SubmergeStyleGAN: Synthetic Underwater Data Generation with Style Transfer for Domain Adaptation

***Abstract*—Underwater computer vision applications are challenged by limited access to annotated underwater datasets. Additionally, convolutional neural networks (CNNs) trained on in-air datasets do not perform well underwater due to the high domain variance caused by degradation impact of water column. This paper proposes an air-to-water dataset generator to create visually plausible underwater scenes out of existing in-air datasets. SubmergeStyleGAN, a generative adversarial network (GAN) designed to model attenuation, backscattering, and absorption, utilizes depth maps to apply range-dependent attenuation style transfer. In this work, the generated attenuated images and their corresponding original pairs are used to train an underwater image enhancement CNN. Real underwater datasets were used to validate the proposed approach by assessing various image quality metrics, including UCIQE, UIQM and CCF, as well as disparity estimation accuracy before and after enhancement. SubmergeStyleGAN exhibits a faster and more robust training procedure compared to existing methods in the literature.**

***Index Terms*—Underwater Perception, Deep Learning, Image Enhancement, Generative Adversarial Network, Style Transfer.**

I. Introduction

Recent advancement in underwater navigation technology has helped achieve some progress in ocean exploration. However, almost 95% of the planet's water bodies remain unexplored. This can be attributed to the shortcomings of underwater sensing technology, which makes underwater navigation a challenging task [1]. Visual perception has established itself as the perfect candidate for low-cost surface vehicle navigation [2, 3]. The same, however, cannot be said for underwater navigation, which still relies heavily on costly acoustic sensors. Underwater imagery suffers from blurred details, color distortion and low contrast [4, 6]. Unmanned underwater vehicles are heavily used in vital applications, such as pipeline repairs, underwater mining, fisheries and military surveillance [5]. Moreover, unmanned explorer-class vehicles are often used to study geological formations, underwater archaeological sites and marine life. Computer vision, despite its many challenges, has a huge role to play in these applications, starting from image processing, feature extraction and up to 3D image reconstruction and autonomous navigation [6].

Light attenuation and back-scattering result in reduced illumination and blurred imagery [4], making it challenging for object detection algorithms to distinguish objects from the background. Moreover, light absorption alters the color spectrum underwater, which negatively impacts the performance of color-based object detection methods. Several methods have been developed to address underwater attenuation by developing detection pipelines which are robust to underwater effects such as the approach proposed by Bazeille et al. [7] which accounts for light modifications occurring along its path between the light source and the camera. In addition to attenuation, light refraction at the interface of different media causes epipoloar lines bending which makes disparity estimation using row matching infeasible. ZhuangS. et al. [8] addressed this problem using an optimized direction-information images. The limited visibility range and degraded image quality underwater make it challenging to obtain robust visual features. Therefore, in case of visual SLAM algorithms, traditional visual feature extraction methods may not be applicable, necessitating the development of novel feature extraction and tracking techniques suitable for underwater scenarios, where feature tracking errors caused by underwater attenuation result in inaccurate pose estimation, affecting the quality and consistency of the generated map. [9].

Lack of annotated underwater datasets poses a significant challenge to machine learning-based methods. The low contrast, color distortion, and noise make manual annotation of underwater objects more challenging, as it requires domain expertise and manual effort [10]. Furthermore, training machine learning models using in-air instances of objects fails to generalize effectively to underwater scenes due to significant domain differences. To address this issue, certain approaches employ synthetic underwater datasets, such as synthetic underwater image dataset (SUID) [11].

This work addresses the challenges associated with the lack of real underwater datasets by proposing a generative adversarial network capable of generating synthetically augmented underwater datasets. SubmergeStyleGAN draws inspiration from [12, 13], which leverages the concept of transferring underwater attenuation styles from specific underwater surveys to in-air images. The proposed approach introduces a more efficient training procedure compared to WaterGAN, by utilizing only pairs of underwater and in-air RGB images. For the proposed approach, a depth map is not required for training, which reduces the training time and effort. Depth maps are only utilized post-inference to achieve a tunable range-dependent. The augmented dataset obtained from this process becomes a valuable resource for training and evaluating computer vision algorithms tailored for underwater environments. Tn this study, the generated dataset will be used to train an image enhancement module.

This paper is organized as follows: section II presents relevant

recent works in the literature, section III gives a thorough explanation of the methodology used to implement and test the proposed approach, section IV displays and discusses thee obtained results, and finally, section V draws the conclusion of this work.

II. Related Work

Traditional image enhancement techniques, such as Histogram Equalization [15], have been used to mitigate degradation impacts by increasing image contrast, or by forcing the average of each colour channel to be gray over the entire image, as in Gray-World assumption. However, these methods lack knowledge of the range-dependent attenuation,leading to photo-metric inconsistencies for the same scene across different viewpoints. This poses a challenge for computer vision algorithms relying on feature matching, such as object detection and stereo matching. Since pixels with comparable colors have a higher likelihood of belonging to the same object, and by understanding the scene's depth, it becomes possible to estimate and correct for the color and contrast changes [16]. Some approaches highlight the correlation between depth estimation and color correction [13, 17]. By leveraging RGB-D images as a sufficient photo-metric and geometrical representation, the light attenuation behavior can be better characterized. Wang and Wu [18] relied on Jaffe-McGlamery model as a range-dependent physical model. However, obtaining the model parameters requires prior knowledge of the full-depth map and specific experiments at a given survey site. This approach suffers from limited generalization due to the model's simplicity and the need to repeat experiments when water characteristics change.

