



Unsupervised underwater image enhancement via content-style representation disentanglement[☆]

Pengli Zhu^{a,b}, Yancheng Liu^a, Yuanquan Wen^a, Minyi Xu^a, Xianping Fu^c, Siyuan Liu^{a,*}

^a The College of Marine Engineering, Dalian Maritime University, Dalian, China

^b The College of Design and Engineering, National University of Singapore, Singapore

^c The College of Information Science and Technology, Dalian Maritime University, Dalian, China

ARTICLE INFO

Keywords:

Underwater image enhancement
Representation disentanglement
Unsupervised learning
Cycle-consistent adversarial translation

ABSTRACT

The absorption and scattering properties of the water medium cause various types of distortion in underwater images, which seriously affects the accuracy and effectiveness of subsequent processing. The application of supervised learning algorithms in underwater image enhancement is limited by the difficulty of obtaining a large number of underwater paired images in practical applications. As a solution, we propose an unsupervised representation disentanglement based underwater image enhancement method (URD-UIE). URD-UIE disentangles content information (e.g., texture, semantics) and style information (e.g., chromatic aberration, blur, noise, and clarity) from underwater images and then employs the disentangled information to generate the target distortion-free image. Our proposed method URD-UIE adopts an unsupervised cycle-consistent adversarial translation architecture and combines multiple loss functions to impose specific constraints on the output results of each module to ensure the structural consistency of underwater images before and after enhancement. The experimental results demonstrate that the URD-UIE technique effectively enhances the quality of underwater images when training with unpaired data, resulting in a significant improvement in the performance of the standard model for underwater object detection and semantic segmentation.

1. Introduction

Various factors distort the quality of underwater images, notably scattering effects that cause blurred details and low contrast, as well as wavelength absorption, resulting in color deviation. Underwater image enhancement (UIE) plays a vital role in perceiving the underwater environment. It aims to remove or alter scene-specific style properties (e.g., chromatic aberration, blur, noise, etc.) while preserving scene-invariant content properties (e.g., texture semantics). This process improves the overall and local characteristics of acquired images, enriches available information, and enhances the capability of image interpretation and recognition.

The UIE methods can be either physical-model based or non-physical model based, as summarized in Table 1. Physical-model based methods (Peng and Cosman, 2017; Song et al., 2020; Zhou et al., 2022c,a; Zhuang et al., 2022) build underwater imaging models (McGlamery, 1980; Jaffe, 1990; Zhou et al., 2023c) based on the principles of underwater image degradation and invert the original image

before degradation by a mathematical process. This kind of methods possesses limited image recovery capability due to model parameter assumptions, i.e., the prior estimates of global atmospheric light and medium transmission rates. Non-physical model based methods (Zhang et al., 2021a; Zhuang et al., 2021; Zhou et al., 2022b; Li et al., 2022; Zhou et al., 2023a,b,d) make pixel-wise contrast, color and sharpness corrections to images based on the characteristics of underwater images. However, the enhancement process might cause information loss and introduce additional artifacts since imaging characteristics and image contents are not taken into account. In addition, these two categories of methods are not applicable to extreme underwater images with multiple distortions, yielding weak robustness in image recovery and superior difficulty in underwater video enhancement.

Recently, deep learning techniques provide an alternative for intelligent UIE, which are driven by either synthetic images (Sun et al., 2019; Anwar et al., 2018) or real-world underwater images (Li et al., 2019; Qi et al., 2021; Zhou et al., 2023b). Additionally, based on the

[☆] This work was supported by the National Natural Science Foundation (NSF) of China under Grants 51709028 and 51979021, Liaoning Provincial Key Research and Development Program under Grant 2017220005 and 2019JH8/10100045, China Scholarship Council (CSC) under Grant 202206570013, Dalian Key Laboratory Construction Project under Grant 84180135.

* Corresponding author.

E-mail address: dmu.s.y.liu@gmail.com (S. Liu).

Table 1
A brief description of the UIE studies.

Reference	Framework	Methodology					Year
		Deep learning	Unsupervised	GAN	Cycle-consistent	Representation disentanglement	
IBLA (Peng and Cosman, 2017)	Physical-model	–	✓	–	–	–	2017
SMBLO (Song et al., 2020)		–	✓	–	–	–	2020
Ucolor (Li et al., 2021)		✓	–	–	–	–	2021
HLRP (Zhuang et al., 2022)		–	✓	–	–	–	2022
UWdepth (Wang et al., 2023b)		✓	✓	–	–	–	2023
WaterGAN (Li et al., 2017)		✓	✓	✓	–	–	2017
UGAN (Fabbri et al., 2018)		✓	–	✓	✓	–	2018
UWGAN (Li et al., 2018)	Non-physical model	✓	–	✓	✓	–	2018
CCBC (Zhang et al., 2021a)		–	✓	–	–	–	2021
Bayesian (Zhuang et al., 2021)		–	✓	–	–	–	2021
FUnIE-GAN (Islam et al., 2020)		✓	✓	✓	✓	–	2021
ACCE-D (Li et al., 2022)		–	✓	–	–	–	2022
TACL (Liu et al., 2022)		✓	✓	✓	✓	–	2022
CVE-Net (Zhou et al., 2023d)		✓	✓	–	–	–	2023
UMGAN (Sun et al., 2023)		✓	✓	✓	✓	–	2023
URD-UIE		✓	✓	✓	✓	✓	2023

learning strategy employed, these methods can be further categorized into two distinct categories: supervised UIE and unsupervised UIE methods. Supervised UIE methods (Anwar et al., 2018; Sun et al., 2019; Li et al., 2019; Hu et al., 2021; Wang et al., 2023b) typically employ various encoder-decoder network architectures and rely on paired data, i.e., distortion-corrupted images and corresponding distortion-free images. By utilizing this paired data, the supervised UIE models are trained in a supervised fashion to enhance various image quality indices such as contrast, sharpness, and color. The effectiveness of these methods relies to some extent on the quality of the paired datasets, which are commonly generated using a priori information. However, the synthesized images can hardly reflect the real conditions of underwater images, leading to limited generalizability in real-world applications.

To alleviate the reliance on paired datasets for supervised UIE methods, unsupervised UIE techniques have started incorporating generative adversarial network (GAN) (Goodfellow et al., 2014) in the field of image transformation, such as WGAN-GP (Gulrajani et al., 2017), DenseGAN (Guo et al., 2019), WaterGAN (Li et al., 2017), UIE-sGAN (Ye et al., 2018), MCycleGAN (Lu et al., 2019), FUnIE-GAN (Islam et al., 2020), UMGAN (Sun et al., 2023). While these unsupervised UIE methods successfully address the issue of dependence on paired datasets, they are primarily based on the fundamental concept of cycle-consistent (Zhu et al., 2017) to reconstruct underwater images with the aid of additional global similarity loss terms, which will lead to the following two problems: (1) changing content features during style transferring, which is contrary to the original intention of UIE; (2) lacking style diversity for generated images (Lee et al., 2020), which cannot reflect the color diversity of underwater objects. To tackle these aforementioned challenges, researchers have proposed several image-to-image translation techniques that employ representation disentanglement learning (Lee et al., 2018; Huang et al., 2018a; Liu et al., 2021). These methods decompose images into domain-invariant content features and domain-specific style features, followed by a collaborative decoding process of content and style features through multi-scale feature fusion algorithms to generate diverse-style outputs.

In this paper, taking inspiration from prior works (Islam et al., 2020; Huang et al., 2018a), we propose an unsupervised representation disentanglement based underwater image enhancement (URD-UIE) method. The key features of our method are summarized as follows:

- (1) We present URD-UIE, a disentangled representation framework for underwater image enhancement that decomposes an underwater image into two distinct components: content that is invariant across domain and style-specific elements (e.g., chromatic aberration, blur, noise, and clarity), which are specific to each domain.

- (2) Our method proposes an unsupervised disentangled cycle translation architecture, which employs representation disentanglement learning to encode content and style information of images separately, and integrates the disentangled information to generate the target distortion-free image, ensuring consistency of image structure before and after enhancement and style diversity of recovered images.
- (3) To accomplish complete representation disentanglement for better UIE performance, we design multiple loss functions for modules of URD-UIE to ensure the reality and diversity of generated images by various combinations of disentangled style and content features. This research has shown that it can drastically improve the performance of subsequent underwater tasks, including object detection and semantic segmentation.

The remaining sections of this paper are structured as follows. In Table 2, we summarize the notations frequently used throughout this paper. Section 2 provides a concise overview of the existing research on unsupervised learning for UIE and representation disentanglement. In Section 3, we provide an outline of the URD-UIE framework, along with its implementation details. The evaluation of our proposed method, including a comparison to state-of-the-art (SOTA) methods and an ablation study, is presented in Section 4. Finally, Section 5 summarizes the conclusions and discussions derived from our study.

2. Related work

2.1. Unsupervised UIE

In recent years, deep learning techniques have made significant advancements in image processing, contributing to various applications (Zhang et al., 2021b; Wang et al., 2022, 2023a; Zhang et al., 2023b,a). However, when it comes to data-driven image enhancement in real-world underwater environments, where distortion-corrupted images and corresponding distortion-free images are scarce, most existing models (Anwar et al., 2018; Sun et al., 2019; Li et al., 2019; Hu et al., 2021; Wang et al., 2023b) become ineffective due to the lack of a large amount of paired image data. To address the challenge of limited paired datasets, unsupervised learning-based methods have been proposed for UIE, leveraging the concept of cycle-consistency, such as CycleGAN (Zhu et al., 2017) and its variants (Li et al., 2017; Ye et al., 2018; Lu et al., 2019; Islam et al., 2020; Li et al., 2021).

Despite the progress made by unsupervised UIE methods, there are several drawbacks that need to be addressed. Firstly, The effectiveness of image enhancement models is intrinsically linked to the quality of the training datasets, specifically the presence of distortion-free

Table 2

List of notations.

