

Perceptual Underwater Image Enhancement With Deep Learning and Physical Priors

Long Chen^{ID}, Zheheng Jiang^{ID}, Lei Tong, Zhihua Liu^{ID}, Aite Zhao, Qianni Zhang,
Junyu Dong^{ID}, Member, IEEE, and Huiyu Zhou^{ID}

Abstract—Underwater image enhancement, as a pre-processing step to support the following object detection task, has drawn considerable attention in the field of underwater navigation and ocean exploration. However, most of the existing underwater image enhancement strategies tend to consider enhancement and detection as two fully independent modules with no interaction, and the practice of separate optimisation does not always help the following object detection task. In this article, we propose two perceptual enhancement models, each of which uses a deep enhancement model with a detection perceptor. The detection perceptor provides feedback information in the form of gradients to guide the enhancement model to generate patch level visually pleasing or detection favourable images. In addition, due to the lack of training data, a hybrid underwater image synthesis model, which fuses physical priors and data-driven cues, is proposed to synthesise training data and generalise our enhancement model for real-world underwater images. Experimental results show the superiority of our proposed method over several state-of-the-art methods on both real-world and synthetic underwater datasets.

Index Terms—Underwater image enhancement, object detection, image synthesis, perceptual loss.

I. INTRODUCTION

UNDERWATER object detection (UOD) is of great importance for underwater applications such as ocean exploring and monitoring and autonomous underwater vehicles [1]. However, underwater images acquired in complicated environments suffer from severe distortion which dramatically degrades image visibility and affects the detection accuracy of UOD tasks.

Manuscript received June 29, 2020; revised September 20, 2020 and October 15, 2020; accepted October 18, 2020. Date of publication October 30, 2020; date of current version August 4, 2021. This article was recommended by Associate Editor J. Hou. (*Corresponding author: Huiyu Zhou*)

Long Chen, Zheheng Jiang, Lei Tong, Zhihua Liu, and Huiyu Zhou are with the School of Informatics, University of Leicester, Leicester LE1 7RH, U.K. (e-mail: lc408@leicester.ac.uk; zj53@leicester.ac.uk; lt228@leicester.ac.uk; zl208@leicester.ac.uk; hz143@leicester.ac.uk).

Aite Zhao is with the College of Computer Science and Technology, Qingdao University, Qingdao 266071, China (e-mail: zhaoaite@qdu.edu.cn).

Qianni Zhang is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: qianni.zhang@eecs.qmul.ac.uk).

Junyu Dong is with the Department of Information Science and Engineering, Ocean University of China, Qingdao 266100, China (e-mail: dongjunyu@ouc.edu.cn).

This article has supplementary downloadable material available at <https://ieeexplore.ieee.org>, provided by the authors.

Color versions of one or more of the figures in this article are available online at <https://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2020.3035108

In recent years, underwater image enhancement (UIE) technologies [1], [2], especially deep learning based approaches, work as a pre-processing operation to boost the detection accuracy of UOD tasks by improving the visual quality of underwater images. However, most of the existing strategies consider UIE and UOD tasks as two separate pipelines, whereas the UIE task is evaluated on the visual quality of images while the UOD task is evaluated on the detection accuracy. Separate optimisation of the two tasks results in inconsistency in the pursuit of image quality and detection accuracy: These two tasks have different optimisation objectives, leading to different optimal solutions. Moreover, current top-performing deep learning based UIE methods, e.g. [3], [4], are normally trained on synthetic images due to the lack of large training data (i.e., pairs of degraded underwater images and high-quality counterparts). The enhancement models trained on synthetic images cannot always be generalised to underwater scenes because the quality of synthetic images cannot be guaranteed by the existing image synthesis (UIS) methods.

To address these two concerns, we firstly propose a hybrid underwater image synthesis model to synthesize realistic training data using data-driven cues and physical priors. This enables our enhancement model to be properly generalised on real-world underwater scenes. Secondly, we propose two detection-perceptual enhancement models, each of which consists of an enhancement model and a detection perceptor. The detection perceptor is well-trained on high-quality in-air images and encodes fine object details and potential detection favouring information of high-quality in-air images. Two perceptual losses are designed to transfer the knowledge encoded in the detection perceptors to the enhancement model in the form of gradients (as inference of updating directions). One of the detection perceptors is named patch detection perceptor with a patch perceptual loss that guides the enhancement model to generate patch level visually pleasing images. The other one is named object-focused detection perceptor with an object-focused perceptual loss which guides the enhancement model to generate detection-favouring images. Fig. 1 shows the object detection results of the same Single Shot MultiBox Detector (SSD) [5] trained on the enhanced results of different UIE algorithms. The deep detectors trained on the enhanced results of two representative UIE algorithms, i.e., Fusion [6] and Retinex [7], often miss detecting “noisy” objects or predict incorrect object categories, while our object-focused perceptual enhancement model (denotes as OursOF) can largely

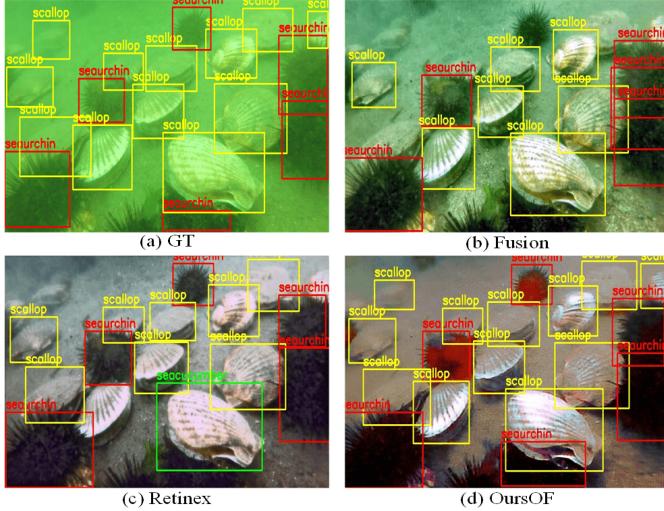


Fig. 1. Object detection results of the Single Shot MultiBox Detector [5] after we have applied different UIE algorithms, including (b) Fusion [6], (c) Retinex [7], and (d) Our proposed object-focused perceptual enhancement model. (a) Raw underwater image with ground-truth annotations.

improve the detection accuracy of the standard deep detector. The gradients from the object-focused perceptual loss highlight the region containing the object and produce clear reddish color around the object in Fig. 1(d). The proposed synthesis and the perceptual enhancement models are integrated into a unified framework named HybridDetectionGAN that can work in an end-to-end style. Our unique contributions can be summarised as follows:

- We propose two detection perceptual enhancement models, each of which consists of an enhancement model and a detection perceptor. The enhancement model can generate visually pleasing images on the patch-level, while the other one can generate detection-favouring images that help improving the detection accuracy. To our knowledge, this is the first practice for underwater image enhancement, aiming to generate detection-favouring rather than visually pleasing images.
- We propose a hybrid underwater image synthesis model, which incorporates both physical priors and data-driven cues. The hybrid synthesis model fully takes into account image characteristics such as color distortion, haze-effects and diversity, enabling our perceptual enhancement models to be generalised to handle real-world underwater scenes.
- The proposed hybrid synthesis and perceptual enhancement models are incorporated into a unified framework named HybridDetectionGAN, and can be jointly optimised in an end-to-end pattern.
- We conduct extensive evaluations and our proposed perceptual enhancement models outperform several state-of-the art UIE algorithms using image quality evaluation metrics and task-related detection accuracy metrics on both synthetic and real-world underwater datasets.

The rest of the paper is organised as follows: Section II summarises the related works. Section III describes the

proposed hybrid synthesis and perceptual enhancement models. Section IV describes the experimental set-up and Section V reports and discusses the experimental results.

II. RELATED WORK

With the success of deep learning in numerous computer vision tasks, researchers have moved their focus from traditional prior-based methods to data-driven methods in the field of UIS and UIE. In this section, we first summarise previous UIS and UIE methods. Then, we discuss the perceptual loss related to our proposed detection perceptual enhancement model.

A. Underwater Image Synthesis Method

Due to the lack of training image pairs, some UIS methods have been proposed to synthesize distorted underwater images from high quality in-air RGB or RGB-D images, and construct underwater-in-air image pairs as training data. These methods can be broadly classified into two categories: Physical model based and deep learning based UIS methods.

1) Physical Model Based Synthesis Methods: Physical model based UIS methods usually use an underwater image formation model to synthesize underwater images. Several works such as [8]–[10] followed the underwater image formation model proposed in [11] and synthesized underwater image datasets using 10 Jerlov water types [12]. The physical model assumes that an underwater image is formed with three optical processes: light absorption, light back-scattering and light forward-scattering. However, the commonly used underwater image formation model [8], [11] neglects the forward-scattering for simplification, which is formulated as:

$$I_w = I_{ab} + I_{bs} \quad (1)$$

where I_w , I_{ab} and I_{bs} are the underwater image, the image with light absorption and the image with light back-scattering, respectively. The light absorption component I_{ab} is the component of light reflected directly by the target object into the camera, which can be modelled as:

$$I_{ab} = I_a T = I_a e^{-\eta^\lambda d}, \quad \lambda \in \{r, g, b\} \quad (2)$$

where I_a is the undistorted in-air image before propagating through water, d is the distance from the camera to the captured scene and its value can be derived from the depth image. η is the wavelength-dependent attenuation coefficient. λ denotes different channels of an image, including red, green, and blue channels. T denotes the transmission map, indicating how much light is kept after the absorption process.

A photon of light travels through water, subject to light scattering which actively builds the characteristics of haze-effects. Denote I_{bs} as the image suffering from light back-scattering, formulated as follows:

$$I_{bs} = B^\lambda (1 - T) = B^\lambda (1 - e^{-\eta^\lambda d}), \quad \lambda \in \{r, g, b\} \quad (3)$$

where B^λ denotes the ambient light, which is a scalar parameter dependent on the light wavelength.

