Signal Processing: Image Communication 89 (2020) 115978

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Diving deeper into underwater image enhancement: A survey

Saeed Anwar [a,b], Chongyi Li [c,*]

[a] Data61, CSIRO, ACT 2601, Australia
[b] College of Engineering and Computer Science, Australian National University, Canberra, ACT 2600 Australia
[c] School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore

ARTICLE INFO

Keywords:
Underwater image enhancement
Deep learning
Convolutional neural networks (CNNs)
Generative adversarial networks (GANs)
Underwater datasets
Underwater evaluation metrics
Survey

ABSTRACT

The powerful representation capacity of deep learning has made it inevitable for the underwater image enhancement community to employ its potential. The exploration of deep underwater image enhancement networks is increasing over time; hence, a comprehensive survey is the need of the hour. In this paper, our main aim is two-fold, (1): to provide a comprehensive and in-depth survey of the deep learning-based underwater image enhancement, which covers various perspectives ranging from algorithms to open issues, and (2): to conduct a qualitative and quantitative comparison of the deep algorithms on diverse datasets to serve as a benchmark, which has been barely explored before.

We first introduce the underwater image formation models, which are the base of training data synthesis and design of deep networks, and also helpful for understanding the process of underwater image degradation. Then, we review deep underwater image enhancement algorithms, and a glimpse of some of the aspects of the current networks is presented, including architecture, parameters, training data, loss function, and training configurations. We also summarize the evaluation metrics and underwater image datasets. Following that, a systematically experimental comparison is carried out to analyze the robustness and effectiveness of deep algorithms. Meanwhile, we point out the shortcomings of current benchmark datasets and evaluation metrics. Finally, we discuss several unsolved open issues and suggest possible research directions. We hope that all efforts done in this paper might serve as a comprehensive reference for future research and call for the development of deep learning-based underwater image enhancement.

1. Introduction

Nowadays, developing, exploring, and protecting the ocean's resources have become the center of interest in the international community. Clear underwater images and videos can provide valuable information of the underwater world, which are essential for numerous engineering and research tasks such as underwater archaeology, underwater surveillance, *etc.* However, the raw underwater images and videos usually suffer from the effects of quality degradation, especially the impact of backscatter in far distances. The issues of quality degradation are mainly introduced by light selective absorption and scattering in water, as well as the use of artificial light in deep water. The degraded underwater images have low contrast and brightness, color deviations, blurry details, and uneven bright speck, which limit their applications in practical scenarios. As an indispensable processing step, underwater image enhancement methods ranging from the conventional techniques (*e.g.*, physical model-based methods, and histogram equalization-based methods) to the data-driven techniques (*e.g.*, convolutional neural networks, and generative adversarial networks) have been attracting increasing attention.

The past few decades have seen the rapid development of deep learning techniques, which have been extensively applied in various computer vision and image processing tasks [1]. Deep learning has significantly improved the performance of high-level vision tasks such as object detection [2] and object recognition [3]. Moreover, the low-level vision tasks, such as image super-resolution [4] and image denoising [5], also benefit from the advantages of deep networks and deliver state-of-the-art performance. Unfortunately, we are unable to observe the appealing performance of deep learning-based underwater image enhancement, although many researchers have attempted to utilize the deep learning techniques to enhance underwater image enhancement.

In this paper, we mainly focus on deep learning methods, which enhance and restore underwater images. Through this exposition, we provide the latest development and comparison of current deep underwater image restoration and enhancement algorithms. Furthermore, we summarize the current issues, analyze the potential reasons, and suggest future research directions. The main contributions of this paper are two-fold:

* Corresponding author.
E-mail address: lichongyi25@gmail.com (C. Li).

https://doi.org/10.1016/j.image.2020.115978
Received 8 January 2020; Received in revised form 22 June 2020; Accepted 8 August 2020
Available online 20 August 2020

- We summarize the deep learning-based underwater image enhancement algorithms, including network architectures, network parameters, training data, loss function, and training configurations. It provides the first comprehensive and in-depth survey for deep learning-based underwater image enhancement to the best of our knowledge, which helps develop more robust and effective deep algorithms.
- We conduct systematic experiments on diverse datasets to qualitatively and quantitatively compare the deep learning-based underwater image enhancement algorithms. Our evaluation and analysis demonstrate the performance of current deep algorithms, point out their limitations, and indicates the bias of existing benchmark datasets and evaluation metrics. As a consequence, we give potential insights for future research directions in this field of study.

The rest of the paper is organized as follows. Section 2 introduces the background of underwater image enhancement and restoration, mainly focusing on the imaging models. Section 3 presents the existing deep learning-based underwater image enhancement algorithms and insights into the network. Section 4 gives the experimental quantitative and qualitative results and analysis, evaluation metrics, and datasets. Section 5 suggests future research directions, and Section 6 concludes this paper.

2. Background

In this section, we mainly introduce the commonly-used physical models for underwater image enhancement, including atmospheric scattering model, simplified underwater image formation model, and revised underwater image formation model. These models are the base of training data synthesis and design of deep networks and also helpful for understanding the process of underwater image degradation.

2.1. Atmospheric scattering model

For an image captured in a scattering medium, only a part of the reflected light from the scene reaches the imaging sensor due to the absorption and scattering effects, typically for hazy image formation. Since underwater images usually have a hazy appearance (similar to the hazy image), the atmospheric scattering model [6] is traditionally used to describe the degradation of the underwater image. The atmospheric scattering model [6] can be characterized as:

$$\mathbf{U}(x) = \mathbf{I}(x)T(x) + B(1 - T(x)), \tag{1}$$

where x denotes the pixel coordinates, $\mathbf{U}(x)$ is the observed image, $\mathbf{I}(x)$ is the haze-free latent image, B is the global atmospheric light which indicates the intensity of ambient light, and $T(x) \in [0, 1]$ is the transmission which represents the percentage of the scene radiance reaching the camera. When the haze is homogeneous, $T(x)$ can be further expressed in an exponential decay term as:

$$T(x) = \exp(-\beta d(x)), \tag{2}$$

where β is the atmospheric attenuation coefficient and $d(x)$ is the distance from the scene to the camera. In this atmospheric scattering model, the scattering is non-selective, and attenuation is independent of wavelengths.

2.2. Simplified model

In fact, there is a significant difference between the atmospheric scattering model and the real-world underwater image formation model. The real-world underwater imaging is far more complicated due to the optical properties of selective attenuation in water. Thus, in the early stage, most physical model-based methods followed a simplified underwater image formulation model provided by [7]. We denote the captured underwater image by $\mathbf{U}_\lambda(x)$, the clear latent image (also known as scene radiance) as $\mathbf{I}_\lambda(x)$, and the homogeneous global background light as B_λ, then the degradation model is given as:

$$\mathbf{U}_\lambda(x) = \mathbf{I}_\lambda(x) \cdot T_\lambda(x) + B_\lambda \cdot \big(1 - T_\lambda(x)\big), \tag{3}$$

where λ presents the wavelength of the RGB channels, and x is a point in the underwater scene. Similarly, $T_\lambda(x)$ is the medium energy ratio, which is the percentage of the scene radiance captured by the camera (the amount of radiance reflected from the point x). This phenomenon causes contrast degradation and color casts. To be precise, $T_\lambda(x)$ is a function of λ and the distance $d(x)$ to the camera from the scene point x, expressed as:

$$T_\lambda(x) = 10^{-\beta_\lambda d(x)} = \frac{E_\lambda\big(x, d(x)\big)}{E_\lambda(x, 0)} = N_\lambda\big(d(x)\big), \tag{4}$$

where β_λ is the medium attenuation coefficient, which is dependent on the wavelength. Furthermore, $E_\lambda(x, 0)$ is the energy of light from the submerged scene before it passes through the transmission medium from a distance $d(x)$ while $E_\lambda\big(x, d(x)\big)$ is the strength of light after absorption by the transmission medium. Moreover, N_λ is the normalized residual energy, which is the ratio of residual energy to the initial energy per unit of distance and is dependent on the wavelength of light. For example, the bluish tone of the most underwater images is due to the fast attenuation of the red wavelength in open water as it possesses a longer wavelength than blue and green ones.

2.3. Revised model

Recent research found that the commonly-used atmospheric scattering model and simplified underwater image formation model ignored some critical components in the process of real-world underwater imaging [8]. Specifically, the attenuation coefficient for backscatter strongly depends on the veiling light. Moreover, unlike the absorption in the atmosphere, the absorption in water should not be neglected. Most importantly, the attenuation coefficients for the direct signal and the scattering signal are different.

Based on the findings mentioned above, Akkaynak & Treibitz [8] proposed a revised underwater image formation model which can be expressed as:

$$\mathbf{U}_\lambda(x) = \mathbf{I}_\lambda(x) e^{-\beta_\lambda^D(\mathbf{v}_D)\cdot z} + B_\lambda^\infty \big(1 - e^{-\beta_\lambda^B(\mathbf{v}_B)\cdot z}\big), \tag{5}$$

where B_λ^∞ is the veiling light, β_λ is the beam attenuation coefficient, D is the direct transmitted light, B is the backscattered light, the vectors $\mathbf{v}_{d(x)}$ and $\mathbf{v}_{b(x)}$ represent the coefficient dependencies. To be more specific, $\mathbf{v}_{d(x)} = \{\mathrm{z}, \rho, \mathrm{E}, \mathrm{S}_\lambda, \beta\}$ and $\mathbf{v}_{b(x)} = \{\mathrm{E}, \mathrm{S}_\lambda, \mathrm{b}, \beta\}$, where z is the range along LOS, ρ is the reflectance, E is the irradiance, S_λ is the sensor spectral response, and b is the beam scattering coefficient. Like the simplified model, $\mathbf{U}_\lambda(x)$ is the observed underwater image, and $\mathbf{I}_\lambda(x)$ is the latent clear underwater image. More details can be found in [8]. Moreover, the coefficient associated with the backscatter varies with the sensor, ambient illumination, and water type. Generally, the coefficient of backscatter is different from the coefficient associated with the direct signal.