Notation	Description
x_1	Distortion-corrupted image
x_2	Distortion-free image
$x_{1 \rightarrow 2}$	Synthetic image in the distortion-free space
$x_{2 \rightarrow 1}$	Synthetic image in the distortion-corrupted space
\hat{x}	Cross-reconstructed image of x
\tilde{x}	Self-reconstructed image of x
c	Disentangled content features of x
s	Disentangled style features of x
\hat{c}_1	Disentangled content features of $x_{1 \rightarrow 2}$
\hat{c}_2	Disentangled content features of $x_{2 \rightarrow 1}$
\hat{s}_1	Disentangled style features of $x_{2 \rightarrow 1}$
\hat{s}_2	Disentangled style features of $x_{1 \rightarrow 2}$
\mathcal{X}_1	Distortion-corrupted domain imageset
\mathcal{X}_2	Distortion-free domain imageset
\mathcal{C}	Domain-invariant content space
\mathcal{S}_1	Distortion-corrupted domain-specific style space
\mathcal{S}_2	Distortion-free domain-specific style space
E^c	Content encoder
E^s	Style encoder
G	Generator
D	Discriminator
\mathcal{L}_{adv}	Adversarial loss
\mathcal{L}_{cyc}	Cycle consistency loss
\mathcal{L}_{rec}	Self-reconstruction loss
\mathcal{L}_{sty}	Style consistency loss
\mathcal{L}_{cont}	Content consistency loss
λ	Loss weights

images with stylistic diversity (Anwar and Li, 2020), as insufficient diversity in the training data can restrict the generalization capability. Secondly, traditional unsupervised UIE methods (Li et al., 2017; Islam et al., 2020) often introduce content feature changes during the style transferring process, which deviates from the original intention of UIE. To partially overcome these limitations, HybrUR (Yan et al., 2021) and UUIE (Shen et al., 2023) recently proposed a joint architecture for underwater visual restoration, which learns from unpaired images, thus alleviating the dependency on paired data to some extent. However, the excessive complexity of the model affects real-time computational performance, making it less suitable for practical applications.

Motivated by the aforementioned challenges, we take inspiration from image-to-image translation (Lee et al., 2018, 2020; Huang et al., 2018a) and propose to approach the underwater image enhancement problem through image style transformation modeling. This alternative perspective aims to improve the underwater visual perception of UIE while mitigating some of the limitations associated with previous approaches.

2.2. Representation disentanglement learning

The objective of representation disentanglement learning is to develop models that can effectively capture the various factors of data variation (Ding et al., 2020). While some works (Cheung et al., 2014; Makhzani et al., 2015; Mathieu et al., 2016) utilize labeled data to separate class-independent and class-related components, recent studies have explored unsupervised training methods such as InfoGAN (Chen et al., 2016) and β -VAE (Higgins et al., 2016). These methods achieve disentanglement by maximizing the cross-information between potential variables and data variation, enabling representation disentanglement without the need for explicit supervision. However, these methods lack control over the model learning process, thereby limiting their ability to generate desired factors.

To address this limitation, ss-InfoGAN (Spurr et al., 2017) introduces tags during training to achieve better control over the learned data-generating elements, while also allowing for the learning of untagged or unknown factors. Another approach to representation disentanglement involves embedding images into distinct latent spaces. For

example, MUNIT (Huang et al., 2018a) decomposes the latent space into separate content and style spaces, while DRIT (Lee et al., 2018) and DRIT++ (Lee et al., 2020) decompose the potential space into a shared content space and a unique attribute space for the source/target domain.

Although several methods (Chen and Pei, 2022; LIU et al., 2022) have recently utilized the framework of representation disentanglement learning and achieved promising results, their impact on subsequent underwater image processing tasks has not been thoroughly examined. Drawing inspiration from prior work on disentangled representation learning (Pang et al., 2021), we aim to disentangle underwater images into content and style features within corresponding latent spaces to facilitate better image enhancement. Furthermore, we integrate UIE with underwater vision perception enhancement to validate the performance improvements in practical underwater tasks, including object detection and semantic segmentation, etc.

3. Unsupervised representation disentanglement based underwater image enhancement (URD-UIE)

We consider the UIE problem as an image-to-image translation task between the distortion-corrupted and distortion-free domains and propose an unsupervised representation disentanglement-based method for UIE, referred to as URD-UIE, as illustrated in Fig. 1. To tackle the challenge of unpaired datasets dependencies, we introduce a cross-cycle consistency loss to facilitate unsupervised representation disentanglement. Specifically, our method begins by taking a pair of non-aligned underwater images and performing a cross-domain translation mapping, which resulted in intermediate outputs obtained by swapping the style features from both images. Subsequently, we reconstruct the original input image pair by reapplying the cross-domain translation mapping. The purpose of this step is to ensure that the reconstructed images maintain consistency with the original ones. To achieve this goal, we employ two encoders to respectively encode each image into content features in a domain-invariant content space and style codes in a domain-specific style space. Furthermore, we construct a generator that utilizes the encoded content features and style codes to produce enhanced images. By incorporating these components and leveraging the cross-cycle consistency loss, we achieve the objective of enhancing underwater images while preserving their original content. In the subsequent section, we will present a comprehensive overview of each component and the learning strategies employed.

3.1. Content-style representation disentanglement

Let \mathcal{X}_1 and \mathcal{X}_2 represent the domains corresponding to distortion-corrupted and distortion-free underwater images, respectively. Our method aims to achieve content-style representation disentanglement between these two domains by utilizing unpaired training images. It is worth emphasizing that even high-quality underwater images may include certain levels of stylistic information in practical scenarios. To address this, we assume that each underwater image can be expressed as a nonlinear combination of content components (e.g., texture, semantics) and style components (e.g., chromatic aberration, blur, noise, and clarity). For two unpaired images ($x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2$), we employ sequential representation disentanglement translation to achieve cross-reconstruction mappings $x_1 \rightarrow \hat{x}_1$ and $x_2 \rightarrow \hat{x}_2$. Our proposed method effectively disentangles the latent spaces of x_1 and x_2 into two distinct spaces: a shared domain-invariant content space denoted by \mathcal{C} , and a domain-specific style space denoted by \mathcal{S}_i for each domain i , as illustrated in Fig. 1(b). Specifically, let us take the mappings $x_1 \rightarrow \hat{x}_1$ as an example. Firstly, we encode x_1 into two latent spaces: the domain-specific style space \mathcal{S}_1 and the domain-invariant content space \mathcal{C} , yielding the content representation c_1 and the disentangled style representation s_1 , respectively. Using these representations, we construct a generator G_1 that generates an intermediate image \hat{x}_1 in

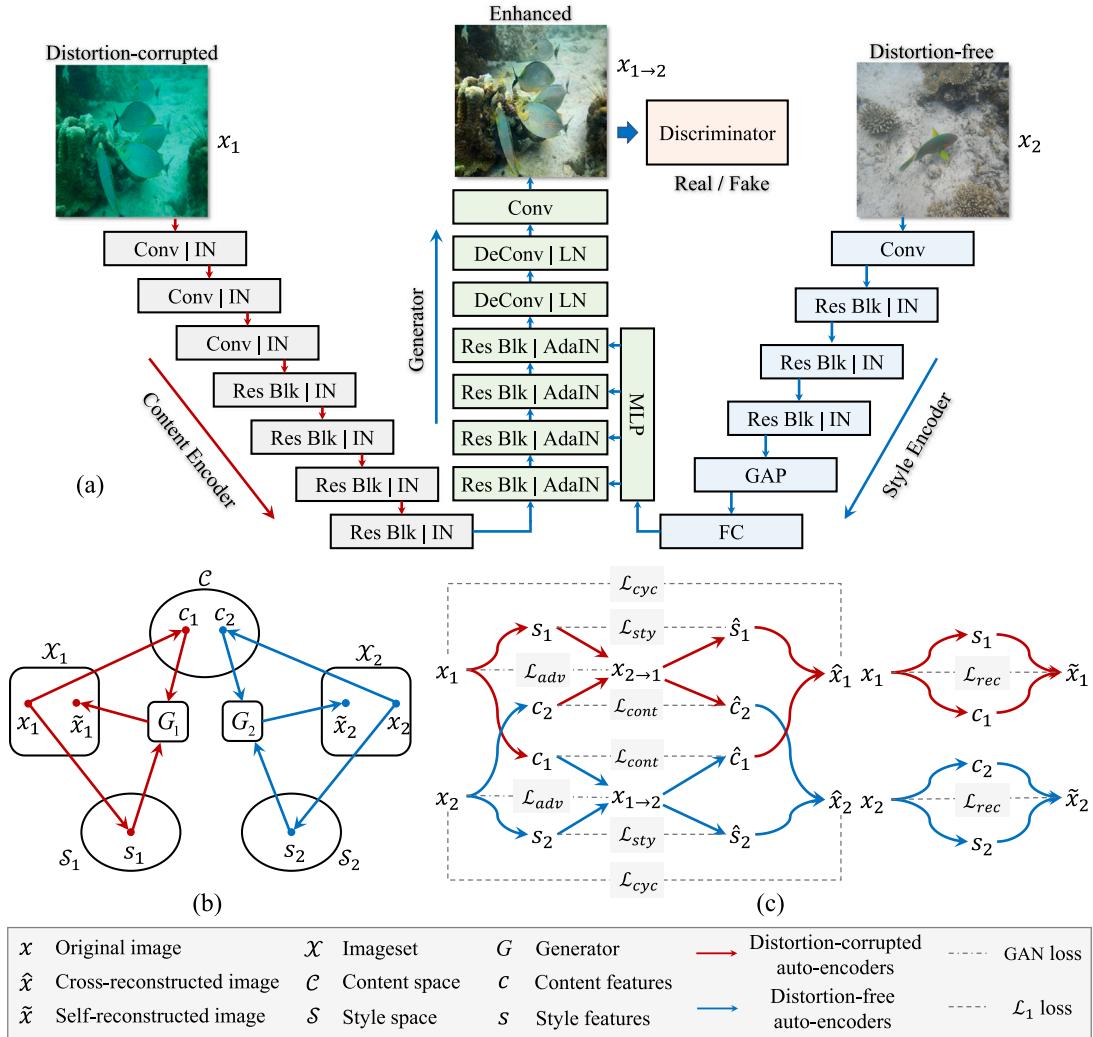


Fig. 1. Overview of URD-UIE. (a) Auto-encoder architecture. (b) Distortion-corrupted and distortion-free domain representation disentanglement. (c) The URD-UIE framework and loss function architecture.

the distortion-corrupted domain \mathcal{X}_1 . The self-reconstruction translation mapping $x_1 \rightarrow \hat{x}_1$ can be summarized as follows: $\{x_1 \in \mathcal{X}_1\} \rightarrow \{s_1 \in S_1, c_1 \in C\} \rightarrow \{\hat{x}_1 \in \mathcal{X}_1\}$. Similarly, we perform mappings $x_2 \rightarrow \hat{x}_2$ using content-style representation disentanglement, i.e., $\{x_2 \in \mathcal{X}_2\} \rightarrow \{s_2 \in S_2, c_2 \in C\} \rightarrow \{\hat{x}_2 \in \mathcal{X}_2\}$. These various combinations of encoding-decoding operations facilitate multiple forms of image style conversion. Moreover, additional constraints can be easily incorporated to enhance the consistency of desired texture and semantic structure information, ensuring the improvement in quality of complex distortion-corrupted underwater images while preserving complete detail information.