2) *Deep Learning Based Synthesis Methods*: Recently, Generative Adversarial Networks (GANs) [3], [4], [13], [14] have been investigated in the underwater image synthesis field due to its successes in image-to-image translation tasks. Li *et al.* [3] treated the UIS task as an image-to-image translation task and exploited a single GAN to synthesize underwater images from in-air RGB-D images. Their generator model can be broken down into three stages: (1) light absorption, which simulates the light absorption process by referring to optical priors; (2) light scattering, which simulates the light scattering process using a convolutional neural network without referring to the optical priors; (3) Vignetting, which produces a shading effect on the image corners, caused by certain camera lenses. However, their weakly-supervised synthesis model have to be trained on unpaired images, making it difficult to simulate the image details such as colors and textures. To alleviate the needs for training image pairs, Fabbri *et al.* [4] applied a two-way Cycle-Consistent Adversarial Networks (CycleGAN) [15], which allows the learning of mutual translation between in-air and underwater images. The CycleGAN includes two generators, where one generator translates in-air images to underwater ones and can be regarded as the synthesis model, whilst the other one translates underwater images into in-air images which can be regarded as the enhancement model.

However, both physical model based and GAN based UIS methods cannot accurately model the degradation progress of underwater imaging, and often result in unsatisfactory synthetic images [1], [16]. The commonly used physical underwater image formation model can only synthesize the scenes of 10 Jerlov water types and considers only two factors in the degradation progress, leading to significant errors in the generated images. Moreover, GAN based methods [17], [18] encounter the model collapse problem that generates images with monotonous colors and frequent artifacts. Their capability to modelling haze-effects is also limited. Different from the previous works, we first improve the physical image formation model, and leverage both physical priors and data-driven cues to a unified hybrid synthesis model to create more realistic underwater images.

B. Underwater Image Enhancement Method

Underwater image enhancement is an indispensable step to improve the visual quality of underwater images and can be categorised into the following three groups: model-free, physical model based, and deep-learning based methods.

1) *Model-Free Enhancement Methods*: Model-free UIE methods [6], [7], [19] aim to adjust image pixel values to improve the visual quality without referring to any physical imaging model. Ancuti *et al.* [6] proposed a fusion-based underwater image enhancement method by fusing a contrast enhanced underwater image and a color corrected image in a multi-scale fusion strategy. Later on, Ancuti *et al.* [20] fused two images derived from a white-balanced version of the underwater image with corresponding weighted maps in a multi-scale way, and important faded features and edges are recovered in the enhanced images. Fu *et al.* [19] presented a two-step approach for underwater image enhancement, which

includes a color correction algorithm based on piece-wise linear transformation and a contrast enhancement algorithm. Fu *et al.* [7] proposed a Retinex-based method for underwater image enhancement, which consists of color correction, layer decomposition and enhancement. Zhang *et al.* [21] extended the standard Retinex-based method by utilizing bilateral and trilateral filters on the three channels of the image in a CIELAB color space. The model-free methods can improve visual quality to some extent, but may accentuate noise, produce artifacts, and introduce color distortion.

2) *Physical Model Based Enhancement Methods*: Physical model based methods [22]–[26] treat underwater image enhancement as an inverse problem of image degradation. These methods usually establish a physical underwater image degradation model, and then estimate the unknown model parameters using various prior assumptions. Finally, high quality images can be retained by inverting this degradation process. Drews *et al.* [22] proposed an underwater dark channel prior (UDCP) by adapting the dark channel prior (DCP) [25] into underwater scenes. Peng *et al.* [23] proposed a generalised dark channel prior (GDCP) for image enhancement, which incorporates adaptive color correction into an image formation model. Galdran *et al.* [24] also proposed a variant of the dark channel prior algorithm, namely Red Channel, which recovers the lost contrast of an underwater image by restoring the colors associated with short wavelengths. Instead of using the DCP [25] prior, Zhou *et al.* [27] exploited the color-line prior to recover the color line of the image patches and well-handled the scattering and absorption problems. Wang *et al.* [28] proposed a novel underwater image restoration method based on adaptive attenuation-curve priors. They also set up the saturation constraint to adjust the transmission to prevent over saturation and reduce noise. Li *et al.* [29] employed a random forest regression model to estimate the medium transmission of the underwater scenes. Peng and Cosman [9] estimated the scene depth via image blurriness and light absorption, and employed the estimated depth to enhance underwater images. Li *et al.* [30] proposed an underwater image enhancement method, based on the minimum information loss principle and histogram distribution priors.

3) *Deep Learning Based Enhancement Methods*: Deep learning based enhancement methods [3], [31]–[33] usually construct deep neural networks and train them using pairs of degraded underwater images and high-quality counterparts. Li *et al.* [3] first synthesized underwater images from RGB-D in-air images, and then trained a two-stage network for underwater image restoration with the synthetic training data. Li *et al.* [31] employed a gated fusion network architecture to learn three confidence maps used to combine the three input images into the final one. To address the lack of paired training data, Ye *et al.* [34] proposed an unsupervised adaptation network to jointly estimate scene depths and correct color from monocular underwater images. Fabbri *et al.* proposed an underwater color transfer model [4] based on CycleGAN [15] without needing paired training data.

However, all of the three category enhancement methods only improve the visibility of images while ignoring their influence on the later high-level tasks such as underwater

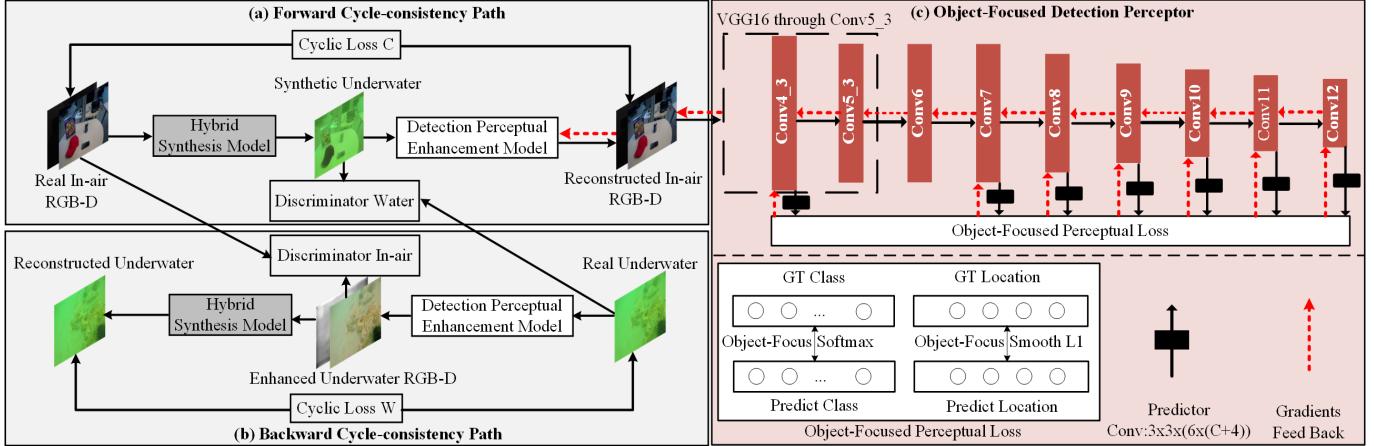


Fig. 2. The overview of our HybridDetectionGAN. It consists of two cycle-consistency paths (a) and (b) to learn the transformation between underwater and in-air domains from unpaired images. An object-focused detection perceptor (c) guides the enhancement model to generate detection favourable images.

object detection [1] and salient object detection [35], [36]. Different from the previous work, we first add the supervision of object detection to the enhancement network that guides the enhancement network to generate detection favourable images rather than visually pleasing images.

III. PROPOSED METHODS

In this section, we first present the overview of our proposed HybridDetectionGAN framework. Then, we describe the proposed hybrid synthesis and perceptual enhancement models. Finally, we introduce the training of the framework.

A. The Overview of the Proposed Framework

As shown in Fig. 2, our HybridDetectionGAN framework exploits two cycle-consistency paths to learn the transformation between underwater and in-air domains from the unpaired images. Particularly, the forward cycle-consistency path starts with real in-air RGB-D images and finishes with reconstructed in-air images (i.e., enhanced underwater images and depth images). The hybrid synthesis model transforms in-air RGB-D images into their underwater counterparts, and the enhancement model transforms underwater images into in-air images. The cyclic/cycle-consistency loss regularises both the synthesis and enhancement models to generate better structural content in the images. Following the enhancement model, a novel object-focused detection perceptor is activated, which is trained with the detection loss on the real in-air images. The enhancement model and the detection perceptor construct the complete detection-perceptual enhancement model. During the training of the enhancement model, detection-favouring information (red dashed arrows in Fig. 2) is given to the enhancement model in the form of gradients via the object-focused perceptual loss, which is consisted of an object-focused Softmax loss and an object-focused SmoothL1 loss. The backward cycle-consistency path starts with real underwater images and finishes with the reconstructed underwater ones. The adversarial process between the enhancement model and the discriminator helps produce realistic in-air images.

B. Hybrid Underwater Image Synthesis Model

The generalisation of the enhancement model highly relies on the quality of the synthetic training data. To develop a robust synthesis model, we incorporate an improved physical model and a data-driven CNN model into a hybrid synthesis model. The improved physical model is able to simulate evident haze-effects and maintains coarse views of underwater images by applying the priors of light absorption and scattering. The CNN model works as a supplement to generate finer details of underwater images by modelling other factors. Fig. 3 shows the overall structure of the hybrid synthesis model. It consists of three branches, i.e., light absorption, light scattering, and CNN branches. The light absorption and scattering branches are built on optical priors, and construct the complete physical model.