In summary, the atmospheric scattering model is suitable for underwater scenarios only in some cases, such as shallow water with low backscatter. Compared to the atmospheric scattering model, the simplified underwater image formation model considers the selective attenuation of different wavelengths, extending the generalization of this model. However, the simplified underwater image formation model assumes the attenuation coefficients are only properties of the water, which is inaccurate because the attenuation coefficients vary with the sensor, ambient illumination, *etc.* Besides, the simplified model ignores the fact that the backscattered light has a different attenuation coefficient from the direct light. Thus, a physically accurate model (*i.e.*, revised underwater image formation model) is proposed, which further

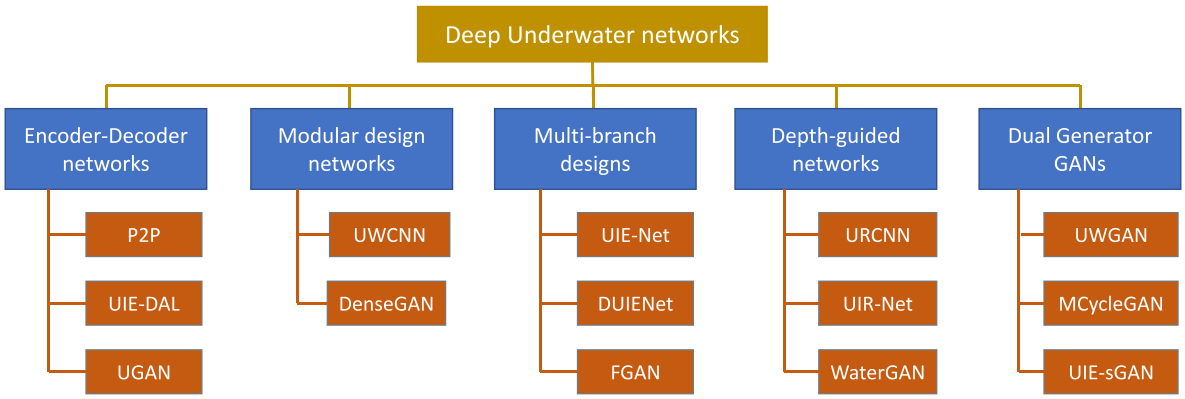

Fig. 1. Categorization of deep underwater networks: The organization of deep networks based on their essential aspects. We have added the references for each network for ease of search.

completes the model of underwater image formation. Nevertheless, such an accurate model has barely received much attention due to its complexity. Most of the deep learning-based underwater image enhancement algorithms still follow the atmospheric scattering model or simplified underwater image formation model to synthesize their training data and design their network architectures.

3. Deep underwater image enhancement algorithms

Deep underwater image enhancement algorithms can ideally be divided into two main categories *i.e.*, CNN-based and GAN-based algorithms. The goal of the CNN algorithms is to be faithful to the original underwater image while the GAN-based algorithms aim to improve the perceptual quality of the images. However, this classification is very naive; therefore, we categorize the networks based on their architectural differences. In Fig. 1, the categorization of deep underwater networks is presented, and in the following sections, we list and provide details for each method into different categories based on essential aspects.

3.1. Encoder–decoder models

The following models benefit from the well-known encoder–decoder architecture to advance the underwater image enhancement research.

3.1.1. P2P network

Recently, Sun et al. [9] suggested the use of pixel-to-pixel (P2P) networks to enhance underwater images. The proposed model is a "symmetric" encoder and decoder network similar to REDNet [10]. The encoder part is composed of three convolutional layers, while the decoder is made from three deconvolutional layers. ReLU follows each network element except the last one.

This model is trained on 3359 images collected from the real-world environment. To simulate the underwater images, the authors pour milk of 30, 50, and 70 ml into 1 m^3 of water to produce low, medium, and high-level degradation, respectively. Finally, out of these, 10,000 images are selected for training and another 2000 images for testing. Moreover, the input to the network is a cropped patch of 66×66. The loss function is ℓ_2 minimized via SGD [11] with an initial learning rate of 10^{-7}.

3.1.2. UIE-DAL

Underwater Image Enhancement using Domain Adversarial Learning (UIE-DAL) [12] aims to learn the agnostic model where it can enhance any underwater-type image. The backbone architecture of the UIE-DAL [12] is the well-known encoder–decoder UNET [13]. The novelty of this work is the incorporation of a neural network classifier, named nuisance classifier, which classifies the latent vector extracted from the encoder.

The authors claim the model to be agnostic, considering that the nuisance classifier is not aware of the underwater type as it receives the latent vector from the encoder, which is agnostic to the features of the underwater types. The UIE-DAL [12] combines three losses *i.e.* ℓ_2, nuisance loss, and adversarial loss. The training is achieved in two steps. First, the only encoder–decoder structure is trained, then a nuisance classier is incorporated in the network.

3.1.3. UGAN

Recently, the Underwater Generative adversarial network, abbreviated as UGAN [14], is proposed to improve the underwater image quality. For discriminator, UGAN chooses WGAN-GP (Wasserstein GAN with gradient penalty) [15] to enforce the soft constraints via the Lipschitz on the gradients norms instead of clipping the gradients in a specific range. The discriminator is fully convolutional and is similar to [16] except batch normalization [17] is not applied to the weights of convolutional layers. Furthermore, the discriminator outputs 32×32 feature matrix similar to PatchGAN [18]. The generator is motivated by CycleGAN [19], comparable to the encoder–decoder network of UNET [13]. The encoder of UGAN [14] is composed of convolutional layers having filter sizes of 4×4 with a stride of two followed by batch normalization [17] and leaky ReLU (slope of 0.2). Similarly, the decoder portion consists of deconvolutional layers followed by ReLU [20] only except the last layer, where TanH is used to restrict distribution between -1 and 1.

The evaluation and training are achieved on the subsets of ImageNet [21]. Moreover, two types of underwater images are collected *i.e.* one set of 6143 images without distortion and another set of 1817 images with distortion. Adam [22] is used as an optimizer with a fixed learning rate of 10^{-4} for 100 epochs. The input to the network is $256 \times 256 \times 3$, while loss is a linear combination of ℓ_1 and Earth-Mover or Wasserstein-1 distance.

3.2. Modular designs

Modular or block designs employ the repetition of the same structure, commonly known as a "block" or a "module", to learn the features. These designs are very successful in computer vision and machine learning tasks. We provide an example of modular or block-based designs for underwater networks below.

3.2.1. UWCNN

To deal with the low contrast and distorted color of the degraded underwater images, Anwar et al. [23,24] proposed a CNN underwater

image enhancement model, called UWCNN.[1] The UWCNN is an end-to-end model trained by the synthetic underwater image datasets, which includes three densely connected building blocks. Furthermore, each basic building block consists of three densely connected convolutional layers. After the three chained building blocks, a convolutional layer is used to learn the difference (residual) between the degraded underwater image and its clean counterpart.

To train the UWCNN [23] model, the authors use the attenuation coefficients of different water types to synthesize various underwater image datasets according to the underwater image formation model resulting in ten types of underwater image datasets which are synthesized by using the RGB-D NYU-v2 dataset [25]. These underwater image datasets simulate the open ocean water types and coastal water types ranging from the clearest to the most turbid. Finally, the authors train ten UWCNN models for the ten types of underwater images. The parameters of the UWCNN model are learned by joint optimizing the ℓ_2 and SSIM loss functions. In the entire UWCNN, the kernel sizes and filter numbers are fixed, *i.e.*, 3×3, and 16, respectively. The learning rate is set to 2×10^{-4}, and ADAM [22] is used for optimization in the TensorFlow framework.

3.2.2. DenseGAN

To enhance the underwater images, Guo et al. [26] introduced a multiscale dense block (MSDB) algorithm, namely, DenseGAN[2] which employs dense connections, residual learning, and multiscale network for underwater image enhancement.

The generator at the start is composed of two convolutional, batch normalization (BN), leaky ReLU (LReLU) sequence then two MSDB blocks followed by the sequence Deconvolutional–BN–LReLU, while at the end there is a deconvolutional layer and a TanH layer. The network architecture of the DenseGAN generator and MSDB are shown in Fig. 2. In each MSDB block, the input features are passed through two branches, where each branch has kernels with different dilations. The features from each branch are concatenated half-way through the MSDB block and fed again into the respective branches. At the end of the MSDB block, the features are concatenated again and passed through a 1×1 convolutional layer. The discriminator network is similar to PatchGAN [18]; however, it is composed of five layers of spectral normalization [28]. Except for the first and last layer, the discriminator is composed of sequences of convolutional–BN–LReLU.

The first two layers of the generator have 7×7 and 3×3 filter size with 64 and 128 feature maps, respectively. The last deconvolution layer outputs the same number of channels as the input. The TanH layer keeps the distribution between −1 and 1. Moreover, slope of the leaky ReLU is fixed at 0.2, and the network is trained via the TensorFlow framework using a learning rate of 10^{-3} with a patch size of $256 \times 256 \times 3$. The ADAM [22] is used for optimization, and batch size is set to 32. The losses employed are GAN loss, ℓ_1, and gradient loss.

3.3. Multi-branch designs

The multiple branch designs aim to either learn different features of the same input at different levels or exploit distinct inputs at separate branches. The following are examples of such networks.

3.3.1. UIE-Net

Wang et al. [29] presented a deep CNN method for the enhancement of underwater images, namely, UIE-Net, which is composed of three subnetworks. The first subnet called sharing network (termed as S-Net) is composed of convolutional layers only. S-Net extracts features from the input image, which is then forwarded to the other two subnets (*i.e.* the branches of the network: the color correction network (CC-Net) and the haze removal network (HR-Net).) CC-Net and HR-Net output color corrected image, and transmission map, respectively. Both CC-Net and HR-Net have the same network structure consisting of four convolutional layers, followed by sigmoid activation. The only difference between CC-Net and HR-Net is the number of output channels *i.e.* three channels and one channel, respectively.

The S-Net has two convolutional layers and a consistent filter size of 5×5, while the CC-Net and HR-Net have four convolutional layers with filter sizes of $1 \times 1, 3 \times 3, 5 \times 5$ and 7×7 to capture contextual information. Fig. 2 shows the underlying network architecture of the UIE-Net. The inputs to the network are 32×32 image patches in the training procedure, and the network is trained on 2×10^5 image patches synthesized from 200 clear images collected from the internet. The initial learning rate is fixed at 5×10^{-3}, which is decreased by half after 5×0^3 until 2.5×10^5.

The loss employed for learning is ℓ_2. Moreover, the authors perform smoothing on the input patches to obtain desirable results. As the last step, the guided image filtering [30] is applied on the transmission map to remove artifacts, if any. It is also to be noted here that UIE-Net is one of the pioneering work in deep learning direction.

3.3.2. DUIENet

More recently, Li et al. [31] constructed a real-world underwater image enhancement dataset, including 950 underwater images, 890 of which have the corresponding reference images. These potential reference images are produced by 12 image enhancement methods, and the final references are selected by 50 volunteers via majority voting.

Inspired by the fusion-based underwater image enhancement method [32], Li et al. [31] proposed a gated fusion CNN trained by the constructed dataset for underwater image enhancement, called DUIENet.[3] First, three input versions are generated by sequentially applying White Balance, Histogram Equalization, and Gamma Correction algorithms to the raw input image. Then, the DUIENet learns three confidence maps, which determine the most essential features remaining in the final result. The DUIENet is a multi-scale FCNN consisting of 14 convolutional layers followed by ReLU except for the last layer (followed by Sigmoid). To reduce the color casts and artifacts introduced by the three pre-processing algorithms, three feature transformation units (FTUs) are used in the DUIENet [31]. The FTU includes three stacked multi-scale convolutional layers. The input of each FTU is the corresponding pre-processed underwater image, and its output is the transformed image. At last, the transformed three inputs are multiplied by the three learned confidence maps, and then the summation of the three products is the enhanced underwater image.

