



## Full length article

Sea-Pix-GAN: Underwater image enhancement using adversarial neural network<sup>☆</sup>

Dhiraj Chaurasia, Prateek Chhikara\*

University of Southern California, Los Angeles, United States of America



## ARTICLE INFO

## Keywords:

Adversarial neural network  
Deep learning  
Image processing  
Underwater imaging

## ABSTRACT

In the last decade, the exploration of deep-sea ecosystems has surged, offering exciting prospects for discovering untapped resources such as medical drugs, food and energy sources, and renewable energy products. Consequently, research in underwater image processing has witnessed substantial growth. However, underwater imaging poses significant challenges, particularly without sophisticated, specialized cameras. Traditional cameras are impacted by absorption and scattering in the aquatic environment, producing hazy images with a blue-green tint. This phenomenon holds implications for marine research and other disciplines that rely on underwater imaging. While hardware advancements have been made over the years, image processing remains a valuable, cost-effective, and practical approach for underwater enhancement. Despite the existence of state-of-the-art techniques for underwater enhancement and restoration, their performance is often inconsistent. While some methods excel in contrast restoration, color restoration remains a pervasive challenge. In this paper, we introduce Sea-Pix-GAN, a Generative Adversarial Network (GAN)-based model that addresses these issues in underwater image enhancement. We redefine the problem as an image-to-image translation task and tailor the objective and loss functions to achieve color, content, and style transfer. The model is trained on a large dataset of underwater scenes, encompassing the diverse color dynamics of underwater subjects. Sea-Pix-GAN demonstrates promising results in restoring color, contrast, texture, and saturation. To validate its effectiveness, we compare the performance of Sea-Pix-GAN quantitatively based on metrics like PSNR, SSIM, and UIQM and qualitatively against several existing techniques.

## 1. Introduction

In recent years, computer vision has emerged as a transformative field with applications spanning across various domains; from real-estate [1], healthcare to autonomous vehicles, computer vision technologies have revolutionized the way we perceive and interact with the world around us [2,3]. However, the enigmatic underwater world has presented unique challenges and opportunities that have recently captured the attention of researchers. The significance of underwater imaging cannot be overstated, as it plays a pivotal role in a wide range of applications [4]. These include ocean resource exploration, marine ecological research, monitoring deep-sea installations (such as underwater robotics and object recognition), and even naval military applications [5]. Underwater imaging serves as a vital medium for acquiring essential information, empowering effective perception and comprehension of underwater environments [6]. The success of exploring underwater applications is heavily reliant on the quality of captured underwater images [7]. Insufficient image quality can lead

to the failure of computer systems employed for visual inspection and analysis [8]. Consequently, enhancing the quality of underwater images has become vital in advancing research and applications in this intriguing domain.

Despite advancements in imaging technology [9], persistent limitations arise from the inherent physical and optical properties of seawater, impeding optimal image capture and quality [10]. The comprehensive investigation of underwater applications hinges on the quality of captured underwater images. Typically, the quality of underwater photos is influenced by various factors, including limited visibility range, non-uniform lighting, unwanted signal-like noise, and color degradation. Water shows strong absorbing properties to most electromagnetic spectrum wavelengths, but it has a narrow transparency window. This window includes the visible spectrum [11,12]. However, the natural presence of particles in the water, including micro phytoplankton, colored dissolved organic matter, and non-algal particles, contributes to light absorption and scattering [13,14]. Therefore, light

<sup>☆</sup> This paper has been recommended for acceptance by Zicheng Liu.

\* Correspondence to: Viterbi School of Engineering, 3650 McClintock Ave, Los Angeles, CA 90089, USA.

E-mail addresses: [dchauras@usc.edu](mailto:dchauras@usc.edu) (D. Chaurasia), [pchhikar@usc.edu](mailto:pchhikar@usc.edu) (P. Chhikara).

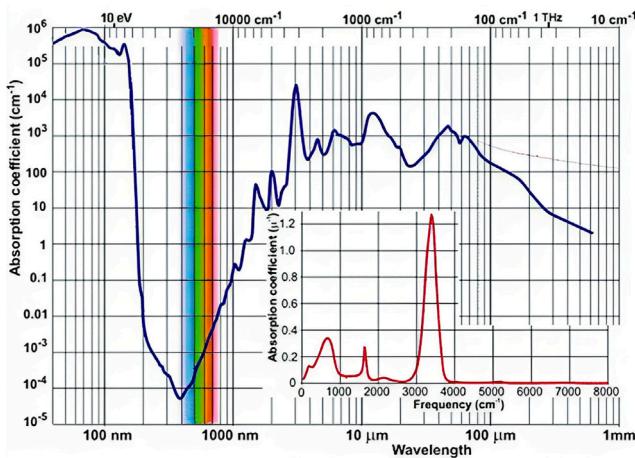


Fig. 1. The normalized Poisson disk radius increases monotonically and converges during disk packing iterations.<sup>1</sup>

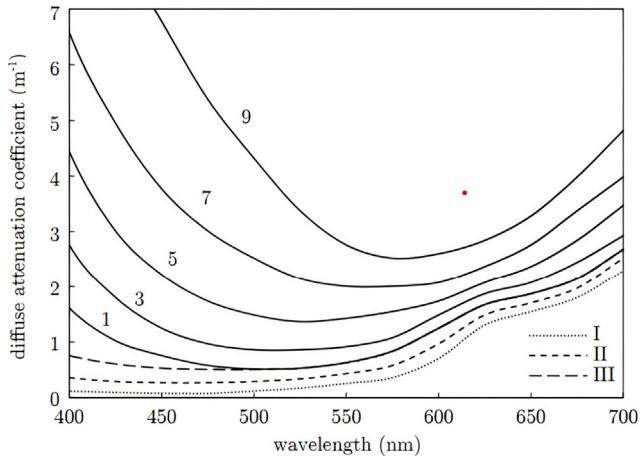


Fig. 2. Jerlov water types, where open ocean waters are labeled I-III, and turbid areas 1-9 [14].

undergoes attenuation, reducing contrast and a blurred image as it travels through water. Absorption impacts visibility while scattering alters the direction of the light path.

Fig. 1 shows a graph of the absorption coefficient of water with respect to the wavelength of light. As light travels further, red wavelengths (625 nm–700 nm) get absorbed 100 times more than blue wavelengths (380 nm–540 nm) [12]. As seen in Fig. 2, blue wavelengths scatter most in the visible spectrum (380 nm–750 nm) [14]. As a consequence, underwater images are tinted with blue and green hues. Moreover, ‘veiling luminance’ originating from the scattering centers produces photon flux much more than that coming from the target subject [15]. These factors make underwater imaging demanding and conventional cameras past a certain optical attenuation length ( $cz$ ) are impractical. Recent hardware advancements in lasers, HMI lighting, and high dynamic range sensors have led to advanced Range-Gated Imaging Systems [16], Laser Field Synchronous Scanning [17], Pulsed Laser Illuminator [18], and many more. However, R&D of hardware technology can be expensive and demanding.

To this end, the development of practical techniques for underwater image enhancement has emerged as a critical area of research within the realm of Image Processing and Computer Vision. This paper presents Sea-Pix-GAN, a novel underwater image enhancement

technique inspired by the U-Net architecture and adversarial neural networks. Sea-Pix-GAN is designed to robustly denoise, enhance, and restore the original colors of underwater images. We emphasize the contributions of this work as follows:

1. We present a robust adversarial neural network capable of learning to enhance underwater images from examples while eliminating the reliance on heuristics and addressing acclimation to various settings.
2. We propose pixel-to-pixel transfer of style and content using an encoder-decoder architecture. We show that incorporating conditional GAN with Euclidean loss in Generator and training it adversarially using a convolutional Markovian PatchGAN discriminator yields consistent results both qualitatively and quantitatively.
3. We have evaluated our approach against existing state-of-the-art image enhancement, restoration, and deep learning models on metrics like PSNR, SSIM, and UIQM. We also present examples to compare the results visually as well as interpret and analyze reconstruction using histograms.

The paper is organized as follows: Section 2 reviews related work in underwater image enhancement, Section 3 presents the proposed model and its training details, Section 4 showcases experimental results and analysis, and finally, Section 5 concludes the findings and outlines future research directions.