1) *Improved Underwater Image Formation Model:* Our underwater image formation model is based on the physical model shown in Eq. (1), and formulated as Eq. (4).

$$I_{sw}^\lambda(x, y) = \sum_{w=0}^W \sum_{h=0}^H \sum_{m=0}^M I_{con}(x + w, y + h, m) \theta_f^\lambda(w, h, m), \quad \lambda \in \{r, g, b\} \quad (4)$$

where

$$I_{con} = I_{ab} + I_{sc} \odot I_{cnn} \quad (5)$$

$I_{sw}(x, y)$ denotes the pixel value of the synthetic underwater image at each point (x, y) . λ denotes different channels of an image, including red, green, and blue channels. θ_f^λ is a $W \times H \times M$ convolutional filter, responsible for converting the outputs of the three branches into the λ -channel of the synthetic underwater image. I_{ab} , I_{sc} and I_{cnn} are the output of the light absorption branch, the light scattering branch and the CNN branch, respectively. I_{con} is the fused output of three branches and formulated as Eq. (5), where $+$ and \odot denote the element-wise add operation and channel-wise concatenation operation respectively. The final synthetic underwater image I_{sw} is achieved through the convolution operation between θ_f and I_{con} in Eq. (4).

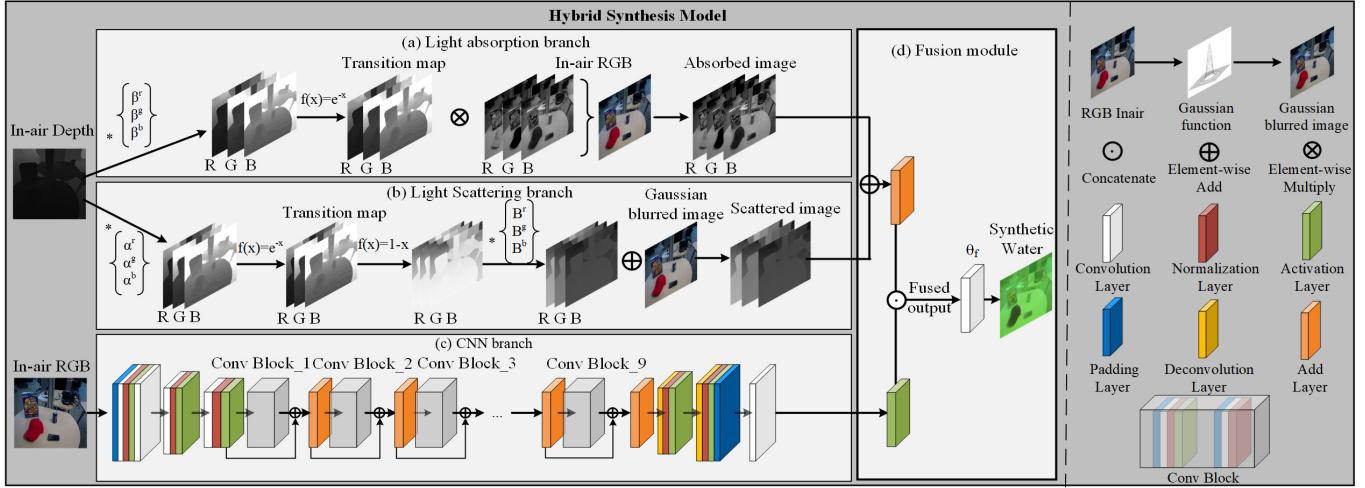


Fig. 3. The overview of the proposed hybrid synthesis model, which consists of (a) light absorption branch, (b) light scattering branch and (c) CNN branch. The outputs of three branches are combined into the final synthetic underwater image shown in (d) fusion module.

Light absorption causes the change of color tones, and the image suffering from light absorption can be described by Eq. (6). Different channels of a RGB image have different absorption coefficient β . The effect of light absorption becomes stronger with increasing object-camera distance d as more energy is absorbed by water. For each pixel on the image, its depth d comes from depth map I_d . To retain the absorption image, we first compute the transition map $T = e^{-I_d\beta^\lambda}$, indicating part of the light has been absorbed during the propagation in the water. Then, we compute each channel of the absorption image I_{ab}^λ using element-wise multiplication operation \otimes between in-air image I_a and T .

$$I_{ab}^\lambda = I_a^\lambda \otimes e^{-I_d\beta^\lambda}, \quad \lambda \in \{r, g, b\} \quad (6)$$

Light scattering is an optical process in underwater imaging, including forward and back scattering [14]. Forward scattering occurs when the light reflected from the object is scattered on its way to the camera, resulting in an effect very similar to Gaussian blurring. However, the commonly used physical model only considers back scattering priors while ignoring forward scattering priors. Hence, we simulate the haze-effects caused by forward scatters using a Gaussian blur function. The scattered image I_{sc} is formulated as

$$I_{sc}^\lambda = I_{bsc}^\lambda + I_{fsc}^\lambda, \quad \lambda \in \{r, g, b\} \quad (7)$$

where

$$I_{bsc}^\lambda = B^\lambda (1 - e^{-I_d\alpha^\lambda}), \quad \lambda \in \{r, g, b\} \quad (8)$$

$$I_{fsc}^\lambda = I_a \Phi(x, y) \quad (9)$$

$$\Phi(x, y) = Ae^{-\frac{x^2+y^2}{q^2}} \quad (10)$$

where A is determined by

$$\int \int Ae^{-\frac{x^2+y^2}{q^2}} dx dy = 1 \quad (11)$$

where I_{bsc} and I_{fsc} are the images suffering from backward and forward scattering, respectively. B and α denote the

ambient light and the backscatter coefficient. $\Phi(x, y)$ denotes the Gaussian function that adds the Gaussian blurring effect onto the in-air image I_a .

Different from the previous physical model based synthesis methods, which use predefined parameters to synthesize the scenes of 10 Jerlov water types, the parameters of our proposed underwater image formation model can be learnt using gradient-based optimisation algorithms and better simulate the characteristics of the targeted underwater images. Moreover, a CNN branch works as a supplement of the physical model to simulate more degraded characteristics of the underwater images. The light absorption and scattering priors of the physical model can be used to simulate coarse color distortion and haze-effects. However, significant errors exist in the resultant images due to image quality degrading. For instance, the existence of artificial lights leads to non-uniform illumination on the images, and the movement of the cameras bring in noise in the captured images. All of these factors have not been modelled in the physical model, and the CNN branch helps us to simulate these factors and generates finer color tones, illumination change and noise. The detailed structure of the CNN branch can be found in Fig. 3.

C. Detection Perceptual Enhancement Model

The perceptual loss has been widely used to improve the quality of images or generate images of interests, and its success in many tasks demonstrates that a high-capacity CNN properly trained has the capability of implicitly encoding fine image details and task-related semantic knowledge (the related work of perceptual loss can be found in Supplementary K). Inspired by these observations, we first train a one-stage deep detector [5] on high quality in-air images, which encodes object details and potential detection favouring images. Then, we add this deep detector as a detection perceptor (its weights are kept fixed to those obtained from training) to provide detection-favouring perceptual information to the enhancement model.

During the adversarial training, the enhanced image goes straight to the detection perceptor. The detection perceptor is an one-stage deep detector [5] as shown in Fig. 2, which first associates 6 default patches of different scales and aspect ratios at each of the seven convolutional layers (we only show 2 default patches on the 9-th convolutional layer in Supplementary A for simplicity). Then, the predictor works with a $C+1$ -dimension class vector and a 4-dimension location vector for each default patch using a 3×3 convolutional filter. C denotes the number of the object classes and 1 denotes the background class. Next, it assigns the ground-truth class and the location for each default patch using the following matching rules: If the Interaction over Union (IoU) between a default patch and its overlapped ground-truth object is larger than 0.5, the class and the location of the object are assigned as those of the default patch. If the default patch does not match any ground-truth object with an IoU larger than 0.5, it is labelled as the background patch and has no ground-truth location. For example, the default patch (red) shown in Supplementary A matches the cereal box (green box) with the IoU of larger than 0.5, so it is an object patch and its ground-truth class and location are labelled as those of the cereal box. The black patch does not match any object, so it is labelled as the background patch and has no target location. Finally, a detection based perceptual loss indicates the discrepancy between the patches of the enhanced image and those of the high quality in-air image, and feeds back the discrepancy to the enhancement model in the form of gradients, based on the enhancement model of continuously updating its parameters. Until no gradient has any impact, when the enhanced images are the same as the in-air images in the detection perceptor space, the object details and detection favouring information of the in-air image encoded in the detection perceptor space have been properly transformed via the enhancement model.

We design two perceptual loss functions having different objectives for two detection perceptors. The first loss is named patch detection perceptual loss that aims to generate patch-level in-air images. The second one is named object-focused detection perceptual loss which aims to generate detection-favouring images that can improve the detection accuracy. Supplementary A shows the overall structure of the patch detection perceptual enhancement and the object-focused detection perceptual enhancement models, each of which consists of an enhancement model and a detection perceptor with a specially designed perceptual loss.

1) Patch Detection Based Perceptual Loss: Patch detection perceptual loss function L_p is an one-stage detection loss [5], which is a weighted sum of classification loss L_{cls} and localization loss L_{loc} .

$$L_p = \frac{1}{N} \sum_{i \in all} L_{cls}(pcls^i, gcls^i) + \frac{1}{N} \sum_{i \notin bg} L_{loc}(ploc^i, gloc^i) \quad (12)$$

where $pcls^i$ and $gcls^i$ denote the predicted and ground-truth class vectors of $C + 1$ dimensions for the i -th default patch. $ploc^i$ and $gloc^i$ denote the predicted and ground-truth

location vector of 4-dimensions for the i -th default patch. The 4-dimension location vector includes the coordinates of center (cx, cy) with width w and height h . all and bg are the set of all the default patches and the patches belonging to the background samples, without any contribution to the location loss because of absent ground-truth location. N and \bar{N} are the numbers of all the default and the object patches. Specially, the classification loss L_{cls} is a softmax loss.

$$L_{cls}(pcls, gcls) = - \sum_{c=1}^{C+1} pcls_c \log(gcls_c) \quad (13)$$

where pre_cls_c and gt_cls_c indicate the c -th element of the predicted and the ground-truth class vectors, respectively. The localisation loss is a smooth L1 loss [5] between the predicted and the ground-truth locations.

$$L_{loc}(ploc, gloc) = \sum_{l=1}^4 smooth_{L1}(ploc_l - gloc_l) \quad (14)$$

L_{cls} encourages the enhancement model to generate images which minimise the class discrepancy between the generated and the ground-truth patches, and L_{loc} encourages the enhancement model to generate the images of minimising the location discrepancy between the generated and the ground-truth patches, thus the patch perceptual detection loss can guide the enhancement model towards generating more realistic patches at accurate locations.