To address the generalization problem and capture more complex models, neural networks can be trained end-to-end [12, 13, 19]. This approach offers the flexibility of repeating the training process in case of changes in water characteristics. Nevertheless, acquiring a sufficiently large dataset of real underwater attenuated images with corresponding ground truth scenes after water removal is impractical. Skinner el al.introduced UWStereoNet [17] that provides a solution which eliminates the need for annotated underwater datasets. UWStereoNet enhances stereo underwater images by utilizing an unsupervised learning approach that incorporates photo-metric wrapping, cyclic reconstruction constraints, and image quality metrics. However, the lack of supervision often leads to low performance due to higher uncertainty and variability in the learned representations or patterns. Skinner et al. introduced WaterGAN [12] that employs generator and discriminator networks within an adversarial training framework to generate synthetic underwater images. These synthetic images are created using both RGB-D in-air images and RGB underwater images as input. The resulting synthetic dataset is then utilized to train a Convolutional Neural Network (CNN) that estimates monocular depth maps and utilizes them for color restoration. While WaterGAN's generator was constructed to incorporate light attenuation, back-scattering and camera-related distortions, the model was simplified to stabilize the GAN's training process. This simplification adversely affects the ability to fine-grain control over specific synthesis parameters, such as depth-dependent attenuation, accordingly, the realism of the generated images and the capability to generate custom underwater scenes that precisely mimic specific scenarios are restricted. Cui et al. at [19] utilized CycleGAN [20] proposed by Zhu et al. for style transfer, utilizing RGB in-air images to generate synthetic underwater images. The generated synthetic underwater images were used to train an underwater disparity estimation network, thus reducing the domain variance between in-air and underwater images and improving the disparity estimation performance. However, CycleGAN does not possess the capacity for fine-tuning to achieve multiple sets of weights of style transfer. Additionally, the absence of explicit annotations in the mapping between two domains can lead to inconsistencies or unexpected results. Although the choice to exclude the depth map information enabled the use of CycleGAN as a data generator, it was not successful in generating data that accurately captures the depth-dependent attenuation which is a crucial characteristic of realistic underwater scenes. This limitation impacts the learning capacity of the disparity estimation network, as it lacks the ability to leverage pixel's attenuation as an additional clue for predicting a pixel's disparity. Ye et al. [13] adopted an adversarial training framework to generate a synthetic underwater dataset using an RGB-D in-air dataset. Additionally, the style transfer approach proposed by Gatys et. al at [14] is incorporated, utilizing content loss for both the generated images and in-air images, while style loss is applied to both the generated images and underwater images. The generated dataset was used to jointly train depth estimation and color correction modules, where the inclusion of style transfer losses enhanced the training convergence compared to WaterGAN.

III. Methods

The objective of this work is to train a Convolutional Neural Network (CNN)-based Image Enhancement module that enhances the quality of underwater images by mitigating the underwater attenuation. The aim is to improve contrast and clarify the features in these images. In an ideal scenario, we would evacuate the water from the images to obtain the ground truth images, which would serve as the ideal training data for the module. However, this approach is not practical. To overcome this challenge, a style transfer module (SubmergeStyleGAN) is employed as a synthetic data generator. This module takes in-air images x_a sampled from distribution X_a, and underwater images x_w sampled from distribution X_w, and generates synthetic images x_g. The generated images x_g must preserve the content (objects, structures, etc.) from the in-air image, while incorporating the desired visual style associated with the underwater image. The generated images x_g, are then used as inputs to an image enhancement module, where the original in-air images, x_a, serve as ground truth references.

A. Style Transfer Module (SubmergeStyleGAN)

The SubmergeStyleGAN shown in Figure 1 is responsible for blending the content of an in-air image x_a with the underwater style of an underwater image x_w. It follows a Generative Adversarial Network (GAN) framework, which involves a minimax game setting. During training , the module includes a generator G trained to apply an underwater attenuation effect to the in-air image. The generated image x_g should appear realistic to the discriminator D when compared to a real underwater image, contributing to the loss term $\mathbb{L}_{GAN}$:

$$\mathbb{L}_{GAN} = \mathbb{E}_{x_w \sim X_w}[\log D(x_w)] + \mathbb{E}_{x_a \sim X_a}[1 - \log D(G(x_a) + x_a)] \quad (1)$$

The proposed approach is innovated by Ye et al. at [13]. However, one key distinction is the exclusion of the in-air depth map during the training process. Instead, the generator is only required to apply the average attenuation style from the corresponding underwater survey uniformly to the in-air images. Depth maps are only utilized during the inference stage to implement a tunable range-dependent attenuation. It was found that by excluding the depth map as an input, the training process becomes more efficient and resilient. This is because the discriminator has no access to the depth map information of the underwater image, consequently, it is unable to recognize the correlation between the attenuation in real underwater images and the corresponding depth maps. As a result, the discriminator will not penalise a high loss if the generator applies attenuation to the in-air images independent of their corresponding depth maps. Accordingly, SubmergeStyleGAN is expected to face challenges in effectively utilizing the provided in-air depth maps. Another advantage of incorporating depth maps during inference, is the ability to easily adapt to different attenuation levels by adjusting a tunable parameter without the need to re-train the entire module, which sets this approach apart from related methods [12, 13]. Furthermore, in case of the absence of the depth maps of the corresponding in-air images, the approach can still generate an acceptable synthetic underwater dataset.