3.2. Auto-encoder architecture

In our URD-UIE method, we propose a novel auto-encoder architecture specifically designed to generate distortion-free images. The auto-encoder illustrated in Fig. 1(a) consists of three key components: (1) a content encoder E^c and a style encoder E^s , responsible for extracting the content and style information from an input underwater image and mapping it to the domain-invariant latent space C and the domain-specific style space S , respectively; (2) a generator G , which utilizes the extracted features to generate an image in the distortion-free domain; and (3) a discriminator D , responsible for discerning between real and generated images within each domain. Next, we will describe the details for each component:

(1) **Content encoder.** The gray section depicted in Fig. 1(a) comprises three strided convolutional layers responsible for downsampling the distortion-corrupted image and four residual blocks (He et al., 2016) to facilitate subsequent processing and generation of content features. All the convolutional layers are activated by ReLU (Glorot et al., 2011), the padding type is reflect (Paszke et al., 2019), and followed by Instance Normalization (IN) (Ulyanov et al., 2017).

(2) **Style encoder.** The blue section of Fig. 1(a) consists of a strided convolutional layer and three residual blocks (He et al., 2016) for downsampling the distortion-free image. All convolutional layers utilize ReLU activation (Glorot et al., 2011) and employ reflect padding (Paszke et al., 2019). The style code is derived through a global average pooling (GAP) layer followed by a fully connected (FC) layer. It is important to note that the IN layer is excluded from the style encoder, as it eliminates the mean and variance of the original features, which contain essential style information (Huang and Belongie, 2017).

(3) **Generator.** As shown in the green part of in Fig. 1(a), the enhanced image is generated by processing the content feature through four residual blocks and two deconvolution layers, followed by one convolutional layer. The activation functions used in these layers are Tanh (Karlik and Olgac, 2011) for the last convolutional layer and ReLU (Glorot et al., 2011) for the rest, with all padding types set to zero (Paszke et al., 2019). To incorporate style codes into the image generation process, we integrate adaptive instance normalization

(AdaIN) (Huang and Belongie, 2017) layers into our image generation process by equipping our residual blocks with parameters that are dynamically generated by a multi-layer perceptron (MLP).

(4) Discriminator. As depicted in the orange segment of Fig. 1(a), we utilize a multi-scale discriminator (Wang et al., 2018) for real/fake classification of each domain, and guide the generators to produce outputs that exhibit both exceptionally realistic details and precise global structures.

3.3. The URD-UIE framework

Fig. 1(c) presents an overview of the URD-UIE network framework, which comprises two disentangled cross-domain translation (distortion-corrupted and distortion-free domain cycles), two self-reconstruction translations, and two adversarial constraints. To achieve the disentangled cross-domain translation, the framework employs two domain translation mappings, i.e. $x_1 \rightarrow x_{1 \rightarrow 2} \rightarrow \hat{x}_1$ and $x_2 \rightarrow x_{2 \rightarrow 1} \rightarrow \hat{x}_2$, for distortion-corrupted and distortion-free domain cycles, respectively. Next, two encoders are employed to disentangle the content and style features of an underwater image, and a generator is utilized to reconstruct the target-domain image by leveraging these extracted features, thus achieving the domain translation mapping. More specifically, the mapping $x_1 \rightarrow x_{1 \rightarrow 2} \rightarrow \hat{x}_1$ is realized by content encoder E_1^c , the style encoder E_1^s , and the generator G_1 , whereas the mapping $x_2 \rightarrow x_{2 \rightarrow 1} \rightarrow \hat{x}_2$ is realized by the content encoder E_2^c , the style encoder E_2^s and the generator G_2 . With any two unpaired images, i.e., the underwater distortion-free image x_1 and the distortion-corrupted image x_2 as inputs, the process of learning to reconstruct the images \hat{x}_1 and \hat{x}_2 by the encoders and the generators, respectively, is represented as follows:

$$\{c_1, s_1\} = \{E_1^c(x_1), E_1^s(x_1)\}, \{c_2, s_2\} = \{E_2^c(x_2), E_2^s(x_2)\}, \quad (1)$$

where c_1 and s_1 represent the disentangled content and style features of x_1 , c_2 and s_2 depict the disentangled content and style features of x_2 , respectively. Based on this, we utilize the generator G_1 to perform a cross-fusion of c_1 and s_2 in the latent space to $x_{1 \rightarrow 2}$ in the distortion-free image space, whereas the distortion-corrupted image $x_{2 \rightarrow 1}$ is generated by c_2 and s_1 :

$$x_{1 \rightarrow 2} = G_1(c_1, s_2), \quad x_{2 \rightarrow 1} = G_2(c_2, s_1). \quad (2)$$

The translated images $x_{1 \rightarrow 2}$ and $x_{2 \rightarrow 1}$ are constrained by x_2 and x_1 , respectively, through discriminators D_2 and D_1 . These discriminators aim to distinguish between real and generated images, and encourage the model to generate images that are as consistent as possible to the real ones. To effectively train the model using unpaired datasets, we apply the cross-domain translation mapping again to reconstruct the original input image pair. By doing this, we can utilize the cross-cycle consistency loss to enforce consistency between the input images x and their reconstructed counterparts \hat{x} . The procedure for extracting content and style features is as follows:

$$\{\hat{c}_1, \hat{s}_1\} = \{E_1^c(x_{1 \rightarrow 2}), E_2^s(x_{2 \rightarrow 1})\}, \quad \{\hat{c}_2, \hat{s}_2\} = \{E_2^c(x_{2 \rightarrow 1}), E_1^s(x_{1 \rightarrow 2})\}. \quad (3)$$

Similarly, the process of content and style feature cross-fusion remains consistent with Eq. (2), which is represented as follows:

$$\hat{x}_1 = G_1(\hat{c}_1, \hat{s}_2), \quad \hat{x}_2 = G_2(\hat{c}_2, \hat{s}_1), \quad (4)$$

where \hat{x}_1 and \hat{x}_2 represent the cross-reconstructed image of x_1 and x_2 , respectively. In addition, we impose constraints on the forward translation mappings through self-reconstruction translations, as illustrated in the right part of Fig. 1(c), to ensure the preservation of image quality when no alteration is expected. To be more specific, we conduct a self-reconstruction process using the disentangled content and style features extracted from x_1 and x_2 , denoted as follows:

$$\tilde{x}_1 = G_1(c_1, s_1), \quad \tilde{x}_2 = G_2(c_2, s_2), \quad (5)$$

where \tilde{x}_1 and \tilde{x}_2 are the self-reconstructed distortion-free and distortion-corrupted image, respectively.

3.4. Loss function

Taking into account the visual perception effect and the preservation of detailed features in the generated image, we have devised five loss functions to achieve the desired output. The relationship between each loss function is illustrated in Fig. 1(c).

3.4.1. Adversarial loss

To enhance the realism of the generated images, we integrate the adversarial loss function (Isola et al., 2017) into both domains to achieve superior outcomes. For the distortion-free domain, the adversarial loss $\mathcal{L}_{adv}^{x_1}$ was defined as:

$$\mathcal{L}_{adv}^{x_1} = \mathbb{E} [\log (1 - D_1(x_{2 \rightarrow 1}))] + \mathbb{E} [\log D_1(x_1)], \quad (6)$$

where D_1 represents a discriminator that aims to distinguish between the translated images $x_{2 \rightarrow 1}$ and the real images in \mathcal{X}_1 . Our objective is to minimize the $\mathcal{L}_{adv}^{x_1}$ to make the enhanced image as similar as possible to the real sample in the distortion-free domain. Similarly, for the distortion-corrupted image domain, the adversarial loss $\mathcal{L}_{adv}^{x_2}$ was defined as:

$$\mathcal{L}_{adv}^{x_2} = \mathbb{E} [\log (1 - D_2(x_{1 \rightarrow 2}))] + \mathbb{E} [\log D_2(x_2)], \quad (7)$$

where D_2 is a discriminator which attempts to distinguish between the translated images $x_{1 \rightarrow 2}$ and real images in \mathcal{X}_2 . Our goal is to minimize the $\mathcal{L}_{adv}^{x_2}$ to make the degraded image as similar as possible to the real sample in distortion-corrupted domain.

3.4.2. Cycle consistency loss

Due to the inability of the unpaired dataset to provide supervised signals, the generated enhanced images may not adequately retain the structural information of the original ones. To effectively tackle this issue, we introduce a cycle consistency loss (Zhu et al., 2017) to enforce a high degree of similarity between the structures of the images before and after enhancement, which is represented as follows:

$$\mathcal{L}_{cyc} = \|x_1 - \hat{x}_1\|_1 + \|x_2 - \hat{x}_2\|_1. \quad (8)$$

3.4.3. Self-reconstruction loss

To ensure the preservation of information and prevent the introduction of new distortions during the encoding and decoding process, it is essential for the UIE model to possess self-reconstruction capability. This capability allows the model to reconstruct the original image using the content and style information extracted by the encoder. To achieve this, we introduce the utilization of self-reconstruction loss \mathcal{L}_{rec} , which is denoted as follows:

$$\mathcal{L}_{rec} = \|x_1 - \hat{x}_1\|_1 + \|x_2 - \hat{x}_2\|_1. \quad (9)$$

3.4.4. Style consistency loss

To ensure that the cross-reconstruction image aligns with the stylistic characteristics of the original image, we introduce a style consistency loss denoted as \mathcal{L}_{sty} , which is represented as follows:

$$\mathcal{L}_{sty} = \|s_1 - \hat{s}_1\|_1 + \|s_2 - \hat{s}_2\|_1. \quad (10)$$

3.4.5. Content consistency loss

To guarantee that the cross-reconstruction image preserves the content information of the original image, we incorporate a content consistency loss denoted as \mathcal{L}_{cont} , which is expressed as follows:

$$\mathcal{L}_{cont} = \|c_1 - \hat{c}_1\|_1 + \|c_2 - \hat{c}_2\|_1. \quad (11)$$

3.4.6. Total loss

All modules are optimized with total loss function \mathcal{L} :

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda_{cyc} \mathcal{L}_{cyc} + \lambda_{rec} \mathcal{L}_{rec} + \lambda_{sty} \mathcal{L}_{sty} + \lambda_{cont} \mathcal{L}_{cont}, \quad (12)$$

where $\mathcal{L}_{adv} = \mathcal{L}_{adv}^{x_1} + \mathcal{L}_{adv}^{x_2}$, λ_{cyc} , λ_{rec} , λ_{sty} and λ_{cont} are the loss weights used for controlling the contributions of the terms in Eq. (12).