## 2. Related work

Image processing is a extensively studied field with notable efforts directed toward enhancing underwater images through various techniques. These techniques can be broadly classified under three broad categories: (a) *image enhancement*, (b) *image restoration*, and (c) *deep learning based*.

**Image enhancement** is a heuristic process whose main objective is to give a better visual representation. The process is subjective, resulting in considerable output variations. In many cases, there is no need for a mathematical degradation model; instead, transformation techniques such as contrast stretching, gaussian stretching, and histogram equalization are commonly employed. Some notable image enhancement techniques are Histogram Equalization (HE) [19], Contrast Limited Adaptive Histogram Equalization (CLAHE) [20], Integrated Colour Model (ICM) [21], Unsupervised Colour Correction Method (UCM) [22], Rayleigh Distribution [23], Relative Global Histogram Stretching (RGHS) [24], and Fusion-based Correction [25]. Both HE and CLAHE are variants of histogram equalization that deal with the adjustment of contrast based on the tonal distribution of the image. CLAHE differs from ordinary HE in the respect that it is in fact a variant of adaptive histogram equalization that computes several histograms corresponding to different sections of the image and prevents overamplified noise by limiting contrast amplification. However, CLAHE suffers from noise amplification in the flat region and ring artifacts at strong edges [26]. Iqbal et al. proposed ICM [21] that followed a two-fold approach based on slide stretching. Contrast stretching is applied to equalize the color contrast and saturation, and intensity stretching of Hue, Saturation, and Intensity (HSI) is used to increase the true color. Iqbal et al. proposed UCM [22] that modified the red and green color channels, gleaned from Von Kries hypothesis,<sup>2</sup> to scale down the color cast and then applied contrast correction on the color model (RGB). We observed that while ICM and UCM improve contrast, they tend to produce darker pictures which cause a loss of details in several areas. In some cases, images and artifacts suffer oversaturation and illumination effects. Moreover, in most cases, the blue-green hue pertains, and other colors still appear washed out. This

<sup>1</sup> [https://water.lsbu.ac.uk/water/water\\_vibrational\\_spectrum.html](https://water.lsbu.ac.uk/water/water_vibrational_spectrum.html).

<sup>2</sup> [https://en.wikipedia.org/wiki/Von\\_Kries\\_coefficient\\_law](https://en.wikipedia.org/wiki/Von_Kries_coefficient_law).

concurs with the discussions of Ghani et al. [27]. RGHS is based on adaptive parameter acquisition that performs contrast correction in green and blue channels and then redistributes histogram with dynamic parameters based on wavelength attenuation of different colors. While RGHS does not suffer from artifacts, we observed that blue-green tinting is still a significant problem. Ancuti et al. [25] proposed a multi-scale fusion-based approach where they fused contrast-improved and color-corrected images. Visibility and global contrast were improved, but some image regions appeared over-enhanced and some under-enhanced.

**Image restoration** techniques focus on mitigating the effects of the sensing environment, specifically attenuation and scattering in our context. This is an objective process where a mathematical degradation model is formulated based on prior knowledge about the scene. Image processing techniques such as inverse filtering,<sup>3</sup> Wiener filtering,<sup>4</sup> and denoising are utilized to achieve restoration. Some notable image restoration techniques are Dark Channel Prior (DCP) [28], Maximum Intensity Prior (MIP) [29], Removal of Water Scattering (RoWS) [30], Low Complexity DCP [31], Underwater Dark Channel Prior (UDCP) [32], Image Restoration Based on Image Blurriness and Light Absorption (IBLA) [33], and Underwater Light Attenuation Prior (ULAP) [34]. As suggested by the names, these methods are based on generating image priors. Image priors are preliminary information about our images that can be used for choosing image processing parameters. The image prior combined with the degradation model can help formulate a restoration algorithm called Image Formation Model (IFM). Kaiming et al. [28] proposed a dark channel before removing haze and used it with the haze imaging model. It is based on the observation that most local patches in haze-free outdoor images contain some pixels which have very low intensities in at least one color channel. However, not all underwater images have a vast variance in attenuation between color channels. DCP has many variations [29–32] that also follow this observation and hence suffer from similar limitations. Our experimentations with DCP and MIP show green channel bleeding, and Low Complexity DCP and UDCP show red channel bleeding in some cases. However, contrast and texture were improved. Peng et al. [33] proposed a depth estimation method based on image blurriness and light absorption. While our experiments showed no color bleeding in IBLA, the colors often appeared washed off and peculiar. Song et al. [34] proposed a prior based on light attenuation and used a learning-based regressor for the model coefficient.

All of the techniques discussed above suffer from some ubiquitous issues. The algorithms fail to recover the lost information about color in almost all cases. The contrast and exposure are erratic in some cases. Generation of artifacts is common. The complexity of the algorithms explodes when high-resolution pictures are provided as input.

**Deep Learning Based** There have been some recent efforts in underwater imaging from deep learning and computer vision angle. The development in computational capacity over the last two decades fueled with a large amount of data spewed out over the period has given rise to a multitude of possibilities [35]. In addition, the inexpensive cloud services have eased the cost and toil of training neural networks. Liu et al. [36] proposed LANet, an adaptive learning attention network for underwater image enhancement through supervised learning. Their work encompasses a multiscale fusion module, a novel parallel attention module (PAM), and an adaptive learning module (ALM), addressing degradation issues by combining spatial information, enhancing illuminated features, and adaptively learning important feature information. Sharma et al. [37] introduced Deep WaveNet with two critical innovations for underwater image restoration. The first is the effective use of receptive field sizes based on color channel traversing, significantly improving UIR performance. The second innovation

involves an attentive skip mechanism for refining multi-contextual features, enhancing the proposed Deep WaveNet framework. Several other deep learning based architectures such as RauneNet,<sup>5</sup> WaterNet [38], UGAN [39], and FunieGAN [40], have also demonstrated good results over previous enhancement and restoration benchmarks. Recently, Ren et al. [41], Cai et al. [42], and Zhao et al. [43] tried to address the limitations of conventional methods that often struggled with erratic estimation errors in image priors due to lack of generalized assumptions. Though they worked well with contrast improvement, they could not address the influence of scattering as they lacked underwater image training sets and made an assumption that all the channels (RGB) have the same transmission. Shin et al. [44] tried to solve this issue by generating numerous patches with several color casts for training the CNN. However, underwater images only have a blue-green tone.

The primary objective of this paper is to address the constrained robustness observed in traditional image processing and learning-based architectures by leveraging the adaptability of Adversarial Neural Networks. Our approach entails employing a pixel-to-pixel transfer with the utilization of a conditional GAN, which departs from the conventional practice of assuming the characteristics of the transmission medium.

### 3. Proposed methodology

#### 3.1. The adversarial network

Sea-Pix-GAN, at its core, is a Generative Adversarial Network (GAN), a type of generative model consisting of two players: the Generator and the Discriminator. These players are trained simultaneously within an adversarial framework, forming a two-player minimax game. In other words, the Generator takes a random noise vector ‘z’ as input and aims to generate an enhanced underwater image ‘y’, denoted as  $G: z \rightarrow y$ . The Generator strives to learn the underlying data distribution and produce images that are indistinguishable from real underwater images. On the other hand, the Discriminator, labeled as D, attempts to distinguish between the original image data and the images generated by the Generator. Specifically, we are using conditional GANs (cGAN) [45], inspired by pix2pix [46], and U-net [47] architectures. We have an input image ‘x’ (raw underwater image) and a noise vector ‘z’. We train the network to learn the mapping  $G: x, z \rightarrow y$ . While we want the Generator to produce more “real” fake images, we also require the Discriminator to be good at recognizing fake images, hence the minimax game. If any of them gets better than the other, the game virtually terminates, and the training reaches saturation. We ought to avoid this condition to achieve better results. Mathematically, the cGAN objective function can be expressed as:

