# A Semi-Supervised Domain-Adaptive Framework for Real-World Underwater Image Enhancement

Junjie Wen, Guidong Yang, Benyun Zhao, Dongyue Huang, Lei Lei, Bo Zhang, Zhi Gao, Xi Chen, and Ben M. Chen, *Fellow, IEEE*

***Abstract*— Underwater optical remote sensing is crucial for geoscience applications but often suffers from image degradation due to complex underwater environments. While learning-based methods have advanced underwater image enhancement (UIE), their efficacy in real-world UIE applications still faces challenges. This limitation arises from training predominantly on synthetic underwater images, resulting in a significant interdomain gap when applied to real-world data. In addition, diverse underwater conditions introduce intradomain challenges, such as color casts and haze, further complicating the UIE process. To address these issues, we propose SSD-UIE, a semisupervised domain-adaptive framework designed to mitigate both interdomain and intradomain gaps. Our approach employs a systematic synthesis pipeline to reduce visual interdomain discrepancies and introduces a large synthetic-real underwater image dataset (LSRUID) to facilitate the training of the framework. The semantic blender is developed to handle semantic interdomain differences, while the intradomain-aware feature extraction (IFE) branch and the feature alignment strategy effectively address intradomain variability. Furthermore, the Dual-Trans Block is introduced to enhance the UIE performance while maintaining computational efficiency. Extensive experiments demonstrate that SSD-UIE outperforms state-of-the-art (SOTA) UIE methods in both qualitative and quantitative evaluations on real-world underwater images. The codes and dataset will be publicly available at https://github.com/RockWenJJ/SSD-UIE.git**

Received 24 December 2024; revised 26 February 2025 and 10 June 2025; accepted 14 July 2025. Date of publication 22 July 2025; date of current version 5 August 2025. This work was supported in part by the Research Grants Council of Hong Kong SAR under Grant 14206821, Grant 14217922, and Grant 14209623; and in part by the InnoHK of the Government of the Hong Kong SAR via Hong Kong Centre for Logistics Robotics. *(Junjie Wen and Guidong Yang contributed equally to this work.) (Corresponding author: Bo Zhang.)*

Junjie Wen is with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong, and also with Guangdong Laboratory of Artificial Intelligence and Digital Economy, Shenzhen 510335, China.

Guidong Yang, Dongyue Huang, Lei Lei, Xi Chen, and Ben M. Chen are with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong.

Benyun Zhao is with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong, and also with the High Performance Robotics Lab, Department of Mechanical Engineering, University of California at Berkeley, Berkeley, CA 94720 USA.

Bo Zhang is with the College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China, and also with Guangdong Laboratory of Artificial Intelligence and Digital Economy, Shenzhen 510335, China (e-mail: zhangbo@szu.edu.cn).

Zhi Gao is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China.

This article has supplementary downloadable material available at https://doi.org/10.1109/TGRS.2025.3590798, provided by the authors.

Digital Object Identifier 10.1109/TGRS.2025.3590798

***Index Terms*— Domain adaptation (DA), underwater image enhancement (UIE), underwater image synthesis.**

## I. Introduction

CLEAR underwater imagery is vital for remote sensing and geoscience applications, such as seafloor mapping and marine monitoring, where autonomous underwater vehicles (AUVs) are extensively used [1], [2]. However, light absorption, backscattering, and suspended particles often degrade image quality, making underwater image enhancement (UIE) essential for accurate data analysis. While learning-based methods have significantly advanced UIE [3], [4], [5], [6], [7], [8], they face two key challenges in real-world applications.

First, the scarcity of paired clear and degraded real underwater images has led researchers to rely on synthetic image pairs. However, UIE methods trained solely on synthetic images struggle to adapt to real underwater scenarios due to the interdomain gap, encompassing both visual and semantic differences. Some studies use transfer learning to generate underwater images from real data [9], but they often fail to capture the physical properties of real underwater scenes. Others rely on simplified underwater image formation models [10], which can introduce significant errors, resulting in inaccurate synthetic images [11]. Although these methods improve visual similarity between synthetic and real underwater images, they fail to address the semantic gap, as synthetic images often lack critical underwater elements such as fish and corals. This highlights the need for a framework that addresses both visual and semantic interdomain gaps.

Second, the complexity of underwater conditions introduces various effects (as shown in Fig. 1), such as color variations from wavelength-dependent light absorption and haziness due to light scattering. This variability, denoted as the intradomain gap, exists in both synthetic and real underwater images. Current UIE methods typically address this gap within either the synthetic or real domain. For example, Berman et al. [12] mitigate the intradomain gap in synthetic images by classifying underwater environments based on Jerlov water types [13]. In contrast, Wang et al. [14] proposed an intradomain adaptation (DA) strategy targeting hard and easy underwater images within the real domain. However, no current method comprehensively addresses the intradomain gap across both domains. Thus, a solution that effectively handles the intradomain gap

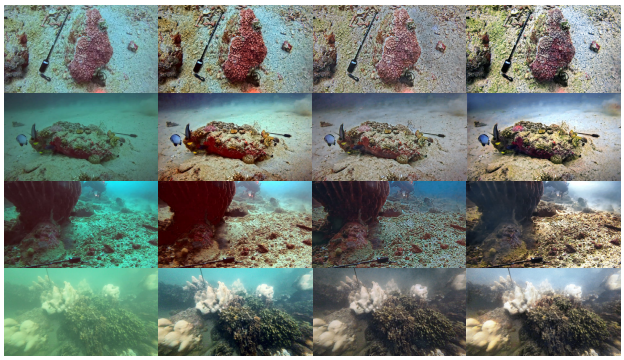



Fig. 1. Real-world enhancement results of our proposed SSD-UIE and other SOTA methods on self-collected underwater scenes[1] with varying illumination, visibility, and color cast. (From left to right) Raw input, Semi-UIR [3], SyreaNet [4], and SSD-UIE. Please zoom in for better visualization.

in both synthetic and real domains is essential for advancing the UIE field.

To address these challenges, we propose SSD-UIE, a semisupervised UIE framework that incorporates novel domain-adaptive strategies to overcome both interdomain and intradomain gaps. For bridging interdomain gaps, SSD-UIE integrates a physically guided underwater image synthesis (PUIS) module, which generates high-quality synthetic images to reduce visual discrepancies, and a semantic blender to minimize semantic gaps. Using the PUIS method, we also introduce the large synthetic-real underwater image dataset (LSRUID) to facilitate the framework's training. To address intradomain gaps in both synthetic and real domains, we propose an intradomain-aware feature extraction (IFE) branch, combined with a feature alignment strategy to handle intradomain discrepancies.

Moreover, we propose a novel UIE network with the introduction of the Dual-Trans Block, which distinguishes our network from existing methods that typically rely on CNNs [15], [16], [17] or Transformers [18]. Unlike traditional networks that either focus on local feature extraction or struggle with high-computational costs due to long-range dependencies, the Dual-Trans Block enables efficient feature extraction while maintaining computational efficiency, further enhancing the performance of our framework.

Our main contributions are summarized as follows.

1) We propose a novel SSD-UIE framework that integrates advanced domain-adaptive strategies to address both interdomain and intradomain challenges in UIE.
2) To bridge the interdomain gap, we introduce a PUIS module to mitigate visual discrepancies and a semantic blender to reduce semantic differences between synthetic and real images. In addition, we develop the LSRUID dataset using the PUIS module.
3) We propose an IFE branch and a feature alignment strategy to effectively address intradomain gaps in both synthetic and real underwater domains.
4) The Dual-Trans Block is introduced to enhance UIE performance while maintaining computational efficiency.
5) Extensive experiments demonstrate the superior performance of SSD-UIE on real-world UIE tasks compared to other state-of-the-art (SOTA) methods.

[1]Refer to the supplementary material for details on the real-world underwater scene image collection.

## II. Related Work

### A. Underwater Image Enhancement

UIE methods are generally divided into traditional and learning-based approaches. Traditional methods include model-free techniques [6], [19], which directly adjust pixel values to improve visual quality, and model-based methods [20], [21], which rely on the underwater image formation model. However, these methods often fail when their assumptions do not accurately reflect specific underwater environments.

Learning-based methods often rely on synthetic underwater images due to the difficulty and cost of acquiring high-quality real underwater image pairs. Some learning-based UIE methods avoid synthetic images by using adversarial learning [22], [23], primarily focusing on perceptual enhancement, which may occasionally lead to inaccurate color restoration. Synthetic data offer the advantage of providing both degraded and clear images, improving the model's ability to recover original colors. For example, Li et al. [15] proposed a lightweight UWCNN that uses underwater scene priors for direct image reconstruction without parameter estimation for different water types [13]. Chen et al. [24] proposed a UIE method that utilizes a multiscale feature fusion network with integrated feature extraction, fusion, and attention reconstruction modules to enhance scene adaptability and visual quality. Li et al. [17] presented a UIE network that employs medium transmission-guided multicolor space embedding to mitigate color casts and low contrast. Cong et al. [25] proposed CECF, which focuses on color correction by learning separate color and content codes to enable controlled adjustments of underwater organisms' colors based on provided guidance. However, training solely with synthetic data may limit the model's effectiveness on real underwater scenes. To address this, some approaches combine synthetic and real underwater images during training. For instance, Huang et al. [3] proposed a semisupervised UIE framework with contrastive regularization and a mean-teacher model, while Wang et al. [14] introduced a UIE framework using a triple-alignment network to minimize domain differences between synthetic and real images, supplemented by rank-based image quality assessments. Nonetheless, the domain gap between synthetic and real underwater images may still compromise the UIE performance.

In this work, we present a learning-based UIE framework using a semisupervised training strategy with both synthetic and real underwater images, incorporating novel domain-adaptive technologies.