The approach proposed by Gatys et al. in [14] is used to expedite the style transfer process from underwater images to in-air images. The generated image features f_g should exhibit Gramians similar to the underwater image features f_w when both images propagate through a pre-trained VGG19 network, contributing to the an additional loss term $\mathbb{L}_{Style}$. Computing the style loss is accomplished by first extracting features that are predominantly influenced by the image's style. The 4th and 5th layers were experimentally found to be the most appropriate layer for capturing the desired style features from underwater images while differentiating them from in-air images. Style loss can be computed as follows:

$$\mathbb{L}_{Style} = \sum_{l \in L_s} w^l \left\| \mathbb{G}^l(f_w) - \mathbb{G}^l(f_g) \right\|_2^2 \quad (2)$$

The use of style loss for underwater dataset generation is innovated by the approach proposed by [13]. One other key difference in the proposed approach, is the exclusion of content loss. This decision is driven by the fact that the original image is forwarded and combined with the output attenuation generated by the network. By incorporating the original image in this way, it is not necessary to learn high-resolution content, eliminating the need for skip connections in the generator's encoder-decoder architecture. This design choice allows the network to focus solely on learning the low-resolution attenuation.

The generator in this module follows the same architecture as the one proposed by CycleGAN architects [20], the discriminator consists of four convolutional layers, each followed by batch normalization, leaky ReLU activation, and average pooling for downsampling. The model progressively increases the number of feature maps in each layer (64 $\rightarrow$ 128 $\rightarrow$ 256 $\rightarrow$ 512) . After the last convolutional block, the output is flattened and fed into three fully connected layers with 1024, 512, and 64 neurons, respectively, activated by leaky ReLU functions. Finally, sigmoid activation is used in the output layer to classify whether the sample is fake or real.

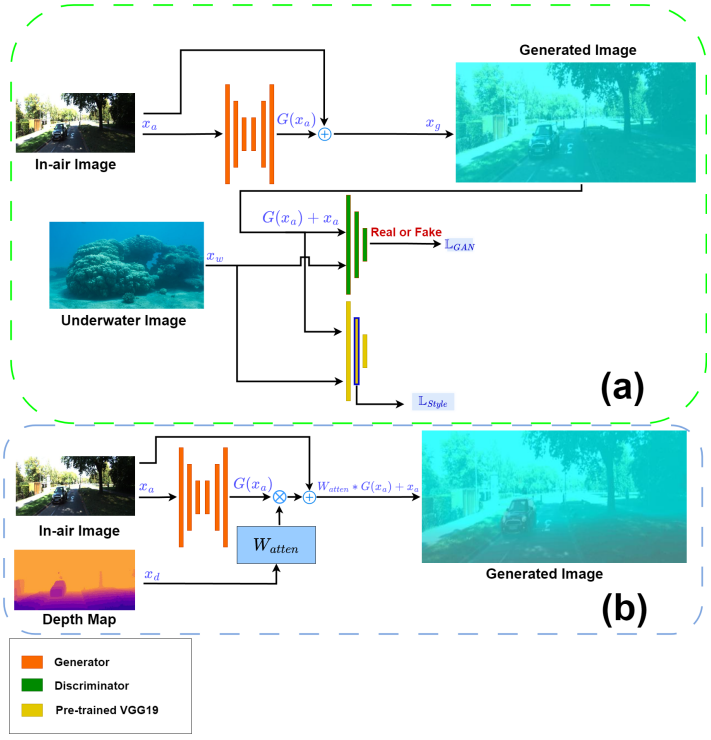

Fig. 1. Architecture of SubmergeStyleGAN for (a) training stage and (b) inference stage

B. Image Enhancement Module

The objective of the image enhancement module (Fig.2) is to reverse the effects of water attenuation, resulting in enhanced images with improved contrast and sharper features. The image enhancement module is trained using the dataset of

attenuated images x_g generated by SubmergeStyleGAN. Additionally, original in-air images x_a are provided as ground truth references. The proposed approach allows the model to learn the necessary adjustments to mitigate the impact of water attenuation effectively.

In underwater environments, color attenuation is proportional to the distance from the camera. This makes depth maps crucial for accurately retrieving the original colors in the scene. Therefore, during the training of the image enhancement module, both the depth maps and their corresponding synthetically generated underwater attenuated images are used. However, certain algorithms like stereo matching and SLAM may only require the sharpening of key visual features in the image, which involves removing only distance-dependent haze and do not necessarily require actual color restoration. Furthermore, some approaches as the approach proposed by Pérez et al. at [21] have been developed to estimate depth maps in underwater images by leveraging the distance-dependent attenuation as a clue. Accordingly, it is reasonable that the additional burden of providing depth maps for image enhancement can be avoided, and the enhancement module can still learn to minimize distance-dependent underwater haze, achieving acceptable results.