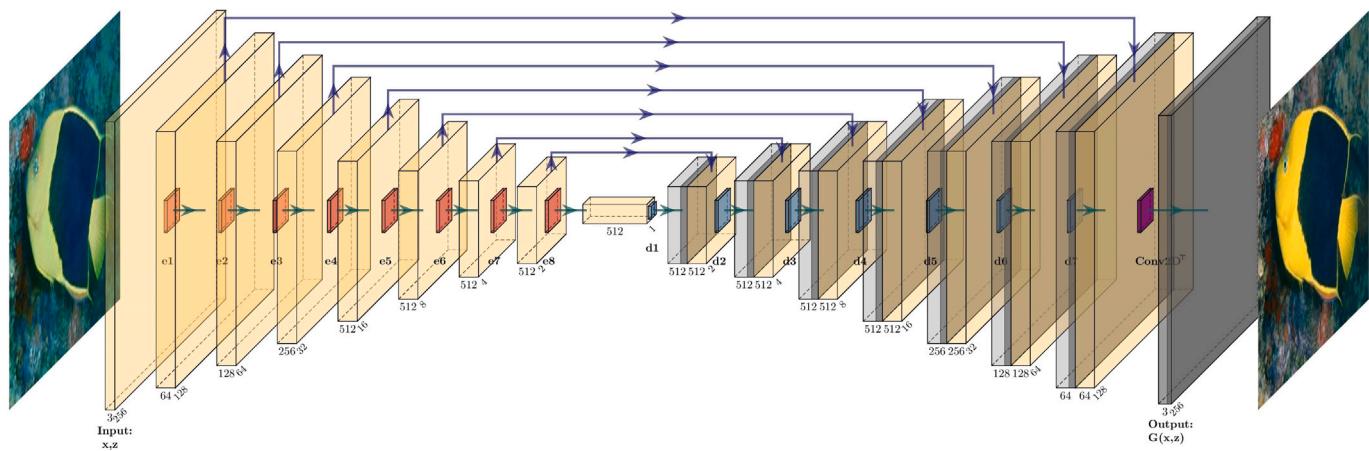
$$\min_G \max_D L_C^{GAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \quad (1)$$

where the generator G tries to minimize this objective against an adversarial D that tries to maximize it. The first term  $\mathbb{E}_{x,y}[\log D(x, y)]$  represents the expected value of the logarithm of the discriminator’s probability  $D(x, y)$  with respect to the joint probability distribution of the raw underwater image x and their corresponding conditional information (enhanced underwater image) y. It encourages the discriminator to output a high probability for real underwater images  $D(x, y)$  and penalizes it when it fails to do so. It quantifies how well the discriminator distinguishes real images from generated ones. The second term,  $\mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]$  represents the expected value of the logarithm of one minus the discriminator’s probability  $(1 - D(x, G(x, z)))$  with respect to the joint probability distribution of the raw underwater image x and the generated underwater image  $G(x, z)$ .

<sup>3</sup> [https://en.wikipedia.org/wiki/Inverse\\_filter](https://en.wikipedia.org/wiki/Inverse_filter).

<sup>4</sup> [https://en.wikipedia.org/wiki/Wiener\\_filter](https://en.wikipedia.org/wiki/Wiener_filter).

<sup>5</sup> <https://github.com/fansuregrin/RAUNE-Net>.



**Fig. 3.** The detailed architecture of Sea-Pix-GAN Generator showing encoder-decoder network with skip connections.

obtained by the generator given noise ‘z’. It encourages the generator to produce images that are classified as “real” by the discriminator. It aims to maximize the probability of the discriminator incorrectly classifying the generated images as real. In other words, it encourages the generator to produce images that are indistinguishable from real ones. The loss function mentioned in Eq. (1) would have been sufficient if style transfer was the objective. However, in underwater imaging, it is of utmost importance for the output image to be close to the ground truth. Hence, incorporating traditional Euclidean loss in the Generator is essential. It follows from [46,48] that L1 loss performs better than L2 with cGANs. L1 loss for the generator G can be mathematically expressed as:

$$L_{L1}(G) = \mathbb{E}_{x,y,z} [|y - G(x, z)|] \quad (2)$$

Hence our final objective function for Sea-Pix-GAN becomes:

$$\begin{aligned} G^* = \min_G \max_D & \mathbb{E}_{x,y} [\log D(x, y)] \\ & + \mathbb{E}_{x,z} [\log(1 - D(x, G(x, z)))] \\ & + \lambda (\mathbb{E}_{x,y,z} [|y - G(x, z)|]) \end{aligned} \quad (3)$$

where  $\lambda$  is a tradeoff parameter. It may be chosen by experimenting with the dataset and algorithm. In this paper, we take  $\lambda = 100$ . The architectures of Generators and Discriminators are inspired by pix2pix [46], U-nets [47], and DCGAN [49]. We have used an encoder-decoder network with dropout layers and skip connections in the Generator. The Discriminator is a PatchGAN responsible for penalizing at the local image patch level. They are discussed in detail in the following subsections.

### 3.2. The generator

If we consider underwater enhancement as an image-to-image translation problem, the underlying structure of each pixel is same at the core. We design our Generator based on this concept. The Generator is a series of Convolution-BatchNorm-ReLU\* as encoder network and a series of Convolution-BatchNormDropout-ReLU\* layers as decoder network with a series of skip connections between mirrored layers  $i$  and  $n-i$ . As suggested in [46,50], and [51], skip connections shows promising results in image-to-image translation.

The encoder and decoder can be viewed as downsampling and upsampling stacks, respectively. We employ a leakyReLU with a slope of 0.2 in the encoder, while the decoder uses the standard ReLU activation function. In downsampling layers of the encoder, 2D Convolutions with a filter size of  $4 \times 4$  and strides of 2 are applied, followed by Batch Normalization and leakyReLU non-linearity. The first block in the encoder is not batch normalized. Whereas in upsampling layers of

decoder, 2D Deconvolutions (Conv2DTranspose) with the same filter size of  $4 \times 4$  and strides of 2 are applied, followed by Batch Normalization. The decoder has a dropout of 50% in the first three blocks only. The input to the network is an RGB channel  $256 \times 256$  image ( $256 \times 256 \times 3$ ). The encoder network (e1-e8) downsamples in the order  $(256 \times 256 \times 3) \rightarrow (128 \times 128 \times 64) \rightarrow (64 \times 64 \times 128) \rightarrow (32 \times 32 \times 256) \rightarrow (16 \times 16 \times 512) \rightarrow (8 \times 8 \times 512) \rightarrow (4 \times 4 \times 512) \rightarrow (2 \times 2 \times 512) \rightarrow (1 \times 1 \times 512)$ . We are finally left with 512 feature maps of dimension  $1 \times 1$ . The decoder network (d1-d7) then makes use of the feature maps and skip connections as shown in Fig. 3 to learn and generate a layer of dimension  $(128 \times 128 \times 128)$ . The decoder upsamples in the order  $(1 \times 1 \times 512) \rightarrow (2 \times 2 \times 1024) \rightarrow (4 \times 4 \times 1024) \rightarrow (8 \times 8 \times 1024) \rightarrow (16 \times 16 \times 1024) \rightarrow (32 \times 32 \times 512) \rightarrow (64 \times 64 \times 256) \rightarrow (128 \times 128 \times 128)$ . Then, deconvolution (Conv2DTranspose) is applied to map the tensor to the original shape of the image, i.e.,  $(256 \times 256 \times 3)$ .

### 3.3. The discriminator

As mentioned earlier, we cannot rely entirely on Euclidean losses because it leads to blurry results, which can be imputed to its inability to capture high frequencies on image-to-image translation problems. However, they are accurate at capturing low frequencies. Moreover, since we have already included an L1 loss in the Generator, the Discriminator can be curated for probing only high frequencies. So, we decided to incorporate a convolutional Markovian PatchGAN. Instead of considering each pixel, the Discriminator evaluates cross-entropy at the level of patches. In simple words, it tells whether the small patches are “real” or not. The inputs to the Discriminator are two images: generated by the Generator  $G(x, z)$  and the target image  $y$ , each of dimension  $256 \times 256 \times 3$ . In Fig. 4, the down layers are identical to the downsampling layers of the encoder network that consists of 2D Convolutions with a filter size of  $4 \times 4$  and strides of 2 followed by Batch Normalization and Leaky-ReLU non-linearity. The Conv2D layers have a filter size of  $4 \times 4$  and strides of 1. The two input images are first concatenated to obtain a tensor of dimension  $256 \times 256 \times 6$ . As shown in Fig. 4, seven convolutional layers are used to transform the tensor into a  $30 \times 30 \times 1$  output. Each block of the  $30 \times 30$  tensor is responsible for classifying a  $70 \times 70$  patch in the original image, as shown in Fig. 5.