### B. Underwater Image Synthesis

Underwater image synthesis plays a vital role in training UIE networks due to the challenges of obtaining clear and degraded underwater image pairs in real-world conditions. Traditional synthesis methods typically rely on a simplified underwater image formation model [26]. While approaches

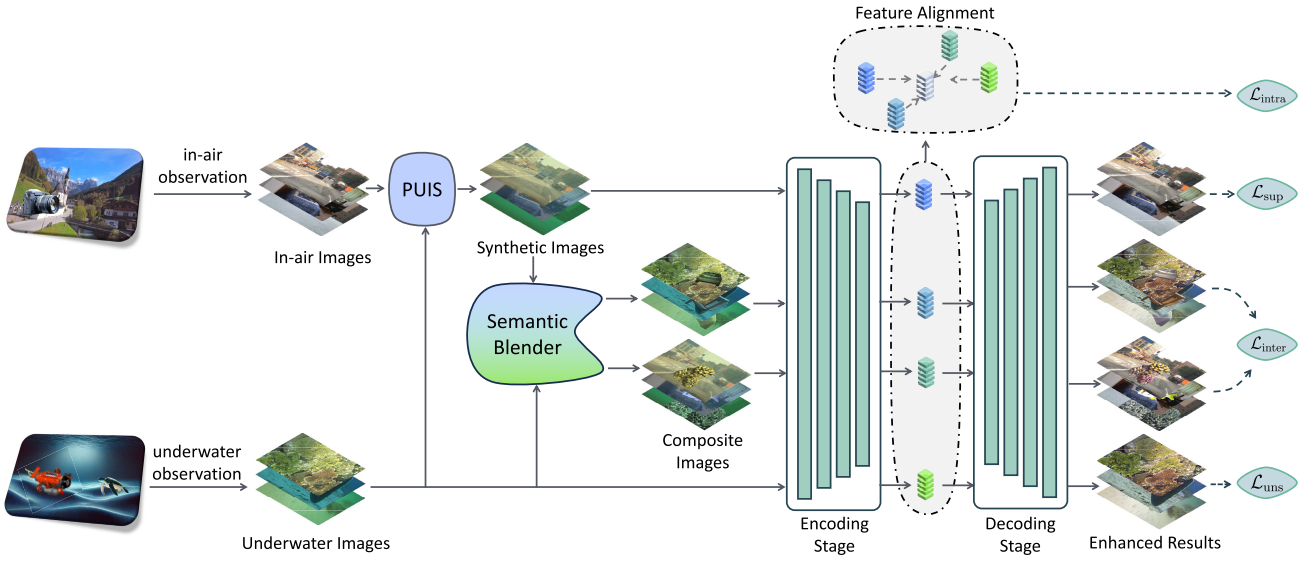



Fig. 2. Illustrated framework of the proposed SSD-UIE: underwater and in-air images are first collected from underwater and in-air observations, respectively. The PUIS module then generates realistic underwater images to mitigate the visual interdomain gap. Next, the semantic blender processes these images to address semantic interdomain discrepancies. The blended outputs, combined with the original images, are fed into the network, where an intra-DA strategy aligns features across diverse underwater environments. The upper-right section, labeled "feature alignment," illustrates how features from different domains are aligned to minimize the intradomain gap. The dashed arrows represent the loss terms that guide the framework's optimization.

such as [27] and [28] use attenuation coefficients from Jerlov water types [13], they struggle to capture the complexity of real underwater environments. Recent efforts have explored learning-based techniques for more realistic underwater image synthesis. Generative adversarial networks (GANs) [29] are commonly used for this purpose due to their strong image generation capabilities, but they often suffer from mode collapse, reducing image diversity [30]. Ye et al. [31] introduced a neural rendering technique using estimated light field maps from real underwater scenes. While this method produces realistic synthetic images, it may not fully adhere to the physical principles for formulating underwater images.

In this work, we propose a method to generate realistic underwater images from in-air images using a revised underwater image formation model and introduce a comprehensive UIE dataset with synthetic and real-world underwater images.

### C. Underwater DA

DA aims to reduce distributional discrepancies between different domains and has been applied across various tasks [9]. As highlighted in [32], DA is particularly critical in UIE due to the unique challenges posed by the underwater environment. Unlike other image restoration tasks such as dehazing or low-light enhancement, UIE faces two key challenges: the scarcity of clear reference underwater images and the variability of underwater conditions. The limited availability of reference images necessitates reliance on synthetic datasets, which often exhibit significant interdomain discrepancies. These discrepancies include both visual and semantic gaps. Visual gaps often appear as color distortions, while semantic gaps arise because synthetic images typically lack underwater elements such as marine organisms. Furthermore, intradomain variability arises from diverse environmental factors, including varying water quality, depth, and lighting conditions. This variability introduces additional challenges, such as inconsistent color casts and haziness, which need to be addressed to improve UIE performance.

Early DA research in UIE primarily focused on mitigating visual differences between synthetic and real underwater images, often using different image synthesis techniques (see Section II-B). More recently, transfer learning has been applied to further bridge these gaps. For instance, Jiang et al. [33] proposed a two-step DA framework that adapts in-air image dehazing methods to real underwater images. Chen and Pei [5] developed a DA framework that separates content and style across synthetic, real, and clean domains. Qiao et al. [34] introduced a knowledge distillation approach, combining semisupervised distillation and self-domain adversarial distillation to resolve intradomain differences and enhance UIE performance. However, these methods typically address either interdomain or intra-DA independently, without fully exploring both simultaneously. Recent studies have attempted to tackle both inter- and intra-DAs. Wang et al. [14] addressed both types of domain gaps, but their intra-DA focuses only on real domain data, classifying samples as easy or hard using a ranking method. Wen et al. [4] also targeted both inter- and intra-DAs but did not effectively address semantic differences between synthetic and real domains.

In this work, we propose a novel DA strategy that bridges both visual and semantic interdomain gaps while effectively handling intradomain differences in both synthetic and real underwater images.

## III. Proposed Method

The overall framework of the proposed SSD-UIE is illustrated in Fig. 2. Underwater and in-air images are first

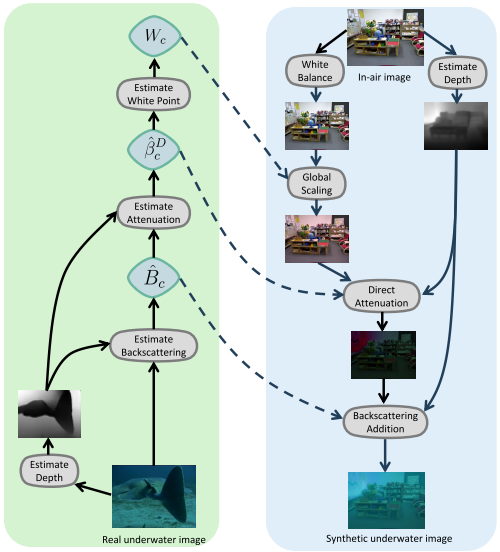



Fig. 3. Illustrated pipeline of our proposed PUIS module, which includes two components (a) params estimator and (b) underwater synthesizer.

collected from their respective observations. The PUIS module (see Section III-A) then generates synthetic underwater images to reduce the visual interdomain gap. Next, both synthetic and real images are processed by the semantic blender (see Section III-B) to address semantic interdomain discrepancy. The blended outputs, along with the original images, are fed into the network. To handle intradomain gaps, an intra-DA strategy is proposed to align features across varying underwater environments. The network architecture is detailed in Section III-D, and the corresponding loss functions are described in Section III-E.

### A. Physically Guided Underwater Image Synthesis

To reduce the visual interdomain gap between synthetic and real underwater images, we propose a PUIS module, which utilizes the revised underwater image formation model [35] to generate more realistic synthetic images. As illustrated in Fig. 3, PUIS consists of two components: the params estimator and the underwater synthesizer.

The params estimator is designed to estimate key parameters of the revised model [35], including backscattering $\hat{B}_c$, direct attenuation $\hat{\beta}_c^D$, and the white point $W_c$. For details on the revised formation model, refer to the supplementary material. This process requires an underwater depth map, which is estimated using SOTA monocular depth estimation methods [36]. The estimated depth map, along with the real underwater image, enables the derivation of parameters relevant to underwater scenes, as shown in Fig. 3(a). Using these estimated parameters, the equation for synthesizing realistic underwater images is expressed as

$$I_c(x) = J_s(x) W_c e^{-\hat{\beta}_c^D(z) z} + \hat{B}_c(x) \tag{1}$$

where $J_s(x)$ is the clear in-air image, $z$ is the depth map, and $I_c(x)$ is the generated synthetic underwater image.

Then, as shown in Fig. 3(b), the proposed underwater synthesizer first preprocesses the in-air image using white balance and estimates the depth map [36]. Subsequently, (1) is applied to generate realistic synthetic underwater images using the preprocessed in-air image and its corresponding depth map. This process involves global scaling, direct attenuation, and backscattering addition, as depicted in Fig. 3(b).

The proposed PUIS method facilitates the development of the LSRUID, comprising 10 000 real underwater images and 50 000 synthetic image pairs. The real underwater images were sourced from the Internet and our own collection, while the in-air images for generating synthetic pairs were obtained online. Sample images from the LSRUID dataset, as shown in Fig. 4, demonstrate that our method produces more realistic synthetic underwater images, significantly reducing the visual interdomain gap. A detailed comparison of this dataset with others is presented in Section IV-B.

### B. Semantic Inter-DA

While generating more realistic underwater images helps reduce the visual interdomain gap, a significant semantic gap still remains. For instance, synthetic underwater images can simulate water-related effects but often lack underwater objects such as fish or corals, while real underwater images do not include terrestrial objects such as cars or bicycles. As a result, networks trained solely on synthetic data may struggle to enhance real-world underwater images due to these semantic discrepancies.

To address this issue, we propose a semantic blender to incorporate semantic information from one domain into the other. As shown in Fig. 5, the interdomain semantic blending process involves three steps: 1) semantic object extraction; 2) C&P; and 3) image harmonization.

*1) Semantic Object Extraction:* In semantic object extraction, the goal is to identify and extract key objects in underwater images from both synthetic and real domains. Leveraging recent advancements in vision-language models (VLMs), we detect objects in unannotated underwater images from both domains. Specifically, we use SAM [37] to detect instance segments, followed by CLIP [38] for classifying their categories. A predefined vocabulary (see supplementary material) specific to in-air and underwater environments ensures the relevance of detected segments. We apply a confidence threshold of 0.9 and limit the extraction to one semantic object per image, ensuring accurate identification of objects that convey the semantic content of the image.

*2) Copy and Paste:* After extracting semantic objects from each domain, we apply a copy and paste (C&P) procedure to blend the semantic information across domains. As shown in Fig. 5, two C&P modules are employed. Each module takes in an image from one domain (synthetic or real) and segmented objects from the other domain. For example, in the left C&P module, segmented real objects are copied and then randomly resized, flipped, and rotated before being pasted onto a synthetic underwater image. This creates a composite image that merges synthetic information with real underwater semantic objects. In addition, masks are generated to conceal pixels from the real domain, facilitating the computation of interdomain loss $\mathcal{L}_{\text{inter}}$ (see Section III-E).

