95*, 103771. [CrossRef]
21. Sarkar, P.; De, S.; Gurung, S.; Dey, P. UICE-MIRNet Guided Image Enhancement for Underwater Object Detection. *Sci. Rep.* **2024**, *14*, 22448. [CrossRef]
22. Liu, Z.; Wang, B.; Li, Y.; He, J.; Li, Y. UnitModule: A Lightweight Joint Image Enhancement Module for Underwater Object Detection. *Pattern Recognit.* **2024**, *151*, 110435. [CrossRef]
23. Pan, T.S.; Huang, H.C.; Lee, J.C.; Chen, C.H. Multi-scale ResNet for real-time underwater object detection. *Signal Image Video Process.* **2021**, *15*, 941–949. [CrossRef]
24. Chen, L.; Liu, Z.; Tong, L.; Jiang, Z.; Wang, S.; Dong, J.; Zhou, H. Underwater object detection using Invert Multi-Class Adaboost with deep learning. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8. [CrossRef]
25. Liu, Y.; An, D.; Ren, Y.; Zhao, J.; Zhang, C.; Cheng, J.; Liu, J.; Wei, Y. DP-FishNet: Dual-path Pyramid Vision Transformer-based underwater fish detection network. *Expert Syst. Appl.* **2024**, *238*, 122018. [CrossRef]
26. Zhang, J.; Zhang, J.; Zhou, K.; Zhang, Y.; Chen, H.; Yan, X. An Improved YOLOv5-Based Underwater Object-Detection Framework. *Sensors* **2023**, *23*, 3693. [CrossRef]

27. Sohn, K.; Berthelot, D.; Carlini, N.; Zhang, Z.; Zhang, H.; Raffel, C.A.; Cubuk, E.D.; Kurakin, A.; Li, C.L. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Volume 33, pp. 596–608.
28. Sohn, K.; Zhang, Z.; Li, C.L.; Zhang, H.; Lee, C.Y.; Pfister, T. A Simple Semi-Supervised Learning Framework for Object Detection. *arXiv* **2020**, arXiv:2005.04757. [CrossRef]
29. Xu, M.; Zhang, Z.; Hu, H.; Wang, J.; Wang, L.; Wei, F.; Bai, X.; Liu, Z. End-to-End Semi-Supervised Object Detection with Soft Teacher. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual, 11–17 October 2021; pp. 3060–3069.
30. Liu, Y.C.; Ma, C.Y.; Kira, Z. Unbiased Teacher v2: Semi-Supervised Object Detection for Anchor-Free and Anchor-Based Detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 9819–9828.
31. Wang, R.J.; Li, X.; Ling, C.X. Pelee: A Real-Time Object Detection System on Mobile Devices. *arXiv* **2019**, arXiv:1804.06882.
32. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]
33. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.

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.