### 3.4. Training the network

The Generator loss function is defined as Eq. (3) in the algorithm proposition section. It can be summarized as:

$$G^* = L_C^{GAN} + \lambda L_{L1} \quad (4)$$

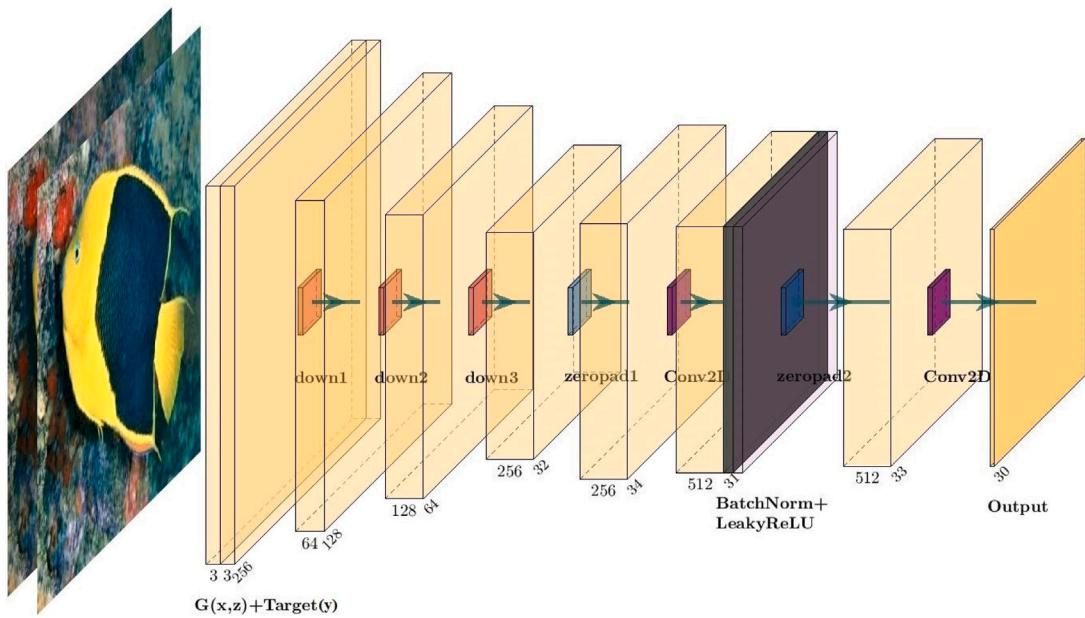


Fig. 4. The detailed architecture of Sea-Pix-GAN Discriminator showing downsampling layers.

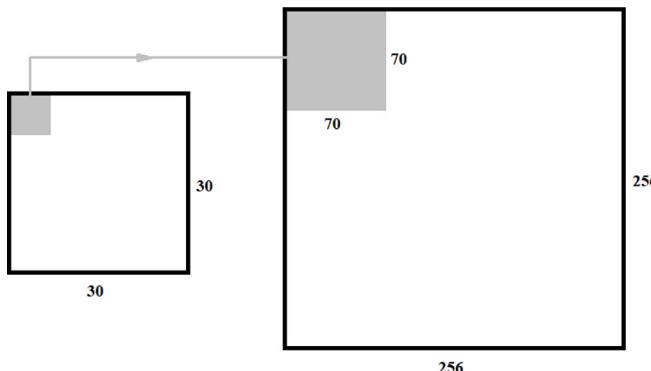


Fig. 5. Markovian PatchGAN with size  $30 \times 30$  responsible for classifying a  $70 \times 70$  patch in the original image.

Fig. 6(a) is the flowchart of training algorithm of the Generator. In simple words,  $L_C^{GAN}$  is the binary cross entropy of the generated image and a tensor of ones. As discussed earlier, mean absolute error ( $L_1$  loss) is also incorporated to capture low frequencies. Lambda ( $\lambda$ ) is the tradeoff factor with which we can experiment. We have chosen  $\lambda = 100$ . Fig. 6(b) illustrates the training algorithm of the Discriminator. The Discriminator takes in both the generated image and the ground truth to evaluate the loss function. Hence, there are two separate losses, the generated loss, and the real loss. The prior is a binary cross-entropy of the generated image  $G(x,z)$  a tensor of zeros. The latter is the binary cross-entropy of the ground truth image and a tensor of 1s. The final Discriminator loss is the sum of these two losses.

For preprocessing, we first resized each image to  $256 \times 256$ , using the nearest neighbor method, to maintain consistency with the network architecture. Then we added random jitter to images using random cropping and random mirroring. All the image tensors were normalized to the range  $[-1, 1]$ . The Generator and Discriminator losses were calculated alternatively. After calculating the gradients of these losses, they were fed to the optimizers that updated weights accordingly. We used the Adam Optimizer with a learning rate of  $2 \times 10^{-4}$  and beta<sub>1</sub> of 0.5 to apply gradients in both the Generator and the Discriminator. The batch size throughout the experiment was 64. We used Google

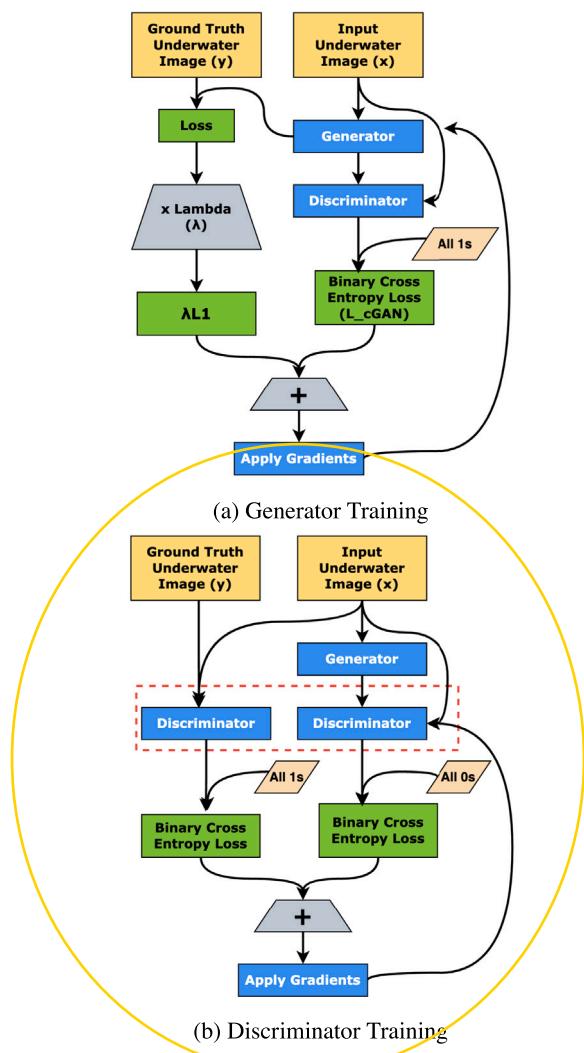
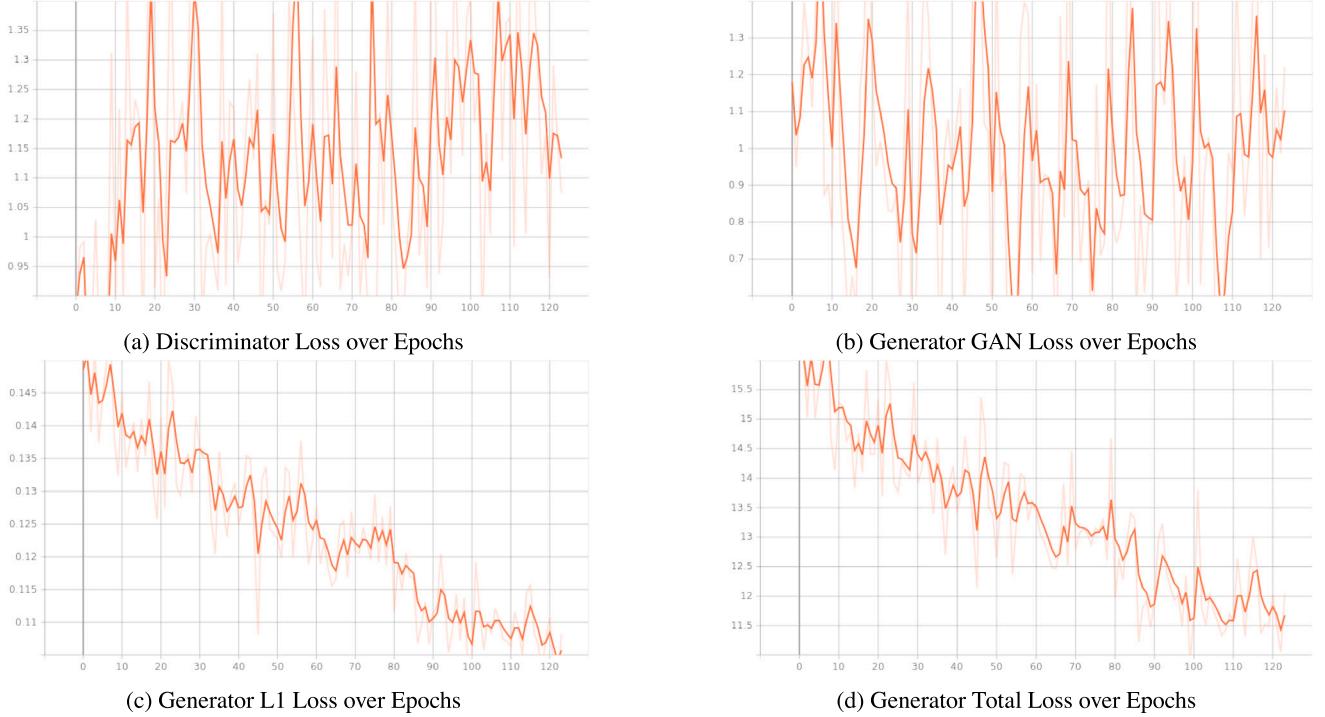