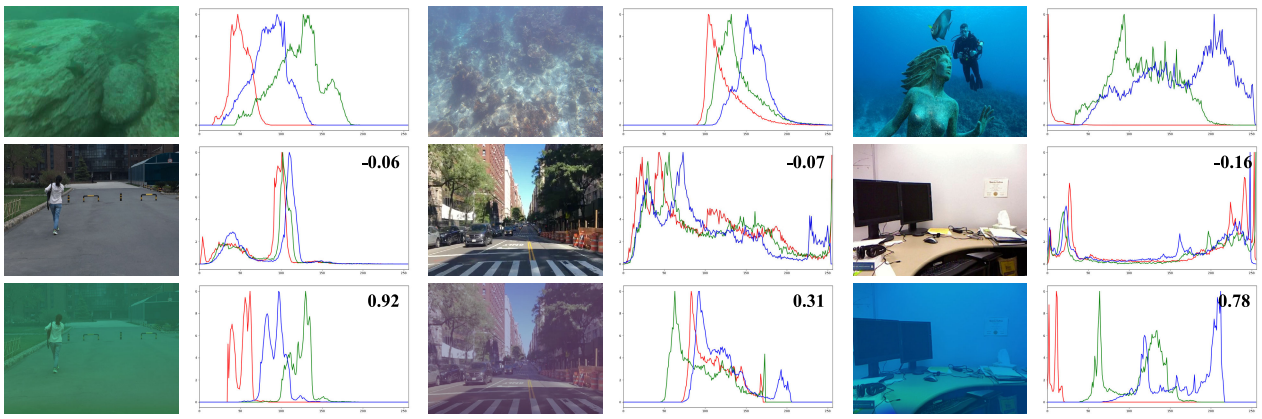



Fig. 4. Sample images from our proposed LSRUID and their corresponding histograms. Row 1: real-world underwater images and their histograms. Row 2: in-air images and their histograms. Row 3: synthetic underwater images and their histograms. The numbers in each histogram indicate the similarity in histogram distribution between the in-air or synthetic images and the real-world underwater images.

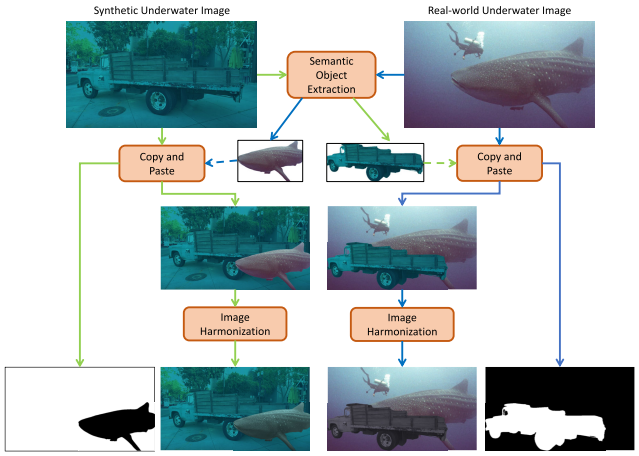



Fig. 5. Pipeline of our proposed semantic blender for inter-DA.

*3) Image Harmonization:* Although composite images contain semantic information from both domains, disparities may arise between the original image and the pasted semantic objects. For example, in Fig. 5, the left composite image shows a synthetic underwater scene with a greenish tone, while the segmented whale has a grayish cast. Similarly, in the right composite image, a grayish underwater scene contrasts with a greenish segmented truck. These inconsistencies between the background and the pasted segments can potentially confuse the network. To resolve this, we apply image harmonization techniques [39] to ensure seamless blending of semantic segments. As shown, harmonization adjusts the whale's tone to bluish, matching the synthetic background, while the truck acquires a grayish tone, fitting the real underwater scene.

After processing with the semantic blender, composite images are generated by integrating synthetic or real backgrounds with semantic objects from the opposite domain, accompanied by masks that exclude real underwater pixels. These composite images are then fed into the network for enhancement, while the masks are used to compute the interdomain loss $\mathcal{L}_{\text{inter}}$ (see Section III-E).

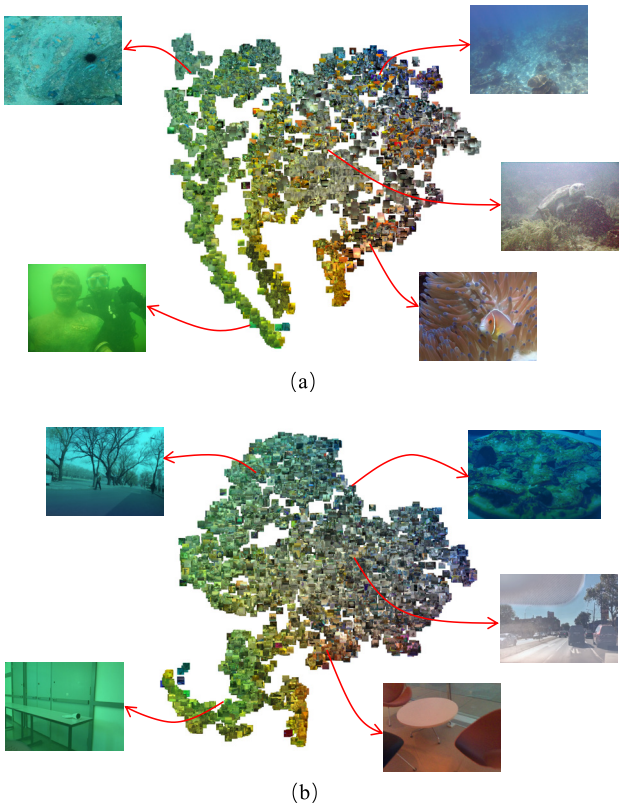



Fig. 6. t-SNE visualizations of PCA feature vectors for real and synthetic underwater images. The visualizations reveal that intradomain discrepancies originate from various water-related effects. (a) Underwater images in the real domain. (b) Underwater images in the synthetic domain.

## C. Intra-DA

Underwater intradomain discrepancies arise from varying water effects, such as differences in light absorption, scattering, and color distortion. As shown in Fig. 6, t-SNE visualizations of PCA feature vectors for real and synthetic underwater images highlight significant intradomain gaps in both real and synthetic domains.

To address these complex intradomain variations, our intra-DA strategy focuses on aligning features that effectively capture these discrepancies. Therefore, we developed an IFE

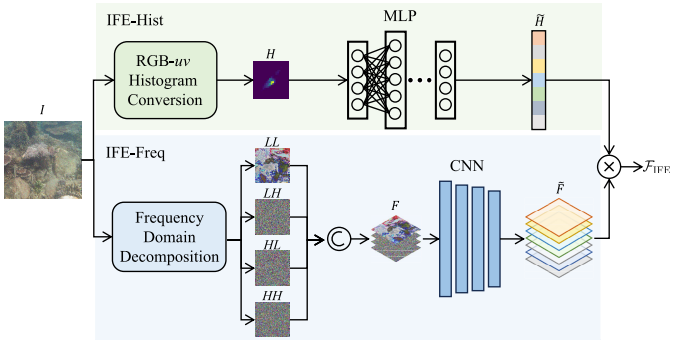



Fig. 7. Illustrated pipeline of the IFE branch. The input image is processed to extract features that integrate both chrominance-based information and frequency-based information.

branch designed to capture features that reflect water-related effects, including the IFE-Hist and IFE-Freq subbranches.

For the IFE-Hist subbranch, as shown in Fig. 7, the input image $I$ is first converted from the traditional RGB color space to a log-chrominance $uv$ space with measures $u$ and $v$ using the following equations:

$$I_{u(i)} = \log I_{\mathrm{G}(i)} - \log I_{\mathrm{R}(i)}$$
$$I_{v(i)} = \log I_{\mathrm{G}(i)} - \log I_{\mathrm{B}(i)} \tag{2}$$

where $I$ represents the pixel value, {R,G,B} denote the RGB color channels, and $i \in \{1, \ldots, N\}$ is the pixel index.

Next, we construct a 2-D histogram $h$ for $I$ from the $uv$ space, where $h(u, v)$ denotes the number of pixels in $I$ whose chrominance is close to $(u, v)$, with histogram counts weighted by each pixel's luminance $I_{y(i)}$

$$h(u, v) = \sum_i I_y(i) \left[ |I_u(i) - u| \leq \frac{\epsilon}{2} \wedge |I_v(i) - v| \leq \frac{\epsilon}{2} \right] \tag{3}$$

where $I_{y(i)} = (I^2_{\mathrm{R}(i)} + I^2_{\mathrm{G}(i)} + I^2_{\mathrm{B}(i)})^{1/2}$, the square bracket $[\cdot]$ denotes an indicator function, and $\epsilon$ represents the histogram bin-width. To enhance the domain discriminative capability of the histogram feature, we calculate the square root of the L1-norm of the histogram counts, resulting in the RGB-$uv$ histogram feature $H$

$$H(u, v) = \sqrt{\frac{h(u, v)}{\sum_{u', v'} h(u', v')}}. \tag{4}$$

The RGB-$uv$ histogram feature $H$ captures chrominance information sensitive to color casts across varying underwater environments. The feature is then processed through a multilayer perceptron (MLP) to produce the output $\tilde{H}$.

For the IFE-Freq subbranch, the original RGB image is transformed into the frequency domain using wavelet decomposition [40], extracting frequency-domain components $\{LL, LH, HL, HH\}$, which represent the fundamental and structural characteristics of underwater scenes. Here, $LL$ corresponds to the low-frequency component, while $LH$, $HL$, and $HH$ represent the horizontal, vertical, and diagonal high-frequency components, respectively. The frequency-domain components $\{LL, LH, HL, HH\}$ are concatenated into a feature map $F$, which is passed through convolutional layers to produce $\tilde{F}$.

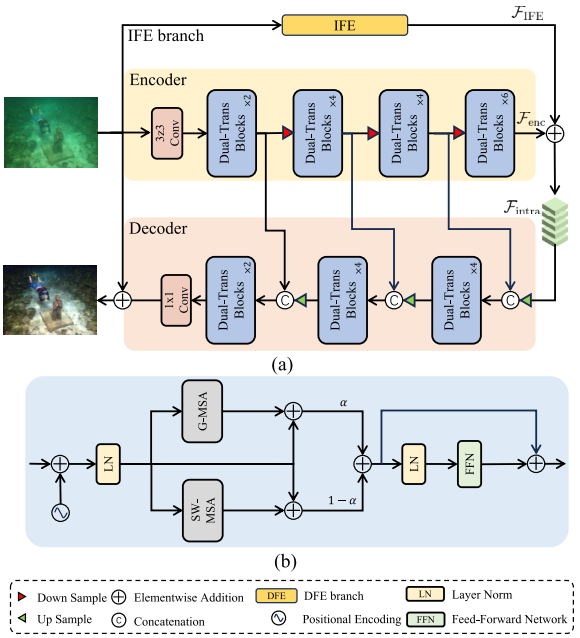



Fig. 8. Illustrated architecture of the network employed in our framework. (a) Network architecture. (b) Details of the Dual-Trans Block.

The intradomain-aware feature $\mathcal{F}_{\mathrm{IFE}}$ is then obtained by elementwise multiplication of $\tilde{H}$ and $\tilde{F}$, effectively combining chrominance- and frequency-based information to capture intradomain variations in underwater environments. As shown in Fig. 8(a), $\mathcal{F}_{\mathrm{IFE}}$ is added to the encoder features $\mathcal{F}_{\mathrm{enc}}$, and the combined features are aligned using the proposed intradomain loss $\mathcal{L}_{\mathrm{intra}}$ (see Section III-E).