2) Object-Focused Detection Based Perception Loss: For underwater object detection applications, many objects look very similar to the background. The complex background may degrade the detection accuracy of the detectors. To deal with this challenge, we propose the object-focused detection based perceptual loss L_{of} :

$$L_{of} = \frac{1}{\bar{N}} \sum_{i \notin bg} L_{cls}(pcls^i, gcls^i) + L_{loc}(ploc^i, gloc^i) \quad (15)$$

Different from the patch detection perceptual loss, the object-focused detection perceptual loss only focuses on feeding back the informations of object patches, while ignoring the background patches, as shown in Supplementary A (the black cross indicates background patches have no information feedback to the enhancement model). From the optimisation perspective, L_{of} is designed to assign ground-truth classes and locations of the object patches on the enhanced image, for achieving detection accuracy of the deep detector trained on the enhanced images.

D. Training of Our Overall HybridDetectionGAN

We first train the standard one-stage deep detector on high quality in-air images. Afterwards, we add the detector after the enhancement model in the forward cycle-consistency path. We then move on to train the synthesis and enhancement models.

1) Training of the Hybrid Synthesis Model: Denoting $G_{\theta_{a2w}}$ as the hybrid synthesis model parameterised by θ_{a2w} and θ_{cnn} as the parameters of the CNN branch, then we have $\theta_{a2w} = \{\alpha, \beta, B, \theta_{cnn}, \theta_f\}$. Denote $G_{\theta_{w2a}}$ as the enhancement model

parameterised by θ_{w2a} . We obtain θ_{a2w} by minimising the loss function L_{a2w} , which is a combination of an adversarial loss L_{adv_w} and a cycle-consistency loss L_{cyc_w} .

$$L_{a2w} = w_1 L_{adv_w} + w_2 L_{cyc_w} \quad (16)$$

The first term is an adversarial loss produced by the discriminator, which is denoted as $D_{\theta_{dw}}$ and parameterised by θ_{dw} . Taking the synthetic underwater image $G_{\theta_{a2w}}(I_a, I_d)$ as the input, the discriminator outputs the estimated probability of the synthetic underwater image treated as a “real” underwater image, denoted as $D_{\theta_{dw}}(G_{\theta_{a2w}}(I_a, I_d))$. By fooling the discriminator with the synthetic underwater image, the adversarial loss is formulated as $L_{adv_w} = -\log D_{\theta_{dw}}(G_{\theta_{a2w}}(I_a, I_d))$, which encourages the hybrid synthesis model to produce more realistic underwater images. The cycle-consistency loss L_{cyc_w} is computed as the L_1 distance between the reconstructed and ground-truth underwater image I_w , i.e., $L_{cyc_w} = ||G_{\theta_{a2w}}(G_{\theta_{a2w}}(I_w)) - I_w||_1$.

Different from the physical-model based synthesis model, our physical parameters α , β and B can better simulate the characteristics of the target underwater images, which are learnt from the training data via the gradient descent optimisation algorithm. The optimisation algorithm iteratively updates the parameter β^λ by

$$\beta^\lambda = \beta^\lambda - \eta \frac{\partial L_{a2w}}{\partial \beta^\lambda} \quad (17)$$

where η is the learning rate. In order to update β^λ , we need to compute $\frac{\partial L_{a2w}}{\partial \beta^\lambda}$, which indicates the gradient of L_{a2w} with respect to β^λ . Denote I_{add}^λ as the output of the physical branch, $I_{add}^\lambda = I_{ab}^\lambda + I_{sc}^\lambda$. I_{ab}^λ , I_{con} and I_{sw} are presented in Eqs. (4)-(6). We can derive $\frac{\partial L_{a2w}}{\partial \beta^\lambda}$ using the chain rule (see Supplementary B) and finally produce β^λ by Eq. (18).

$$\begin{aligned} \beta^\lambda &= \beta^\lambda + \eta(w_1 \frac{\partial L_{adv_w}}{\partial I_{sw}} + w_2 \frac{\partial L_{cyc_w}}{\partial I_{sw}}) \\ &\quad (I_a^\lambda \otimes e^{-I_d \beta^\lambda} I_d \frac{\partial I_{sw}}{\partial I_{con}} \frac{\partial I_{con}}{\partial I_{add}^\lambda}) \end{aligned} \quad (18)$$

Similarly, we update α and B through the chain rule, and have the final parameters when the models converge.

$$\begin{aligned} \alpha^\lambda &= \alpha^\lambda - \eta(w_1 \frac{\partial L_{adv_w}}{\partial I_{sw}} + w_2 \frac{\partial L_{cyc_w}}{\partial I_{sw}}) \\ &\quad (B^\lambda e^{-I_d \alpha^\lambda} I_d \frac{\partial I_{sw}}{\partial I_{con}} \frac{\partial I_{con}}{\partial I_{add}^\lambda}) \end{aligned} \quad (19)$$

$$\begin{aligned} B^\lambda &= B^\lambda - \eta(w_1 \frac{\partial L_{adv_w}}{\partial I_{sw}} + w_2 \frac{\partial L_{cyc_w}}{\partial I_{sw}}) \\ &\quad ((1 - e^{-I_d \alpha^\lambda}) \frac{\partial I_{sw}}{\partial I_{con}} \frac{\partial I_{con}}{\partial I_{add}^\lambda}) \end{aligned} \quad (20)$$

We obtain θ_{dw} by optimising the loss function L_{d_w} , which encourages the discriminator to identify the difference between the synthetic and real underwater images:

$$L_{d_w} = -\log D_{\theta_{dw}}(I_w) - \log(1 - D_{\theta_{dw}}(G_{\theta_{a2w}}(I_a, I_d))) \quad (21)$$

2) Training of the Detection Perceptual Enhancement Model: We obtain θ_{w2a} by optimising the loss function L_{w2a} , which is a weighted combination of adversarial loss L_{adv_a} , cycle-consistency loss L_{cyc_a} , and perceptual loss L_{of} (or L_{patch}).

$$L_{w2a} = w_1 L_{adv_a} + w_2 L_{cyc_a} + w_3 L_{of} \quad (22)$$

Denote $D_{\theta_{da}}$ as the discriminator parameterised by θ_{da} . Taking the enhanced RGB-D underwater image $G_{\theta_{w2a}}(I_w)$ as input, the discriminator outputs the estimated probability of the enhanced image as a real in-air image, denotes as $D_{\theta_{da}}(G_{\theta_{w2a}}(I_w))$. To fool the discriminator with the enhanced underwater image, the adversarial loss L_{adv_a} is formulated as $L_{adv_a} = -\log D_{\theta_{da}}(G_{\theta_{w2a}}(I_w))$. L_{cyc_a} is computed as the L_1 distance between the reconstructed and ground-truth in-air images, $L_{cyc_a} = ||G_{\theta_{w2a}}(G_{\theta_{a2w}}(I_a)) - I_a||_1$. L_{of} is the object-focused perceptual loss that encourages the enhancement model to generate detection-favouring outcomes.

We obtain θ_{da} by optimising L_{d_a} , which encourages the discriminator to address the difference between the enhanced underwater images and the real in-air images.

$$L_{d_a} = -\log D_{\theta_{da}}(I_a) - \log(1 - D_{\theta_{da}}(G_{\theta_{w2a}}(I_w))) \quad (23)$$

3) How the Detection Perceptor Influence the Enhancement Model in the Form of Gradients: During the training of the enhancement model, the optimisation algorithm iteratively updates the enhancement model’s parameter θ_{w2a} by

$$\begin{aligned} \theta_{w2a} &= \theta_{w2a} - \eta \frac{\partial L_{w2a}}{\partial \theta_{w2a}} \\ &= \theta_{w2a} - \eta(w_1 \frac{\partial L_{adv_a}}{\partial \theta_{w2a}} + w_2 \frac{\partial L_{cyc_a}}{\partial \theta_{w2a}} + w_3 \frac{\partial L_{of}}{\partial \theta_{w2a}}) \end{aligned} \quad (24)$$

$$\frac{\partial L_{of}}{\partial \theta_{w2a}} = \frac{1}{N} \sum_{i \notin bg} \frac{\partial L_{cls}(pcls^i, gcls^i)}{\partial \theta_{w2a}} + \frac{\partial L_{loc}(ploc^i, gloc^i)}{\partial \theta_{w2a}} \quad (25)$$

In each iteration, the detection perceptor feeds the gradients $\eta w_3 \frac{\partial L_{of}}{\partial \theta_{w2a}}$ back to the enhancement model. From Eqs. (24) and (25), we can see that the enhancement model continuously updates its parameter θ_{w2a} to minimise $L_{cls}(pcls^i, gcls^i)$ and $L_{loc}(ploc^i, gloc^i)$, equivalently maximising the class and location prediction accuracy of the object patches. Thus the gradients $\eta w_3 \frac{\partial L_{of}}{\partial \theta_{w2a}}$ help the enhancement model to generate the images with accurate object detection in the following process.

IV. EXPERIMENTAL SETUP

To demonstrate the effectiveness of the proposed method, we conduct comprehensive evaluations on both the unpaired ChinaMM-MultiView and paired OUC datasets. In this section, we first introduce the experimental datasets and the evaluation metrics. Then, we describe the implementation details.