Fig. 6. Training of Generator and Discriminator modules.



**Fig. 7.** Loss curves for Generator and Discriminator, x-axis represents epochs and y-axis represents the loss.

Colab Pro for training our model. It offers NVIDIA Tesla P100 PCI-E with 16 GB VRAM and 24.4 GB System Memory. It allowed us to train at a pace of about 100 s/epoch. We trained for a total of 150 epochs, given the diversity of the dataset.

### 3.5. Dataset

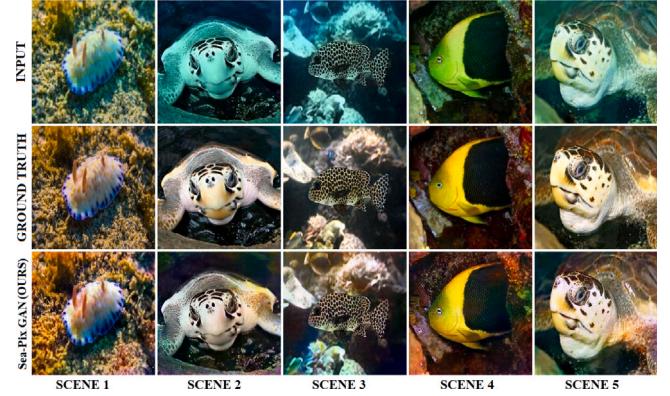
We used the Enhancing Underwater Visual Perception (EUVP) Dataset [40]. It comprises an extensive collection of paired and unpaired underwater images. They used seven different cameras, including multiple GoPros, Aqua AUV's uEye cameras, low-light USB cameras, and Trident ROV's HD camera, to capture images for the dataset. The dataset was thoughtfully designed to incorporate a large assortment of natural variability. The dataset is composed of three subsets Underwater Dark, Underwater ImageNet, and Underwater Scenes having 5500, 3700, and 2185 training pairs, respectively. We split the collective 11 385 pairs in the ratio of 4:1:1 for training, validation, and testing.

## 4. Results and analysis

We first discuss the graphs plotted during training to illustrate the learning process. Then, we evaluate the performance of the proposed Sea-Pix-GAN model both qualitatively and quantitatively. After analyzing the performance, we compare it to several state-of-the-art enhancement and restoration techniques discussed in Section 2.

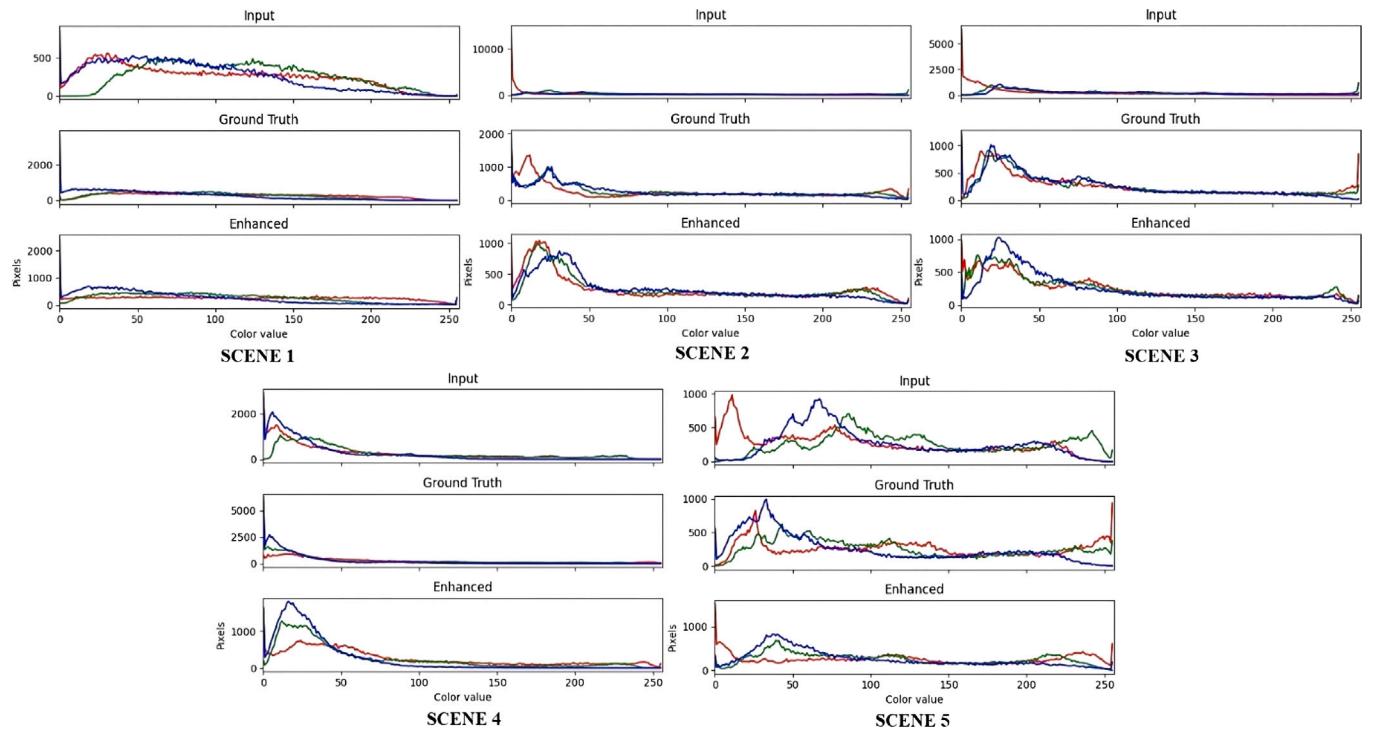
### 4.1. The training progress

Being an Adversarial Neural Network, Sea-Pix-GAN's loss function is different than conventional Neural Networks. While we would want the losses to decrease in conventional Neural Networks, in conditional GANs, we want the Generator and Discriminator to be adversaries. We do not want any player defeating the other for optimal training. This idea is reflected in the graphs of Discriminator Loss and Generator GAN Loss  $L_C^{GAN}$ , Eq. (1), as seen in Figs. 7(a) and 7(b), respectively. The Discriminator loss quantifies its ability to differentiate between real and generated images. Over epochs, we expect this loss



**Fig. 8.** Images generated by our model compared to ground truth.

to rise, indicating that the Generator is becoming increasingly proficient at creating enhanced underwater images. Fig. 7(a) illustrates that, although the Discriminator loss exhibits fluctuations due to the concurrent training of both networks, the local minima in the loss graph steadily ascend. This observation suggests that while the Discriminator learns from its mistakes, the task of distinguishing generated samples from real ones becomes progressively challenging. Conversely, the Generator GAN loss assesses the Generator's efficacy in deceiving the Discriminator into accepting the generated underwater images as genuine. Over time, we anticipate this loss to decline, signifying the Generator's improving performance. Fig. 7(b) demonstrates that, despite loss fluctuations due to the joint training of the Generator and Discriminator, the local minima in the Generator loss graph consistently decrease. This trend indicates that the Generator is learning to produce superior enhanced underwater images. This philosophy helps us keep improving the performance of both networks over epochs of training. However, as in Fig. 7(c), the L1 Loss  $L_{L1}$ , Eq. (2), goes down over the process, which is responsible for low frequency noise. This ensures that the images generated by the Generator are closer to the ground truth.