Table 3
The parameters of experimental configuration.

Parameter	Value
λ_{cyc}	10.0
λ_{rec}	10.0
λ_{sty}	1.0
λ_{cont}	1.0
Batch size	64
Cropped image size	256 × 256
CPU	Intel Xeon Silver 4210
GPU	NVIDIA RTX A6000 (48 GB RAM)
Framework	PyTorch 1.12.0
System	Ubuntu 20.04
CUDA & cuDNN	CUDA 11.6 and cuDNN 8.3.3

4. Experimental results and analysis

To validate the effectiveness of the URD-UIE method, we conducted experimental comparisons with traditional and learning-based methods, and analyzed the results from both subjective and objective perspectives.

4.1. Dataset and training details

In this paper, the unpaired dataset from EUVP (Islam et al., 2020) is chosen as the training set, containing a total of 3195 low quality images and 3140 high quality images, which are separated by volunteers based on attributes such as color, contrast and sharpness, and this unpaired dataset supports the modeling of human preferences for underwater image quality perception. To evaluate the proposed method, a testing set of 500 severely distortion-corrupted underwater images was obtained from online sources. The majority of these images exhibit characteristics of overblue or overgreen characteristics, providing a challenging evaluation scenario.

During the training process, URD-UIE was implemented using PyTorch (Paszke et al., 2019). Both training and evaluation were conducted on a machine equipped with a CPU (Intel Xeon Silver 4210) and three GPUs (NVIDIA RTX A6000 48 GB). The Adam (Kingma and Ba, 2014) optimizer was employed with a learning rate of 1×10^{-4} . In the process of training, we used $\lambda_{cyc} = 10$, $\lambda_{rec} = 10$, $\lambda_{sty} = 1$ and $\lambda_{cont} = 1$ for the corresponding losses, E , G and D are optimized in an alternating manner. The specific training parameters of the URD-UIE method are shown in Table 3.

The image dimensions are cropped to a uniform size of 256 × 256, and the maximum number of training iterations is 100 000. In addition, to guarantee the fairness of experimental comparisons, all unsupervised methods used the same unpaired dataset and configurations for training and testing, while the optimal model provided by the source code was used for testing of the supervised methods.

4.2. Qualitative analysis

To qualitatively analyze the performance of URD-UIE method, we selected images with different severities of chromatic aberration distortion from the testing set and then compared them with IBLA (Peng and Cosman, 2017), U LAP (Song et al., 2018), RGHS (Huang et al., 2018b), SMBLO (Song et al., 2020), UGAN (Fabbri et al., 2018), CycleGAN (Zhu et al., 2017), UWGAN (Li et al., 2018) and FUNIE-GAN (Islam et al., 2020), Ucolor (Li et al., 2021), TACL (Liu et al., 2022), where the literatures (Peng and Cosman, 2017; Song et al., 2020; Huang et al., 2018b; Song et al., 2018) are the traditional method and the literatures (Islam et al., 2020; Zhu et al., 2017; Li et al., 2021; Fabbri et al., 2018; Li et al., 2018; Liu et al., 2022) are the learning-based methods.

The qualitative comparison of enhanced results for authentic underwater images using different methods are presented in Fig. 2. The analysis leads to several key observations. Firstly, IBLA lacks a color

correction function and fails to correct blue-green distortion, rendering it ineffective in addressing the issue of color bias in underwater images. SMBLO exhibits significant enhancement effects, primarily attributed to its complex restoration mechanism. However, it tends to excessively increase contrast and brightness, resulting in overly bright visuals or even overexposure, as seen in images 4 and 6. U LAP, while providing enhancement effects that fall between the aforementioned methods, struggles to effectively recover the original colors in severely chromatic distortion-corrupted images. Among the traditional comparison methods, RGHS demonstrates the most favorable enhancement effect. Nonetheless, it proves suboptimal for recovering images with partial distortion, as evident in images 1 and 7. In contrast, learning-based methods yield superior visual enhancements. Both UGAN and FUNIE-GAN exhibit color overcompensation issues in local regions, as observed in image 3, and are not adept at handling severe chromatic distortion, as seen in images 1 and 5. Additionally, UGAN tends to oversaturate bright objects in the scene, as depicted in image 2, while FUNIE-GAN employs a lightweight model that significantly improves processing speed but struggles with correcting color styles in unpaired training models. CycleGAN, due to its reliance on a relatively single loss function constraint, falls short in recovering chromatic aberrations effectively. Regarding background enhancement in underwater images, UWGAN outperforms the aforementioned methods by successfully restoring the background features associated with dewatering characteristics. However, some local objects still exhibit color distortions, such as the human statue in image 3 and the turtle's back in image 5, which display varying degrees of greenish coloration. Benefiting from the advantages of self-supervised learning, Ucolor and TACL showcase the most impressive enhancement effects among all the comparison methods. However, some of the enhanced images display unnatural colors, as seen in image 2 of Ucolor, and slight oversaturation, as evident in images 3 and 5 of TACL. In comparison to all the traditional and learning-based methods mentioned above, our method surpasses the physical-based methods in visual enhancement without using scene depth or a priori water body information, outperforms existing unsupervised models and rivals supervised methods in the case of training without paired datasets.

To demonstrate the generalizability and robustness of the URD-UIE method, we conducted experiments on synthetic images. From Fig. 3, all input images are artificially generated with significant color distortion, which challenges the enhancement performance of the model. In this case, the traditional methods i.e. IBLA, SMBLO, U LAP and RGHS almost struggled to effectively restore the color aberration. In contrast, partial learning-based methods such as UGAN, FUNIE-GAN, Ucolor, and TACL showcased their superiority in terms of distortion recovery. Among all the compared methods, Ucolor and TACL emerged as the top performers. However, it should be noted that some of the enhanced images failed to fully recover the original colors, and the enhancement effect of FUNIE-GAN exhibited over-compensation in color, resulting in less realistic outputs (e.g., image 1, 2, and 3 displaying varying degrees of reddishness). Similarly, UGAN showed insensitivity toward synthetic underwater images, with the generated images exhibiting a more uniform style. On the other hand, CycleGAN produced results that were largely consistent with authentic underwater images, but the recovery of chromatic aberration was not prominent. After careful comparison, our method demonstrated a more realistic enhancement effect for synthetic images, with richer details and color information, and excellent performance in terms of visual sensory effects.

4.3. Quantitative analysis

To quantitatively evaluate and analyze the performance of URD-UIE, we employed generalized no-referenced underwater image quality evaluation metrics: Underwater Image Quality Measure (UIQM) (Yang and Sowmya, 2015), Underwater Color Image Quality Evaluation

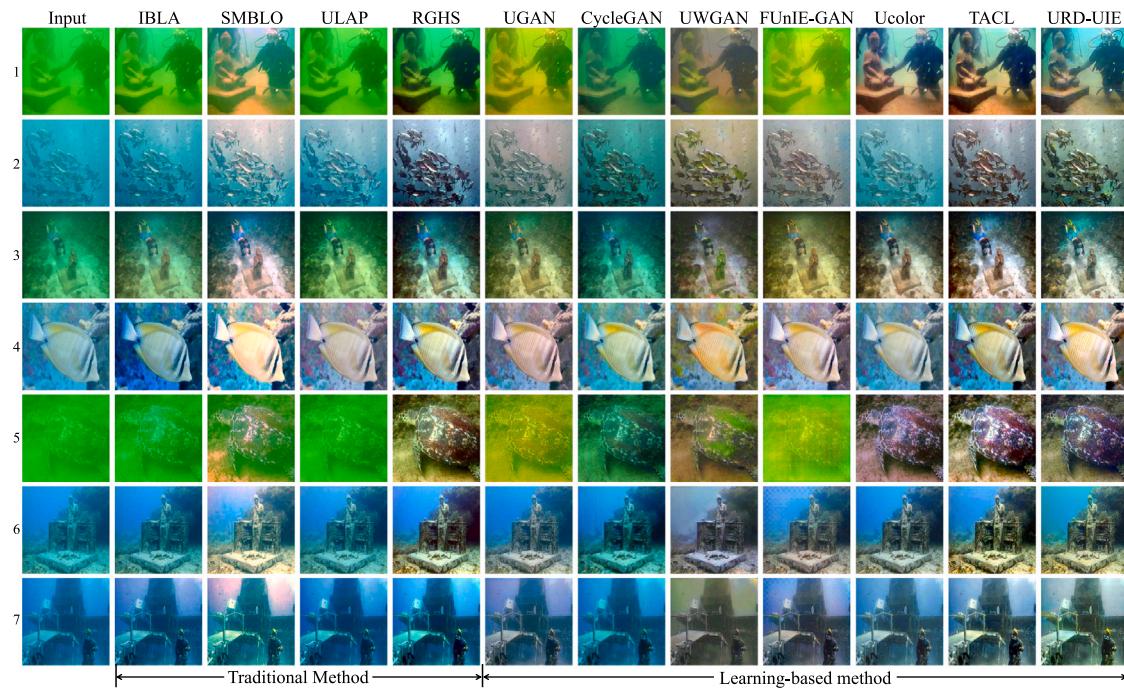


Fig. 2. Qualitative comparison of UIE for authentic underwater images using different methods.