A. Datasets

The unpaired ChinaMM-MultiView dataset is constructed by collecting images from an underwater image dataset ChinaMM [1] and an in-air image dataset MultiView [37]. ChinaMM is a public competition dataset for evaluating UIE algorithms. The owner of the dataset has publicly released the train set of 2,071 images and the validation set of 676 images. This dataset provides bounding box annotations and contains three object categories: seacucumber, seaurchin and scollap. The resolution of each image is 720×405 pixels. The in-air dataset Multiview consists of 14,179 training images and 1,206 testing images which are captured in the in-door scenes with high quality. This dataset provides RGB images (640×480 pixels), depth images and bounding box annotations. It contains five object categories: bowl, cap, coffee mug, cereal box and soda can. To construct the unpaired ChinaMM-MultiView dataset, we randomly choose 2,071 images as the training set and 676 images as the testing set of MultiView.

The paired OUC dataset [38] provides underwater images, high quality reference images and bounding box annotations. The training set contains 2,500 image pairs where the testing set contains 1,198 image pairs. The dataset does not provide depth images which are needed by our hybrid synthesis model, so we apply the technology reported in [13] to obtaining depth maps for all the reference images.

The unpaired Berman-MultiView dataset is constructed by collecting images from the well-known underwater dataset of Berman *et al.* [39] and the in-air image dataset MultiView [37]. The Berman dataset [39] provides 114 high resolution TIF images ($5,474 \times 3,653$ pixels), raw images, camera calibration files, and the reconstructed scene distance maps that can be downloaded from the project's webpage: http://csms.haifa.ac.il/profiles/tTreibitz/datasets/ambient_forwardlooking/index.html.

B. Evaluation Metrics

We conduct extensive experiments to quantitatively and qualitatively evaluate the proposed hybrid synthesis and detection perceptual models. For the qualitative evaluations, we directly present the resultant images. For the quantitative evaluations, we apply several commonly used full-reference image quality evaluation metrics, where ground-truth or references are available. The full-reference metrics include the widely used Mean Square Error (MSE), Structural Similarity (SSIM) and Peak Signal-to-Noise Ratio (PSNR), and Patch-based Contrast Quality Index (PCQI) [40]. For the experiments with no reference image, we apply two non-reference image quality evaluation metrics, including underwater image quality measure (UIQM) [41] and underwater color image quality evaluation (UCIQE) [42]. UIQM is a linear combination of three components, i.e., Underwater Image Colorfulness Measure (UICM), Underwater Image Sharpness Measure (UISM), and Underwater Image Contrast Measure (UIConM). In the experiments, we also reveal the values of these three components for detailed discussion. In addition to the image quality evaluation metrics, we train

the deep detectors using the enhanced images by different UIE algorithms and use the mean Average Precision (mAP) as a detection task-specific evaluation metric to evaluate different UIE algorithms.

C. Implementation Details

All the experiments are conducted on a server with an Intel Xeon CPU @ 2.40GHz and 2 parallel Nvidia Tesla P100 GPUs. We implement the proposed HybridDetectionGAN framework using the Keras framework. We train the detection perceptor using the Adam optimiser [43] with 120 epochs and an initial learning rate of 1e-3. The learning rate is decreased by a factor of 10 after 60 epochs. We train HybridDetectionGAN for 200 epochs. The initial learning rate of the hybrid synthesis model, the perceptual enhancement model and two discriminators are 2e-4, and after 100 epochs, we apply a linear decay of the learning rate for all four components. The source code is available at: <https://github.com/LongChenCV/HybridDetectionGAN>.

V. RESULTS AND DISCUSSION

In this section, we present and discuss the experimental results and findings. We first conduct ablation experiments to investigate the influence of different components of our proposed HybridDetectionGAN framework. Then, we compare our method against several state-of-the-art methods on the two datasets. Finally, we investigate how these UIE algorithms influence the deep detectors in the following process.

A. Ablation Studies

The proposed HybridDetectionGAN framework integrates a hybrid underwater image synthesis model and a detection perceptual underwater enhancement model. We conduct ablation experiments in order to evaluate them on both unpaired ChinaMM-MultiView and paired OUC datasets.

1) Ablation Studies of the Hybrid Synthesis Model: We conduct ablation experiments to investigate how the CNN, scattering and absorption branches and the complete physical branch (consisting of the scattering and absorption branches) influence the synthetic image results.

Fig. 4 presents the qualitative comparison of the synthesis models with different component settings on MultiView dataset. We observe that our complete hybrid synthesis model produces the visual appearance most similar to the real underwater images of ChinaMM. The synthesis models without the scattering branch have a limited capability of modelling the haze-effects. For example, the resultant images of CNN-Only and CNN+Absorption synthesis models are relatively clear in spite of severe color distortion. After having incorporated the scattering prior in these two models, we clearly witness the haze-effects on the images. Both CNN-Only and Physical-Only synthesis models do not generate diverse results. This is because the former one usually runs into the classical mode collapse problem that only produces outcomes of a single mode, e.g., all the synthetic images are of the same color tone. The latter also generates underwater images in a

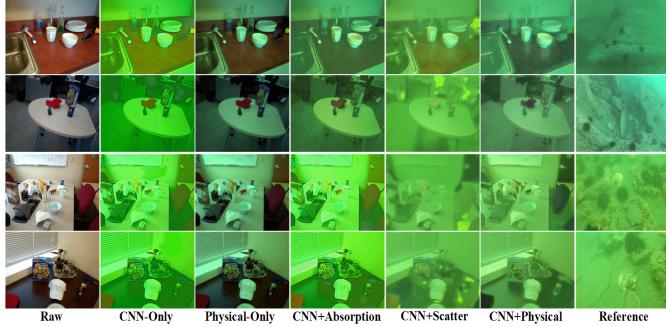


Fig. 4. Qualitative comparison of synthesis models with different component settings on MultiView. From left to right are raw in-air images, the results of the synthesis models with only the CNN branch, only the physical branch, CNN and absorption branches, and CNN and scattering branches, complete hybrid synthesis model, and reference underwater images from ChinaMM. Best viewed in the digital form.

TABLE I
QUANTITATIVE COMPARISON OF THE SYNTHESIS MODELS WITH DIFFERENT COMPONENTS ON THE OUC DATASET

Model	CNN	Absorption	Scatter	MSE	PSNR	SSIM	PCQI
Synthesis Models	✓			0.2777	23.7386	0.7664	0.6627
	✓	✓		0.1360	26.8716	0.8963	0.9398
	✓		✓	0.1289	27.0851	0.9046	0.9364
		✓	✓	0.2882	23.5787	0.7659	0.6600
	✓	✓	✓	0.0978	28.2895	0.9131	0.9518

monotonous style. Once the physical model has been trained, only one fixed set of parameters have been optimised, leading to optimal results within a specific environment. When we integrate the CNN and absorption branches, diverse results are obtained. The absorption prior helps CNN to generate images with different color tones whilst avoiding the artifact problem. For example, artifacts frequently occur in the resultant images of the synthesis models without the absorption branch. Supplementary C presents the qualitative comparison of the synthesis models with different component settings on the OUC dataset.

In addition to the qualitative comparison, we also use four full-reference image quality evaluation metrics to evaluate the synthesis models supported by the reference images in the OUC dataset. From Table I, we observe the superiority of the complete hybrid synthesis model over the other models as to four metrics. This indicates the synthetic underwater images of the hybrid model are the closest ones to the reference images. After having removed the absorption branch, the values of the four metrics decrease due to the existence of frequent artifacts. Removing the scattering branch also decreases the values of the four metrics due to the haze-effects. The synthesis model with the physical model only generates the images with color distortion and the worst quantitative scores.

2) **Ablation Studies of the Enhancement Model:** We first investigate the influence of the quality of the training data, i.e., the synthetic underwater images, on the enhancement model. Then, we analyse how the detection perceptor affects the enhancement model.

The influence of the quality of the synthetic underwater images on the enhancement model. We divide the synthetic underwater images into four categories: (A) Synthetic underwater images with incorrect color tones generated by

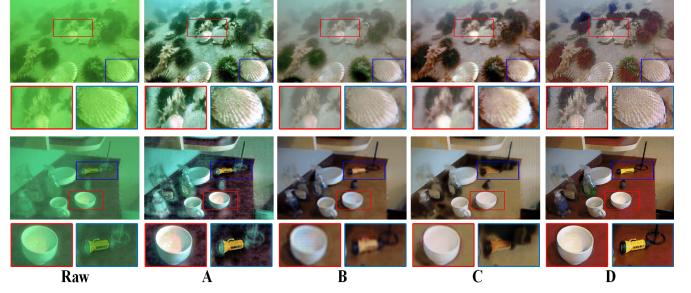


Fig. 5. Qualitative comparison of the enhancement models trained on different synthetic underwater images on ChinaMM (top row) and MultiviewUnderwater (bottom row). From left to right are the raw underwater images, the results of the enhancement model trained on (A) synthetic images with incorrect color tones, (B) synthetic images without evident haze-effects, (C) synthetic images with artifacts, and (D) synthetic images with pleasing appearance.

the Physical-Only synthesis model; (B) Synthetic underwater images without evident haze-effects generated by the CNN+absorption synthesis model; (C) Synthetic underwater images with artifacts generated by the CNN+scattering synthesis model; (D) Synthetic images with pleasing appearance generated by the hybrid synthesis model.

For the unpaired ChinaMM-MultiView dataset, we examine different enhancement models on two sub-datasets, i.e., the synthetic underwater dataset, MultiviewUnderwater, which is generated by our hybrid synthesis model with the RGB-D in-air images of MultiView, and the real-world underwater dataset, ChinaMM. Fig. 5 shows the qualitative comparison of the enhancement models trained on different synthetic underwater images on ChinaMM and MultiviewUnderwater. We observe that the enhancement model trained on (A) (synthetic images with incorrect color tones) cannot correct the color casts due to the lack of learning on color transformation between the underwater and in-air images. The enhancement model trained on (B) (haze-free synthetic underwater images) leaves evident haze-effects on its results while the enhancement model trained on (C) (synthetic underwater images with artifacts) further aggravates the artifacts problem in the resultant images. By contrast, the enhancement model trained on (D) (visually pleasing synthetic images) performs much better in removing haze-effects and artifacts. However, the results of the enhancement model trained on (D) still suffer from minor artifacts and haze-effects. This is because the deep enhancement model is constructed using fully convolutional layers, which have a limited ability to remove artifacts and haze-effects.