### D. Network Architecture

As shown in Fig. 8(a), the proposed network follows a standard encoder–decoder architecture with encoding and decoding stages. During encoding, the input image is processed through two branches: the IFE branch (see Section III-C), which extracts intradomain-aware features $\mathcal{F}_{\mathrm{IFE}}$, and the encoder branch, which captures global semantics and local details $\mathcal{F}_{\mathrm{enc}}$. These features are combined via elementwise addition. The decoding stage mirrors the encoder, progressively reconstructing a clear image by integrating features from both branches. Unlike conventional UIE networks that use CNN or Transformer blocks [41], [42], both the encoder and decoder in our framework are built on the proposed Dual-Trans Block.

While CNNs effectively provide local connectivity through convolutional operations, they struggle to capture long-range pixel dependencies. In contrast, Transformers [41] excel at modeling long-range dependencies via the self-attention (SA) mechanism. However, the computational complexity of SA increases quadratically with input spatial resolution, limiting the applicability of Transformers in real-world UIE. To address this challenge and leverage the strengths of Transformers while maintaining computational efficiency, we propose the Dual-Trans Block. As illustrated in Fig. 8(b), this block comprises two parallel SA branches: shifted-window multihead SA (SW-MSA) and global multihead SA (G-MSA).

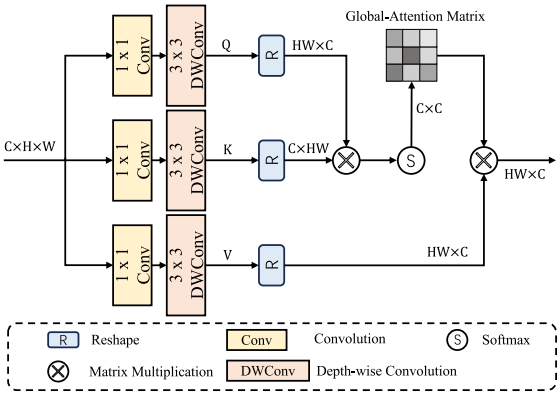



Fig. 9. Detailed architecture of G-MSA.

The SW-MSA architecture, based on [42], applies the SA mechanism within partitioned windows, achieving linear computational complexity. However, its focus on smaller spatial windows limits its ability to model long-range dependencies effectively. To overcome this limitation, we incorporate a parallel G-MSA, as detailed in Fig. 9. Unlike SW-MSA, G-MSA operates on global features of dimensions $C \times H \times W$. It begins with $1 \times 1$ convolutions followed by $3 \times 3$ depthwise convolutions to generate the feature maps $Q$, $K$, and $V$. These maps are reshaped into 2-D features, and the SA mechanism is applied to $Q$ and $K$, producing a global attention matrix of size $C \times C$. This matrix is then multiplied by $V$. The computational cost of G-MSA is linear with respect to the input feature size, expressed as

$$\Omega(\text{G-MSA}) = 3HWC^2 + 27HWC + 2HWC^2. \tag{5}$$

As illustrated in Fig. 8(b), the outputs of SW-MSA and G-MSA are first added to their respective inputs using shortcut connections and then modulated by a learnable parameter $\alpha$ to balance the importance of global and local features. Given an input $x$ to the Dual-Trans Block, the final output $\hat{x}$ of the block is computed as follows:

$$\begin{aligned} f_1 &= \text{G-MSA}(\text{LN}(x)) + \text{LN}(x) \\ f_2 &= \text{SW-MSA}(\text{LN}(x)) + \text{LN}(x) \\ f_3 &= \alpha f_1 + (1-\alpha) f_2 \\ \hat{x} &= f_3 + \text{FFN}(\text{LN}(f_3)) \end{aligned} \tag{6}$$

where $\text{LN}(\cdot)$ represents layer normalization and $\text{FFN}(\cdot)$ denotes a feed-forward network.

### E. Loss Design

Fig. 2 illustrates that the total loss $\mathcal{L}_{\text{total}}$ is a combination of the following losses:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{sup}} + \lambda_2 \mathcal{L}_{\text{uns}} + \lambda_3 \mathcal{L}_{\text{intra}} + \lambda_4 \mathcal{L}_{\text{inter}} \tag{7}$$

where $\mathcal{L}_{\{\text{sup,uns,intra,inter}\}}$ denote the supervised, unsupervised, intradomain, and interdomain losses, respectively, and $\lambda_{i\{i=1,\ldots,4\}}$ represent the corresponding balanced weights.

*1) Supervised Loss $\mathcal{L}_{sup}$:* The supervised loss $\mathcal{L}_{\text{sup}}$ is designed to guide the network training using synthetic underwater image pairs. Since these synthetic pairs include both degraded images and clear reference images, full supervision is available. The supervised loss $\mathcal{L}_{\text{sup}}$ is defined as

$$\mathcal{L}_{\text{sup}} = \left|\left|I - \hat{I}\right|\right| + \mathcal{L}_{\text{SSIM}}\left(I, \hat{I}\right) + \mathcal{L}_{\text{VGG}}\left(I, \hat{I}\right) \tag{8}$$

where $I$ represents the clear reference, $\hat{I}$ denotes the enhanced result, $\mathcal{L}_{\text{SSIM}}(\cdot)$ is the SSIM loss [43], and $\mathcal{L}_{\text{VGG}}(\cdot)$ refers to the perceptual loss [44].

*2) Unsupervised Loss $\mathcal{L}_{uns}$:* For real underwater images, where ground-truth references are unavailable, we employ an unsupervised loss $\mathcal{L}_{\text{uns}}$ to guide the enhancement process. Since enhanced real-world images should adhere to principles characteristic of clear natural images, the unsupervised loss $\mathcal{L}_{\text{uns}}$ is formulated based on the "gray-world" assumption [45]

$$\mathcal{L}_{\text{uns}} = \frac{1}{3}\sum_{c}\left(\left(\frac{1}{N}\sum_{i,j}\hat{I}_c(i,j)\right) - 0.5\right)^2 \tag{9}$$

where $c$ denotes the color channel, $N$ represents the number of pixels, and $(i, j)$ indicates the pixel position.

*3) Intradomain Loss $\mathcal{L}_{intra}$:* To mitigate intradomain disparities, we align $\mathcal{F}_{\text{intra}} = \mathcal{F}_{\text{enc}} + \mathcal{F}_{\text{IFE}}$ across various underwater environments within either synthetic or real domains. The intradomain loss $\mathcal{L}_{\text{intra}}$ is designed to minimize the covariance distances between features originating from distinct intradomains, which is expressed as

$$\mathcal{L}_{\text{intra}} = \frac{1}{B}\sum\left|\left|C\left(\mathcal{F}^i_{\text{intra}}\right) - C\left(\mathcal{F}^j_{\text{intra}}\right)\right|\right|^2_F \tag{10}$$

where $C(\cdot)$ denotes the feature covariance matrix, $||\cdot||^2_F$ represents the squared Frobenius norm, $B$ is the batch size, and $i \neq j$ indicates the feature index.

*4) Interdomain Loss $\mathcal{L}_{inter}$:* The interdomain loss $\mathcal{L}_{\text{inter}}$ is designed to facilitate the training of composite images generated by the semantic blender. This loss leverages both the composite images and their corresponding masks. As detailed in Section III-B, unmasked pixels originate from the synthetic domain, enabling full supervision, while masked pixels correspond to the real domain, lacking ground-truth references. To bridge the synthetic-real domain gap, $\mathcal{L}_{\text{inter}}$ employs a triplet contrastive learning approach. The key idea is that the enhanced color intensity in unsupervised masked regions should closely resemble the enhanced intensity in supervised unmasked regions while remaining distinct from the original intensity of the masked regions. Thus, $\mathcal{L}_{\text{inter}}$ is defined as

$$\mathcal{L}_{\text{inter}} = \frac{1}{3}\sum_{c}\left(\frac{\left|\left|\text{AVE}\left(\text{M}\left(\hat{I}_c\right)\right) - \text{AVE}\left(\text{UM}\left(\hat{I}_c\right)\right)\right|\right|}{\left|\left|\text{AVE}\left(\text{M}\left(\hat{I}_c\right)\right) - \text{AVE}\left(\text{M}\left(\tilde{I}_c\right)\right)\right|\right|}\right) \tag{11}$$

where $\hat{I}$ represents the enhanced result, $\tilde{I}$ denotes the original input, and $c$ refers to the color channel. $\text{AVE}(\cdot)$ denotes the mean average operation, while $\text{M}(\cdot)$ and $\text{UM}(\cdot)$ represent the masked and unmasked regions, respectively.

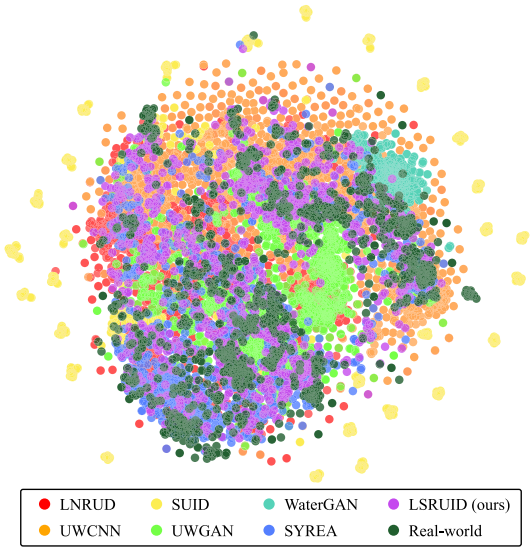



Fig. 10. Visualization results of t-SNE on different synthesized underwater datasets and real underwater datasets.

TABLE I
COMPARISON BETWEEN THE LSRUID AND OTHER UIE DATASETS

| Dataset Name | Dataset Type | Synthetic Pairs | Real-world Images | Avg. Resolution |
|---|---|---|---|---|
| UIEB [10] | Real | × | 890 | 868 × 595 |
| RUIE [46] | Real | × | 4,230 | 400 × 300 |
| EUVP [47] | Real | × | 6,676 | 276 × 258 |
| LSUI [18] | Real | × | 4,279 | 447 × 309 |
| UWCNN [15] | Synthetic | 13,041 | × | 620 × 460 |
| SUID [48] | Synthetic | 900 | × | 840 × 541 |
| UWGAN [29] | Synthetic & Real | 11,136 | 4,242 | 383 × 297 |
| WaterGAN [29] | Synthetic & Real | 24,534 | 7,000 | 372 × 283 |
| LNRUD [31] | Synthetic & Real | 50,000 | 5,000 | 405 × 382 |
| SYREA [4] | Synthetic & Real | 21,826 | 5,340 | 579 × 472 |
| LSRUID (ours) | Synthetic & Real | 50,000 | 10,000 | 797 × 574 |

## IV. EXPERIMENTS

### A. Implementation Details

Our proposed SSD-UIE is trained using PyTorch on two NVIDIA Tesla V100 GPUs. The ADAM optimizer with an initial learning rate of 0.0001 is employed, and the learning rate is adjusted using cosine annealing until convergence. The balanced weights from $\lambda_1$ to $\lambda_4$ are empirically set as 1, 0.1, 0.5, and 1, respectively. The training process begins by initially cropping input images to a size of 256 × 256 with a batch size of 16. Subsequently, we gradually scale up the image size while reducing the batch size. The network training undergoes two training stages: 1) the initial stage involves training exclusively using synthetic underwater images without any DA strategies for 50 epochs and 2) the second stage fine-tunes the network using both synthetic and real-world underwater images with our proposed DA strategies for an additional 50 epochs.