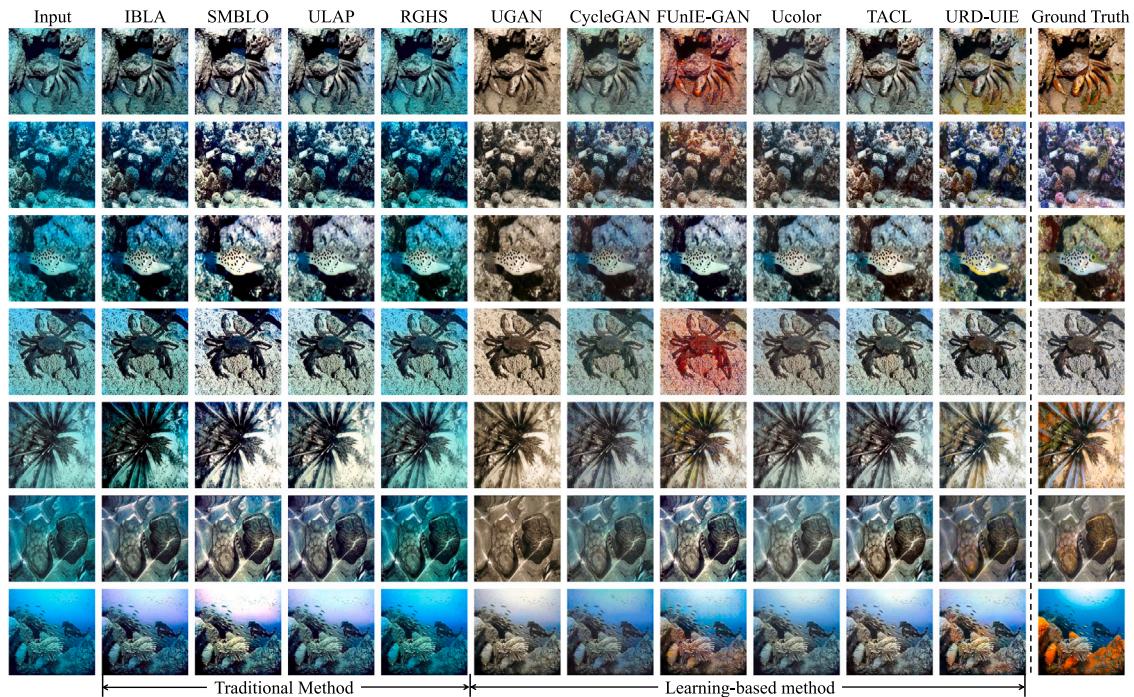


Fig. 3. Qualitative comparison of UIE for synthetic underwater images using different methods.

(UCIQE) (Panetta et al., 2015), Underwater Image Colorfulness Measure (UICM), Underwater Image Sharpness Measure (UISM), and Underwater Image Contrast Measure (UIConM). UIQM, which is based on the HSV model, takes into account colorfulness, sharpness, and contrast. UCIQE utilizes the contrast, chroma, and saturation of the CIELab color space, as human perception is strongly influenced by chroma variance in degraded underwater images. UICM quantifies the colorfulness attribute, while UISM measures sharpness by examining grayscale edges. UISM stands for the sharpness measurement via the gray-scale edges, and the contrast is measured via UIConM by applying

Logarithmic AME by Entropy (logAMEE) (Panetta et al., 2010). Higher values for each evaluation metric indicate superior visual quality of the image. Quantitative comparisons on UIE of authentic and synthetic underwater images are presented in Tables 4 and 5, respectively.

From Table 4, it is evident that URD-UIE outperforms all other methods in most metrics, except for SMBLO in UICM. This superiority can be attributed to the additional enhancement provided by SMBLO, which optimizes the gain factor for white balance color correction, thereby further improving the color properties of UICM. The comparative results of the quantitative analysis for synthetic underwater images,

Table 4
Quantitative comparison of UIE on authentic underwater images.

Method	UICM↑	UISM↑	UIConM↑	UIQM↑	UCIQE↑
Original image	3.1494	5.0460	0.2046	2.3105	4.0049
IBLA (Peng and Cosman, 2017)	5.1639	4.9656	0.1268	2.0654	5.0761
SMBLO (Song et al., 2020)	9.0997	5.4342	0.1696	2.4677	5.7153
ULAP (Song et al., 2018)	5.2072	5.3459	0.1959	2.4259	5.1606
RGHS (Huang et al., 2018b)	6.8919	5.9577	0.2386	2.8066	5.8338
UGAN (Fabri et al., 2018)	4.5587	6.8883	0.2282	2.9785	4.9390
CycleGAN (Zhu et al., 2017)	4.4869	5.5061	0.2750	2.7357	4.8816
UWGAN (Li et al., 2018)	4.4339	6.5132	0.2921	3.0926	5.2140
FUNIE-GAN (Islam et al., 2020)	4.8336	6.7341	0.2425	2.9919	5.5505
Ucolor (Li et al., 2021)	5.5892	7.0390	0.2942	3.2595	4.3341
TACL (Liu et al., 2022)	7.4388	7.2061	0.2894	3.3705	5.4433
URD-UIE	7.4684	7.4451	0.2963	3.4470	6.5943

Table 5
Quantitative comparison of UIE on synthetic underwater images.

Method	UICM↑	UISM↑	UIConM↑	UIQM↑	UCIQE↑
Original image	3.5520	6.7553	0.2178	2.8736	3.7441
IBLA (Peng and Cosman, 2017)	4.4442	5.1289	0.1686	2.2426	4.5644
SMBLO (Song et al., 2020)	5.6213	3.5913	0.1473	1.7458	4.9824
ULAP (Song et al., 2018)	4.1366	4.5891	0.1679	2.0723	4.3071
RGHS (Huang et al., 2018b)	4.3669	6.2063	0.1833	2.6113	4.5288
UGAN (Fabri et al., 2018)	2.0707	6.9960	0.2856	3.4106	2.4744
CycleGAN (Zhu et al., 2017)	3.3540	6.4171	0.2653	2.9383	4.1682
FUNIE-GAN (Islam et al., 2020)	4.5406	7.4654	0.2694	3.3553	4.9329
Ucolor (Li et al., 2021)	3.4604	7.5886	0.2540	3.4114	3.5341
TACL (Liu et al., 2022)	3.8589	7.5844	0.2617	3.4179	3.3670
URD-UIE	4.7153	7.6115	0.2697	3.4376	5.0752

as presented in [Table 5](#), consistently support the findings observed in authentic underwater images. Based on the comprehensive quantitative analysis above, it is proved that URD-UIE effectively restores the actual colors, enhances contrast and sharpness, and preserves the inherent texture structure features.

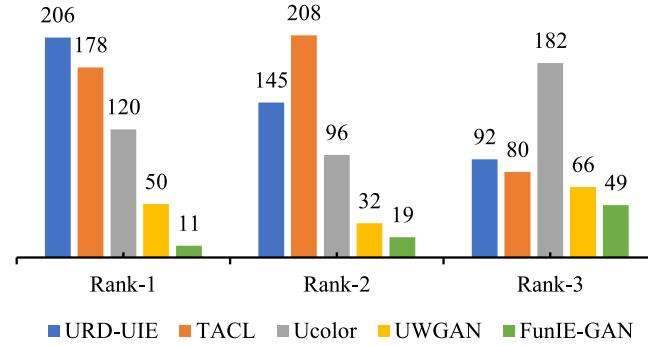
Besides the qualitative analysis of authentic and synthetic images, we also measured the computational efficiency of URD-UIE by calculating the Frames Per Second (FPS), which calculates the number of enhanced images per second. For a batch images with size 256×256 , the FPS value of URD-UIE was approximately 13, which satisfies the real-time requirements for underwater video enhancement (UVE). From experiments conducted on UVE, URD-UIE achieved significant results in chromatic aberration correction, deblurring and contrast improvement. Our real-time underwater enhancement video is available at <https://youtu.be/VktJ8PQA188>.

4.4. User study

We conducted a user study with 20 volunteers who are familiar with underwater images. The volunteers were presented with different sets of images (one for each enhancement model) and asked to rank their top three selections. We received a total of 600 responses. [Fig. 4](#) summarizes the rank-1, rank-2, and rank-3 times of the top 5 models, which indicates that the users preferred the images enhanced by URD-UIE, TACL, Ucolor, UWGAN, and FUNIE-GAN compared to the other models. Our user study results demonstrate that our method is superior in terms of subjective visual enhancement effects, which is further corroborated by our qualitative and quantitative analyses.

4.5. Improved visual perception

As depicted in [Fig. 5](#), we conducted additional experiments to provide a quantitative explanation for the effectiveness of URD-UIE enhanced images in underwater visual perception. To increase the difficulty of the experiment, we utilized object detection and semantic segmentation data of underwater farmed products (i.e., holothurian, echinus, starfish and scallops) with severe distortion for validation.



[Fig. 4](#). Rank-n times plot for the top five models.

[Fig. 6](#) presents a comparative analysis of underwater object detection, from which we can see that the enhanced image exhibits a greater number of selected bounding boxes and achieves higher detection accuracy. Furthermore, the comparison of underwater semantic segmentation, illustrated in [Fig. 7](#), clearly demonstrates the superior segmentation performance of the enhanced image.

As shown in [Tables 6](#) and [7](#), we have summarized the quantitative results of underwater object detection using different methods, i.e., YOLOX (Ge et al., 2021), Centernet (Zhou et al., 2019), and Faster R-CNN (Girshick, 2015). Additionally, we have presented the results of underwater semantic segmentation using DeepLabV3 (Chen et al., 2017), PSPNet (Zhao et al., 2017), and U-net (Ronneberger et al., 2015). From [Tables 6](#) and [7](#), we can observe significant improvements in the average precision (AP) for object detection and the intersection over union (IoU) for semantic segmentation across various underwater object categories, i.e., holothurian, echinus, starfish, and scallop. Specifically, the mean average precision (mAP) for object detection in enhanced images has increased by 55.1%, 125.0% and 139.1% using YOLOX (Ge et al., 2021), Centernet (Zhou et al., 2019) and Faster R-CNN (Ren et al., 2015) respectively. Furthermore, the mean intersection over union (mIOU) for semantic segmentation in enhanced images

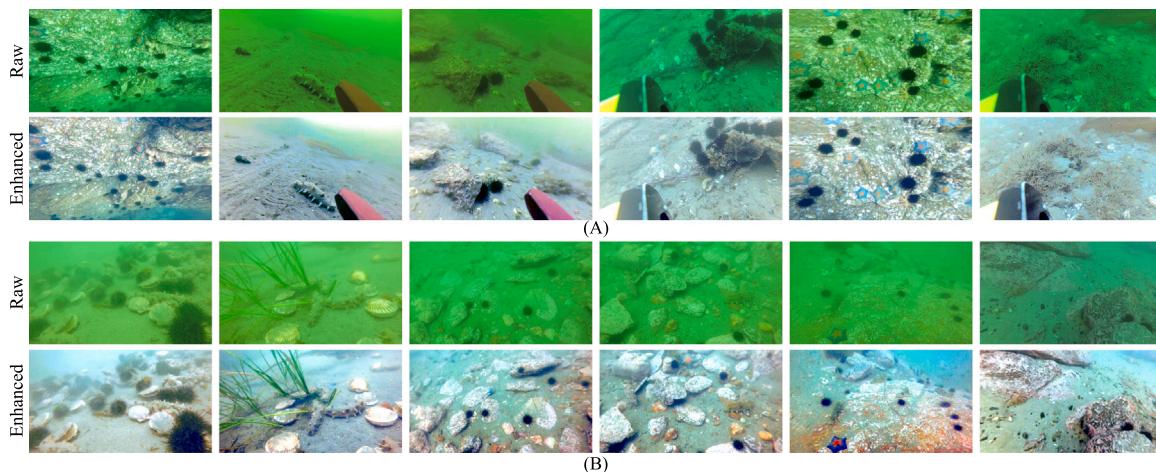


Fig. 5. Underwater image enhancement effect for object detection and semantic segmentation.