Tables II reports the quantitative scores of the enhancement models trained on different synthetic underwater images on the testing sets of MultiviewUnderwater and ChinaMM. We also list the full-reference scores of the ground-truth in-air images on MultiviewUnderwater and the non-reference scores of the ground-truth underwater images of ChinaMM as references. In terms of image quality evaluation metrics, the enhancement model trained on (D) visually pleasing images performs best while the one trained on (A) images with color distortion achieves the worst scores. The color cast in the results of the latter enhancement model leads to the decreasing scores of

TABLE II

QUANTITATIVE COMPARISON OF THE ENHANCEMENT MODELS TRAINED WITH DIFFERENT SYNTHETIC UNDERWATER IMAGES ON THE MULTIVIEWUNDERWATER AND CHINAMM DATASETS

Enhancement models	MultiviewUnderwater					ChinaMM		
	MSE	PSNR	SSIM	PCQI	mAP	UCIQE	UIQM	mAP
Model trained on A	1.3061	15.5623	0.1906	0.5294	75.9	23.2144	2.3864	72.2
Model trained on B	1.0058	17.0883	0.2971	0.5549	76.3	25.2220	2.7138	74.0
Model trained on C	0.9249	18.4655	0.3831	0.5425	72.1	25.6417	3.1529	71.1
Model trained on D	0.7153	20.5488	0.5632	0.5677	77.5	27.3206	3.8196	76.5
GT in-air image	0.0000	Inf	1.0000	1.0000	79.9	-	-	-
GT underwater image	-	-	-	-	-	21.3083	1.4410	68.6

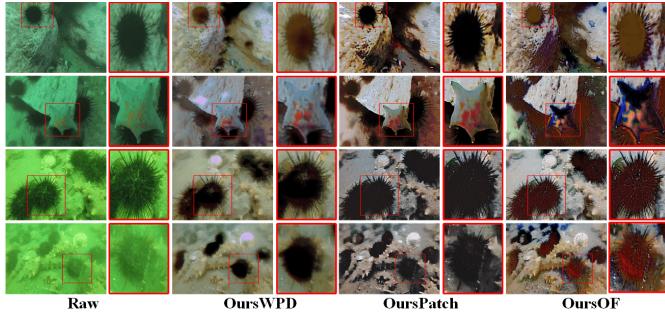


Fig. 6. Qualitative comparison of the enhancement models with different detection perceptor settings on ChinaMM. From left to right are the raw underwater images, the results of OursWDP, OursPatch, and OursOF, respectively.

the image quality metrics. In terms of mAP, the model trained on (C) images with artifacts achieves the lowest score even though it has relatively higher quantitative scores for the image quality evaluation metrics. The artifacts smear the details of the images or objects, deteriorating the detection accuracy more than the incorrect color casts and haze-effects. The enhancement models trained on the visually pleasing images obtain the best scores of image quality evaluation metrics and mAP on the two datasets. In summary, the enhancement model trained on more realistic synthetic underwater images can learn more accurate mappings between the images in underwater and in-air domains and be better generalised on the real-world underwater dataset.

The influence of the detection perceptor on the enhancement model. We compare three enhancement models with different detection perceptor settings, i.e., enhancement model without detection perceptor (denoted as OursWDP), enhancement model with a patch detection perceptor (denoted as OursPatch), and enhancement model with an object-focused detection perceptor (denoted as OursOF).

Fig. 6 and Supplementary D Fig. 3 present the qualitative comparison of the enhancement models with different detection perceptor settings on ChinaMM and MultiviewUnderwater, respectively. We observe that without a detection perceptor, the enhanced results of OursWDP on the two datasets still contain artifacts and haze-effects. OursPatch removes artifacts and restores the details of the image patches such as color tones, visibility, and saturation on the two datasets. This is because the patch detection perceptor trained on the high quality in-air images can properly learn potential attributes of high visual quality. These potential attributes, in the form of gradients, help restore the details of the image patches. We notice that

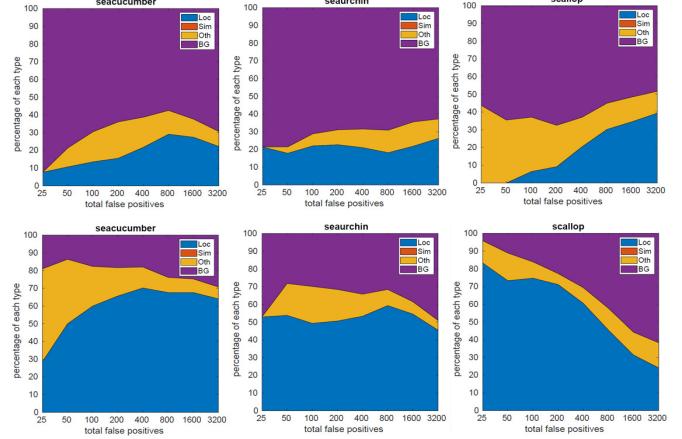


Fig. 7. The distribution of the top-ranked false positive measures for images of ChinaMM. The false positive measures include localisation errors (Loc), confusion with similar categories (Sim), with others (Oth), or with background (BG). The top row shows the results of OursPatch and the bottom row shows the results of OursOF.

TABLE III

QUANTITATIVE COMPARISON OF THE ENHANCEMENT MODELS WITH DIFFERENT DETECTION PERCEPTOR SETTINGS ON THE MULTIVIEWUNDERWATER AND CHINAMM DATASETS

Models	MultiviewUnderwater					ChinaMM		
	MSE	PSNR	SSIM	PCQI	mAP	UCIQE	UIQM	mAP
OursWDP	0.7153	20.5488	0.5632	0.5677	77.5	27.3206	3.8196	76.5
OursPatch	0.2453	33.3680	0.9374	0.8441	79.9	32.5976	4.8720	79.3
OursOF	0.4683	26.1421	0.6364	0.6741	86.7	28.3269	4.2194	83.9

the detection perceptor is only trained on the in-air images of MultiView, and the object categories of MultiView are different from those of ChinaMM, however, the detection perceptor helps the enhancement model to improve the quality of the images of ChinaMM. This indicates that a well-trained detection perceptor not only encodes category-dependent attributes but also category-agnostic attributes such as high quality edges, textures and colors. Compared to OursPatch, OursOF seems to generate sharp objects with high contrast between the objects and the background. We also reveal the detection results after having applied different enhancement models on ChinaMM in Supplementary D Fig. 4, from which we can see the detector trained on the enhanced results of OursOF achieves the best detection results.

Table III reports the quantitative results of the enhancement models with different detection perceptor settings on MultiviewUnderwater and ChinaMM. On MultiviewUnderwater, OursPatch achieves the best full-reference scores, and the corresponding deep detector obtains almost the same mAP as the one trained on the ground-truth in-air images of MultiView. This quantitative performance attributes to its enhanced results similar to the ground-truth in-air images. The deep detector trained on the enhanced results of OursOF achieves the best mAP. We have similar experimental results on ChinaMM, where OursPatch achieves the best UCIQE score (32.5976) and UIQM score (4.8720), whilst OursOF achieves the best mAP (83.9). We believe that the best detection accuracy is due to the reduction of the disturbing background. To verify this assumption, we use the detection tool of [44] to analyse the false positives of the two detectors trained on the results of



Fig. 8. Qualitative comparison of different UIS algorithms on the MultiView dataset. From left to right are raw in-air images of MultiView, results of Physical [8], WaterGAN [3], CycleGAN [15], OursHybrid, and real underwater images from ChinaMM as the references. Best viewed in the digital form.

OursPatch and OursOF. Fig. 7 and Supplementary F show the distribution of the top-ranked false positive measures for each category on the testing sets of ChinaMM and MultiviewUnderwater. The former detector cannot well distinguish the objects with complex background while the latter one largely reduces the background errors. The qualitative and quantitative comparisons of the enhancement models with different detection perceptor settings on the paired OUC dataset can be found in Supplementary E.

B. Comparison With State-of-the-Art Methods

In this subsection, we first compare our hybrid synthesis model with three state-of-the-art UIS algorithms. Then, we compare our two detection based perceptual enhancement models with eleven state-of-the-art UIE algorithms.

Comparison with state-of-the-art UIS methods. We compare our hybrid synthesis model (denoted as OursHybrid) with three state-of-the-art UIS algorithms, including physical model based UIS method (denoted as Physical) [8], CycleGAN [15] and WaterGAN [3]. Physical [8] applied the physical underwater image formation model and 10 groups of pre-defined parameters to synthesise 10 Jerlov type underwater images from the RGB-D in-air images, and the synthetic dataset contains 10 types of underwater images with various color distortions and haze-effects. We choose the images most similar to the ChinaMM-style underwater images in the qualitative comparison. WaterGAN and CycleGAN are deep learning based UIS methods, we tune their parameter settings to generate satisfactory results.