With the constructed dataset, the authors selected 800 pairs of images randomly to generate the training set. These images are resized to 112×11, and data augmentation is used to obtain seven additional versions of the original 800 pairs of training data. The rest 90 pairs of images are treated as the testing set. To reduce the artifacts induced by pixel-wise loss functions, the authors minimize the perceptual loss (layer relu5_4 of the pre-trained VGG19 network [33]).

[1] Code at https://github.com/saeed-anwar/UWCNN.

[2] The authors' term the model as UWGAN; however, Li et al. [27] proposed a model with the same name earlier. To avoid confusion, we call it DenseGAN due to its dense connections.

[3] Code at https://li-chongyi.github.io/proj_benchmark.html.

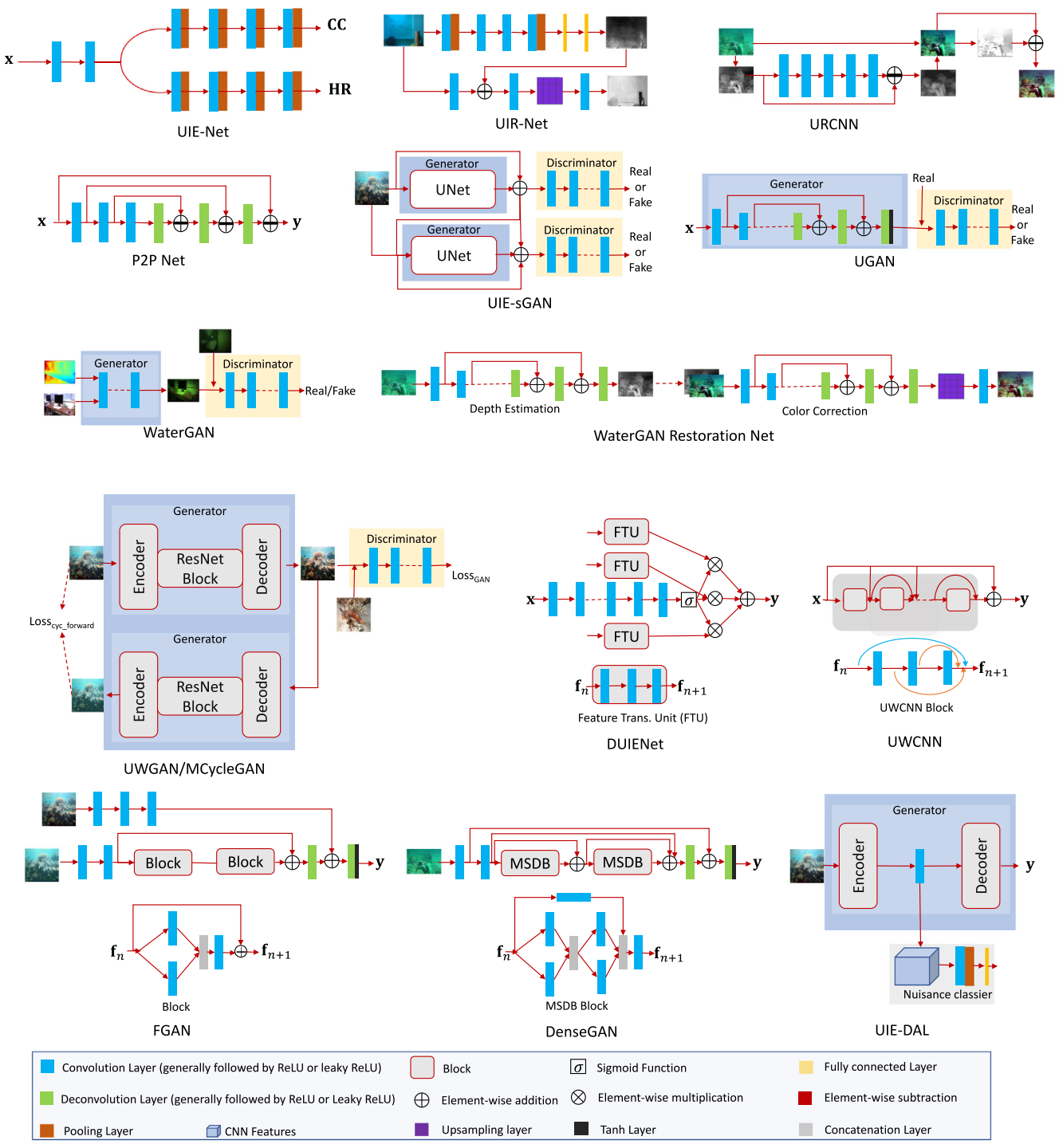

Fig. 2. Network architectures: A glimpse of network architectures used for underwater image enhancement using CNNs and GANs. Best viewed with zoom-in on a digital display.

3.3.3. FGAN

Fusion generative adversarial network, abbreviated as FGAN [34], takes multiple inputs and passes them through different branches in the same network. In the end, the features are summed before the loss of the generator. The architecture of FGAN [34] is similar to DenseGAN with slight modifications in the block's architecture. The generator with the fundamental block structure is shown in Fig. 2. The discriminator is composed of five convolutional layers employing spectral normalization [28]. The discriminator is similar to PatchGAN [18].

A batch-mode learning method with a batch size of 16 is applied. The RGB images of size 256 × 256 are used as inputs. Further, the learning rate is set to 10^{-3}. The loss function is a combination of relativistic GAN loss [35], adversarial loss, and ℓ_2 loss.

3.4. Depth-guided networks

Depth map or transmission map plays a vital role in restoring the underwater image, which is related to the degradation induced by scattering. Therefore, it is a natural choice to predict the depth map or transmission map of the underwater image to improve the performance of enhancement and restoration. We list the depth-guided networks next.

3.4.1. URCNN

Underwater residual convolutional neural network (URCNN) [36] is proposed by Hou et al. which aims to learn the transmission map. The URCNN, in the first, uses a convolutional layer followed by ReLU to extract features. The batch normalization and ReLU succeed the second Conv layer. This pattern is repeated until the reconstruction layer, where only the convolutional layer is employed to output the transmission map. A global skip connection is used to enforce residual learning. The output transmission map is used to refine the input image.

The network architecture of the URCNN is a modified version of VGG [33], and the input to the network is 180×180 transmission map instead of the original image. The underwater images are generated from randomly selected 1000 NYU dataset [25] images. Furthermore, a total of 1800 images are generated for training and 200 images for testing using random medium attenuation coefficient and background light. The initial learning rate is selected to be 10^{-1} and reduced to 10^{-4} for 60 epochs. The depth of the network is 25 layers, with each layer having 64 feature maps and a filter size of 3×3. Similar to [29], the loss used for learning is ℓ_2. Similar to [29], the loss used for learning is ℓ_2.

3.4.2. UIR-Net

Cao et al. [37] lately developed a deep network for underwater image restoration inspired by classical methods where the transmission map and the background light are estimated and computed independently. Consequently, two different network architectures were proposed *i.e.* the light network (BL-Net) and the transmission map network (TM-Net) while collectively, the network is called UIR-Net [37]. The background light network (BL-Net) is simple and consists of five layers. The initial three layers are convolutional with BN and pooling. The last two layers are fully connected ones. The output of this BL-Net is thresholded to constrain it in the range of [0,1]. The transmission map network (TM-Net) is more complicated and is based on [38], consisting of two subnets, *i.e.*, coarse-global subnet, and refine subnet. The coarse subnet is made of five convolutional layers, with the first two convolutional layers having pooling and batch normalization. The last layers of the coarse-global subnet are fully connected ones. The refined subnet has three convolutional layers and an upsampling layer that lies before the final convolutional layer. The output of this network is the depth map. Using depth maps, the transmission maps are computed. As a last preprocessing step, the guided filter [30] is applied to further refine the maps.

The loss for the BL-Net is Euclidean, while for the TM-Net is a scale-invariant minimum square error (MSE) adopted from Eigen et al. [38]. Similar to [29], UIR-Net [37] use NYU-v2 dataset [25] to generate 12,000 synthetic underwater images using a total of 29 different underwater ambient lights. The BL-Net is initialized randomly, while TM-Net utilizes the weights from VGG [33].

3.4.3. WaterGAN

WaterGAN [39], as the name indicates, is a generative adversarial network, which manipulates RGB-D images to simulate underwater images for color correction. The authors present a two-part solution where the first part in the pipeline is the WaterGAN [39], and the second part is the image restoration network, composed of a depth estimation network and a color correction network. The WaterGAN has two systems: a generator G and discriminator D. The generator is a noise vector, which is projected, reshaped, and passed through several convolutional and deconvolutional layers which output a synthetic image. The discriminator distinguishes between real image (from another dataset) and synthetic (generated by generator). The generator aims to create images that the discriminator classifies as real.

The underwater images generated by [39] are passed through an image restoration network. The network is inspired by an encoder–decoder architecture, particularly, pixel-wise dense learning, and SegNet [40]. The SegNet uses a non-parametric upsampling layer, which benefits from the max-pooling index information in the encoder. Furthermore, the authors incorporate the skipping layers in the encoder–decoder architecture to compensate for the loss of high frequencies due to pooling operation.

The authors collect 7000 images from Michigan's Marine Hydrodynamics Laboratory. Another 6500 images are collected from Port Royal, Jamaica. Similarly, 6083 images are gathered from the coral reef system, Australia [41]. Besides, four Kinect datasets *i.e.* the B3DO [42], the UW RGB-D [43], the NYU [25] and the Microsoft 7-scenes [44], are utilized to form 15,000 underwater images via WaterGAN, out of which 12,000 are used for training and 3000 for testing. The depth estimation network is trained separately at a fixed learning rate of 10^{-6} while the color correction network is initially trained with an input resolution of 128×128 having a learning rate 10^{-6}. After that, the authors refined the color correction network with input images of 512×512 resolution, reducing the base learning rate to 10^{-7}. The ℓ_2 loss is utilized for depth estimation and color correction networks, and further, as a post-processing step, the images are normalized *i.e.* [0,1].

3.5. Dual generator GANs

The dual generator GANs algorithms for underwater image enhancement employ multiple generators to predict the improved image. Currently, the trend is to use two generators with one discriminator or two generators with two discriminators; either the aim is to share the features between the generators or use the prediction of one generator as an input to the other generator. Examples of the dual generator GANs are the following.