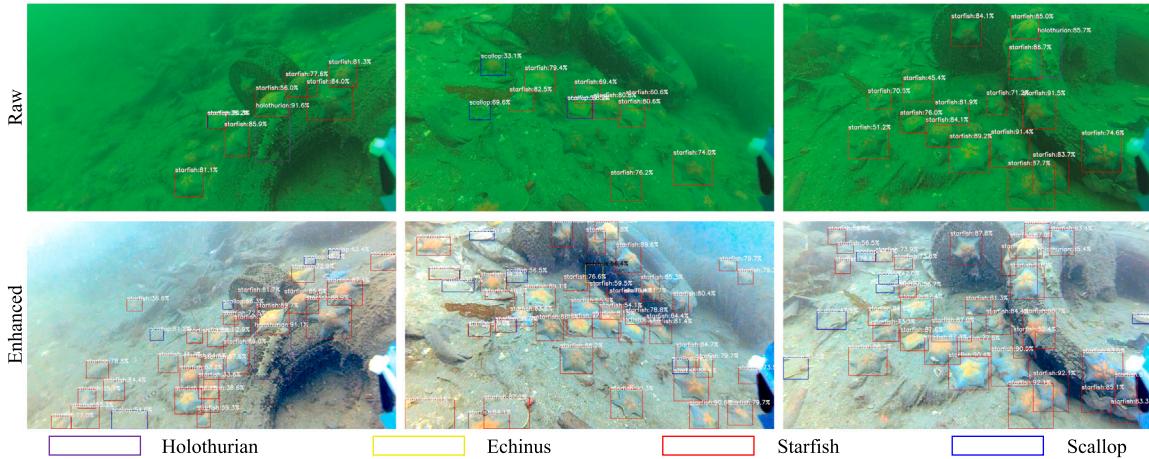


Fig. 6. Comparison of underwater object detection for underwater images without and with enhancement.

Table 6
Quantitative comparison of URD-UIE effect for object detection.

Category	YOLOX (Ge et al., 2021)			Centernet (Zhou et al., 2019)			Faster R-CNN (Ren et al., 2015)		
	Original	Enhanced	Improvement	Original	Enhanced	Improvement	Original	Enhanced	Improvement
Holothurian	0.3377	0.7177	112.6%	0.1627	0.7067	334.4%	0.1456	0.6759	364.2%
Echinus	0.7359	0.8588	16.7%	0.2381	0.8928	274.9%	0.2289	0.8653	278.0%
Starfish	0.6423	0.8369	30.3%	0.3861	0.8905	130.6%	0.3504	0.8421	140.3%
Scallop	0.3126	0.7331	134.5%	0.6622	0.7708	16.4%	0.5575	0.6824	22.4%
mAP	0.5071	0.7866	55.1%	0.3623	0.8152	125.0%	0.3206	0.7664	139.1%

has improved by 10.8%, 10.8% and 5.8% using DeepLabV3 (Chen et al., 2017), PSPNet (Zhao et al., 2017) and U-net (Ronneberger et al., 2015) respectively. These results clearly indicate that URD-UIE plays a crucial role in enhancing underwater visual perception performance.

To further quantitatively evaluate and analyze the enhancement performance of URD-UIE for subsequent underwater image processing tasks, we conducted a comparative analysis between our proposed method URD-UIE and the best traditional and learning-based methods (i.e., RGHS and TACL) used in previous qualitative and quantitative evaluations. The quantitative comparison of various object detection and semantic segmentation algorithms using enhanced images obtained from different UIE methods is presented in Tables 8 and 9, respectively. These tables clearly illustrate that URD-UIE outperforms other

methods in terms of mAP for object detection and mIoU for semantic segmentation, indicating its superiority for improved underwater visual perception, thus highlighting its potential for improving underwater image processing tasks.

4.6. Ablation experiments

To further investigate the impact of different loss functions on the performance of URD-UIE, we conducted comprehensive ablation experiments. In order to ensure a fair analysis of the effect of each component, consistent data and training iterations were maintained across all experiments. Fig. 8 illustrates the qualitative analysis of the ablation experiments. Specifically, M1 represents the model trained

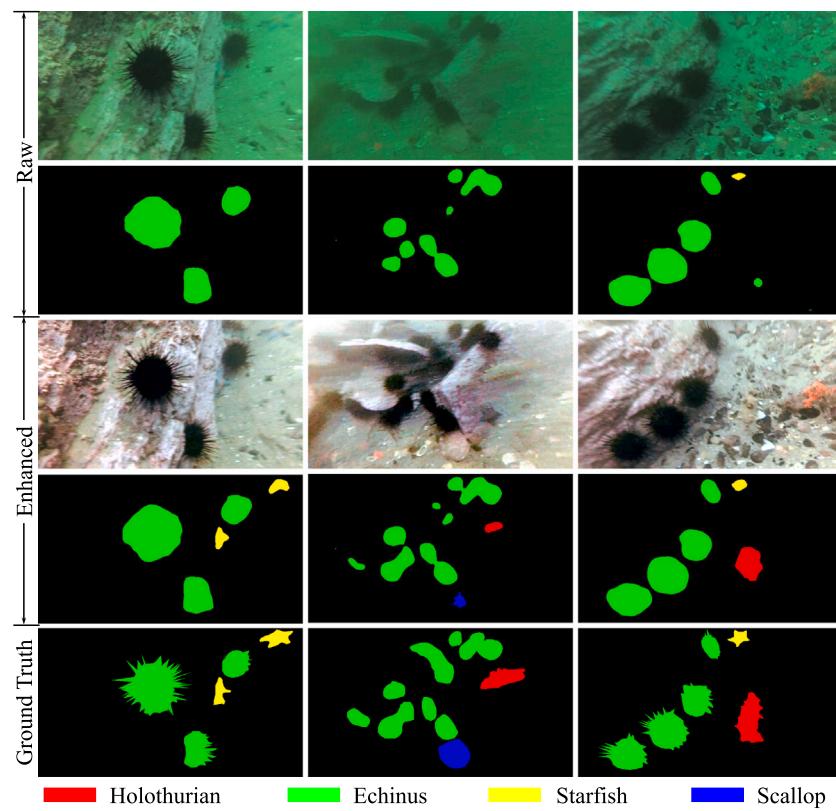


Fig. 7. Comparison of underwater semantic segmentation of underwater images without and with enhancement.

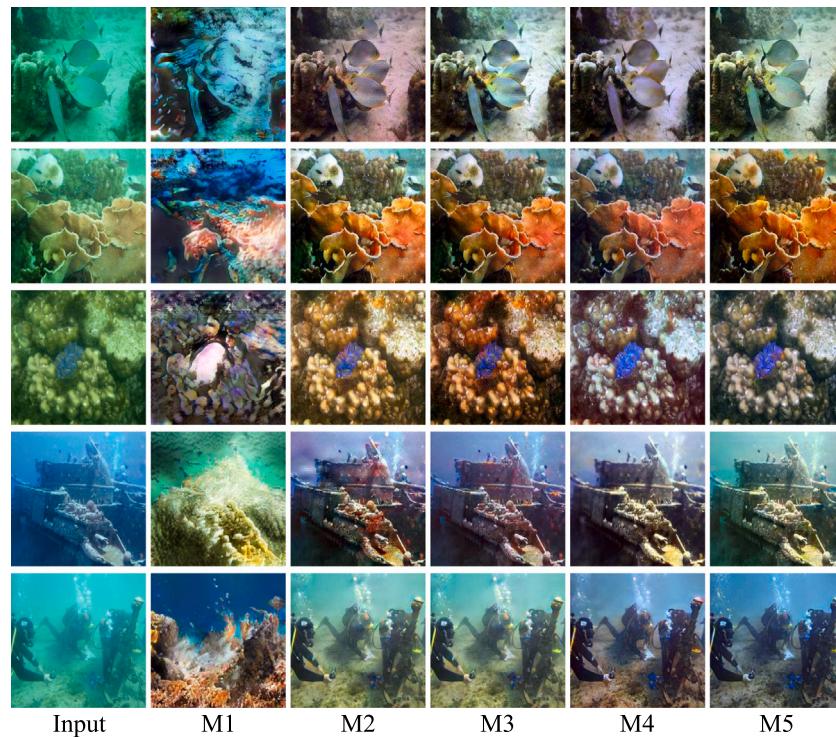


Fig. 8. Qualitative comparison of UIE using different losses.

Table 7

Quantitative comparison of URD-UIE effect for semantic segmentation.

Category	DeepLabV3 (Chen et al., 2017)			PSPNet (Zhao et al., 2017)			U-net (Ronneberger et al., 2015)		
	Original	Enhanced	Improvement	Original	Enhanced	Improvement	Original	Enhanced	Improvement
Background	0.9742	0.9797	0.6%	0.9856	0.9867	0.1%	0.9886	0.9893	0.1%
Holothurian	0.4783	0.5039	5.4%	0.2587	0.3407	31.7%	0.3759	0.4114	9.4%
Echinus	0.6709	0.6821	1.7%	0.5119	0.5992	17.1%	0.6436	0.6677	3.7%
Starfish	0.5137	0.6192	20.5%	0.4184	0.4651	11.2%	0.5976	0.6369	6.6%
Scallop	0.4649	0.6732	44.8%	0.3011	0.3508	16.5%	0.3971	0.4725	19.0%
mIOU	0.6204	0.6876	10.8%	0.4952	0.5485	10.8%	0.6005	0.6356	5.8%

Table 8

Quantitative comparison of object detection algorithms using enhanced images for different UIE methods.

Category	YOLOX (Ge et al., 2021)			Centernet (Zhou et al., 2019)			Faster R-CNN (Ren et al., 2015)		
	RGHS	TACL	URD-UIE	RGHS	TACL	URD-UIE	RGHS	TACL	URD-UIE
Holothurian	0.6846	0.6731	0.7177	0.6496	0.6294	0.7067	0.6254	0.6102	0.6759
Echinus	0.8205	0.8385	0.8588	0.8739	0.8583	0.8928	0.8079	0.8289	0.8653
Starfish	0.8257	0.8176	0.8369	0.8288	0.8186	0.8905	0.8128	0.7895	0.8421
Scallop	0.7158	0.7093	0.7331	0.7053	0.6979	0.7708	0.6039	0.6241	0.6824
mAP	0.7617	0.7596	0.7866	0.7644	0.7511	0.8152	0.7125	0.7132	0.7664

Table 9

Quantitative comparison of semantic segmentation algorithms using enhanced images for different UIE methods.