In this approach, both image enhancement methods are employed: one with depth maps and one without. The evaluation of each method involves stereo matching performance metrics, assessing the feature strength in both cases. This comparison reveals the trade-offs between full color restoration using depth maps and feature sharpening alone without using depth maps, offering insights into their suitability for different scenarios. The module is trained using L_2 per-pixel difference between the enhanced image and its in-air counterpart.
The proposed module follows an encoder-decoder architecture. The encoder begins with an initial convolutional layer, producing 64 output channels, and is followed by four dense blocks as proposed by Huang et al. [22]. Each dense block comprises three convolutional layers with a growth rate of 12, and transition blocks are inserted after each dense block to reduce the concatenated input channels. The first transition block reduces the channels to 128, the second to 256, and the third to 512. After the final dense block, batch normalization and ReLU activation are applied, followed by a 1x1 convolutional layer to compress the feature maps to 512 channels. The decoder consists of three transposed convolutional layers that upsample the feature maps, starting with 64, 128, and 512 channels. Each decoder layer is accompanied by batch normalization and ReLU activation. The Atrous Spatial Pyramid Pooling (ASPP) module proposed by Chen et al. [23] is applied after the third decoder layer, retaining 512 channels to effectively capture multi-scale contextual information. The module concludes with a convolutional layer with 512 input channels and 3 output channels, corresponding to RGB color channels, to generate the final output.

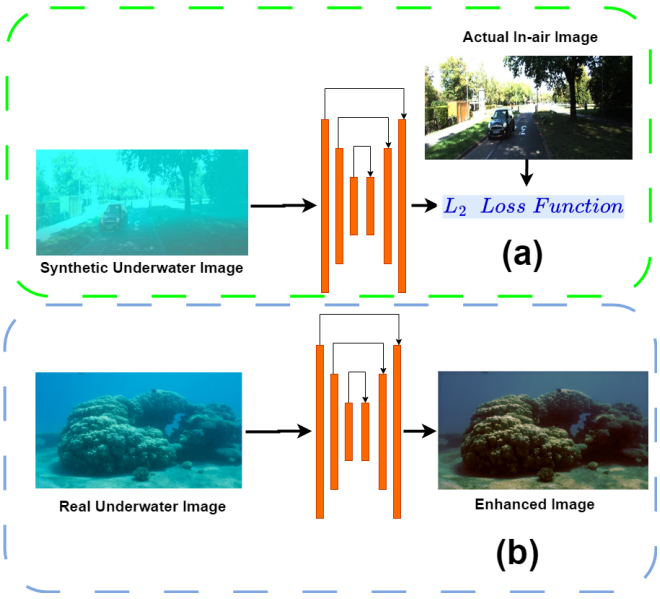

Fig. 2. Architecture of the image enhancement module for both (a) training stage and (b) inference stage

IV. Experiments and Results

A. Training

1) SubmergeStyleGAN: The training procedure involved augmenting the stereo KITTI2015 dataset with 250 underwater images from a custom dataset collected from a swimming pool, using a batch size of 4 and images of size 320×240.. The Adam solver was employed with parameters β_1 = 0.5 and β_2 = 0.999, and the learning rate was set to 2×10^{-4}. For the style loss components in Equation 2, w^l is set to 500 for both the 4$^{\text{th}}$ and 5$^{\text{th}}$ layers. The model underwent training for 15 epochs, and the style loss was applied only after the 2$^{\text{nd}}$ epoch. During inference, to apply range-dependent attenuation, the maximum depth map was set to 15 m for numerical stability, and the attenuation weight W_{atten} was set to 30. Figure 3 shows an example for an in-air image and its corresponding transformed synthetic underwater-style images created through the style transfer process from a real underwater image.

2) Image Enhancement Module: The training procedure involved utilizing the synthetic underwater dataset generated by the style transfer module, with its corresponding in-air images as ground truth. During training, a batch size of 4 and image dimensions of 640×480 were used. The Adam solver was employed with parameters β_1 = 0.5 and β_2 = 0.999, and the learning rate was set to 2×10^{-4}. The model underwent training for 40 epochs. It is important to note that two versions of the model were trained: one that takes the input image's depth map and one that doesn't. This was done to examine the model's behavior in case the requirement of a depth map is computationally heavy or impractical.

B. SubmergeStyleGAN's Evaulation:

To justify the selection of the 4th and 5th layers of p trained VGG-19 as appropriate style feature extractors the analysis, a comparison of Gramian matrices derived fr features extracted from a sample from the testing pool data was conducted. These matrices were contrasted with matri... obtained from other sets of images, including 100 underwater images from the testing pool, 100 in-air KITTI images [24], and 100 generated synthetic underwater images (samples from the mentioned datasets are shown in Figure 3). The comparison was based on the mean squared difference, quantifying the dissimilarity of style features. The distribution of the style difference depicted in Figure 4 demonstrate that the style features extracted from layers 4th and 5th exhibit a low mean square difference regarding both underwater images and synthetic underwater images. This suggests that these layers effectively capture common style features present in underwater scenes while distinguishing them from in-air images. Conversely, a higher mean square difference is observed regarding in-air KITTI images, indicating that the style features extracted from these layers differ significantly from the style characteristics found in in-air images. As we delve deeper into the pre-trained VGG19 model, reaching layers 8 and 13, the distributions corresponds to synthetic underwater images and in-air images tend to become more similar, as shown in Figure 4. This convergence in distributions occurs because the higher-level features present in the deeper layers of VGG19 are primarily focused on capturing content-related information rather than style. Hence, the 4th and 5th layers are confirmed to be appropriate for capturing and distinguishing style features in underwater images

.