3.5.1. UWGAN

Based on the GANs [45], Li et al. [27] proposed a weakly supervised color transfer method for underwater image color correction, called UWGAN.[4] The UWGAN model relaxes the need for paired underwater images for training and allows the underwater images to be regarded in unknown locations, which benefits from adversarial learning. Following the CycleGAN [19], the UWGAN model adopts a cycle structure which includes a forward network and a backward network to learn the mapping functions between a source domain (*i.e.*, underwater) and a target domain (*i.e.*, air). The purpose of such a cycle structure is to capture the unique characteristics of one image collection and figure out how these characteristics could be translated into the other image collection.

The generators used in the UWGAN [27] have the same architecture as [46]. For the discriminators, the UWGAN uses 70×70 PatchGANs [18]. To train the network, 3800 underwater images and 3800 high-quality air images are collected and are resized to 256×256. The final loss function is the linear combination of three-loss functions, including adversarial loss, cycle consistency loss, and SSIM loss. The adversarial loss is to match the distribution of generated images with that of the target domain. The cycle consistency loss is to prevent the learned mappings from contradicting each other. The SSIM loss is to preserve the content and structure of source images.

3.5.2. MCycleGAN

To restore underwater images, Lu et al. [47] proposed a Multiscale Cycle Generative Adversarial Network (MCycleGAN), which is a variant of the CycleGAN network [19]. The authors incorporate the multiscale SSIM loss into the CycleGAN [19] to improve the image restoration task. The aim is to transfer the underwater style to the recovered style image.

As a first step, the dark channel prior (DCP) [49] is used to obtain the transmission map of a turbid underwater image. Additionally, the transmission maps provide depth information in the form of three binary filters. The turbid underwater images are forwarded through

[4] Code at https://li-chongyi.github.io/proj_Emerging_water.html.

Table 1
Network specifics: Essential parameters of underwater image enhancement and restoration networks. The losses *i.e.*, ℓ_{gan}, ℓ_c, ℓ_W, ℓ_{nui}, ℓ_r and ℓ_g represents adversarial, consistency, Wasserstein, nuisance, relativistic and gradient losses, respectively. The "–" means information is not available.

Methods	Network parameters								
	Patch size	Network depth	Feature maps	Variable kernels	Blocks	Residual learning	Skip connections	Framework	Loss
UIE-Net [29]	32 × 32	7	16–20	✓				–	ℓ_2
UIR-Net [37]	224 × 224	8	96–384	✓				–	ℓ_2
P2P Net [9]	66 × 66	6	96–384	✓		✓	✓	Caffe	ℓ_2
UIE-sGAN [48]	256 × 256	16	64–512			✓	✓	TensorFlow	ℓ_{gan}, ℓ_c
WaterGAN [39]	512 × 512	42	128–512			✓	✓	Caffe	ℓ_2
UGAN [14]	256 × 256	9	64–512	✓		✓	✓	TensorFlow	ℓ_1, ℓ_W
UWCNN [23]	310 × 230	10	32		✓	✓	✓	TensorFlow	ℓ_2, ℓ_{SSIM}
URCNN [36]	180 × 180	25	64			✓	✓	MatConvNet	ℓ_2
UWGAN [27]	256 × 256	18	64–256	✓	✓			TensorFlow	ℓ_{gan}, ℓ_c, ℓ_{SSIM}
DUIENet [31]	112 × 112	8	32–128	✓				TensorFlow	$\ell_{perceptual}$
MCycleGAN [47]	256 × 256	24	64–128	✓	✓	✓	✓	TensorFlow	ℓ_{gan}, ℓ_c, ℓ_{MSSIM}
DenseGAN [26]	256 × 256	10	64–512	✓	✓	✓	✓	TensorFlow	ℓ_2, ℓ_{gan}, ℓ_g
FGAN [34]	256 × 256	8	64–256	✓	✓	✓	✓	TensorFlow	ℓ_2, ℓ_{gan}, ℓ_r
UIE-DAL [12]	256 × 256	27	64–512			✓	✓	–	ℓ_2, ℓ_{gan}, ℓ_{nui}

the generator network. The turbid and generated clear underwater images are split into R, G, and B channels. The channels are then subjected to different sizes of sliding windows to compute the SSIM loss between the turbid and generated images. Furthermore, the SSIM maps are multiplied with corresponding filters and added together, which results in the multiscale SSIM map for final loss computation. As a final step, both the real-world underwater image and the computed ones are passed through the discriminator.

CycleGAN [19] inspired the generator and discriminator of MCycleGAN [47]. More specifically, the generator is adapted from image superresolution by Johnson et al. [46], which consists of nine ResNet blocks with training images of size 256 × 256 while the discriminator is based on 70 × 70 PatchGANs [50,51] to differentiate between real and fake image patches. The loss function is a union of the adversarial loss, the cycle-consistent loss, and the multiscale SSIM loss. The dataset is composed of 1037 turbid underwater images collected from ImageNet [21] and Jiao Zhou Bay, out of which 837 are retained as a training dataset, and the rest 200 are reserved for testing. ADAM [22] is used as an optimizer adopting a fixed learning rate of 0.0002 until convergence.

3.5.3. UIE-sGAN

Yu et al. [48] proposed an underwater image enhancement system using stacked conditional generative adversarial networks, abbreviated as UIE-sGAN. The proposed network architecture consists of two subnetworks *i.e.* haze detection subnetwork and color correction subnetwork. Each subnetwork has a generator and discriminator, and the color correction subnetwork is stacked on the haze detection subnetwork. For the haze detection subnet, the generator is similar to UNET [13] consisting of seven convolutional layers and seven deconvolutional layers, both followed by BN and leaky ReLU except the first convolutional layer where only leaky ReLU is employed and the last deconvolutional layer where TanH nonlinear function is realized. While the discriminator is made of four convolutional layers where the initial layer has leaky ReLU purely, and the subsequent ones have batch normalization and leaky ReLU followed by a sigmoid layer. The output of the haze detection network is a haze mask. The structure of the haze detection subnet and the color-correction subnet are identical except that the color-correction subnet takes the haze mask and RGB images as input and outputs a color corrected underwater image.

The UIE-sGAN [48] has three losses *i.e.* the adversarial loss for each network and a consistency loss. The training is accomplished by using WaterGAN [39] to generate underwater images from the NYU-v2 dataset [25]. Out of 1449 images, 1200 are held for training while the network is evaluated on the remaining ones. The images are resized to 286 × 286 and then cropped to 256 × 256 and further applying data augmentation. The network is optimized using ADAM by fixing the learning rate as 5×10^{-5}.

3.6. Network specifics

After reviewing current deep learning-based underwater image enhancement algorithms, we emphasize the different aspects of the above-mentioned deep models. First, we summarize the network specifics of different models in Table 1 and then further analyze network loss, depth, parameters, and input patch size.

Network Loss Network loss plays an integral part in learning the task underhand. Here, we discuss the losses employed in deep underwater image enhancement. The most popular type of loss functions are to minimize the per-pixel error between the ground-truth image and the predicted image, commonly known as ℓ_1 and ℓ_2. For example, the UIE-Net [29], UIR-Net [37], P2P Net [9], and URCNN [36] only use ℓ_2 to optimize their networks. Usually, other losses such as SSIM, gradient *etc.*, are combined with the ones mentioned earlier to improve the performance of the networks, *e.g.* UWCNN [23]. On the other hand, GANs rely on adversarial loss and perceptual loss to enhance the perceptual quality of the enhanced images, such as DenseGAN [26], UWGAN [27], *etc.*

Network Depth and Parameters The network depth and the number of parameters are related. The deeper the network, the more the number of parameters. Unlike other image classification [52] and enhancement tasks [53] where the network depth has exponentially increased and even consists of hundreds of convolutional layers, the underwater image enhancement networks are still very shallow composed of less than 45 layers (deepest network is the WaterGAN [39] with 42 layers); hence comprised of very less number of parameters.[5]

Input Patch Size Contrary to low-level vision tasks, most of the underwater image enhancement algorithms operate on full-size images. The reason may be to incorporate the wavelength dissipation of red, green, and blue channels. Furthermore, some algorithms reduce the image to predefined size, which requires upsampling as a post-processing step, such as MCycleGAN [47], DenseGAN [26], and UWGAN [27].

4. Experimental settings

4.1. Real-world underwater image datasets

Due to the limitations of synthetic underwater image datasets (*e.g.*, inaccurate formation models, hard assumptions, insufficient images, specific scenes, *etc.*), we mainly introduce the real-world underwater image datasets in this section.

[5] As most of the network models are not publicly available, a fair comparison to determine the exact number of parameters is not possible.

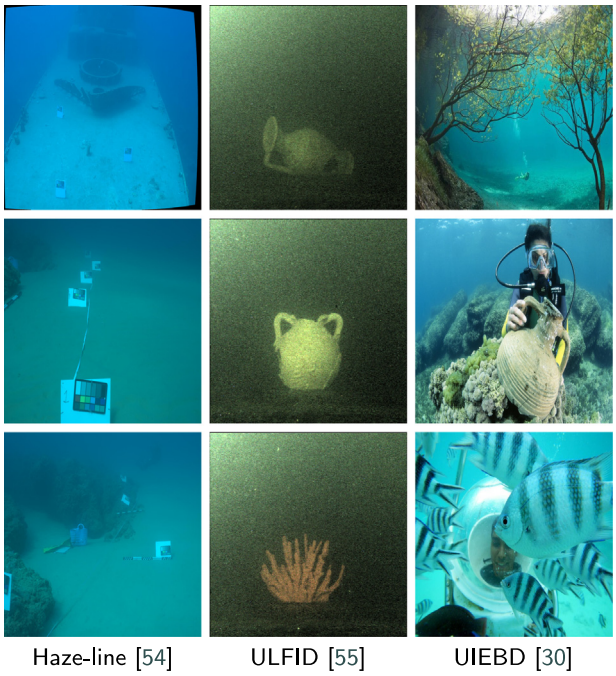

Fig. 3. Representative images: Three sample images from Haze-line [54], ULFID [55], and UIEBD [31] datasets to show the diversity of the underwater images.

- **Fish4Knowledge** [56] is funded by the European Union Seventh Framework program for the study of marine ecosystems, which provides a video and fish analysis dataset (about 200 Tb in size).[6]
- **ULFID:** Underwater Light Field Image Dataset [55] contains several underwater light field images in pure water and hazy conditions, as well as images taken in the air for reference.[7]
- **MARIS:** Marine Autonomous Robotics for InterventionS [57] is to advance the development of cooperating AUVs for undersea intervention in the offshore industry, in search-and-rescue tasks, and in various flavors of scientific exploration. This project provides several underwater images and videos captured by underwater stereo vision system.[8]
- **Haze-line Dataset** [54] collected a dataset of images taken in different locations with varying water properties, showing color charts in the scenes (about 33GB in size). Moreover, the 3D structure of the scene was calculated based on stereo imaging.[9]
- **UIEBD:** Underwater Image Enhancement Benchmark Dataset [31] includes 950 real-world underwater images, 890 of which have the corresponding reference images where each reference image is selected from 12 enhanced results. The rest 60 underwater images that cannot obtain satisfactory references are treated as challenging data. The UIEBD [31] contains a large range of image resolution and spans diverse scene/main object categories.[10]

The existing real-world underwater image datasets usually have monotonous content and limited quality degradation types. Moreover, these datasets did not provide the corresponding ground truth images because it is impractical to simultaneously obtain the degraded underwater image and the ground-truth of the same scene. The UIEBD [31] provides the corresponding reference images which can be considered for full-reference image quality assessment. We conduct experimental quantitative and visual comparisons on this dataset. Besides, to validate the generalization of current deep algorithms, we also present the visual results of different methods on another two datasets *i.e.*, Haze-line dataset [54] and ULFID [55]. Some representative samples of these three datasets are given in Fig. 3.