Fig. 8 shows the qualitative comparison of different UIS algorithms on MultiView. It is evident that the resultant images of our proposed hybrid synthesis model are very similar to the reference images of ChinaMM in terms of diversity, color casts and haze-effects. In contrast, the results of WaterGAN suffer from insufficient haze-effects due to the lack of light scattering prior, even though they apply a shallow convolutional network to simulating the light scattering process. However, in practice, without referring to the optical property of light scattering, the ability of CNN to simulate haze-effects is limited. In addition, the results of WaterGAN suffer from unrealistic color distortion even though it has used the light

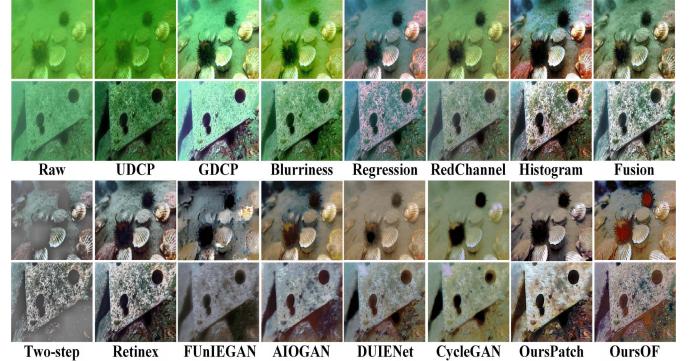


Fig. 9. Qualitative comparison of different UIE algorithms on the ChinaMM dataset. From left to right are raw underwater images, results of UDCP [22], GDPC [23], Blurriness [9], Regression [29], RedChannel [24], Histogram [30], Fusion [6], Two-step [19], Retinex [7], FUnIEGAN [2] and AIOGAN [48], DUIENet [31], CycleGAN [15], OursPatch and OursOF. The top raw is greenish and the bottom one is bluish.

absorption prior. This is mainly because the weakly supervised WaterGAN is trained with only an adversarial loss on the unpaired images, which cannot provide sufficient supervision information to simulate color distortion. Apart from the adversarial loss, CycleGAN adds a pixel-wise cycle-consistency loss as an additional constraint to supervising the training, and the cycle-consistency loss to minimise the discrepancy between the generated and the ground-truth images in the pixel level that leads to more realistic color distortion. Nevertheless, evident artifacts without haze-effects appear in the results of CycleGAN because the CNN structure has the limited capability to simulate haze-effects and remove artifacts. It is also worth noting that WaterGAN and CycleGAN generate underwater images in a monotonous style due to the model collapse problem, restricting their generalisation capability to yield diverse real-world underwater images. By integrating the physical prior into CNN, OursHybrid can generate images with diverse properties. Physical is able to generate artifact-free results but cannot simulate realistic color distortions as it uses the fixed parameters defined by the Jerlov water type. The qualitative and quantitative comparison of different UIS methods on OUC can be found in Supplementary G.

Comparison with state-of-the-art UIE methods. We compare our two detection perceptual enhancement models with eleven state-of-the-art UIE algorithms, including six physical model based methods (i.e., UDCP [22], GDPC [23], Blurriness [9], Regression [29], RedChannel [24] and Histogram [30]), three model-free methods (i.e., Fusion [6], Two-step [19], and Retinex [7]), and two deep learning based methods (i.e., DUIENet [31] and CycleGAN [15]) on three datasets (i.e., synthetic MultiviewUnderwater, real-world ChinaMM and OUC). To investigate the performance of deep learning based UIE algorithms on the real-world datasets, we add another two deep learning based UIE algorithms (i.e., FUnIEGAN [2] and AIOGAN [48]) as comparison methods on real-world dataset ChinaMM.

Fig. 9 and Supplementary H Fig. 8 show the qualitative comparison of different UIE algorithms on ChinaMM and MultiviewUnderwater, respectively. For the two datasets, most

TABLE IV
NON-REFERENCE IMAGE QUALITY AND DETECTION ACCURACY EVALUATIONS ON THE CHINAMM DATASET

Metrics	Raw	UDCP	GDCP	Blurriness	Regression	RedChannel	Histogram	Fusion	Twostep	Retinex	FUNIEGAN	AIOGAN	DUIENet	CycleGAN	OursPatch	OursOF
UCIQE	21.3083	28.6184	33.6328	30.5865	29.3877	30.8712	33.3443	31.7698	15.1238	28.447	30.4528	30.8534	31.5588	30.7922	32.5976	32.0967
UIQM	1.4410	3.0184	2.6468	3.6984	3.7616	3.3028	4.6728	4.0696	2.6728	4.7306	3.6066	3.3731	2.7021	3.7036	4.8720	4.4621
UICM	-80.1429	-56.7266	-53.7373	-58.1656	-21.9885	-31.2248	0.4558	-22.2500	-4.3928	-0.4811	-34.2738	-40.8772	-39.4952	6.9885	13.9967	12.4050
UISM	6.7799	6.7033	6.7830	6.7155	6.7060	6.6955	6.7210	6.6525	5.7978	6.6751	6.7427	6.6010	6.6242	6.5753	6.7326	6.5311
UICONM	0.4669	0.7380	0.6039	0.9385	0.6716	0.6170	0.7482	0.7643	0.3033	0.7755	0.7221	0.7206	0.5201	0.4376	0.6962	0.6107
mAP	68.6	71.6	72.7	77.3	71.6	73.7	76.8	75.5	58.6	78.8	73.6	72.8	71.6	67.8	79.3	83.9



Fig. 10. Qualitative comparison of different UIE algorithms on the Berman dataset. From left to right are raw underwater images, results of Contrast [39], Drews *et al.* [22], Peng and Cosman [9], Ancuti *et al.* [45], Ancuti *et al.* [46], Emberton *et al.* [47], Ancuti *et al.* [20], Berman *et al.* [39] and OursPatch.

of the physical model based UIE algorithms cannot deal with severe color distortions. Among them, Histogram performs relative better for color distortions which benefits from the histogram prior that it uses. But it generates over-saturation and excessive contrast in some image regions. UDCP and Blurriness even aggravate the color distortion due to the limitations of the priors used in these two methods. Regression tends to introduce bluish color casts on account of its inaccurate color correction algorithm, and GDCP over enhances the brightness that results in the loss of image details. RedChannel improves little on color distortion. Among the model-free algorithms, Retinex can effectively remove color distortion and produce more natural scenes while Fusion improves little on color distortion. Twostep over-enhances the contrast of the underwater images and generates unnatural results. The deep learning based methods FUNIEGAN, AIOGAN, DUIENet and CycleGAN can deal with color casts, however, both of them still leave evident haze on the resultant images. In terms of haze removal, most of the physical model based methods are able to remove haze-effects to some extent, benefiting from the use of light scattering priors. Among the model-free methods, Retinex and Fusion effectively remove the haze-effects on the underwater images while Twostep contributes little towards haze removal. Among all the UIE methods, OursPatch achieves the best qualitative performance in terms of color tones, visibility, saturation and contrast. The qualitative comparison of different UIE algorithms on OUC can be found in Supplementary H Fig. 10.

In addition to qualitative evaluations, we also report the quantitative results of different UIE algorithms. The best values are marked in bold. The quantitative results on the testing set of MultiviewUnderwater and OUC are presented

in Supplementary I, from which we can see that OursPatch achieves the lowest MSE score and the highest scores of PSNR, SSIM and PCQI. A higher PSNR score and a lower MSE score denote that the result is closer to the reference image in terms of image content, while a higher SSIM score denotes that the result is more similar to the reference image in terms of image structure and texture. Table IV gives the quantitative scores of non-reference metrics, i.e., UIQM, UCIQE, and three components of UIQM (UICM, UISM and UIConM). The following algorithms performs best over one single metric: GDCP achieves the best UCIQE score, and Blurriness achieves the best UIConM score. However, the results of both suffer from serious color casts as shown in Fig. 9. There exists certain discrepancy between the qualitative images and the quantitative scores in some cases which has also been verified in [1]. For mAP, the Precision/Recall curves in Fig. 11 and Supplementary H Fig. 11 show that OursOF (black curve) performs best across all the categories on ChainMM and MultiviewUnderwater, suggesting that the interaction between the enhancement model and the detection perceptor brings significant performance improvement for the following deep detector.

To demonstrate the generalisation ability of our proposed method, we also compare our proposed method with several representative UIE algorithms (Contrast [39], Drews *et al.* [22], Peng and Cosman [9], Ancuti *et al.* [45], Ancuti *et al.* [46], Emberton *et al.* [47], Ancuti *et al.* [20], and Berman *et al.* [39]) provided on the webpage of the Berman dataset (stated above). The Berman dataset only contains 114 images, which are not sufficient to train OursPatch. We conduct data augmentation by randomly cropping 114 high resolution images ($5,474 \times 4,653$ pixels) into 2,280 patches (512×512 pixels) and use the 2,280 patches to train OursPatch. Supplementary H Fig. 9 shows that OursPatch trained on 114 images cannot generate satisfactory results, and data augmentation helps greatly improve the visual quality of the enhanced results. The Berman dataset does not provide high quality reference images and bounding box annotations, so we only present qualitative comparison of OursPatch and the contrast UIE algorithms in Fig. 10. For fair comparison, we directly download the enhanced results of the contrast UIE algorithms from the webpage above. We observe that OursPatch performs much better in restoring the colors of the color charts and sands.

C. The Influences of UIE Algorithms on the Detection Task

Previous works seem to suggest that UIE algorithms will bring improvements of image quality, which further boosts

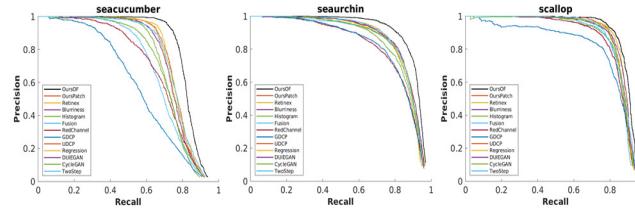


Fig. 11. Precision/Recall curves of deep detectors trained on the results of different UIE methods on ChinaMM.

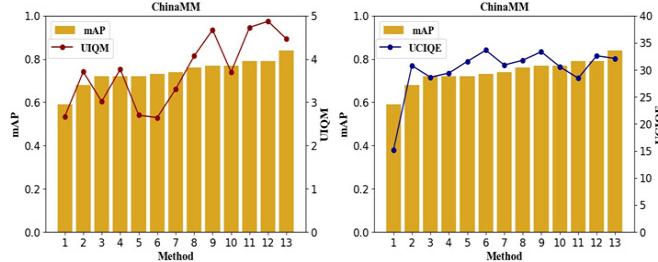


Fig. 12. Image quality evaluation metrics and mAP on ChinaMM. The histogram represents the mAP and the polyline represents different image quality evaluation metrics. Numbers 1 to 13 refer to thirteen UIE algorithms ordered according to increasing mAP values.

the performance of the following high-level detection tasks. We conduct the following analysis to investigate whether or not UIE brings improvement of detection accuracy. We compare the quantitative scores of the raw underwater and the enhanced underwater images by different UIE algorithms on the three underwater datasets. It is observed that not all the UIE algorithms increase the quantitative scores of the raw underwater images. Deep learning based methods can stably improve image quality on both full-reference and non-reference metrics, while other methods only work for some special image quality evaluation metrics.