The effectiveness of SubmergeStyleGAN is evaluated by comparing it to WaterGAN [12]. The evaluation involves assessing both the stability of the training procedure and the quality of the generated synthetic images. In Figure 5, we observe that the losses of the generator and the discriminator in SubmergeStyleGAN show stronger indications of convergence. This suggests that the training process is more stable and effective compared to WaterGAN. Furthermore, when generating synthetic underwater images, WaterGAN encounters difficulties in accurately capturing the true colors of the underwater style. Additionally, challenges were faced during the adjustment of WaterGAN's tuning parameters to obtain more accurate results that represent the true characteristics of underwater attenuation. On the other hand, SubmergeStyleGAN demonstrates superior performance in preserving and representing the authentic colors associated with underwater environments.

C. Image Enhancement Module's Evaulation

In the evaluation, three non-reference metrics, namely UCIQE [25], UIQM [26], and CCF [27], are utilized, which are commonly employed for assessing the quality of underwater images. The UCIQE score provides an indication of the balance among chroma, saturation, and contrast in the output. A

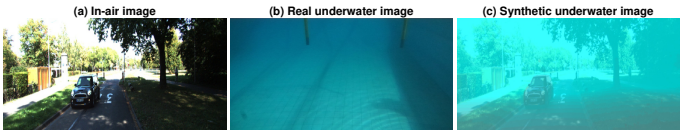

Fig. 3. Style transfer from a real underwater image to an in-air image using SubmergeStyleGAN (a) in-air image (b) real underwater image (c) synthetic underwater image.

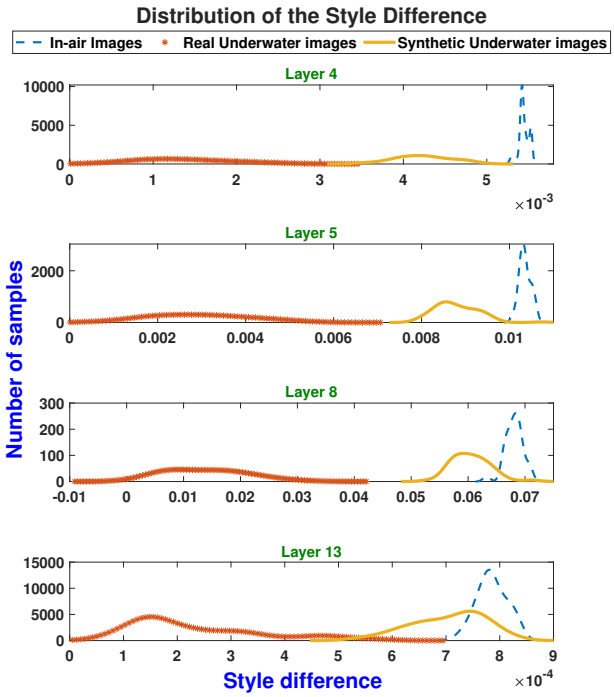

Fig. 4. The distribution of style differences between an underwater image taken from a testing pool dataset and samples from three distinct datasets: a testing pool dataset, in-air images from Kitti [24] dataset, and the synthetic underwater version of Kitti dataset. The style features were extracted from

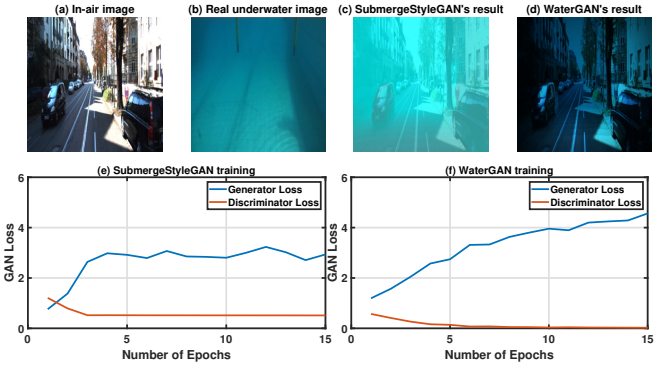

Fig. 5. Comparison between SubmergeStyleGAN and WaterGAN [12] in two key aspects: the quality of generated synthetic data and the training loss curves that provided insights into the convergence behavior and stability of the training process for each method.