Category	DeepLabV3 (Chen et al., 2017)			PSPNet (Zhao et al., 2017)			U-net (Ronneberger et al., 2015)		
	RGHS	TACL	URD-UIE	RGHS	TACL	URD-UIE	RGHS	TACL	URD-UIE
Background	0.9768	0.9756	0.9797	0.9843	0.9831	0.9867	0.9859	0.9878	0.9893
Holothurian	0.5446	0.3938	0.5039	0.2568	0.2759	0.3407	0.4203	0.4324	0.4114
Echinus	0.6578	0.6376	0.6821	0.4719	0.4622	0.5992	0.6310	0.6328	0.6677
Starfish	0.4437	0.5101	0.6192	0.4381	0.4090	0.4651	0.5896	0.5841	0.6369
Scallop	0.5532	0.5075	0.6732	0.2897	0.2735	0.3508	0.4247	0.4203	0.4725
mIOU	0.6352	0.6049	0.6876	0.4882	0.4807	0.5485	0.6103	0.6115	0.6356

Table 10

Quantitative comparison of UIE using different losses.

Method	UICM↑	UISM↑	UIConM↑	UIQM↑	UCIQE↑
M1 (\mathcal{L}_{adv} only)	7.1518	6.7757	0.2492	3.0990	5.1204
M2 (M1 with \mathcal{L}_{cyc})	6.2135	6.2551	0.2691	3.3277	5.5503
M3 (M2 with \mathcal{L}_{con})	6.6370	6.8340	0.2732	3.1522	5.9856
M4 (M3 with \mathcal{L}_{sty})	6.8642	7.3723	0.2928	3.4530	6.2030
M5 (M4 with \mathcal{L}_{rec})	7.4684	7.4451	0.2963	3.4470	6.5943

solely with \mathcal{L}_{adv} , while M2 corresponds to the model trained with both \mathcal{L}_{adv} and \mathcal{L}_{cyc} . Furthermore, M3 and M4 denote the models trained by additionally incorporating \mathcal{L}_{con} and \mathcal{L}_{sty} into M2, respectively. Building upon these variants, M5 is the model trained with all loss functions.

As depicted in Fig. 8, it is evident that the generated image does not perfectly align with the structure of original image, leading to a significant loss of detailed information when employing M1 with only \mathcal{L}_{adv} . In contrast, the results generated by M2 exhibit more comprehensive preservation of structural information, primarily attributed to the contribution of \mathcal{L}_{cyc} in retaining fine structural details in the images. While M3 accurately retains the vast majority of the structural information of original image due to the content consistency constraints, there are certain areas where the colors deviate, resulting in a mismatch with reality. By incorporating the style consistency loss, the images generated by M4 demonstrate improved color balance. Finally, the M5 model, incorporating the self-reconstruction mechanism, achieves optimal results in terms of texture details and subjective human visual perception. This model effectively enhances image sharpness and contrast, while also providing more natural and realistic color reproduction. To validate our qualitative observations in Fig. 8, we present the quantitative results of the ablation experiment in Table 10, which are consistent with our findings.

5. Conclusion and discussion

In this paper, we propose an unsupervised underwater image enhancement method based on representation disentanglement (URD-UIE). Our study makes the following contributions: Firstly, we analyze the content and style information of different distortion-corrupted underwater images and define a variety of combined content and style information constraints applicable to multiple types of images, and obtain an unsupervised content-style representation disentanglement learning mechanism by combining image cyclic consistency transformation, so as to achieve the complete separation of image content and style information. Then, we design a network structure with content-style dual encoding and multi-scale fusion decoding, and incorporate multiple specific loss functions to establish a general quality enhancement model for underwater images. Qualitative and quantitative experiments demonstrate that the URD-UIE method can effectively correct the image distortion and improve the image quality for different distortion-corrupted underwater images, especially for those with the distortion of severe chromatic aberrations. The performance improvements of object detection and semantic segmentation using enhanced images demonstrate that URD-UIE can contribute to subsequent underwater image processings. However, it is important to acknowledge certain limitations in our work. Specifically, our method generalizes mixed distortion types of underwater images to a single style code, which restricts its ability to enhance images with different distortion types. In future research, we will focus on addressing this limitation and exploring methods to improve the quality of underwater images with diverse distortion types.

CRediT authorship contribution statement

Pengli Zhu: Writing – original draft, Conceptualization, Methodology, Experiment. Yancheng Liu: Conceptualization, Methodology, Project administration. Yuanquan Wen: Validation, Writing – review. Minyi Xu: Writing – review, Validation, Data curation. Xianping Fu:

Writing – review, Validation, Data curation. **Siyuan Liu:** Conceptualization, Methodology, Writing – review, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article

Acknowledgments

We thank the editor, the associate editor, and the anonymous reviewers, who provided critical comments and helpful suggestions for this paper.

References

- Anwar, Saeed, Li, Chongyi, 2020. Diving deeper into underwater image enhancement: A survey. *Signal Process., Image Commun.* 89, 115978.
- Anwar, Saeed, Li, Chongyi, Porikli, Fatih, 2018. Deep underwater image enhancement. arXiv preprint [arXiv:1807.03528](https://arxiv.org/abs/1807.03528).
- Chen, Xi, Duan, Yan, Houthooft, Rein, Schulman, John, Sutskever, Ilya, Abbeel, Pieter, 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In: *Advances in Neural Information Processing Systems*. Vol. 29.
- Chen, Liang-Chieh, Papandreou, George, Schroff, Florian, Adam, Hartwig, 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint [arXiv:1706.05587](https://arxiv.org/abs/1706.05587).
- Chen, Yu-Wei, Pei, Soo-Chang, 2022. Domain adaptation for underwater image enhancement via content and style separation. *IEEE Access* 10, 90523–90534.
- Cheung, Brian, Livezey, Jesse A, Bansal, Arjun K, Olshausen, Bruno A, 2014. Discovering hidden factors of variation in deep networks. arXiv preprint [arXiv:1412.6583](https://arxiv.org/abs/1412.6583).
- Ding, Zheng, Xu, Yifan, Xu, Weijian, Parmar, Gaurav, Yang, Yang, Welling, Max, Tu, Zhuowen, 2020. Guided variational autoencoder for disentanglement learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7920–7929.
- Fabri, Cameron, Islam, Md Jahidul, Sattar, Junaed, 2018. Enhancing underwater imagery using generative adversarial networks. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 7159–7165.
- Ge, Zheng, Liu, Songtao, Wang, Feng, Li, Zeming, Sun, Jian, 2021. Yolox: Exceeding yolo series in 2021. arXiv preprint [arXiv:2107.08430](https://arxiv.org/abs/2107.08430).
- Girshick, Ross, 2015. Fast r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 1440–1448.
- Glorot, Xavier, Bordes, Antoine, Bengio, Yoshua, 2011. Deep sparse rectifier neural networks. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings*, pp. 315–323.
- Goodfellow, Ian, Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron, Bengio, Yoshua, 2014. Generative adversarial nets. *Advances in neural information processing systems* 27.
- Gulrajani, Ishaan, Ahmed, Faruk, Arjovsky, Martin, Dumoulin, Vincent, Courville, Aaron, 2017. Improved training of wasserstein gans. arXiv preprint [arXiv:1704.00028](https://arxiv.org/abs/1704.00028).
- Guo, Yecal, Li, Hanyu, Zhuang, Peixian, 2019. Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE J. Ocean. Eng.* 45 (3), 862–870.
- He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, Sun, Jian, 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770–778.
- Higgins, Irina, Matthey, Loic, Pal, Arka, Burgess, Christopher, Glorot, Xavier, Botvinick, Matthew, Mohamed, Shakir, Lerchner, Alexander, 2016. Beta-vae: Learning basic visual concepts with a constrained variational framework.
- Hu, Junkang, Jiang, Qiuping, Cong, Runmin, Gao, Wei, Shao, Feng, 2021. Two-branch deep neural network for underwater image enhancement in hsv color space. *IEEE Signal Process. Lett.* 28, 2152–2156.
- Huang, Xun, Belongie, Serge, 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 1501–1510.
- Huang, Xun, Liu, Ming-Yu, Belongie, Serge, Kautz, Jan, 2018a. Multimodal unsupervised image-to-image translation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 172–189.
- Huang, Dongmei, Wang, Yan, Song, Wei, Sequeira, Jean, Mavromatis, Sébastien, 2018b. Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition. In: *International Conference on Multimedia Modeling*. Springer, pp. 453–465.
- Islam, Md Jahidul, Xia, Youya, Sattar, Junaed, 2020. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* 5 (2), 3227–3234.
- Isola, Phillip, Zhu, Jun-Yan, Zhou, Tinghui, Efros, Alexei A, 2017. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1125–1134.
- Jaffe, Jules S., 1990. Computer modeling and the design of optimal underwater imaging systems. *IEEE J. Ocean. Eng.* 15 (2), 101–111.
- Karlik, Bekir, Olgac, A. Vehbi, 2011. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Int. J. Artif. Intell. Exp. Syst.* 1 (4), 111–122.
- Kingma, Diederik P., Ba, Jimmy, 2014. Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- Lee, Hsin-Ying, Tseng, Hung-Yu, Huang, Jia-Bin, Singh, Maneesh, Yang, Ming-Hsuan, 2018. Diverse image-to-image translation via disentangled representations. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 35–51.
- Lee, Hsin-Ying, Tseng, Hung-Yu, Mao, Qi, Huang, Jia-Bin, Lu, Yu-Ding, Singh, Maneesh, Yang, Ming-Hsuan, 2020. Drift++: Diverse image-to-image translation via disentangled representations. *Int. J. Comput. Vis.* 128 (10), 2402–2417.
- Li, Chongyi, Anwar, Saeed, Hou, Junhui, Cong, Runmin, Guo, Chunle, Ren, Wenqi, 2021. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.* 30, 4985–5000.
- Li, Chongyi, Guo, Jichang, Guo, Chunle, 2018. Emerging from water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Process. Lett.* 25 (3), 323–327.
- Li, Chongyi, Guo, Chunle, Ren, Wenqi, Cong, Runmin, Hou, Junhui, Kwong, Sam, Tao, Dacheng, 2019. An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389.
- Li, Xinjie, Hou, Guojia, Li, Kunqian, Pan, Zhenkuan, 2022. Enhancing underwater image via adaptive color and contrast enhancement, and denoising. *Eng. Appl. Artif. Intell.* 111, 104759.
- Li, Jie, Skinner, Katherine A, Eustice, Ryan M, Johnson-Roberson, Matthew, 2017. Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot. Autom. Lett.* 3 (1), 387–394.
- LIU, Yancheng, DONG, Zhangwei, ZHU, Pengli, LIU, Siyuan, 2022. Unsupervised underwater image enhancement based on feature disentanglement. *J. Electron. Inf. Technol.* 44, 1–10.
- Liu, Risheng, Jiang, Zhiying, Yang, Shuzhou, Fan, Xin, 2022. Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Trans. Image Process.* 31, 4922–4936.
- Liu, Siyuan, Thung, Kim-Han, Qu, Liangqiong, Lin, Weili, Shen, Dinggang, Yap, Pew-Tian, 2021. Learning MRI artefact removal with unpaired data. *Nat. Mach. Intell.* 3 (1), 60–67.
- Lu, Jingyu, Li, Na, Zhang, Shaoyong, Yu, Zhibin, Zheng, Haiyong, Zheng, Bing, 2019. Multi-scale adversarial network for underwater image restoration. *Opt. Laser Technol.* 110, 105–113.
- Makhzani, Alireza, Shlens, Jonathon, Jaitly, Navdeep, Goodfellow, Ian, Frey, Brendan, 2015. Adversarial autoencoders. arXiv preprint [arXiv:1511.05644](https://arxiv.org/abs/1511.05644).
- Mathieu, Michael F., Zhao, Junbo Jake, Zhao, Junbo, Ramesh, Aditya, Sprechmann, Pablo, LeCun, Yann, 2016. Disentangling factors of variation in deep representation using adversarial training. In: *Advances in Neural Information Processing Systems*. Vol. 29.
- McGlamery, B.L., 1980. A computer model for underwater camera systems. In: *Ocean Optics VI*. Vol. 208. International Society for Optics and Photonics, pp. 221–231.
- Panetta, Karen, Agaian, Sos, Zhou, Yicong, Wharton, Eric J., 2010. Parameterized logarithmic framework for image enhancement. *IEEE Trans. Syst. Man Cybern. B* 41 (2), 460–473.
- Panetta, Karen, Gao, Chen, Agaian, Sos, 2015. Human-visual-system-inspired underwater image quality measures. *IEEE J. Ocean. Eng.* 41 (3), 541–551.
- Pang, Yingxue, Lin, Jianxin, Qin, Tao, Chen, Zhibo, 2021. Image-to-image translation: Methods and applications. *IEEE Trans. Multimed.* 24, 3859–3881.
- Paszke, Adam, Gross, Sam, Massa, Francisco, Lerer, Adam, Bradbury, James, Chanan, Gregory, Killeen, Trevor, Lin, Zeming, Gimelshein, Natalia, Antiga, Luca, et al., 2019. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* 32, 8026–8037.
- Peng, Yan-Tsung, Cosman, Pamela C., 2017. Underwater image restoration based on image blurriness and light absorption. *IEEE Trans. Image Process.* 26 (4), 1579–1594.
- Qi, Qi, Zhang, Yongchang, Tian, Fei, Wu, QM Jonathan, Li, Kunqian, Luan, Xin, Song, Dalei, 2021. Underwater image co-enhancement with correlation feature matching and joint learning. *IEEE Trans. Circuits Syst. Video Technol.*
- Ren, Shaoqing, He, Kaiming, Girshick, Ross, Sun, Jian, 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems*. Vol. 28.
- Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas, 2015. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 234–241.