Fig. 12 and Supplementary J Fig. 12 illustrate mAP and image quality evaluation metrics on ChinaMM and Multi-viewUnderwater datasets, from which we investigate how the detection accuracy is related to different image quality evaluation metrics. There is no strong correlation between the mAP and the image quality evaluation metrics on the two datasets. For MultiViewUnderwater, Regression receives the best MSE and PCQI scores among the six physical model based methods, however, its detection accuracy is the worst among these methods. For ChinaMM, CycleGAN can greatly improve the UCIQE and UIQM scores, but its mAP (67.8%) is even lower than that of the raw underwater images (68.4%). On the OUC dataset, both GDCP and UDCP decrease the MSE and PSNR scores of the raw underwater images, but their mAP are even higher than these of the high quality reference images (86.6% mAP). Therefore, certain discrepancies exist between the image quality evaluation metrics and the detection accuracy. mAP may be biased to some image quality metrics such as high UICONM and UICM. For example, Bluriness, Histogram, Fusion and Retinex have top four UICONM scores, and achieve the top four detection accuracy among the non-deep learning based methods. OursPatch and OursOF receive significantly higher UICM scores and rank top two in terms of mAP. The detection task tends to favour the results

with high contrasts (high UICONM) between the objects and the background, or that with the over-enhanced objects. One possible explanation lies in that high contrast suppresses the complicated background while bright color protrudes the objects. For illustration, we show the object detection results of two underwater images from MultiviewUnderwater and ChinaMM in Supplementary J.

We believe there is a gap between the image quality evaluation metrics for the low-level enhancement task and the accuracy metric for the high-level detection task. Underwater image enhancement task is usually evaluated using the image quality evaluation metrics. However, the objective of the low-level enhancement task typically differs from that of the high-level object detection task so that the enhancement algorithm can hardly recover features favoured by the high-level detection task. The interaction between the enhancement and detection tasks are important indeed.

VI. CONCLUSION

In this article, we have proposed two detection-perceptual enhancement models, i.e., patch detection-based perceptual enhancement and object-focused detection based perceptual enhancement models. With the help of two detection perceptors, our patch detection-based perceptual enhancement model can generate high quality in-air images with patch-level details, and our object-focused detection-based perceptual enhancement model can generate images which improves the detection accuracy of the following deep detectors. Moreover, to advance the generalisation ability of deep UIE algorithms, we have proposed a hybrid underwater image synthesis model to synthesise more realistic training images, and the enhancement model trained on them can learn more robust translation between the underwater and high quality in-air images, and generalise well on the real-world underwater scenes.

The generalisation ability of GAN based UIE algorithms highly depends on the diversity of the training data, and we are trying to combine Bayesian estimation and GAN in our future work to address the mode collapse problem. This combination helps us to explore multiple modes (i.e., in-air to multiple water type images).

REFERENCES

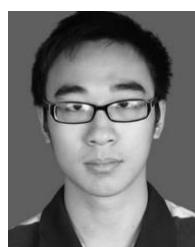
- [1] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Jan. 3, 2020, doi: 10.1109/TCSVT.2019.2963772.
- [2] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.
- [3] 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.
- [4] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 7159–7165.
- [5] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, pp. 21–37.
- [6] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 81–88.
- [7] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4572–4576.

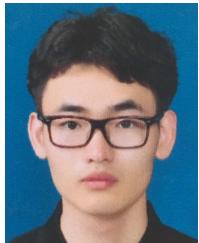
- [8] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107038.
- [9] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.
- [10] Y. Wang, J. Zhang, Y. Cao, and Z. Wang, "A deep CNN method for underwater image enhancement," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1382–1386.
- [11] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012.
- [12] N. G. Jerlov, *Marine Optics*, vol. 14. Amsterdam, The Netherlands: Elsevier, 1976.
- [13] H. Gupta and K. Mitra, "Unsupervised single image underwater depth estimation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 624–628.
- [14] T. Łuczynski, Y. Petillot, and S. Wang, "Seeing through water: From underwater image synthesis to generative adversarial networks based underwater single image enhancement," in *Proc. ICRA Workshop*, 2019.
- [15] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
- [16] M. Hou, R. Liu, X. Fan, and Z. Luo, "Joint residual learning for underwater image enhancement," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 4043–4047.
- [17] J. Y. Zhu *et al.*, "Toward multimodal image-to-image translation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 465–476.
- [18] H. You, Y. Cheng, T. Cheng, C. Li, and P. Zhou, "Bayesian CycleGAN via marginalizing latent sampling," 2018, *arXiv:1811.07465*. [Online]. Available: <http://arxiv.org/abs/1811.07465>
- [19] X. Fu, Z. Fan, M. Ling, Y. Huang, and X. Ding, "Two-step approach for single underwater image enhancement," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Nov. 2017, pp. 789–794.
- [20] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.
- [21] S. Zhang, T. Wang, J. Dong, and H. Yu, "Underwater image enhancement via extended multi-scale retinex," *Neurocomputing*, vol. 245, pp. 1–9, Jul. 2017.
- [22] 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.
- [23] Y. T. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, Mar. 2018.
- [24] 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.
- [25] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
- [26] D. Berman, T. Treibitz, and S. Avidan, "Diving into haze-lines: Color restoration of underwater images," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Sep. 2017, vol. 1, no. 2, pp. 1–12.
- [27] Y. Zhou, Q. Wu, K. Yan, L. Feng, and W. Xiang, "Underwater image restoration using color-line model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 3, pp. 907–911, Mar. 2019.
- [28] Y. Wang, H. Liu, and L.-P. Chau, "Single underwater image restoration using adaptive attenuation-curve prior," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 3, pp. 992–1002, Mar. 2018.
- [29] C. Li, J. Guo, C. Guo, R. Cong, and J. Gong, "A hybrid method for underwater image correction," *Pattern Recognit. Lett.*, vol. 94, pp. 62–67, Jul. 2017.
- [30] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.
- [31] C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.
- [32] X. He, H. Xu, X. Ji, and R. Xu, "Underwater image enhancement using stacked generative adversarial networks," in *Proc. Pacific Rim Conf. Multimedia*, Cham, Switzerland: Springer, Sep. 2018, pp. 514–524.
- [33] C. Guo *et al.*, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1780–1789.
- [34] X. Ye *et al.*, "Deep joint depth estimation and color correction from monocular underwater images based on unsupervised adaptation networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3995–4008, Nov. 2020.
- [35] M. Jian, Q. Qi, J. Dong, Y. Yin, W. Zhang, and K.-M. Lam, "The OUC-vision large-scale underwater image database," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 1297–1302.
- [36] C. Li, R. Cong, J. Hou, S. Zhang, Y. Qian, and S. Kwong, "Nested network with two-stream pyramid for salient object detection in optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9156–9166, Aug. 2019.
- [37] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view RGB-D object dataset," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 1817–1824.
- [38] L. Chen *et al.*, "A benchmark dataset for both underwater image enhancement and underwater object detection," 2020, *arXiv:2006.15789*. [Online]. Available: <http://arxiv.org/abs/2006.15789>
- [39] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 2, 2020, doi: [10.1109/TPAMI.2020.2977624](https://doi.org/10.1109/TPAMI.2020.2977624).
- [40] S. Wang, K. Ma, H. Yeganeh, Z. Wang, and W. Lin, "A patch-structure representation method for quality assessment of contrast changed images," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2387–2390, Dec. 2015.
- [41] 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.
- [42] 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.
- [43] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic gradient descent," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–15.
- [44] D. Hoiem, Y. Chodpathumwan, and Q. Dai, "Diagnosing error in object detectors," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, Oct. 2012, pp. 340–353.
- [45] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer, R. Garcia, and A. C. Bovik, "Multi-scale underwater descattering," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 4202–4207.
- [46] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, L. Neumann, and R. Garcia, "Color transfer for underwater dehazing and depth estimation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 695–699.
- [47] S. Emberton, L. Chittka, and A. Cavallaro, "Underwater image and video dehazing with pure haze region segmentation," *Comput. Vis. Image Understand.*, vol. 168, pp. 145–156, Mar. 2018.
- [48] P. M. Uplavikar, Z. Wu, and Z. Wang, "All-in-one underwater image enhancement using domain-adversarial learning," in *Proc. CVPR Workshops*, May 2019, pp. 1–8.

Long Chen is currently pursuing the Ph.D. degree with the School of Informatics, University of Leicester, U.K. His research interests include computer vision and machine learning.



Zheheng Jiang is currently pursuing the Ph.D. degree with the School of Informatics, University of Leicester, U.K.





Lei Tong is currently pursuing the M.Phil. degree with the School of Informatics, University of Leicester, U.K.



Qianni Zhang is currently a Senior Lecturer with the School of Electronic Engineering and Computer Science, Queen Mary University of London. Her research interest includes medical image analysis and understanding.



Zhihua Liu is currently pursuing the M.Phil. degree with the School of Informatics, University of Leicester, U.K. He is a Researcher with the Biomedical Image Processing Laboratory (BIPL).



Junyu Dong (Member, IEEE) is currently a Professor and the Head of the Department of Computer Science and Technology. His research interests include machine learning, big data, computer vision, and underwater image processing.



Aite Zhao is currently a Lecturer with the College of Computer Science and Technology, Qingdao University, China.



Huiyu Zhou is currently a Professor with the School of Informatics, University of Leicester, U.K. His research work has been or is being supported by the U.K. EPSRC, MRC, EU, Royal Society, Leverhulme Trust, Puffin Trust, Invest NI, and industry.