**Fig. 9.** RGB Histogram comparison of input (noisy), ground truth (reference) and enhanced images (generated by Sea-Pix-GAN). The solid blue, green and red curves denote the corresponding color pixel counts in the histogram.

The gradual decrease in L1 Loss is also reflected in the total loss  $G^*$ , Eq. (3), as implied by Fig. 7(d).

#### 4.2. Qualitative analysis

We used five scenes for sample demonstration from the test set (images not seen during training) of our dataset. Fig. 8 shows the input images, the ground truths, and the images generated by our trained Sea-Pix-GAN model in three rows, respectively (top to bottom). As evident from Fig. 8, our model generated images that have significantly improved color, contrast, and sharpness. The blue-green tint is corrected, and the images are comparable to the ground truth. This fact is supported by the RGB histograms of the images corresponding to each scene. Fig. 9 shows the histograms of the 5 test scenes as labeled. The x-axis is the color value, and the y-axis is the distributed pixel count. Each of the five diagrams contains three different histograms in three consecutive rows corresponding to Input, Ground Truth, and Enhanced Images. The histograms show discrepancies between input images and corresponding ground truth in all three channels (RGB). As evident from the second and third rows of each scene, the enhanced images have significantly fewer discrepancies in all three channels. The histograms of Scene 2 and Scene 4 show that even though enhanced images have a higher pixel count for certain color values, the pattern is still aligned.

To better benchmark the performance, we conducted an experiment in which we compared Sea-Pix-GAN with established as well as recent underwater imaging techniques. We considered six enhancement algorithms: Histogram Equalization (HE) [19], Contrast Limited Adaptive Histogram Equalization (CLAHE) [20], Integrated Colour Model (ICM) [21], Unsupervised Colour Correction Method (UCM) [22], Rayleigh Distribution [23], and RGHS [24] and seven restoration algorithms: Dark Channel Prior (DCP) [28], Maximum Intensity Prior (MIP) [29], Removal of Water Scattering (RoWS) [30], Low Complexity DCP [31], Underwater Dark Channel Prior (UDCP)

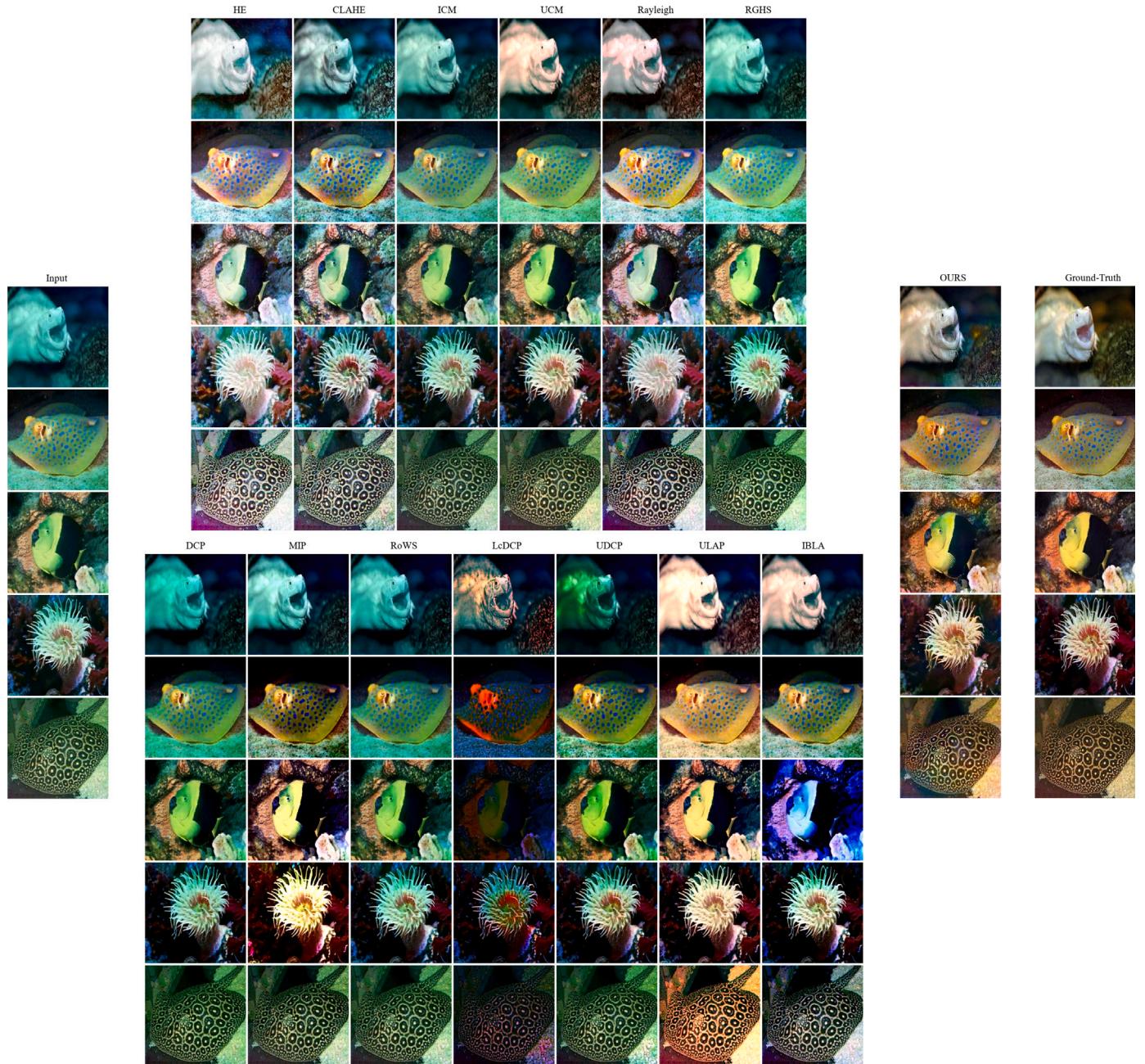
[32], Image Restoration Based on Image Blurriness and Light Absorption (IBLA) [33], and Underwater Light Attenuation Prior (ULAP) [34]. Further we compared our approach with six deep learning based approaches: LANet [36], Deep WaveNet [37], RAUNE-Net, WaterNet [38], UGAN [39], FunieGAN [40].<sup>6</sup> We chose five images at random from our EUVP test dataset [40] and fed them to all the models. Fig. 10 shows the obtained results. It can be observed that Sea-Pix-GAN struggles significantly less in enhancing across a variety of settings. HE suffers from washed colors and low contrast. While CLAHE, ICM, RGHS, DCP, MIP, and RoWS seems to improve contrast in general, the scattering of blue wavelength is still prominent. LcDCP struggles with global brightness and often produces erratic colors. UDCP shows a green hue bleeding effect. ULAP seems to perform consistently in comparison, but global contrast and color dynamics are erroneous. IBLA seems to suffer from the same problem as LcDCP in certain scenarios. In contrast, Sea-Pix-GAN seems to produce well-balanced images. The colors are accurate, the objects have decent contrast, and there is significantly less blue-green hue.

#### 4.3. Quantitative analysis

We continued the experiment by benchmarking our Sea-Pix-GAN against the enhancement, restoration, and deep learning-based models based on three different metrics: Peak signal-to-noise ratio (PSNR) [52], Structural Similarity Index (SSIM) [52,53], and Human-Visual-System-Inspired Underwater Image Quality Measures (UIQM) [54]. PSNR in dB is computed as:

$$PSNR(x, y) = 10 \log_{10} \frac{MAX^2}{MSE(x, y)} dB, \quad (5)$$

<sup>6</sup> Implementations of RAUNE-Net, WaterNet, UGAN, and FunieGAN are taken from <https://github.com/fansuregrin/RAUNE-Net>.