[6] http://groups.inf.ed.ac.uk/f4k/index.html.
[7] https://github.com/kskin/data.
[8] http://rimlab.ce.unipr.it/Maris.html.
[9] http://csms.haifa.ac.il/profiles/tTreibitz/datasets/ambient_forwardlooking/index.html.
[10] https://li-chongyi.github.io/proj_benchmark.html.

4.2. Evaluation metrics

Evaluations performed for underwater image enhancement can be broadly categorized into automatic evaluation metrics and human visual system (HVS). The automatic evaluations are performed using six metrics, out of these, four are also most widely used in image enhancement and restoration problems *i.e.* PSNR, MSE, and SSIM [58], and PCQI [59] while the other two are specific for underwater image enhancement *i.e.* UCIQE [60] and UIQM [61]. Next, to make the article inclusive, we describe all the evaluation metrics and then detail their limitations and reliability. Moreover, we also provide the report which details the human visual evaluation and its importance.

4.2.1. Automatic evaluation metrics

- **MSE and PSNR:** We begin our discussion with Mean Square Error (MSE) as the signal measure. The MSE aims to provide a quantitative score that represents the similarity or distortion between the two signals. Usually, one of the signals is the original signal, and the other one is recovered from some distortion or contamination. Mathematically, the MSE between the two signals can be expressed as:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2, \tag{6}$$

where x and y are two signals, in this case, images and x_i and y_i are the pixels at ith location. Similarly, N are the number of pixels. Furthermore, in the image processing literature, peak signal to noise ratio (PSNR) measure is computed from MSE as:

$$\mathrm{PSNR} = 10\log_{10}\frac{L^2}{\mathrm{MSE}} \tag{7}$$

where L is the dynamic range of image pixel intensities (*i.e.*, 255 for image). The usage of MSE and PSNR has many attractive features *e.g.* (1) it is simple, (2) all norms are valid distance metrics, (3) it has a clear physical meaning, and (4) these are excellent metrics in the context of optimization. The mentioned measures assume that the signal fidelity is independent of the relationship between (1) the original signal, (2) the distorted and original signal, and (3) the signs of the error signal. Unfortunately, none of them even roughly holds in the context of measuring the visual perception of image fidelity [62]. In the next section, we discuss alternatives to these measures.

- **SSIM:** Another commonly used measure is the Structural SIMilarity (SSIM) index. The main ideas of SSIM were presented by Wang & Bovik [63] and formulated in [64,65]. Let us consider that x and y are the patches taken from the two different images but locations to be compared against each other. Then SSIM takes three measures into account, which are the similarity of the patch (1) luminance $l(x, y)$, (2) contrasts $c(x, y)$, and (3) the local structures $s(x, y)$. As pointed out in [65], these similarities are expressed and computed using simple statistics and are combined to produce local SSIM as:

$$\begin{aligned}\mathrm{SSIM} &= l(x,y)\cdot c(x,y)\cdot s(x,y),\\ &= \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2+\mu_y^2+C_1}\right)\left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2+\sigma_y^2+C_2}\right)\left(\frac{\sigma_{xy}+C_3}{\sigma_x+\sigma_y+C_3}\right),\end{aligned} \tag{8}$$

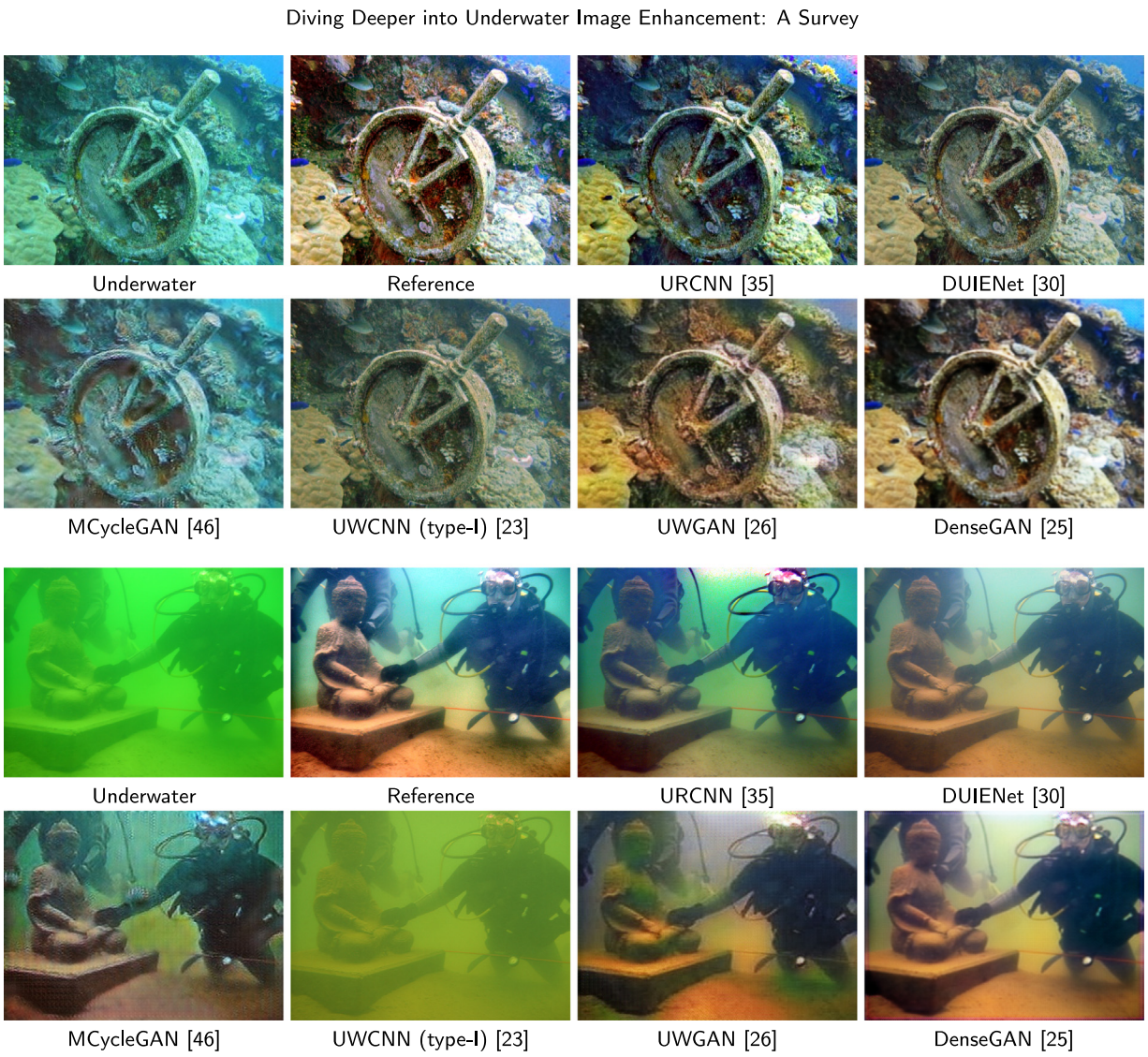

Fig. 4. Visual comparison of greenish images: Comparisons of different methods on the greenish underwater samples from UIEBD [31]. Here, UWCNN-type-I represents the model trained by synthetic type-I training data.

where μ_x and μ_y are means while σ_x and σ_y are standard deviations of the patches x and y, respectively. Similarly, σ_{xy} cross-correlation of the patches after removing their means. The constants C_1, C_2 and C_3 stabilize the terms to avoid near-zero divisions.

- **PCQI:** Patch-based contrast quality index (PCQI) [59] relies on patch-based approach as contrary to relying on global statistics. The PCQI depends on three independent quantities of an image patch *i.e.* mean, signal strength, and structure. Mathematically, a patch-based contrast image quality index (PCQI) is given by:

$$\text{PCQI} = q_i(x, y) \cdot q_c(x, y) \cdot q_s(x, y), \tag{9}$$

where $q_i(x, y)$ is to compare mean intensity, $q_c(x, y)$ is to determine the structural distortion and $q_s(x, y)$ is the contrast change. PCQI is mathematically expensive as compared to other metrics. Next, we discuss quantitative measures, which are more specific to underwater image enhancement.

- **UCIQE:** Underwater color image quality evaluation abbreviated as UCIQE [60], is based chroma, contrast, and saturation of CIELab and is defined as:

$$\text{UCIQE} = C_1 \times \sigma_c + C_2 \times con_l + C_3 \times \mu_s, \tag{10}$$

where σ_c, con_l, and μ_s are the standard deviation of chroma, the contrast of luminance, and the mean of saturation. It is to be noted here that human perception has a good correlation with the variance of chroma for underwater images.

- **UIQM:** UIQM [61] stands for underwater image quality measure and is different from earlier defined evaluation metrics. The UIQ employs the HVS model only and does not require a reference image; hence, it is a better candidate for evaluating underwater images. UIQM is dependent on three attribute measures the underwater images, which are (1) image colorfulness measure (UICM), (2) sharpness measure (UISM), and (3) contrast measure (UIConM). Following is the formulation of UIQM:

$$\text{UIQM} = c_1 \times \text{UICM} + c_2 \times \text{UISM} + c_3 \times \text{UIConM}, \tag{11}$$

where c_1, c_2 and c_3 are the parameters which are application dependent, *e.g.*, more weight should be given to c_1 for underwater color correct while c_2 for increasing visibility in the underwater scene.

Fig. 5. Qualitative comparisons on bluish images: The results of various CNN-based and GAN-based methods on the sample underwater images from UIEBD [31].

4.2.2. Human visual system

Due to the lack of real ground-truth data, human subjects are used to evaluate the quality of the predicted images in an attempt to incorporate the perceptual measures. These human inputs may either be crowd-sourced or specialist persons in different competitions. However, none of these methods have shown any significant advantage over the mathematical measure. In other words, mathematically defined measures are still attractive due to the following reasons.

- They are simple to calculate and computationally inexpensive normally.
- They are independent of distinct individuals and observing conditions.