higher UCIQE score suggests that the resulting image exhibits a better balance of these color-related elements. A higher UIQM score indicates that the output aligns more consistently with human visual perception. Additionally, the CCF metric assesses the colorfulness, contrast, and fog density of the image, offering insights into these specific visual attributes. The evaluation was conducted on samples from three different datasets: SQUID dataset [28], Roboflow Aquarium dataset [29] and a custom dataset collected from a testing pool.
The image enhancement module trained with depth maps was evaluated using enhanced samples from the SQUID dataset, generated by state-of-the-art image enhancement techniques, such as fusion-based enhancement by Ancuti et al. [31], image enhancement using depth estimation by Peng et al. [32], and image enhancement using transmission estimation by Drews et al. [34]. The SQUID dataset samples and the outputs of these techniques can be found at [30], as shown in Figure 7 and Table I. Additionally, the performance of the image enhancement module that doesn't require depth maps was evaluated on the Aquarium dataset and a custom dataset collected from a test pool. This module was compared against various state-of-the-art techniques, including BAL [35], fusion-based enhancement by Ancuti et al. [31], and UWCNN [36]. The techniques used in these comparisons can be generated at [37], as illustrated in Figures 8 and 9, and Tables II and III.

It is observed from Tables I,II and III that the best performers in terms of UCIQE and UIQM metrics do not align with the subjective pairwise comparisons, despite the fact that both UCIQE and UIQM claim to account for human visual perception. Moreover, the analysis indicates that UCIQE tends to give higher scores to images with greater contrast, yet it does not adequately account for color shifts and artifacts in the results (as observed in the outcomes of [34] and [32] shown in Figure 7 and Table I). The visual evaluation results do not always align precisely with the quantitative scores obtained from non-reference metrics. This discrepancy is attributed to a gap between the objective quantitative scores and the subjective visual quality perceived by humans. In essence, the current image quality evaluation metrics designed for underwater images have limitations in certain cases.

The feature sharpening capabilities of the two versions of the image enhancement module, with a focus on feature enhancement without full color restoration as discussed in Section III-B, were assessed using a STereo TRansformer (STTR) model. This model was initially trained on the Kitti dataset [24] and further fine-tuned on the original underwater SQUID dataset [28]. The results, depicted in Figure 6 and Table IV, showed a slight improvement when using depth maps for image enhancement. This indicates that for specific applications where emphasizing features is essential, the use of depth maps may not be necessary and does not significantly enhance the overall performance.

TABLE I
THE EVALUATION THAT COMPARES THE PROPOSED APPROACH WITH DIFFERENT STATE-OF-THE-ART METHODS ON SAMPLES FROM SQUID DATASET [28], USING THE UIQM [26], UCIQE [25], AND CCF [27] METRICS.

Method	UCIQE	UIQM	CCF
Raw Images	0.40	0.24	19.50
Proposed method	0.53	0.57	25.30
Peng et al., [32]	0.56	0.52	17.96
Ancuti et al., [33]	0.57	0.48	13.68
Drews et al., [34]	0.65	0.72	27.53

TABLE II
THE EVALUATION THAT COMPARES OUR APPROACH WITH DIFFERENT STATE-OF-THE-ART METHODS ON SAMPLES FROM ACQUARIAM DATASET [29], USING THE UIQM [26], UCIQE [25], AND CCF [27] METRICS.

Method	UCIQE	UIQM	CCF
Raw Images	0.51	0.58	23.30
Proposed method	0.55	0.71	31.46
Peng et al., [35]	0.60	0.70	41.53
Ancuti et al., [31]	0.61	0.67	20.54
Li et al., [36]	0.50	0.59	15.87

TABLE III
THE EVALUATION THAT COMPARES OUR APPROACH WITH DIFFERENT STATE-OF-THE-ART METHODS ON SAMPLES FROM A CUSTROM DATASET COLLECTED FROM A TESTING POOL, USING THE UIQM [26], UCIQE [25], AND CCF [27] METRICS.

Method	UCIQE	UIQM	CCF
Raw Images	0.45	0.54	20.20
Proposed method	0.56	0.69	37.09
Peng et al., [35]	0.59	0.77	45.02
Ancuti et al., [31]	0.58	0.68	15.62
Li et al., [36]	0.45	0.54	10.80

Fig. 6. Visual comparison on a sample from SQUID dataset [28], from left to right, shown raw underwater image, enhanced Image without and with depth map.

TABLE IV
STTR [38] MODEL'S PERFORMANCE COMPARED WITH THREE INPUT STEREO IMAGES SCENARIOS: ORIGINAL UNDERWATER IMAGES, ENHANCED WITHOUT AND WITH DEPTH MAPS.

Input Images	3 px Error	EPE
Raw Underwater Images	15.49	0.99
Enhanced Images (w/o depth information)	10.75	0.65
Enhanced Images (w/ depth information)	9.9	0.89