**Fig. 10.** Qualitative comparison of Sea-Pix-GAN with enhancement models: HE [19], CLAHE [20], ICM [21], UCM [22], Rayleigh Distribution [23], RGHS [24] and restoration models: DCP [28], MIP [29], RoWS [30], LcDCP [31], UDPC [32], IBLA [33], and ULAP [34].

where MAX=255 (maximum value of a pixel in a general RGB image) and MSE is the mean squared error between pixel values of the input(x) and reference images(y). It quantifies the reconstruction quality of noisy input. SSIM is computed as:

$$SSIM(x, y) = \frac{(2\mu_a\mu_b + c_1)(2\sigma_{AB} + c_2)}{(\mu_A^2 + \mu_B^2 + c_1)(\sigma_A^2 + \sigma_B^2 + c_2)}, \quad (6)$$

where  $\mu_A$  and  $\mu_B$  are mean values,  $\sigma_{AB}$  is the covariance, and  $\sigma_A$  and  $\sigma_B$  are the standard deviations. All of them are the functions of x and y. UIQM is computed as:

$$UIQM = 0.0282 \times UICM + 0.2953 \times UISM + 3.3753 \times UIConM, \quad (7)$$

where UICM is Underwater Image Colorfulness Measure, UISM is Underwater Image Sharpness Measure, and UIConM is Underwater Image

Contrast Measure. Evident from the names, it incorporates color, sharpness, and contrast of the image. Unlike PSNR and SSIM, it does not need a reference image.

Due to the high computational complexity of some models, we limited our evaluation to 500 randomly selected paired images from the test dataset. The mean values  $\pm$  one standard deviation for the metrics are presented in Table 1. The results in Table 1 demonstrate the consistent performance of our model across all three metrics. While certain models may slightly outperform Sea-Pix-GAN in specific metrics, none excel in all of them. For instance, ICM surpasses Sea-Pix-GAN by 0.257% in terms of PSNR, but Sea-Pix-GAN outperforms ICM by 3.947% and 7.984% in SSIM and UIQM, respectively. Additionally, our model exhibits the least variability in scores, indicating its stability. Notably, Sea-Pix-GAN achieves the highest UIQM of 2.84, which is in close proximity to the UIQM of the ground truth (2.89). Based on these

**Table 1**

Quantitative comparison of Sea-Pix-GAN with enhancement and restoration techniques based on average PSNR, SSIM and UIQM on 500 paired test set.

	Model	PSNR ( $\mu \pm \sigma$ )	SSIM ( $\mu \pm \sigma$ )	UIQM ( $\mu \pm \sigma$ )
Enhancement	HE	16.81 ± 4.32	0.63 ± 0.15	2.82 ± 0.30
	CLAHE	19.11 ± 2.01	0.69 ± 0.08	2.81 ± 0.34
	ICM	<b>23.36 ± 3.41</b>	0.76 ± 0.08	2.63 ± 0.43
	UCM	21.80 ± 2.99	0.77 ± 0.08	2.74 ± 0.38
	Rayleigh distribution	19.35 ± 3.34	0.67 ± 0.11	2.83 ± 0.35
	RGHS	21.97 ± 3.71	0.72 ± 0.08	2.34 ± 0.37
Restoration	DCP	21.83 ± 2.16	0.66 ± 0.09	2.02 ± 0.46
	MIP	19.11 ± 3.89	0.60 ± 0.15	1.67 ± 0.64
	RoWS	22.38 ± 2.09	0.68 ± 0.08	2.09 ± 0.49
	Low complexity DCP	14.00 ± 4.02	0.48 ± 0.15	1.51 ± 0.49
	UDCP	18.03 ± 3.58	0.62 ± 0.09	1.94 ± 0.44
	IBLA	19.95 ± 4.79	0.59 ± 0.15	1.80 ± 0.55
	ULAP	20.74 ± 2.47	0.67 ± 0.09	1.91 ± 0.55
Deep learning based	LANet	21.97 ± 3.73	0.76 ± 0.12	2.74 ± 0.36
	Deep WaveNet	22.49 ± 3.35	<b>0.79 ± 0.11</b>	2.83 ± 0.24
	RAUNE-Net	21.99 ± 2.97	0.76 ± 0.09	2.80 ± 0.22
	WaterNet	22.33 ± 3.15	0.75 ± 0.10	2.66 ± 0.27
	UGAN	21.24 ± 2.59	0.75 ± 0.08	2.82 ± 0.20
	FunieGAN	20.49 ± 2.33	0.70 ± 0.09	2.78 ± 0.23
Ours	<b>Sea-Pix-GAN</b>	23.30 ± 1.68	<b>0.79 ± 0.09</b>	<b>2.84 ± 0.20</b>

findings, we confidently assert that Sea-Pix-GAN consistently and commendably enhances underwater images.

## 5. Conclusions and outlook

Sea-Pix-GAN is a model of simplicity and elegance, effectively learning to enhance and restore the color, contrast, and texture of hazy underwater images through an engaging mini-max game with a virtual image administrator. Our model's objective function incorporates style transfer guided by L1 loss, while our Discriminator PatchGAN accurately captures high-frequency details, leading to precise image translation. The experimental results obtained from an unseen dataset strongly indicate that the generated images faithfully represent the ground truth. This conclusion finds additional support from histogram analysis and quantitative comparisons. Considering the scale of our training, our method performs commendably well. As we continue to train Sea-Pix-GAN with a more diverse dataset, its performance will undoubtedly improve further. Moreover, the model's computational efficiency sets it apart from existing techniques, rendering it suitable for real-time video enhancement. In the future, we aspire to scale this project to enhance high-resolution underwater images. We firmly believe that Sea-Pix-GAN's computational efficiency, cost-effectiveness, and robustness will significantly contribute to the advancement of underwater research.

## Declaration of competing interest

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

## Data availability

The data is open source and can be downloaded from the respective authors website.

## References

- [1] Prateek Chhikara, Anil Goyal, Chirag Sharma, RE-tagger: A light-weight real-estate image classifier, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2022, pp. 627–630.
- [2] Chenggang Yan, Lixuan Meng, Liang Li, Jiehua Zhang, Zhan Wang, Jian Yin, Jiayong Zhang, Yaoqi Sun, Bolun Zheng, Age-invariant face recognition by multi-feature fusion and decomposition with self-attention, *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* 18 (1s) (2022) 1–18.
- [3] Chenggang Yan, Biao Gong, Yuxuan Wei, Yue Gao, Deep multi-view enhancement hashing for image retrieval, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (4) (2020) 1445–1451.
- [4] Peixian Zhuang, Jiamin Wu, Fatih Porikli, Chongyi Li, Underwater image enhancement with hyper-laplacian reflectance priors, *IEEE Trans. Image Process.* 31 (2022) 5442–5455.
- [5] Smitha Raveendran, Mukesh D. Patil, Gajanan K. Birajdar, Underwater image enhancement: a comprehensive review, recent trends, challenges and applications, *Artif. Intell. Rev.* 54 (2021) 5413–5467.
- [6] Weidong Zhang, Peixian Zhuang, Hai-Han Sun, Guohou Li, Sam Kwong, Chongyi Li, Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement, *IEEE Trans. Image Process.* 31 (2022) 3997–4010.
- [7] Chenggang Yan, Yiming Hao, Liang Li, Jian Yin, Anan Liu, Zhendong Mao, Zhenyu Chen, Xingyu Gao, Task-adaptive attention for image captioning, *IEEE Trans. Circuits Syst. Video Technol.* 32 (1) (2021) 43–51.
- [8] Chenggang Yan, Tong Teng, Yutao Liu, Yongbing Zhang, Haoqian Wang, Xiangyang Ji, Precise no-reference image quality evaluation based on distortion identification, *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* 17 (3s) (2021) 1–21.
- [9] Chenggang Yan, Zhisheng Li, Yongbing Zhang, Yutao Liu, Xiangyang Ji, Yongdong Zhang, Depth image denoising using nuclear norm and learning graph model, *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* 16 (4) (2020) 1–17.
- [10] Lintao Peng, Chunli Zhu, Liheng Bian, U-shape transformer for underwater image enhancement, *IEEE Trans. Image Process.* (2023).
- [11] Rod Nave, HyperPhysics, Georgia State University, Dept. of Physics and Astronomy, 2000-, Atlanta, Ga., 2017.
- [12] Krzysztof B. Beć, Christian W. Huck, Breakthrough potential in near-infrared spectroscopy: Spectra simulation. A review of recent developments, *Front. Chem.* 7 (2019) 48.
- [13] Curtis D. Mobley, Charles D. Mobley, Light and Water: Radiative Transfer in Natural Waters, Academic Press, 1994.
- [14] Laura J. Johnson, Optical Property Variability in the Underwater Optical Wireless Channel (Ph.D. thesis), University of Warwick England, 2015.
- [15] Frank Michael Caimi, Underwater effects, in: Katsushi Ikeuchi (Ed.), Computer Vision: A Reference Guide, Springer US, Boston, MA, 2014, pp. 831–837.
- [16] Patrizio Mariani, Iñaki Quinceo, Karl H. Haugolt, Yves Chardard, Andre W. Visser, Chris Yates, Giuliano Piccinno, Giancarlo Reali, Petter Risholm, Jens T. Thielemann, Range-gated imaging system for underwater monitoring in ocean environment, *Sustainability* 11 (1) (2019) 162.
- [17] Houde Wu, Ming Zhao, Wenhui Xu, Underwater de-scattering imaging by laser field synchronous scanning, *Opt. Lasers Eng.* 126 (2020) 105871.
- [18] Yves Lutz, Nicolas Metzger, Near infrared laser illuminator for very long-range flash active imaging applications, in: *Illumination Optics II*, Vol. 8170, International Society for Optics and Photonics, 2011, p. 81700C.
- [19] Komal Vij, Yaduvir Singh, Enhancement of images using histogram processing techniques, *Int. J. Comp. Tech. Appl.* 2 (2) (2009) 309–313.