Furthermore, it is thought that viewing conditions play an influential role in the human perception of image quality. However, if there are multiple viewing conditions, a method dependent on viewing conditions may produce different estimations that may be inconvenient to utilize. Moreover, it may also be specific to the user observation, and it becomes the responsibility of each to compute the viewing conditions and provide the output to the measurement systems. On the other hand, a method independent of viewing conditions computes a single quantity that provides a general idea about the image quality. Besides, the experience of volunteers significantly affects human visual perception. The volunteers who understand what the degrading effects of attenuation and backscatter are, and what it looks like when either is improperly corrected can provide more reliable subjective scores of image quality.

4.3. Benchmark results

The benchmark results for each technique[11] on UIEBD [31] dataset are reported in Table 2. The quantitative experiments are conducted on UIEBD [31] because it is, to the best of our knowledge, the only one dataset which provides the corresponding reference images for image quality assessment. The results by using reference images can provide realistic feedback on the quality of enhanced results to some extent. Moreover, in the case of multiple variants of the same algorithm, all the results are reported. We encourage the readers to consult the original paper for a detailed analysis of each variant of the same model.

The results are presented via the metrics mentioned earlier. It is to be noted here that the PSNR, SSIM, PCQI, UCIQE, and UIQM, the higher, the better while the MSE, the lower, the better. Also, to be fair amidst all the methods under consideration, we resize the output of the network where the predicted image is a scaled-down version of the underwater scene input. From Table 2, DUIENet [31] results are the best among the competitors while the UWCNN [23] performs

[11] The results are reported for the methods having the source code or executables available, or the respected authors agreed to provide the results on the dataset.

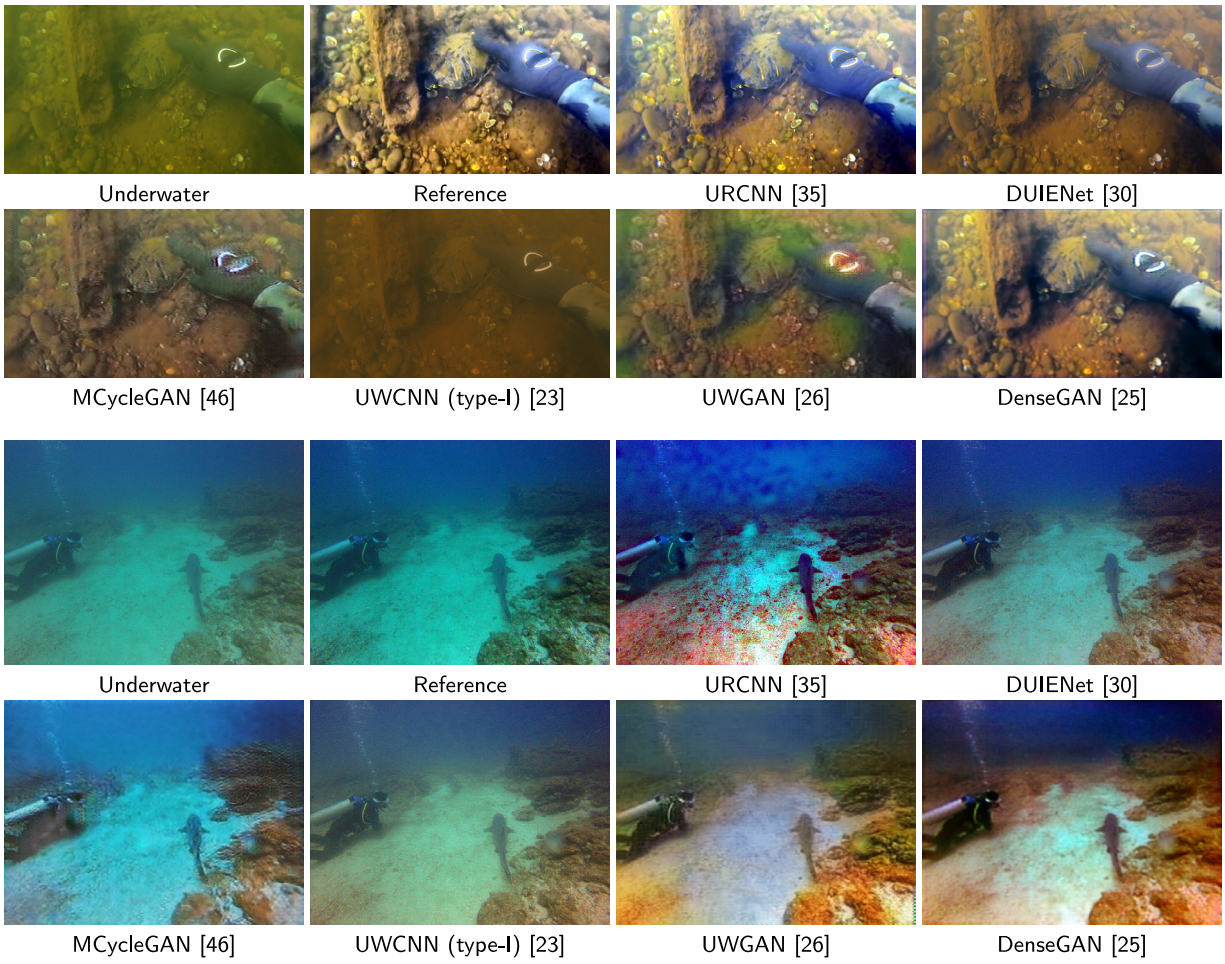

Fig. 6. The low and high backscatter images: The challenging images to remove the backscatter. The images are selected from UIEBD [31] dataset. The top image shows the low backscatter, while the bottom image illustrates the high backscatter.

worst due to training on the synthesized underwater images which are different from the images in the UIEBD [31]. However, it is challenging to state the superiority of one method against the others due to many factors involved, for example, the number of parameters, the depth of network, training images, patch size, number of channels and loss function, *etc.* To compare fairly, most of these determinants should be kept consistent. To further validate the performance of different deep algorithms, we conduct qualitative comparisons on diverse underwater images from different datasets in the next section.

4.4. Qualitative comparisons

We present the visual results on UIEBD [31], Haze-line [54] and ULFID [55] in Figs. 4–8. The ground-truth images for Haze-line [54] and ULFID [55] are not available; hence, we furnish the visual results only for both the datasets.

- **Greenish tone images:** In Fig. 4, we present the visual comparisons of greenish underwater images from UIEBD [31] for the state-of-the-art CNN-based and GAN-based methods. The GAN-based models aim to improve the perceptual quality, while CNN models are more focused on the PSNR values of the enhanced images. One can notice that the outputs of GAN methods are generally different in the tone compared to CNN methods, as the later is more faithful to the original underwater image colors. This also contributes to the higher PSNR for the CNN methods compared to GAN methods, as shown in Table 2. It is to be noted that in Fig. 4, we only show one of the variants in case of the same algorithm for the limited space.

Table 2
Quantitative results: The best results are highlighted with red color while the blue color represents the second best.

Method	UWE dataset					
	PSNR ↑	MSE ↓	SSIM ↑	PCQI ↑	UCIQE ↑	UIQM ↑
Original	17.36	1768.90	0.6168	1.1118	0.5196	1.1571
MCycleGAN [47]	18.33	1132.21	0.6138	0.4521	0.5196	1.1471
URCNN [36]	15.94	2195.89	0.5972	1.0936	0.5196	1.5332
UWGAN [27]	16.06	1853.70	0.2945	0.6000	0.5921	1.1099
DUIENet [31]	19.29	1012.20	0.8093	0.9844	0.5720	1.2963
DenseGAN [26]	17.56	1363.60	0.4239	0.6697	0.6291	1.0952
UWCNN_type-1 [23]	13.03	3930.80	0.4795	1.0310	0.4876	1.1319
UWCNN_type-3 [23]	13.58	3297.40	0.5482	1.0146	0.4771	1.1035
UWCNN_type-5 [23]	13.29	3427.20	0.5102	0.9223	0.4303	1.0122
UWCNN_type-7 [23]	13.30	3372.60	0.4287	0.8693	0.4533	1.0385
UWCNN_type-9 [23]	10.58	6164.80	0.2598	0.4958	0.3636	0.7775
UWCNN_type-I [23]	15.00	2345.00	0.5306	1.0890	0.4954	1.1294
UWCNN_type-II [23]	13.46	3654.10	0.4509	1.0631	0.4766	1.1048
UWCNN_type-III [23]	14.24	2920.20	0.4945	1.0486	0.4739	1.0333

- **Bluish tone images:** Fig. 5 shows the visual comparisons on two bluish images from UIEBD [31] consisting of a ray and statues. The bluish tone is ubiquitous in underwater images and difficult to be completely removed by current algorithms. DUIENet [31] and UWCNN [23] render the best outcomes; however, the results still have a bluish tone, especially in far distances (more severe backscatter). By contrast, the UWGAN [27] and DenseGAN [26]