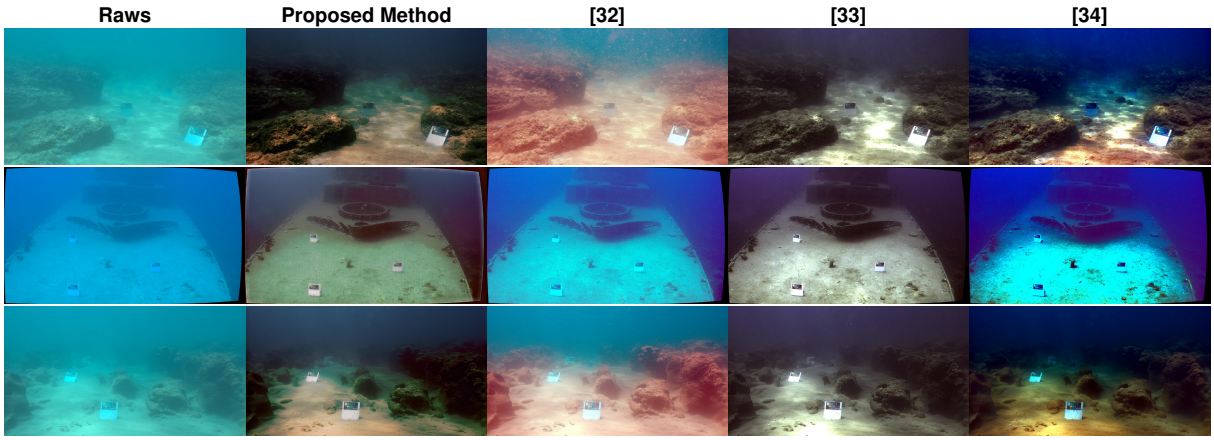

Fig. 7. Visual comparison on samples from SQUID dataset [28], from left to right, shown raw underwater images, the porposed method, image enhancement using depth estimation by Peng et al., (2015) [32], color transfer by Ancuti et al., (2017) [33], and image enhacement using transimission estimation by Drews et al., (2013) [34].

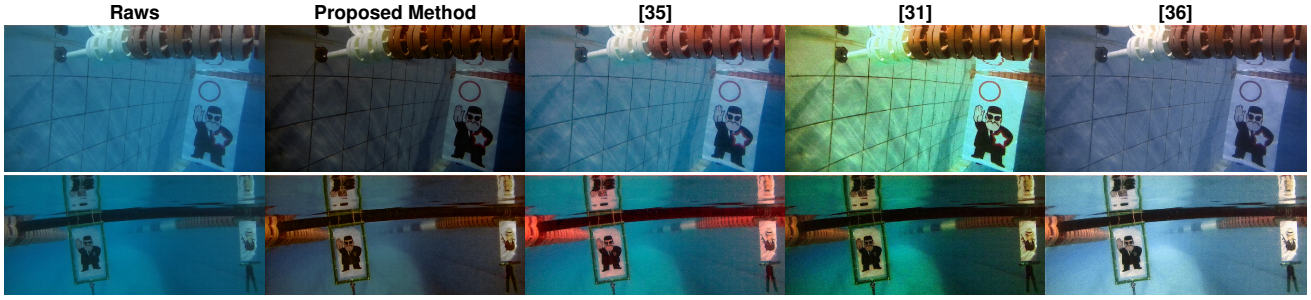

Fig. 8. Visual comparison on samples from Aquarium dataset [29], from left to right, shown raw underwater images, the proposed method, BAL [35], fusion-based by Ancuti et al., (2018) [31] and UWCNN [36]

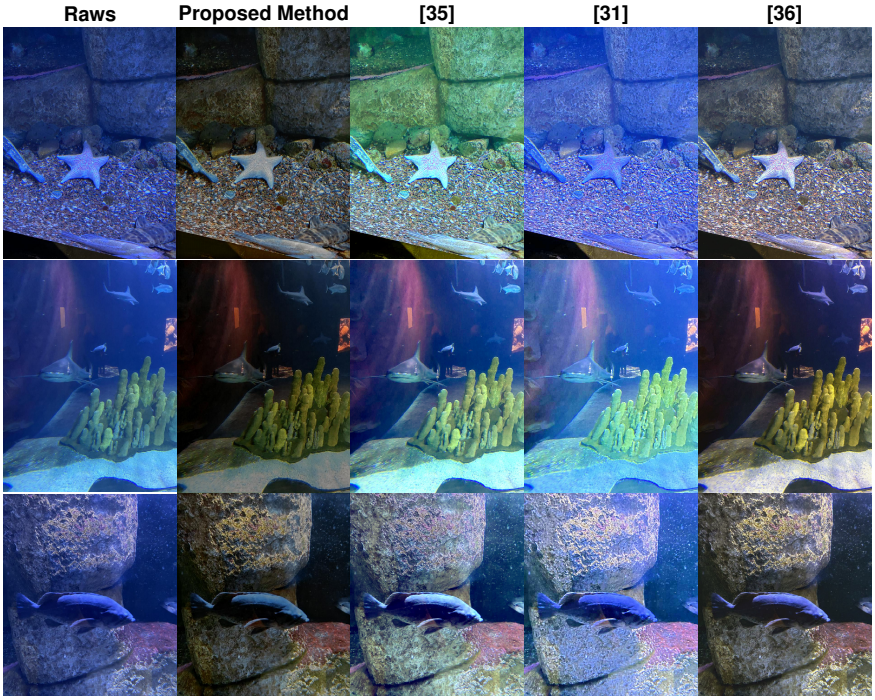

Fig. 9. Visual comparison on samples from a custom dataset collected from a testing pool, from left to right, shown raw underwater images, the proposed method, BAL [35], fusion-based by Ancuti et al., (2018) [31] and UWCNN [36]