The training dataset consists of both synthetic and real underwater images, all sourced from our LSRUID dataset (as detailed in Section IV-B). For the testing set, we utilize real-world underwater images from UIEB [10], EUVP [47], and RUIE [46] datasets. It is important to note that the images used for testing are not included in the LSRUID dataset's training set to avoid data leakage.

TABLE II
QUANTITATIVE COMPARISON OF SYNTHETIC UNDERWATER IMAGES

| Synthesized Images | IR ↑ | FID ↓ | KID ↓ | KLD ↓ |
|---|---|---|---|---|
| UWCNN [15] | 12.35% | 323.34 | 76.55 | 11.47 |
| SUID [48] | 19.25% | <u>266.05</u> | 65.16 | 12.93 |
| LNRUD [31] | 62.54% | 321.76 | 66.02 | **9.61** |
| SYREA [4] | <u>82.45%</u> | 288.33 | <u>58.24</u> | 12.79 |
| WaterGAN [29] | 9.37% | 306.15 | 73.51 | 14.57 |
| UWGAN [49] | 14.26% | 320.35 | 71.10 | 10.35 |
| LSRUID (ours) | **88.34%** | **256.88** | **51.16** | <u>9.69</u> |

**Note:** The best and second-best are in **bold** and <u>underline</u>, respectively.

### B. Comparison of UIE Datasets

To facilitate training of the framework, we developed the LSRUID dataset, comprising 50 000 synthetic underwater image pairs generated using the PUIS method (see Section III-A). The in-air images were sourced from [50], [51], and [52], with depth map estimated via a monocular depth estimation algorithm [36]. In addition, we compiled 10 000 real-world underwater images, comprising both Internet-sourced and team-captured samples. The Internet-sourced images were obtained by searching for high-resolution underwater content using general and specific keywords (e.g., "underwater scenery" and "coral reefs"). The team-captured images were acquired using DJI Action 2 and GoPro Hero 12 cameras at various locations in Hong Kong and the Philippines under diverse underwater conditions. For more details on the data collection, please refer to the supplementary material.

Fig. 4 presents sample images from LSRUID, illustrating that our synthetic images closely align with real underwater scenes. The histograms further quantify this alignment, showing that the histogram distributions of the synthetic images closely match those of the real underwater images, effectively reducing the visual interdomain gap. Table I compares LSRUID with other UIE datasets, showing it as the most comprehensive dataset for both real and synthetic underwater images.

We further validated our PUIS synthesis method through t-SNE visualizations, which show greater overlap between synthetic and real underwater images in LSRUID compared to other datasets (see Fig. 10). This overlap is quantified by the intersection ratio (IR), calculated based on the overlap of t-SNE clusters between synthetic and real images

$$\mathrm{IR} = \frac{|R \cap S|}{|R|} \times 100\% \tag{12}$$

where $R$ denotes the real image region, $S$ represents the synthetic image region, $\bigcap$ denotes the intersection operation, and $|\cdot|$ indicates the area of the respective region. As shown in Table II, LSRUID achieves an IR of 88.34%, outperforming other datasets such as SYREA [4] (82.45%) and LNRUD [31] (62.54%), indicating better alignment with real images. In addition, we evaluate LSRUID using other metrics: the Frechet inception distance (FID) [53], the kernel inception distance (KID) [54], and the Kullback–Leibler divergence (KLD) [55]. Our proposed LSRUID achieves the lowest FID of 256.88, indicating that its synthetic images closely match the feature distribution of real-world images. In addition, the

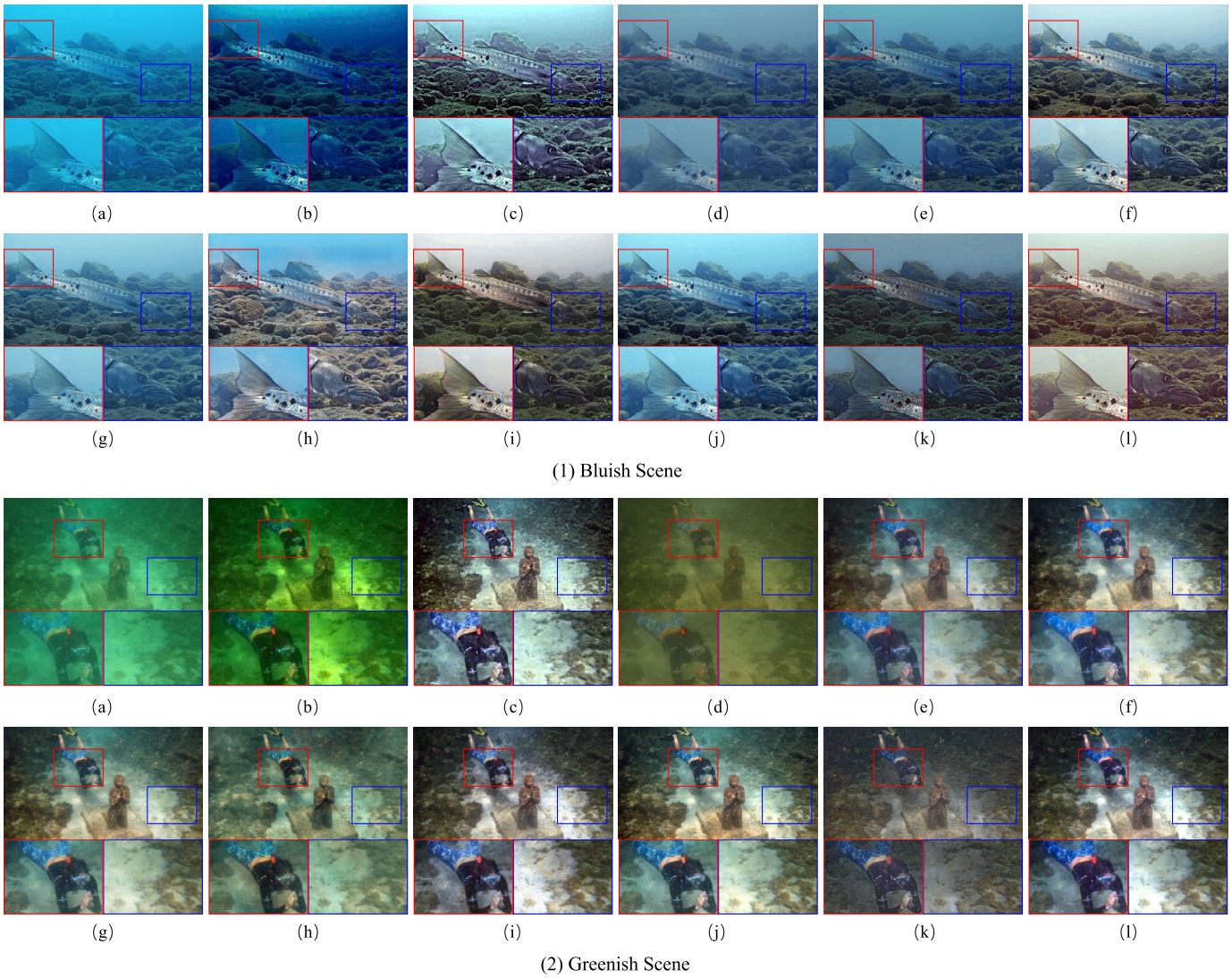



Fig. 11. Visual comparison of the performances of our proposed SSD-UIE and other SOTA UIE methods on real-world images of (1) bluish and (2) greenish scenes in the UIEB dataset [10]. The results from (a) to (l) correspond to (a) raw image, (b) UDCP [21], (c) WWPF [6], (d) UWCNN [10], (e) CECF [25], (f) GLNet [16], (g) UColor [17], (h) UShapeTrans [18], (i) TUDA [5], (g) Semi-UIR [3], (k) SyreaNet [4], and (l) SSD-UIE (ours). Zoom in for a better view.

lower KID (51.16) and KLD (9.69) values suggest that the synthetic images in LSRUID better preserve the distribution of real images and exhibit less divergence.

In Section IV-E, we provide a quantitative comparison by training the same network with different synthetic datasets and evaluating its performance on real underwater images, further validating the effectiveness of our synthesis method in bridging the visual interdomain gap.

### C. Visual Comparison on Real Underwater Images

We evaluate the UIE performance of our SSD-UIE method against several traditional and learning-based UIE techniques on various real underwater image datasets. The compared methods include traditional approaches such as UDCP [21] and WWPF [6] and learning-based methods such as UWCNN [10], CECF [25], GLNet [16], UColor [17], UshapeTrans [18], SyreaNet [4], Semi-UIR [3], and TUDA [5]. For fair comparison, all learning-based models were retrained using our LSRUID dataset with their publicly available codes.

Fig. 11 presents a visual comparison of various UIE methods on bluish and greenish underwater images from the UIEB dataset [10]. While UDCP [21] reduces haze, it struggles with bluish and greenish conditions, resulting in reduced illumination. WWPF [6] enhances image details but introduces artifacts, affecting image quality and naturalness. UWCNN [10] tends to darken images, especially in low-brightness areas, and is less effective at removing bluish/greenish tones. CECF [25], GLNet [16], and UColor [17] improve brightness and remove greenish casts but fall short in addressing bluish casts. UshapeTrans [18] struggles with consistent processing of foreground and background objects, particularly in scenes with varying depths. TUDA [5] and Semi-UIR [3] enhance image quality but are less effective in restoring fine details and correcting bluish/greenish casts. SyreaNet [4] restores fine details well but often results in


Fig. 12. Visual comparison of our proposed SSD-UIE and other SOTA UIE methods on real-world images in EUVP [47] (rows 1 and 2) and RUIE [46] (rows 3 and 4) datasets. Zoom in for a better view.