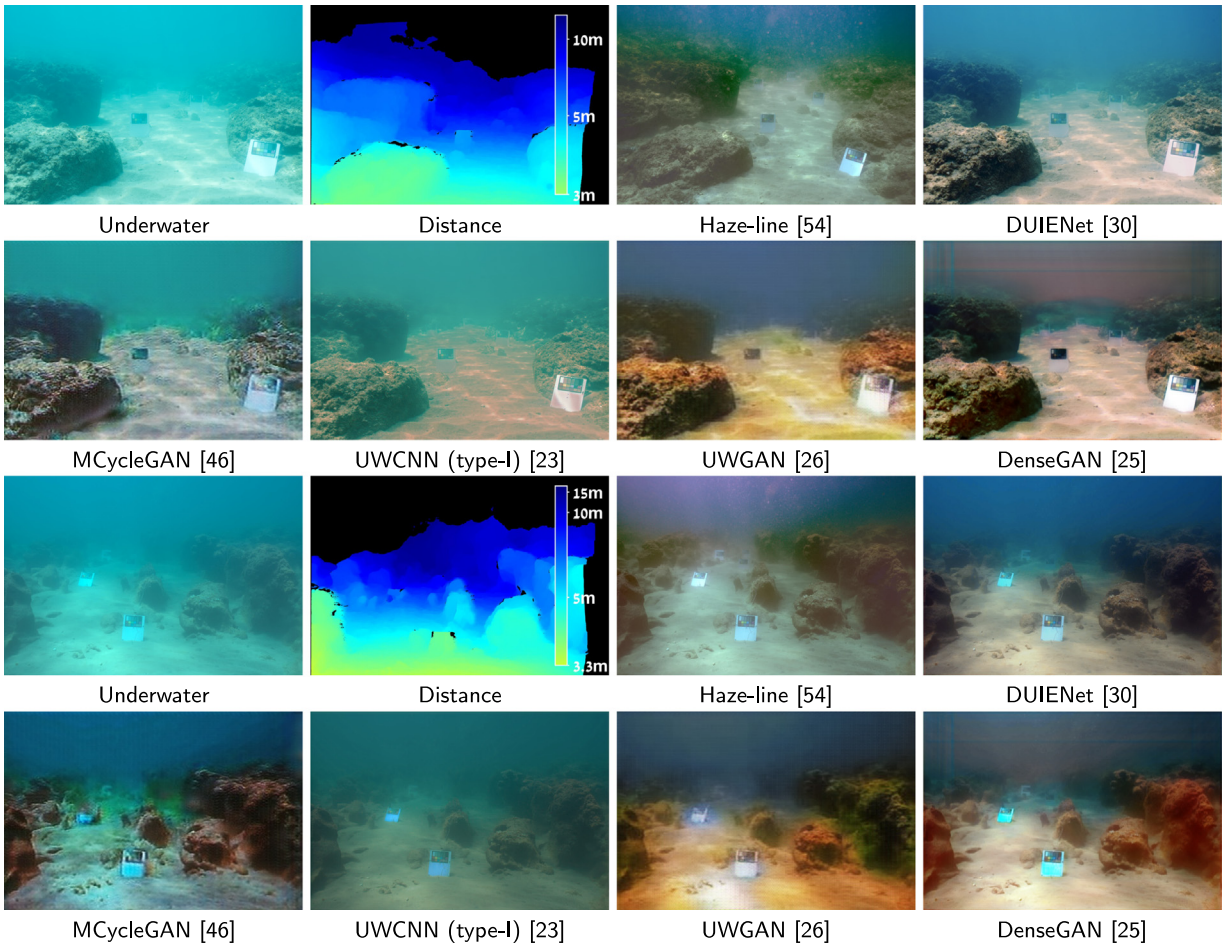

Fig. 7. Visual comparisons on Haze-line [54]: The Haze-line dataset provides an accurate distance based on the stereo. To be fair to the authors of Haze-line [54], we have also included the results of the best performer (*i.e.*, Haze-line [54], a conventional method) on this dataset.

introduces obvious artificial colors mainly inducing by the shortcomings of their unpaired training data.

- **Low and high backscatter images:** Backscatter is a challenging problem faced during the underwater imaginary. The leading causes of backscattering are the strobes or the internal flash, which lights up the particles in the water present between the subject and the camera lens. This phenomenon can also be observed behind the subject, lighting up the open water. With a dark background, backscattering is more natural to recognize. Here, we present two images in Fig. 6 on low and high backscatter from [31]. The first image in Fig. 6 is an example of low backscatter, while the bottom one is of high backscatter. We can visually observe that the URCNN [36] has over-exposed the images while the UWGAN [27] created some artificial colors. In addition, the low backscatter is relatively easier to be removed than the high backscatter. For the high backscatter image, none of the methods can produce visually pleasing results, and current methods even introduce annoying artifacts and color casts. It should also be regarded here that UWCNN [23] can produce good results if the model matches the type of water.

- **Haze-line** [54] **images**: The visual comparisons for underwater images from Haze-line dataset [54] is provided in Fig. 7. This dataset only provides the depth maps reconstructed from the stereo images; however, no ground-truth images are available for computing the evaluation metrics. The images in this dataset are challenging since most of the images have bluish tones and high backscatter. UWGAN [27] and DenseGAN [26] provide visually promising results, but both have created false colors, and this is also the case with DUIENet [31] and MCycleGAN [47] networks. It is obvious that all deep algorithms fall behind the performance of a conventional method [54], which mismatches the progress of deep learning in other low-level visual tasks.

- **ULFID** [55] **images**: As the last example, we show the images with severe degradations from ULFID [55] in Fig. 8. The ground-truth images for this dataset are not feasible to evaluate the models; hence, we only present the visual results. Although the deep algorithms can remove the greenish tone from the images; however, all of them fail to furnish clear images and even amplify the noise. This dataset is an excellent example that the underwater image enhancement still requires concerted efforts to progress, and the noise in underwater images should be paid more attention in future studies.

5. Future and emerging directions

Underwater image enhancement is a classical research area and has improved a lot in recent years, mainly due to the rapid development of deep learning techniques. The performance is still lacking in many aspects compared to other image enhancement techniques like image super-resolution, deblurring, and dehazing. There is ample room to advancement the underwater image enhancement direction. Here, in the following paragraphs, we present the list of some of the potential future directions.

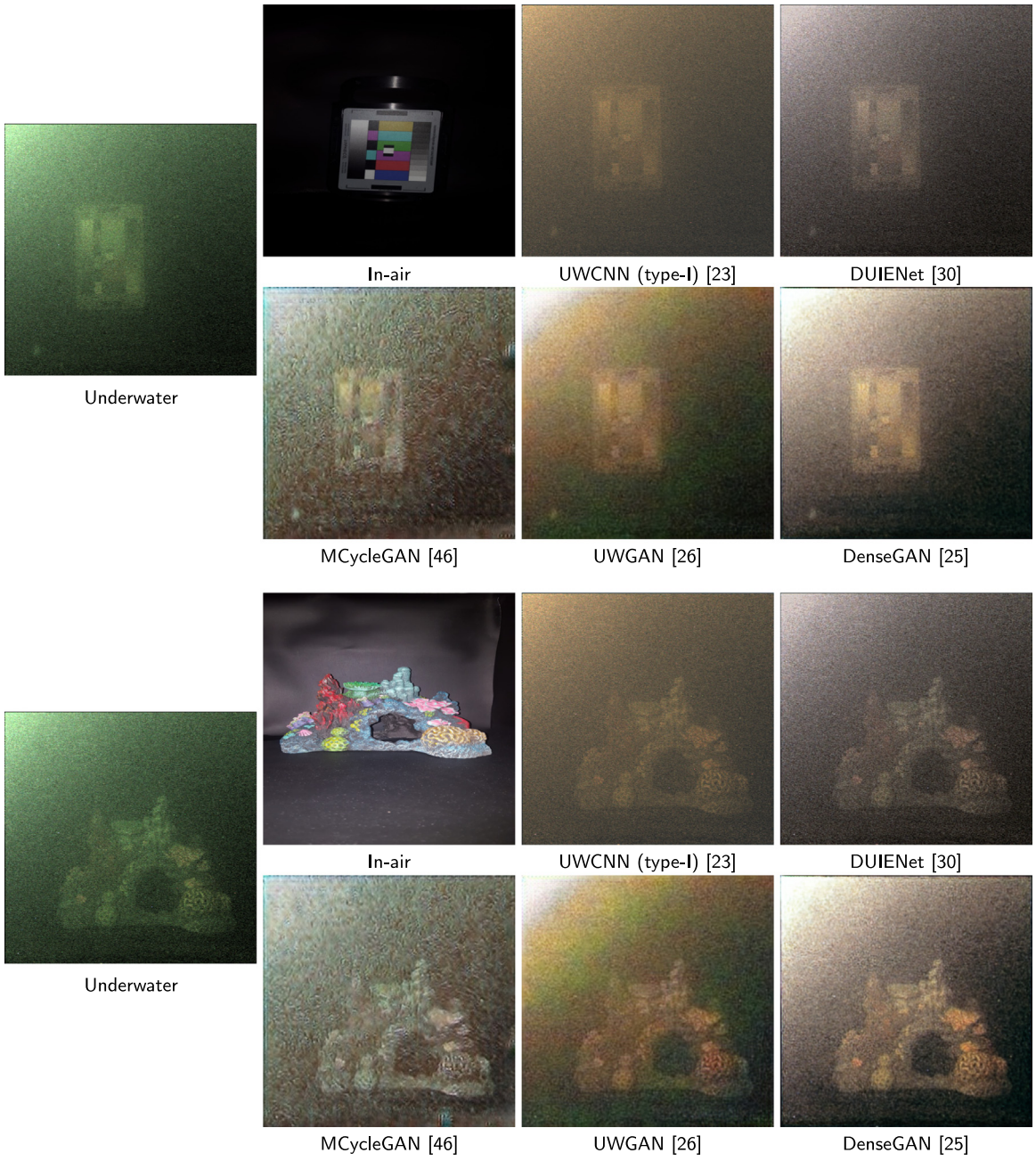

Fig. 8. Images from ULFID [55]: A challenging dataset where all the methods fail to provide clean results.

- **Datasets:** Underwater image enhancement methods usually employ synthetic images for training due to a lack of representative real-world underwater images and their corresponding ground-truth images. Although there are limited datasets available that have underwater and their reference images; however, these datasets consist of a finite number of images and are typically used as test images rather than training the models. A true effort in this direction may improve underwater image enhancement models and provide realistic feedback on the image quality of enhanced results using different methods.

- **Objective functions and evaluation metrics:** Current algorithms predominantly employ objective functions common to image enhancement techniques. Although these functions produce some favorable results; however, none of them incorporate the underwater physical model properties. Likewise, the available evaluation metrics to underwater images are limited and have failure cases, which keeps the field of underwater image enhancement at a standstill. For example, the visual results shown in Figs. 4–8 do not match the quantitative results in Table 2. Therefore, more specialized objective functions and evaluation metrics are required to advance the underwater image enhancement research.

- **Prior knowledge:** The human perception of the scene depends on the extensive domain or prior knowledge. When experts describe the image quality, they do not solely rely on the content of the visuals; instead, they also use their domain knowledge. An exciting venue to explore is to augment the current techniques with prior or domain knowledge [66]. This has shown an increase in the performance in areas like visual question answering and would likely improve underwater image enhancement.

- **Unsupervised learning:** Many methods generate synthetic data to train their models due to the lack of dataset, which has underwater images and their ground-truth images. Although these models exhibit promising results for synthetic underwater scenes; however, they fail on real-world underwater images. To deal with the lack of data, a possible research direction could be unsupervised learning, also known as zero-shot or few-shot learning. This capability may lead to promising results, but the zero-shot problem itself is not trivial. A more realistic scenario would be to employ the present limited datasets, few-shot learning, where the network learns from a few available images. The development of unsupervised learning is an open research problem.

- **Real vs. Synthetic:** Existing algorithms use diverse physical (mathematical) models to generate underwater images. The distribution of the generated underwater scenes may not be conferred to the real-world scenes; therefore, the models trained on artificially produced datasets lack generalization capability. A more thorough and exhaustive effort is required to generate artificial datasets, and one solution may be to use GAN-based networks to transfer style from underwater images to the simulated scenes. Even though minimal work [39] has been done in this direction, there is still a lot of scope of improvement.

6. Conclusion

We presented the first comprehensive literature survey on CNNs and GANs for underwater image enhancement. To the best of our knowledge, we have included all the deep learning-based methods, which deal with underwater image enhancement, including those which are available on arxiv.[12] Moreover, we provided and reviewed the datasets, which can be used for training and testing the algorithms. We also discussed the details of the evaluation metrics with their limitations. Using all the metrics, we compared the performance on the benchmark dataset. We also presented the visual comparisons to illustrate the varying difficulty and the robustness of the algorithms. As a final step, we reviewed the limitations and provided future research areas to advance the underwater image enhancement.