- Shen, Zhen, Xu, Haiyong, Jiang, Gangyi, Yu, Mei, Du, Baozhen, Luo, Ting, Zhu, Zhongjie, 2023. Pseudo-retinex decomposition-based unsupervised underwater image enhancement and beyond. *Digit. Signal Process.* 137, 103993.
- Song, Wei, Wang, Yan, Huang, Dongmei, Liotta, Antonio, Perra, Cristian, 2020. Enhancement of underwater images with statistical model of background light and optimization of transmission map. *IEEE Trans. Broadcast.* 66 (1), 153–169.
- Song, Wei, Wang, Yan, Huang, Dongmei, Tjondronegoro, Dian, 2018. A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration. In: *Pacific Rim Conference on Multimedia*. Springer, pp. 678–688.
- Spurr, Adrian, Aksan, Emre, Hilliges, Otmar, 2017. Guiding infogan with semi-supervision. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 119–134.
- Sun, Xin, Liu, Lipeng, Li, Qiong, Dong, Junyu, Lima, Estanislau, Yin, Ruiying, 2019. Deep pixel-to-pixel network for underwater image enhancement and restoration. *IET Image Process.* 13 (3), 469–474.
- Sun, Boyang, Mei, Yupeng, Yan, Ni, Chen, Yingyi, 2023. UMGAN: Underwater image enhancement network for unpaired image-to-image translation. *J. Mar. Sci. Eng.* 11 (2), 447.
- Ulyanov, Dmitry, Vedaldi, Andrea, Lempitsky, Victor, 2017. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 6924–6932.
- Wang, Qi, Liu, Zhaoying, Zhang, Ting, Alasmary, Hisham, Waqas, Muhammad, Halim, Zahid, Li, Yujian, 2023a. Deep convolutional cross-connected kernel mapping support vector machine based on SelectDropout. *Inform. Sci.*.
- Wang, Tian, Liu, Zhaoying, Zhang, Ting, Hussain, Syed Fawad, Waqas, Muhammad, Li, Yujian, 2022. Adaptive feature fusion for time series classification. *Knowl.-Based Syst.* 243, 108459.
- Wang, Ting-Chun, Liu, Ming-Yu, Zhu, Jun-Yan, Tao, Andrew, Kautz, Jan, Catanzaro, Bryan, 2018. High-resolution image synthesis and semantic manipulation with conditional gans. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 8798–8807.
- Wang, Junting, Ye, Xiufen, Liu, Yusong, Mei, Xinkui, Hou, Jun, 2023b. Underwater self-supervised monocular depth estimation and its application in image enhancement. *Eng. Appl. Artif. Intell.* 120, 105846.
- Yan, Shuaizheng, Chen, Xingyu, Wu, Zhengxing, Wang, Jian, Lu, Yue, Tan, Min, Yu, Junzhi, 2021. Hybrur: A hybrid physical-neural solution for unsupervised underwater image restoration. *arXiv preprint arXiv:2107.02660*.
- Yang, Miao, Sowmya, Arcot, 2015. An underwater color image quality evaluation metric. *IEEE Trans. Image Process.* 24 (12), 6062–6071.
- Ye, Xincheng, Xu, Hongcan, Ji, Xiang, Xu, Rui, 2018. Underwater image enhancement using stacked generative adversarial networks. In: *Pacific Rim Conference on Multimedia*. Springer, pp. 514–524.
- Zhang, Weidong, Dong, Lili, Zhang, Tong, Xu, Wenhai, 2021a. Enhancing underwater image via color correction and bi-interval contrast enhancement. *Signal Process., Image Commun.* 90, 116030.
- Zhang, Ting, Gao, Zihang, Liu, Zhaoying, Hussain, Syed Fawad, Waqas, Muhammad, Halim, Zahid, Li, Yujian, 2023a. Infrared ship target segmentation based on adversarial domain adaptation. *Knowl.-Based Syst.* 265, 110344.
- Zhang, Ting, Waqas, Muhammad, Fang, Yu, Liu, Zhaoying, Halim, Zahid, Li, Yujian, Chen, Sheng, 2023b. Weakly-supervised butterfly detection based on saliency map. *Pattern Recognit.* 109313.
- Zhang, Ting, Waqas, Muhammad, Shen, Hao, Liu, Zhaoying, Zhang, Xiangyu, Li, Yujian, Halim, Zahid, Chen, Sheng, 2021b. A neural network architecture optimizer based on DARTS and generative adversarial learning. *Inform. Sci.* 581, 448–468.
- Zhao, Hengshuang, Shi, Jianping, Qi, Xiaojuan, Wang, Xiaogang, Jia, Jiaya, 2017. Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2881–2890.
- Zhou, Jingchun, Pang, Lei, Zhang, Dehuan, Zhang, Weishi, 2023a. Underwater image enhancement method via multi-interval subhistogram perspective equalization. *IEEE J. Ocean. Eng.*
- Zhou, Jingchun, Sun, Jiaming, Zhang, Weishi, Lin, Zifan, 2023b. Multi-view underwater image enhancement method via embedded fusion mechanism. *Eng. Appl. Artif. Intell.* 121, 105946.
- Zhou, Xingyi, Wang, Dequan, Krähenbühl, Philipp, 2019. Objects as points. *arXiv preprint arXiv:1904.07850*.
- Zhou, Jingchun, Yang, Tongyu, Chu, Weishen, Zhang, Weishi, 2022a. Underwater image restoration via backscatter pixel prior and color compensation. *Eng. Appl. Artif. Intell.* 111, 104785.
- Zhou, Jingchun, Yang, Tongyu, Chu, Weishen, Zhang, Weishi, 2022b. Underwater image restoration via backscatter pixel prior and color compensation. *Eng. Appl. Artif. Intell.* 111, 104785.
- Zhou, Jingchun, Yang, Tongyu, Zhang, Weishi, 2023c. Underwater vision enhancement technologies: A comprehensive review, challenges, and recent trends. *Appl. Intell.* 53 (3), 3594–3621.
- Zhou, Jingchun, Zhang, Dehuan, Zhang, Weishi, 2022c. Auto color correction of underwater images utilizing depth information. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5.
- Zhou, Jingchun, Zhang, Dehuan, Zhang, Weishi, 2023d. Cross-view enhancement network for underwater images. *Eng. Appl. Artif. Intell.* 121, 105952.
- Zhu, Jun-Yan, Park, Taesung, Isola, Phillip, Efros, Alexei A, 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 2223–2232.
- Zhuang, Peixian, Li, Chongyi, Wu, Jiamin, 2021. Bayesian retinex underwater image enhancement. *Eng. Appl. Artif. Intell.* 101, 104171.
- Zhuang, Peixian, Wu, Jiamin, Porikli, Fatih, Li, Chongyi, 2022. Underwater image enhancement with hyper-laplacian reflectance priors. *IEEE Trans. Image Process.* 31, 5442–5455.