TABLE III
QUANTITATIVE COMPARISON OF NONREFERENCE METRICS FOR UIE METHODS ACROSS VARIOUS REAL-WORLD UNDERWATER DATASETS

| Dataset | UIEB [10] | | | | EUVP [47] | | | | RUIE [46] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Methods | UIQM ↑ | UCIQE ↑ | URanker ↑ | TMix ↑ | UIQM ↑ | UCIQE ↑ | URanker ↑ | TMix ↑ | UIQM ↑ | UCIQE ↑ | URanker ↑ | TMix ↑ |
| Raw Image | 1.117 | 0.487 | 0.154 | 0.743 | 0.802 | 0.469 | 0.236 | 0.744 | 0.571 | 0.415 | -0.874 | 0.326 |
| UDCP [21] | 1.297 | 0.576 | 0.595 | 0.729 | 0.854 | 0.513 | 0.411 | 0.752 | 1.034 | 0.451 | 0.317 | 0.426 |
| WWPF [6] | 1.437 | 0.605 | 2.141 | 1.407 | 1.318 | 0.583 | 1.589 | 1.156 | 1.184 | 0.544 | 1.648 | 0.648 |
| UWCNN [10] | 0.987 | 0.457 | -0.197 | 0.359 | 0.751 | 0.438 | 0.147 | 0.497 | 0.618 | 0.441 | -0.157 | 0.407 |
| CECF [25] | 1.387 | 0.601 | 2.422 | 1.346 | 1.109 | 0.591 | 1.614 | 1.225 | 1.028 | 0.543 | 1.869 | 0.639 |
| GLNet [16] | 1.599 | **0.619** | 1.893 | 1.260 | 1.319 | 0.544 | 1.215 | 1.242 | 1.401 | 0.541 | 1.215 | 0.702 |
| UColor [17] | 1.372 | 0.558 | 1.547 | 1.286 | 1.252 | 0.497 | 1.269 | 1.021 | 1.360 | 0.480 | 0.925 | 0.620 |
| UShapeTrans [18] | 1.581 | 0.544 | 1.869 | 1.357 | 0.987 | 0.538 | 1.384 | 0.631 | 1.058 | 0.537 | 1.675 | 0.319 |
| SyreaNet [4] | **1.607** | 0.536 | 1.328 | 1.200 | 1.248 | 0.525 | 1.379 | 0.569 | 1.363 | 0.481 | 0.925 | 0.620 |
| Semi-UIR [3] | 1.510 | 0.603 | 1.834 | 1.358 | **1.409** | 0.581 | **1.723** | 1.192 | 1.024 | 0.536 | 1.391 | 0.723 |
| TUDA [5] | 1.544 | 0.593 | **2.701** | 1.352 | 1.281 | 0.588 | 1.530 | 0.893 | **1.436** | **0.566** | 1.299 | 0.437 |
| SSD-UIE (ours) | 1.569 | 0.613 | 2.624 | **1.732** | 1.336 | **0.592** | 1.657 | **1.321** | 1.347 | 0.558 | **2.179** | **0.929** |

**Note:** The best and second-best are in **bold** and underline, respectively.

darker images, reducing clarity and visual appeal. In contrast, our proposed SSD-UIE significantly enhances visual quality, excels in detail restoration, and effectively eliminates both bluish and greenish casts.

Fig. 12 shows the enhancement results on the EUVP [47] and RUIE [46] datasets. The performance of various UIE methods on these datasets is consistent with the results observed on the UIEB [10] dataset. The proposed SSD-UIE outperforms other UIE methods by significantly enhancing the visual quality of underwater images while also effectively restoring details and correcting color casts.

### *D. Quantitative Comparison on Real Underwater Images*

To quantitatively assess the performance of our proposed SSD-UIE method on real-world underwater images, we utilize a range of nonreference metrics that evaluate different aspects of image quality. These include underwater image quality measure (UIQM) [56], which evaluates the overall quality of underwater images by considering factors such as sharpness, contrast, and brightness; underwater color image quality evaluation (UCIQE) [57], which specifically assesses the color quality of underwater images by measuring the preservation of color fidelity and the reduction of color distortion; URanker [58], a learning-based metric that ranks the quality of underwater images based on both global and local degradation; and TMix [59], another rank learning-based metric that combines high- and low-quality underwater images. As shown in Table III, our SSD-UIE achieves SOTA results across various real underwater image datasets, demonstrating its superior performance in enhancing underwater image quality.

In addition, a subjective user study was conducted to assess the enhancement quality of various methods. Ten enhanced underwater images were randomly selected from each method across all test datasets. A total of 35 participants rated the

TABLE IV
USER STUDY OF UIE METHODS ON REAL UNDERWATER DATASETS

| Dataset | UIEB [10] | EUVP [47] | RUIE [46] | Average |
|---|---|---|---|---|
| Raw Image | 4.36 | 4.12 | 3.54 | 4.01 |
| UDCP [21] | 3.22 | 3.08 | 2.78 | 3.03 |
| WWPF [6] | 5.38 | 5.85 | 4.95 | 5.39 |
| UWCNN [10] | 3.01 | 2.94 | 2.51 | 2.82 |
| CECF [25] | 6.88 | 7.16 | 6.13 | 6.72 |
| GLNet [16] | 6.78 | 7.38 | 5.37 | 6.51 |
| UColor [17] | 6.63 | 7.03 | 5.78 | 6.48 |
| UShapeTrans [18] | 6.81 | 7.26 | 5.26 | 6.44 |
| SyreaNet [4] | 6.50 | 6.82 | 5.17 | 6.16 |
| Semi-UIR [3] | 6.94 | **7.82** | 6.09 | 6.95 |
| TUDA [5] | 7.03 | 7.24 | 5.98 | 6.75 |
| SSD-UIE (ours) | **7.35** | 7.41 | **6.31** | **7.02** |

**Note:** The best and second-best are in **bold** and underline, respectively.

TABLE V
AVERAGE RGB ANGULAR ERROR FOR DIFFERENT UIE METHODS ON DIFFERENT SITES IN THE SEA-THRU DATASET [35]

| Sites | D1 | D2 | D3 | D4 | D5 | All |
|---|---|---|---|---|---|---|
| Raw Image | 28.84 | 32.79 | 35.39 | 30.44 | 33.23 | 33.10 |
| UDCP [21] | 35.11 | 40.48 | 43.13 | 35.56 | 40.26 | 38.33 |
| WWPF [6] | 28.92 | 31.82 | 34.17 | 30.38 | 33.24 | 32.54 |
| UWCNN [10] | 29.38 | 35.12 | 30.05 | 28.15 | 30.13 | 31.42 |
| CECF [25] | 18.94 | 25.13 | 26.45 | 16.31 | 22.68 | 21.90 |
| GLNet [16] | 18.30 | 22.96 | 27.03 | 15.43 | 18.99 | 19.12 |
| UColor [17] | 17.38 | 21.71 | 25.21 | 17.66 | 22.15 | 20.06 |
| UShapeTrans [18] | **10.93** | 14.77 | **15.74** | 12.35 | 16.26 | 15.07 |
| SyreaNet [4] | 13.24 | 15.30 | 20.18 | 13.08 | 15.43 | 15.44 |
| Semi-UIR [3] | 20.27 | 24.34 | 27.85 | 18.96 | 22.40 | 21.39 |
| TUDA [5] | 14.43 | 16.98 | 19.09 | 12.20 | 15.82 | 14.92 |
| SSD-UIE (ours) | 12.80 | **12.08** | 16.46 | **9.87** | **11.62** | **13.15** |

**Note:** The best and second-best are in **bold** and underline, respectively.

images on a scale of 0–10, with higher scores indicating better quality. The results, as shown in Table IV, highlight the superior performance of our proposed method compared to other UIE methods.

While our SSD-UIE method performs well in nonreference metrics and user evaluations, these primarily reflect visual appeal. To thoroughly evaluate its effectiveness in color restoration and correcting water-related casts, we use the average RGB angular error $\bar{\psi}$ [35], which measures the angular difference between enhanced and reference color patches from a color chart. This metric is calculated using images from the Sea-Thru dataset [35], which includes color charts from five distinct underwater sites. The results in Table V demonstrate the superior color recovery of our method across different sites. Fig. 13 provides visual examples, highlighting its effectiveness in restoring true colors, especially for distant objects.

### E. Ablation Study

*1) Effectiveness of LSRUID:* Since our proposed framework employs semisupervised training using both synthetic and real underwater images from our LSRUID dataset, we evaluate their respective impacts separately.

We first replace the synthetic underwater images in our framework with those from other synthetic underwater datasets to compare their effectiveness in reducing the visual interdomain gap. We evaluate the trained networks using the

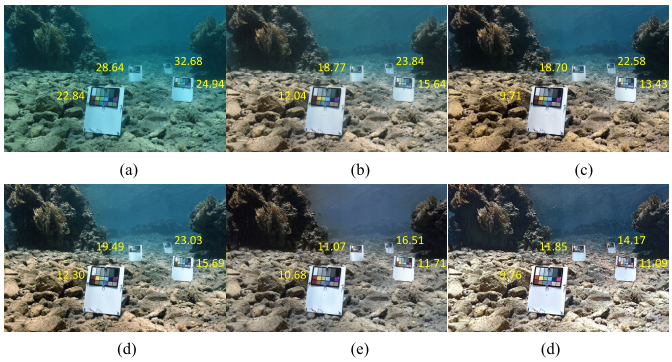



Fig. 13. Example of the RGB angular error $\bar{\psi}$ (the lower the better) of various UIE methods. (a) Raw input. (b) UColor. (c) GLNet. (d) Semi-UIR. (e) TUDA. (d) SSD-UIE (ours).

TABLE VI
QUANTITATIVE COMPARISON FOR NETWORKS TRAINED WITH DIFFERENT SYNTHETIC UNDERWATER IMAGES

| Synthetic Images | UIQM ↑ | UCIQE ↑ | URanker ↑ | TMix ↑ | $\bar{\psi}$ ↓ |
|---|---|---|---|---|---|
| -w. UWCNN [15] | 0.902 | 0.455 | 0.771 | 0.729 | 20.55 |
| -w. SUID [48] | 0.986 | 0.501 | 1.254 | 0.896 | 17.24 |
| -w. LNRUD [31] | 1.144 | 0.574 | 1.606 | 0.965 | 19.38 |
| -w. SYREA [4] | 1.258 | 0.543 | 1.732 | 0.958 | 15.04 |
| -w. WaterGAN [29] | 0.941 | 0.483 | 0.831 | 0.797 | 18.25 |
| -w. WGAN [49] | 1.025 | 0.538 | 1.264 | 0.893 | 17.48 |
| -w. LSRUID (ours) | **1.417** | **0.587** | **2.153** | **1.327** | **13.15** |

**Note:** The best and second-best are in **bold** and underline, respectively.

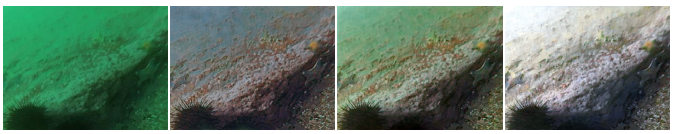



Fig. 14. Enhancement results when training with different synthetic underwater datasets.

nonreference metrics on real underwater images from the UIEB [10], EUVP [47], and RUIE [46] datasets, as well as the average RGB angular error $\bar{\psi}$ on the Sea-Thru dataset [35]. As shown in Table VI, the framework trained with synthetic images from our LSRUID dataset consistently outperforms those trained with other synthetic datasets across all evaluated metrics, highlighting the significant impact that domain differences have on model performance. The visual examples in Fig. 14 further demonstrate the effectiveness of our synthesis method in reducing the visual interdomain gap, making synthetic images more closely resemble real underwater images.