The deep learning-based underwater image enhancement methods still follow the development of deep learning ranging from CNNs to GANs. Most of the current models are the modifications of existing network architectures such as the encoder–decoder network and CycleGAN. The significant difference is the training data (*i.e.*, underwater images). Besides, there is no network architecture or loss function well-designed for underwater image enhancement tasks, resulting in the unstable and visually displeasing results. In most cases, the deep learning-based methods fall behind state-of-the-art conventional methods. More importantly, almost all models use synthetic data for networks' training. The synthetic training data limit the generalization of models. Thus, the development of deep learning-based underwater image enhancement has a long way to go.

According to our survey, the underwater research progress is hindered by the lack of purposely built evaluation metrics and large training datasets. The current metrics are taken from the image enhancement while the training datasets are synthetically generated. One approach to develop evaluation metrics is to incorporate underwater image properties. Similarly, more realistic datasets can be created using the GANs.

CRediT authorship contribution statement

Saeed Anwar: Conceptualization, Methodology, Software, Data curation, Writing - original draft. **Chongyi Li:** Visualization, Investigation, Software, Validation, Writing - reviewing and editing.

[12] At the time of submission.

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.

References

[1] Y. LeCun, Y. Bengio, G. Hinton, Nature (2015).
[2] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2012) 1397–1409.
[3] K. He, X. Zhang, S. Ren, et al., Deep residual learniing for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[4] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, P. Han, Hierarchical features driven residual learning for depth map super-resolution, IEEE Trans. Image Process. 28 (5) (2018) 2545–2557.
[5] K. Zhang, W. Zuo, S. Gu, Learning deep cnn denoiser prior for image restoration, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[6] H. Koschmieder, Theorie der horizontalen sichtweite, in: Beitrage Zur Physik Der Freien Atmosphare, 1924.
[7] J. Chiang, Y. Chen, Underwater image enhancement by wavelength compensation and dehazing, IEEE Trans. Image Process. 21 (4) (2012) 1756–1769.
[8] D. Akkaynak, T. Treibitz, A revised underwater image formation model, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[9] X. Sun, L. Liu, Q. Li, J. Dong, E. Lima, R. Yin, Deep pixel to pixel network for underwater image enhancement and restoration, IET Image Process. (2018).
[10] X. Mao, C. Shen, Y.-B. Yang, Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections, in: Advances in Neural Information Processing Systems, 2016, pp. 2802–2810.
[11] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE (1998).
[12] P. Uplavikar, Z. Wu, Z. Wang, All-in-one underwater image enhancement using domain-adversarial learning, 2019, arXiv preprint arXiv:1905.13342.
[13] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.
[14] C. Fabbri, M.J. Islam, J. Sattar, Enhancing underwater imagery using generative adversarial networks, 2018, arXiv preprint arXiv:1801.04011.
[15] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of wasserstein gans, in: Advances in Neural Information Processing Systems, 2017, pp. 5767–5777.
[16] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, 2015, arXiv preprint arXiv:1511.06434.
[17] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: ICML, 2015.
[18] C. Li, M. Wand, Precomputed real-time texture synthesis with markovian generative adversarial networks, in: IEEE European Conference on Computer Vision (ECCV), 2016.
[19] Y. Zhu, T. Park, A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: IEEE International Conference on Computer Vision, 2017.
[20] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: ICML, 2010.
[21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009.
[22] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: ICLR, 2014.
[23] S. Anwar, C. Li, F. Porikli, Deep underwater image enhancement, 2018, arXiv preprint arXiv:1807.03528.
[24] C. Li, S. Anwar, F. Porikli, Underwater scene prior inspired deep underwater image and video enhancement, Pattern Recognit. 98 (2020) 107038.
[25] N. Silberman, D. Hoiem, P. Kohli, R. Fergus, Indoor segmentation and support inference from rgbd images, in: IEEE European Conference on Computer Vision (ECCV), 2012.
[26] Y. Guo, H. Li, P. Zhuang, Underwater image enhancement using a multiscale dense generative adversarial network, IEEE J. Ocean. Eng. (2019).
[27] C. Li, J. Guo, C. Guo, Emerging from water: Underwater image color correction based on weakly supervised color transfer, IEEE Signal Process. Lett. 25 (3) (2018) 323–327.
[28] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, 2018, arXiv preprint arXiv:1802.05957.
[29] Y. Wang, J. Zhang, Y. Cao, Z. Wang, A deep cnn method for underwater image enhancement, in: 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 2017, pp. 1382–1386.

[30] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2012) 1397–1409.
[31] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, IEEE Trans. Image Process. 29 (2019) 4376–4389.
[32] C. Aucuti, C.O. Ancuti, P. Bekaert, Enhancing underwater images and videos by fusion, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012.
[33] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: ICLR, 2014.
[34] H. Li, J. Li, W. Wang, A fusion adversarial underwater image enhancement network with a public test dataset, 2019, arXiv e-prints, arXiv:1906.06819, arXiv:1906.06819.
[35] A. Jolicoeur-Martineau, The relativistic discriminator: a key element missing from standard gan, 2018, arXiv preprint arXiv:1807.00734.
[36] M. Hou, R. Liu, X. Fan, Z. Luo, Joint residual learning for underwater image enhancement, in: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, 2018, pp. 4043–4047.
[37] K. Cao, Y.-T. Peng, P.C. Cosman, Underwater image restoration using deep networks to estimate background light and scene depth, SSIAI (2018).
[38] D. Eigen, C. Puhrsch, R. Fergus, Depth map prediction from a single image using a multi-scale deep network, in: Advances in Neural Information Processing Systems, 2014, pp. 2366–2374.
[39] J. Li, K.A. Skinner, R.M. Eustice, M. Johnson-Roberson, Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images, IEEE Robot. Autom. Lett. 3 (1) (2017) 387–394.
[40] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.
[41] O. Pizarro, A. Friedman, M. Bryson, S.B. Williams, J. Madin, A simple, fast, and repeatable survey method for underwater visual 3d benthic mapping and monitoring, Ecol. Evol. (2017).
[42] A. Janoch, S. Karayev, Y. Jia, J.T. Barron, M. Fritz, K. Saenko, T. Darrell, A category-level 3d object dataset: Putting the kinect to work, in: Consumer Depth Cameras for Computer Vision, 2013.
[43] K. Lai, L. Bo, D. Fox, Unsupervised feature learning for 3d scene labeling, in: ICRA, 2014.
[44] J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi, A. Fitzgibbon, Scene coordinate regression forests for camera relocalization in rgb-d images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013.
[45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.
[46] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: IEEE European Conference on Computer Vision (ECCV), 2016.
[47] J. Lu, N. Li, S. Zhang, Z. Yu, H. Zheng, B. Zheng, Multi-scale adversarial network for underwater image restoration, Opt. Laser Technol. (2019).
[48] X. Ye, H. Xu, X. Ji, R. Xu, Underwater image enhancement using stacked generative adversarial networks, in: Pacific Rim Conference on Multimedia, Springer, 2018, pp. 514–524.
[49] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2010) 2341–2353.
[50] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[51] C. Ledig, Z. Wang, W. Shi, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[52] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 770–778.
[53] S. Anwar, N. Barnes, Densely residual laplacian super-resolution, 2019, arXiv preprint arXiv:1906.12021.
[54] D. Berman, D. Levy, S. Avidan, T. Treibitz, Underwater single image color restoration using haze-lines and a new quantitative dataset, 2018, arXiv preprint arXiv:1811.01343.
[55] K.A. Skinner, M. Johnson-Roberson, Underwater image dehazing with a light field camera, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.
[56] B.J. Boom, J. He, S. Palazzo, P.X. Huang, H.-M. Chou, F.-P. Lin, C. Spampinato, R.B. Fisher, A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage, Ecol. Inform. (2014).
[57] F. Oleari, F. Kallasi, D.L. Rizzini, J. Aleotti, S. Caselli, An underwater stereo vision system: from design to deployment and dataset acquisition, in: OCEANS 2015-Genova, IEEE, 2015, pp. 1–6.
[58] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment, in: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, 2003.
[59] S. Wang, K. Ma, H. Yeganeh, Z. Wang, W. Lin, A patch-structure representation method for quality assessment of contrast changed images, IEEE Signal Process. Lett. 22 (12) (2015) 2387–2390.
[60] M. Yang, A. Sowmya, An underwater color image quality evaluation metric, IEEE Trans. Image Process. 24 (12) (2015) 6062–6071.
[61] K. Panetta, C. Gao, S. Agaian, Human-visual-system-inspired underwater image quality measures, IEEE J. Ocean. Eng. 41 (3) (2015) 541–551.
[62] Z. Wang, A.C. Bovik, Mean squared error: Love it or leave it? A new look at signal fidelity measures, IEEE Signal Process. Mag. (2009).
[63] Z. Wang, A.C. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9 (3) (2002) 81–84.
[64] Z. Wang, A.C. Bovik, Modern image quality assessment, Synth. Lect. Image Video Multimedia Process. 2 (1) (2006) 1–156.
[65] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[66] Q. Wu, P. Wang, C. Shen, A. Dick, A. van den Hengel, Ask me anything: Free-form visual question answering based on knowledge from external sources, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4622–4630.

Saeed Anwar is a Research Scientist at ICV (Imaging and Computer Vision) of Cyber–Physical Systems in CSIRO and an adjunct Lecturer in the Research School of Engineering, Australian National University (ANU). He has received his Ph.D. from Australian National University (ANU) and Data61/CSIRO. His research interests include computer vision, deep learning, low-level vision, image enhancement with commercial applications in video surveillance, 3D vision, and medical systems. Dr. Anwar presented Oral papers at ICCV and BMVC. He won several professional prizes. He authored more than 20 publications. He also reviews for many prestigious venues such as TPAMI, TIP, CVPR, ICCV, ECCV, ICIP, WACV, and many others. He has also worked in different roles in five different universities and various computer vision and robotics industries. Further information can be found at https://saeed-anwar.github.io/.

Chongyi Li received his Ph.D. degree with the School of Electronic Information Engineering, Tianjin University, Tianjin, China in June 2018. From 2016 to 2017, he took his one year study at the Research School of Engineering, Australian National University (ANU) as a visiting Ph.D. student supported by the CSC. Now, he is a postdoctoral research fellow at the Department of Computer Science, City University of Hong Kong. His current research focuses on image processing, computer vision, and deep learning, particularly in the domains of image restoration and enhancement, such as images captured under the bad weather (hazy, foggy, sandy, dusty, rainy, snowy day) and special circumstances (underwater, weak illumination). He also focuses on other low-level vision problems, such as image/depth super-resolution reconstruction, image deblurring, image denoising, and multi-exposure image fusion. Further information can be found at https://li-chongyi.github.io/.