- [20] Karel Zuiderveld, Contrast limited adaptive histogram equalization, in: *Graphics Gems*, Academic Press, 1994, pp. 474–485.
- [21] Kashif Iqbal, Rosalina Abdul Salam, Azam Osman, Abdullah Zawawi Talib, Underwater image enhancement using an integrated colour model, *IAENG Int. J. Comput. Sci.* 34 (2) (2007).
- [22] Kashif Iqbal, Michael Odetayo, Anne James, Rosalina Abdul Salam, Abdullah Zawawi Hj. Talib, Enhancing the low quality images using unsupervised colour correction method, in: 2010 IEEE International Conference on Systems, Man and Cybernetics, IEEE, 2010, pp. 1703–1709.
- [23] Ahmad Shahrizan Abdul Ghani, Nor Ashidi Mat Isa, Underwater image quality enhancement through composition of dual-intensity images and Rayleigh-stretching, *SpringerPlus* 3 (1) (2014) 1–14.
- [24] Dongmei Huang, Yan Wang, Wei Song, Jean Sequeira, Sébastien Mavromatis, Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition, in: International Conference on Multimedia Modeling, Springer, 2018, pp. 453–465.
- [25] Cosmin Ancuti, Codruța Orniana Ancuti, Tom Haber, Philippe Bekaert, Enhancing underwater images and videos by fusion, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 81–88.
- [26] A. Vishwakarma, A. Mishra, Color image enhancement techniques: A critical review, 2012.
- [27] Ahmad Shahrizan Abdul Ghani, Nor Ashidi Mat Isa, Underwater image quality enhancement through Rayleigh-stretching and averaging image planes, *Int. J. Nav. Archit. Ocean Eng.* 6 (4) (2014) 840–866.
- [28] Kaiming He, Jian Sun, Xiaou Tang, Single image haze removal using dark channel prior, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1956–1963.
- [29] Nicholas Carlevaris-Bianco, Anush Mohan, Ryan M. Eustice, Initial results in underwater single image dehazing, in: *Oceans 2010 Mts/IEEE Seattle*, IEEE, 2010, pp. 1–8.
- [30] Liu Chao, Meng Wang, Removal of water scattering, in: 2010 2nd International Conference on Computer Engineering and Technology, Vol. 2, 2010, pp. V2–35–V2–39.
- [31] Hung-Yu Yang, Pei-Yin Chen, Chien-Chuan Huang, Ya-Zhu Zhuang, Yeu-Horng Shiau, Low complexity underwater image enhancement based on dark channel prior, in: 2011 Second International Conference on Innovations in Bio-Inspired Computing and Applications, IEEE, 2011, pp. 17–20.
- [32] Paul Drews, Erickson Nascimento, Filipe Moraes, Silvia Botelho, Mario Campos, Transmission estimation in underwater single images, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 825–830.
- [33] Yan-Tsung Peng, Pamela C. Cosman, Underwater image restoration based on image blurriness and light absorption, *IEEE Trans. Image Process.* 26 (4) (2017) 1579–1594.
- [34] Wei Song, Yan Wang, Dongmei Huang, Dian Tjondronegoro, A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration, in: Pacific Rim Conference on Multimedia, Springer, 2018, pp. 678–688.
- [35] Prateek Chhikara, Nikhil Jain, Rajkumar Tekchandani, Neeraj Kumar, Data dimensionality reduction techniques for Industry 4.0: Research results, challenges, and future research directions, *Softw. - Pract. Exp.* 52 (3) (2022) 658–688.
- [36] Shiben Liu, Huijie Fan, Sen Lin, Qiang Wang, Naida Ding, Yandong Tang, Adaptive learning attention network for underwater image enhancement, *IEEE Robot. Autom. Lett.* 7 (2) (2022) 5326–5333.
- [37] Prasen Sharma, Ira Bisht, Arijit Sur, Wavelength-based attributed deep neural network for underwater image restoration, *ACM Trans. Multimed. Comput. Commun. Appl.* 19 (1) (2023) 1–23.
- [38] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, Dacheng Tao, An underwater image enhancement benchmark dataset and beyond, *IEEE Trans. Image Process.* 29 (2019) 4376–4389.
- [39] Cameron Fabbri, Md Jahidul Islam, Junaed Sattar, Enhancing underwater imagery using generative adversarial networks, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, pp. 7159–7165.
- [40] Md Jahidul Islam, Youya Xia, Junaed Sattar, Fast underwater image enhancement for improved visual perception, *IEEE Robot. Autom. Lett.* 5 (2) (2020) 3227–3234.
- [41] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, Ming-Hsuan Yang, Single image dehazing via multi-scale convolutional neural networks, in: Bastian Leibe, Jiri Matas, Nicu Sebe, Max Welling (Eds.), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, 2016, pp. 154–169.
- [42] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, Dacheng Tao, Dehazenet: An end-to-end system for single image haze removal, *IEEE Trans. Image Process.* 25 (11) (2016) 5187–5198.
- [43] Xi Zhao, Keyan Wang, Yunsong Li, Jiaoqiao Li, Deep fully convolutional regression networks for single image haze removal, in: 2017 IEEE Visual Communications and Image Processing (VCIP), IEEE, 2017, pp. 1–4.
- [44] Young-Sik Shin, Younggun Cho, Gaurav Pandey, Ayoung Kim, Estimation of ambient light and transmission map with common convolutional architecture, in: *OCEANS 2016 MTS/IEEE Monterey*, IEEE, 2016, pp. 1–7.
- [45] Mehdi Mirza, Simon Osindero, Conditional generative adversarial nets, 2014, arXiv preprint [arXiv:1411.1784](https://arxiv.org/abs/1411.1784).
- [46] P. Isola, J. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [47] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
- [48] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, Alexei A. Efros, Context encoders: Feature learning by inpainting, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2536–2544.
- [49] Alec Radford, Luke Metz, Soumith Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, 2015, arXiv preprint [arXiv:1511.06434](https://arxiv.org/abs/1511.06434).
- [50] Cameron Fabbri, Md Jahidul Islam, Junaed Sattar, Enhancing underwater imagery using generative adversarial networks, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, pp. 7159–7165.
- [51] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, Yung-Yu Chuang, Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6306–6314.
- [52] A. Hore, D. Ziou, Image quality metrics: PSNR vs. SSIM, in: 2010 20th International Conference on Pattern Recognition, IEEE, 2010.
- [53] Jim Nilsson, Tomas Akenine-Möller, Understanding ssim, 2020, arXiv preprint [arXiv:2006.13846](https://arxiv.org/abs/2006.13846).
- [54] Karen Panetta, Chen Gao, Sos Agaian, Human-visual-system-inspired underwater image quality measures, *IEEE J. Ocean. Eng.* 41 (3) (2015) 541–551.