We then substitute the real underwater images in the LSRUID training set with images from other underwater datasets while keeping the synthetic images from our LSRUID dataset. The results presented in Table VII indicate that networks trained and tested on the same real-world dataset perform well but show reduced performance when evaluated on other datasets. In contrast, the real underwater images from our LSRUID dataset exhibit superior generalization across various real underwater datasets, attributed to its diverse and comprehensive content coverage.

TABLE VII

QUANTITATIVE COMPARISON FOR NETWORKS TRAINED WITH DIFFERENT REAL-WORLD UNDERWATER IMAGES

| Real-World Images | UIEB [10] | | EUVP [47] | | RUIE [46] | |
|---|---|---|---|---|---|---|
| | UIQM ↑ | TMix ↑ | UIQM ↑ | TMix ↑ | UIQM ↑ | TMix ↑ |
| -w. UIEB [10] | 1.538 | **1.741** | 1.023 | 0.978 | 0.954 | 0.785 |
| -w. EUVP [47] | 1.215 | 1.214 | **1.341** | 1.297 | 0.872 | 0.751 |
| -w. RUIE [46] | 1.149 | 1.325 | 0.975 | 0.896 | 1.324 | 0.916 |
| -w. LSRUID (ours) | **1.569** | 1.732 | 1.336 | **1.321** | **1.347** | **0.929** |

**Note:** The best and second-best are in **bold** and underline, respectively.

TABLE VIII

ABLATION STUDY OF VARIOUS DA STRATEGIES

| Ablations | | $Ab_1$ | $Ab_2$ | $Ab_3$ | $Ab_4$ | $Ab_5$ | $Ab_6$ | Full |
|---|---|---|---|---|---|---|---|---|
| semantic *inter*-DA | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| *intra*-DA | -w. IFE-Hist | ✓ | ✓ | ✓ | | | | ✓ |
| | -w. IFE-Freq | ✓ | ✓ | | ✓ | | | ✓ |
| | -w. $\mathcal{L}_{\text{intra}}$ | ✓ | | ✓ | ✓ | ✓ | | ✓ |
| UCIQE ↑ | | 0.538 | 0.546 | 0.553 | 0.559 | 0.551 | 0.542 | **0.587** |
| TMix ↑ | | 1.225 | 1.294 | 1.310 | 1.307 | 1.304 | 1.287 | **1.327** |
| $\bar{\psi}$ ↓ | | 14.08 | 14.86 | 14.56 | 14.98 | 15.01 | 15.55 | **13.15** |

**Note:** The best and second-best are in **bold** and underline, respectively.

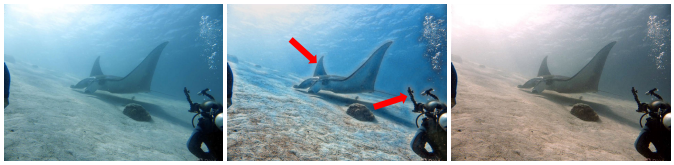



Fig. 15. Enhancement results with and without semantic inter-DA strategy.

*2) Effectiveness of Semantic Inter-DA Strategy:* To evaluate the effectiveness of the semantic inter-DA strategy, it is important to note that the interdomain loss $\mathcal{L}_{\text{inter}}$ is specifically designed to function in conjunction with the semantic blender, and their effects cannot be tested independently. Therefore, we assess the impact of the semantic inter-DA strategy by removing both the semantic blender and $\mathcal{L}_{\text{inter}}$. The results, shown in model $Ab_1$ in Table VIII, indicate that the absence of the inter-DA strategy primarily affects nonreference metrics, while the performance in color restoration is only slightly diminished. Fig. 15 demonstrates that without the inter-DA strategy, the network struggles to effectively handle foreground objects in real underwater scenarios, particularly noticeable in the presence of halos around underwater objects.

*3) Effectiveness of Intra-DA Strategy:* Given that the intra-DA strategy introduces the IFE branch (including IFE-Hist and IFE-Freq) and the intradomain loss $\mathcal{L}_{\text{intra}}$, we assess the impact of each module individually. The models $Ab_2$ to $Ab_6$ in Table VIII provide quantitative comparisons for each component of the intra-DA strategy. The results indicate that omitting either the IFE branch or $\mathcal{L}_{\text{intra}}$ significantly reduces the network's color restoration capabilities, highlighting the importance of each module in dealing with intra-DA gap. In addition, the t-SNE visualizations in Fig. 16 show that with our proposed intra-DA strategy, the framework could effectively align the features of underwater images from different underwater scenarios, leading to a significant reduction in the intradomain gap.

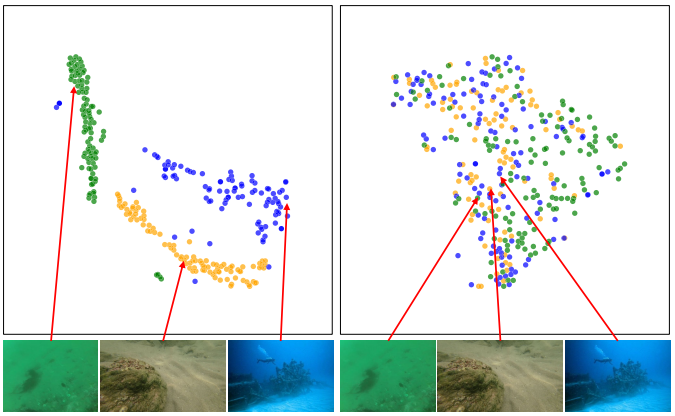



Fig. 16. t-SNE visualization of underwater images from different environments before and after intra-DA. (a) Before inter-DA. (b) After inter-DA.

TABLE IX

EXPERIMENTS ON THE EFFECTIVENESS OF DUAL-TRANS BLOCK

| Ablations | UIQM ↑ | UCIQE ↑ | URanker ↑ | TMix ↑ | $\bar{\psi}$ ↓ |
|---|---|---|---|---|---|
| -w/o. SW-MSA | 1.386 | 0.509 | 1.978 | 1.291 | 14.59 |
| -w/o. G-MSA | 1.374 | 0.497 | 2.021 | 1.267 | 15.22 |
| -w. CNN block | 1.251 | 0.453 | 1.556 | 1.254 | 16.48 |
| Ours | **1.417** | **0.587** | **2.153** | **1.327** | **13.15** |

**Note:** The best and second-best are in **bold** and underline, respectively.

TABLE X

PARAMETER COUNTS AND INFERENCE TIME OF VARIOUS METHODS ACROSS DIFFERENT INPUT SIZES

| Methods | Parameters (M)↓ | Inference Speed (ms) ↓ | | |
|---|---|---|---|---|
| | | 128 × 128 | 256 × 256 | 512 × 512 |
| UshapeTrans [18] | 31.59 | - | 23.08 | - |
| SyreaNet [4] | 29.03 | **7.29** | 17.29 | 63.25 |
| Semi-UIR [3] | **1.67** | 15.81 | **17.20** | **55.48** |
| TUDA [5] | 5.93 | 40.80 | 41.70 | 93.06 |
| SSD-UIE (ours) | 7.80 | 13.97 | 23.18 | 84.56 |

**Notes:** 1. The best and second-best are in **bold** and underline, respectively.
2. - indicates that the input size is not supported.

*4) Effectiveness of Dual-Trans Block:* To evaluate the effectiveness of our proposed Dual-Trans Block, we conduct a comparison by removing either SW-MSA or G-MSA components and by replacing the Dual-Trans Block with a conventional CNN block. The results, presented in Table IX, demonstrate that our Dual-Trans Block outperforms the other ablation models. In addition, we assess the computational performance of our framework during inference by measuring the inference time for different input sizes. As illustrated in Table X, the inference time of our proposed SSD-UIE scales linearly with input size. A comparison of SSD-UIE's inference time with other SOTA methods is also provided, highlighting its competitive computational efficiency across varying input sizes.

## V. CONCLUSION

In this study, we proposed SSD-UIE, a novel semisupervised framework that effectively addresses both interdomain and intradomain challenges in real-world UIE. To tackle the interdomain gap, we developed the PUIS module to generate high-quality synthetic underwater images and alleviate visual discrepancies between synthetic and real data, complemented

by the establishment of the comprehensive LSRUID dataset. In addition, the semantic blender was introduced to address semantic interdomain differences. For intradomain challenges, the IFE branch and feature alignment strategy were designed to adapt to diverse underwater conditions. The Dual-Trans Block was also introduced, enhancing performance while ensuring computational efficiency. Extensive experiments demonstrate that SSD-UIE outperforms SOTA methods, with ablation studies validating the contribution of each component. Future work will explore extending the framework to additional underwater vision tasks, including object detection and 3-D reconstruction, broadening its applicability in geoscience and remote sensing.

## REFERENCES


[1] S.-H. Cho, H.-K. Jung, H. Lee, H. Rim, and S. K. Lee, "Real-time underwater object detection based on DC resistivity method," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6833–6842, Nov. 2016.

[2] F. Zhang, H. Bian, and M. Wei, "Contrastive learning ideas in underwater terrain image matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5923212, doi: 10.1109/TGRS.2022.3222500.

[3] S. Huang, K. Wang, H. Liu, J. Chen, and Y. Li, "Contrastive semi-supervised learning for underwater image restoration via reliable bank," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 18145–18155.

[4] J. Wen et al., "SyreaNet: A physically guided underwater image enhancement framework integrating synthetic and real images," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2023, pp. 5177–5183.

[5] Y.-W. Chen and S.-C. Pei, "Domain adaptation for underwater image enhancement via content and style separation," 2022, *arXiv:2202.08537*.

[6] W. Zhang et al., "Underwater image enhancement via weighted wavelet visual perception fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2469–2483, Apr. 2024.

[7] G. Wang et al., "RUE-net: Advancing underwater vision with live image enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5642814, doi: 10.1109/TGRS.2024.3404662.

[8] F. Xiao, J. Liu, Y. Huang, E. Cheng, and F. Yuan, "Neuromorphic computing network for underwater image enhancement and beyond," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5644417, doi: 10.1109/TGRS.2024.3473020.

[9] Y. Liu et al., "From synthetic to real: Image dehazing collaborating with unlabeled real data," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 50–58.

[10] C. Li et al., "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.

[11] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6723–6732.

[12] D. Berman, T. Treibitz, and S. Avidan, "Diving into haze-lines: Color restoration of underwater images," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, vol. 1, no. 2, Sep. 2017, p. 2.

[13] N. G. Jerlov, "Classification of sea water in terms of quanta irradiance," *ICES J. Mar. Sci.*, vol. 37, no. 3, pp. 281–287, Sep. 1977.

[14] Z. Wang, L. Shen, M. Xu, M. Yu, K. Wang, and Y. Lin, "Domain adaptation for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 1442–1457, 2023.

[15] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107038.

[16] X. Fu and X. Cao, "Underwater image enhancement with global–local networks and compressed-histogram equalization," *Signal Process., Image Commun.*, vol. 86, Aug. 2020, Art. no. 115892.

[17] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, 2021.

[18] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 3066–3079, 2023.

[19] A. S. A. Ghani and N. A. M. Isa, "Automatic system for improving underwater image contrast and color through recursive adaptive histogram modification," *Comput. Electron. Agricult.*, vol. 141, pp. 181–195, Sep. 2017.

[20] A. Galdran, D. Pardo, A. Picón, and A. Alvarez-Gila, "Automatic red-channel underwater image restoration," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 132–145, Jan. 2015.

[21] P. Drews Jr., E. do Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 825–830.

[22] R. Du, W. Li, S. Chen, C. Li, and Y. Zhang, "Unpaired underwater image enhancement based on CycleGAN," *Information*, vol. 13, no. 1, p. 1, Dec. 2021.

[23] Z. Jiang, Z. Li, S. Yang, X. Fan, and R. Liu, "Target oriented perceptual adversarial fusion network for underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 10, pp. 6584–6598, Oct. 2022.

[24] R. Chen, Z. Cai, and W. Cao, "MFFN: An underwater sensing scene image enhancement method based on multiscale feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4205612.

[25] X. Cong, J. Gui, and J. Hou, "Underwater organism color fine-tuning via decomposition and guidance," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, Mar. 2024, pp. 1389–1398.

[26] Y. Bahat and M. Irani, "Blind dehazing using internal patch recurrence," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, May 2016, pp. 1–9.

[27] S. Anwar, C. Li, and F. Porikli, "Deep underwater image enhancement," 2018, *arXiv:1807.03528*.

[28] C. Desai, R. A. Tabib, S. S. Reddy, U. Patil, and U. Mudenagudi, "RUIG: Realistic underwater image generation towards restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 2181–2189.

[29] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2018.

[30] Y. Shi, L. Han, L. Han, S. Chang, T. Hu, and D. Dancey, "A latent encoder coupled generative adversarial network (LE-GAN) for efficient hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5534819, doi: 10.1109/TGRS.2022.3193441.

[31] T. Ye, S. Chen, Y. Liu, Y. Ye, E. Chen, and Y. Li, "Underwater light field retention: Neural rendering for underwater imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 487–496.

[32] X. Cong, Y. Zhao, J. Gui, J. Hou, and D. Tao, "A comprehensive survey on underwater image enhancement based on deep learning," 2024, *arXiv:2405.19684*.

[33] Q. Jiang, Y. Zhang, F. Bao, X. Zhao, C. Zhang, and P. Liu, "Two-step domain adaptation for underwater image enhancement," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108324.

[34] N. Qiao, C. Sun, L. Dong, and Q. Ge, "Semi-supervised feature distillation and unsupervised domain adversarial distillation for underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 7671–7682, Aug. 2024.

[35] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1682–1691.

[36] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," 2021, *arXiv:2103.13413*.

[37] A. Kirillov et al., "Segment anything," 2023, *arXiv:2304.02643*.

[38] A. Radford et al., "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn.*, vol. 139, 2021, pp. 8748–8763.

[39] W. Cong et al., "DoveNet: Deep image harmonization via domain verification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8394–8403.

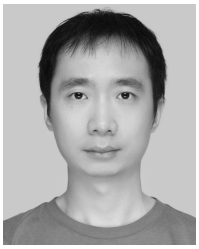
[40] J. Yoo, Y. Uh, S. Chun, B. Kang, and J.-W. Ha, "Photorealistic style transfer via wavelet transforms," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9036–9045.

[41] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[42] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.

[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[44] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Jun. 2016, pp. 694–711.
[45] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.
[46] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4861–4875, Dec. 2020.
[47] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.
[48] G. Hou, X. Zhao, Z. Pan, H. Yang, L. Tan, and J. Li, "Benchmarking underwater image enhancement and restoration, and beyond," *IEEE Access*, vol. 8, pp. 122078–122091, 2020.
[49] N. Wang, Y. Zhou, F. Han, H. Zhu, and J. Yao, "UWGAN: Underwater GAN for real-world underwater color restoration and dehazing," 2019, *arXiv:1912.10269*.
[50] R. Raturi, "Adapting deep features for scene recognition utilizing places database," in *Proc. 2nd Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, vol. 27, Apr. 2018, pp. 184–189.
[51] Z. Li and N. Snavely, "MegaDepth: Learning single-view depth prediction from internet photos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2041–2050.
[52] M. Cordts et al., "The cityscapes dataset," in *Proc. CVPR Workshop Future Datasets Vis.*, Jan. 2015.
[53] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2017.
[54] M. Bińkowski, D. J. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," 2018, *arXiv:1801.01401*.
[55] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, Mar. 1951.
[56] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.
[57] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
[58] C. Guo et al., "Underwater ranker: Learn which is better and how to be better," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 702–709.
[59] Z. Fu, X. Fu, Y. Huang, and X. Ding, "Twice mixing: A rank learning based quality assessment approach for underwater image enhancement," *Signal Process., Image Commun.*, vol. 102, Mar. 2022, Art. no. 116622.



**Junjie Wen** received the B.Sc. degree in automotive engineering from Dalian University of Technology, Dalian, China, in 2013, and the M.Sc. degree in mechanical engineering from Tsinghua University, Beijing, China, in 2016. He is currently pursuing the Ph.D. degree in mechanical and automation engineering with The Chinese University of Hong Kong, Hong Kong, China.

His research interests include image restoration, restoration-oriented detection, and vision-based navigation.

**Guidong Yang** received the B.Eng. degree in mechanical and automation engineering from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2018, the M.Eng. degree in vehicle engineering from SJTU, and the M.Sc. degree in mechanical engineering from Politecnico di Milano, Milan, Italy, In 2021. He is currently pursuing the Ph.D. degree in mechanical and automation engineering with The Chinese University of Hong Kong, Hong Kong, China.

His research interests include image enhancement, multiview stereo, and unmanned aerial vehicle systems.

**Benyun Zhao** received the M.Sc. degree in mechanical and automation engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, China, in 2021, where he is currently pursuing the Ph.D. degree in mechanical and automation engineering.

He was a Research Assistant at CUHK and Hong Kong Centre for Logistics Robotics (HKCLR), Hong Kong, from 2021 to 2022. He is also a Visiting Student at HiPeRLab, University of California at Berkeley, Berkeley, CA, USA. His research interests include object detection, semantic segmentation, and 3-D scene understanding.

**Dongyue Huang** received the joint B.E. degree in mechanical engineering from Queen's University Belfast, Belfast, U.K., and Guangdong University of Technology, Guangzhou, China, in 2019, and the M.Sc. and Ph.D. degrees in mechanical and automation engineering from The Chinese University of Hong Kong, Hong Kong, in 2021 and 2024, respectively.

He is currently an Assistant Researcher at Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS), Shenzhen, China. His research interests include the design and control of unmanned aerial vehicle multimodal systems.

**Lei Lei** received the B.E. degree in naval architecture and ocean engineering from the Department of Ship Engineering, Harbin Engineering University, Harbin, China, in 2016, the M.E. degree in mechanical engineering from the Department of Mechanical Science and Engineering, Huazhong University of Science and Engineering, Wuhan, China, in 2019, and the Ph.D. degree from the Department of Systems Engineering, City University of Hong Kong, Hong Kong, in 2024.

He is currently a Post-Doctoral Fellow at the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong. His research interests include underwater unmanned aerial vehicle systems and ocean big data.

**Bo Zhang** received the Ph.D. degree in aircraft design from Northwestern Polytechnical University, Xi'an, China, in 2013.

He is currently a Research Professor (International Doctoral Supervisor) with the College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China, where he works as the Deputy Director and the Technical Leader of Shenzhen City Joint Laboratory of Autonomous Unmanned Aerial Vehicle Systems and Intelligent Manipulation (AUSIM), Shenzhen, and leads the Geospatial-Wide Intelligent Perception Platform (GIPP) Group, Guangdong Laboratory of Artificial Intelligence and Digital Economy, Shenzhen. His research interests include cooperative intelligent robots, autonomous unmanned aerial vehicle systems, and ocean wave energy converter arrays.

**Zhi Gao** received the B.E. and Ph.D. degrees from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In 2008, he joined the Interactive and Digital Media Institute, National University of Singapore (NUS), Singapore, as a Research Fellow and the Project Manager. In 2014, he joined the Temasek Laboratories, NUS (TL@NUS), Singapore, as a Research Scientist and the Principal Investigator. He is currently a Full Professor with the School of Remote Sensing and Information Engineering, Wuhan University. He has published more than 90 academic articles, which have been published in IJCV, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, and other top journals.

Dr. Gao received the Prestigious "National Plan for Young Talents" Award and Hubei Province Funds for Distinguished Young Scientists. In addition, he is a "Chutian Scholar" Distinguished Professor in Hubei. He serves as an Associate Editor for *Unmanned Systems* journal.

**Xi Chen** is currently a Research Assistant Professor of mechanical and automation engineering with The Chinese University of Hong Kong (CUHK), Hong Kong. He has over ten years of experience in sustainable building technology related to urban energy systems, renewable applications in buildings, and built environment modeling and has led or managed multiple research projects, including ARC, MOST, RGC, and consultancy projects with the local government and industry. He has published over 40 articles in peer-reviewed international journals and co-authored a book in green building and renewable application areas.

Dr. Chen has been awarded the DECRA Fellow by Australian Research Council and a Fulbright Scholar at the Lawrence Berkeley National Laboratory. In addition, he serves as an Editorial Board Member for *Buildings*, *Energies*, and *Advances in Applied Energy*.

**Ben M. Chen** (Fellow, IEEE) was an Assistant Professor with the Department of Electrical Engineering, State University of New York at Stony Brook, Stony Brook, NY, USA, from 1992 to 1993. He was a Provost's Chair Professor with the Department of Electrical and Computer Engineering, National University of Singapore (NUS), Singapore, before joining CUHK, in 2018. He is currently a Professor of mechanical and automation engineering with The Chinese University of Hong Kong (CUHK), Hong Kong, China. He has authored/co-authored hundreds of journal and conference articles and a dozen research monographs in control theory and applications, unmanned aerial vehicle systems, and financial market modeling. His research interests include unmanned aerial vehicle systems and their applications.

Dr. Chen is a fellow of the Academy of Engineering, Singapore. He served on the editorial boards for a dozen international journals, including *Automatica* and IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He is serving as an Editor-in-Chief for *Unmanned Systems* and an Editor for *International Journal of Robust and Nonlinear Control*.