V. Conclusion

In Conclusion, this paper tackles the scarcity of annotated underwater datasets in underwater computer vision applications, and the limitations of CNNs trained on in-air data due to high domain variance caused by underwater degradation. To overcome these challenges, the paper proposes an innovative air-to-water dataset generator that synthesizes realistic underwater scenes from in-air datasets. SubmergeStyleGAN, effectively models attenuation, backscattering, and absorption and using depth maps for range-dependent attenuation style transfer. Using the generated synthetic underwater images and their originals, an underwater image enhancement CNN is trained. Evaluations show that SubmergeStyleGAN outperforms WaterGAN in generating realistic underwater images and achieving more efficient training. The image enhancement module is evaluated on three real underwater datasets, showing promising results through visual evaluation and various image quality metrics, such as UCIQE, UIQM, and CCF. Also, the concept of enhancing underwater images without relying on computationally expensive depth maps is explored, aiming to improve features without restoring the exact image colors. To evaluate this approach, the paper establishes a benchmark that measures the accuracy of disparity estimation, serving as an indicator of the features' strength, which reveals a slight accuracy difference between image enhancement with and without using depth maps.

References

[1] Sun, Kai, Weicheng Cui, and Chi Chen. "Review of underwater sensing technologies and applications." Sensors 21.23 (2021): 7849.

[2] Sun, Yao, et al. "Visual perception based situation analysis of traffic scenes for autonomous driving applications." 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2020.

[3] Shi, Weijing, et al. "Algorithm and hardware implementation for visual perception system in autonomous vehicle: A survey." Integration 59 (2017): 148-156

[4] Lu, Huimin, et al. "Underwater optical image processing: a comprehensive review." Mobile networks and applications 22 (2017): 1204-1211.

[5] Bonin, F., Antoni Burguera, and Gabriel Oliver. "Imaging systems for advanced underwater vehicles." Journal of Maritime Research 8.1 (2011): 65-86.

[6] González-Sabbagh, Salma P., and Antonio Robles-Kelly. "A survey on underwater computer vision." ACM Computing Surveys (2023).

[7] Bazeille, Stéphane, Isabelle Quidu, and Luc Jaulin. "Color-based underwater object recognition using water light attenuation." Intelligent service robotics 5 (2012): 109-118.

[8] Zhuang, Sufeng, et al. "A dense stereo matching method based on optimized direction-information images for the real underwater measurement environment." Measurement 186 (2021): 110142.

[9] Köser, Kevin, and Udo Frese. "Challenges in underwater visual navigation and SLAM." AI technology for underwater robots (2020): 125-135.

[10] Panetta, Karen, et al. "Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with GAN." IEEE Journal of Oceanic Engineering 47.1 (2021): 59-75.

[11] Guojia Hou, Xin Zhao, Zhenkuan Pan, Huan Yang, Lu Tan, Jingming Li, June 29, 2020, "SUID: Synthetic Underwater Image Dataset", IEEE Dataport, doi: https://dx.doi.org/10.21227/agdr-y109.

[12] Li, Jie, et al. "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images." IEEE Robotics and Automation letters 3.1 (2017): 387-394.

[13] Ye, Xinchen, et al. "Deep joint depth estimation and color correction from monocular underwater images based on unsupervised adaptation networks." IEEE Transactions on Circuits and Systems for Video Technology 30.11 (2019): 3995-4008.

[14] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "Image style transfer using convolutional neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[15] Hummel, R., "Image enhancement by histogram transformation", ieht.rept, 1975.

[16] Zhou, Jingchun, et al. "Auto color correction of underwater images utilizing depth information." IEEE Geoscience and Remote Sensing Letters 19 (2022): 1-5.

[17] Skinner, Katherine A., et al. "Uwstereonet: Unsupervised learning for depth estimation and color correction of underwater stereo imagery." 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019.

[18] Wang, Yan, and Bo Wu. "Fast clear single underwater image." 2010 International Conference on Computational Intelligence and Software Engineering. IEEE, 2010.

[19] Cui, Jiadi, et al. "Underwater depth estimation for spherical images." Journal of Robotics 2021 (2021): 1-12.

[20] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.

[21] Pérez, Javier, et al. "Recovering depth from still images for underwater dehazing using deep learning." Sensors 20.16 (2020): 4580.

[22] Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[23] Chen, Liang-Chieh, et al. "Rethinking atrous convolution for semantic image segmentation." arXiv preprint arXiv:1706.05587 (2017).

[24] Geiger, Andreas, et al. "Vision meets robotics: The kitti dataset." The International Journal of Robotics Research 32.11 (2013): 1231-1237.

[25] Yang, Miao, and Arcot Sowmya. "An underwater color image quality evaluation metric." IEEE Transactions on Image Processing 24.12 (2015): 6062-6071.

[26] Panetta, Karen, Chen Gao, and Sos Agaian. "Human-visual-system-inspired underwater image quality measures." IEEE Journal of Oceanic Engineering 41.3 (2015): 541-551.

[27] Wang, Yan, et al. "An imaging-inspired no-reference underwater color image quality assessment metric." Computers and Electrical Engineering 70 (2018): 904-913.

[28] Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz. (2021). SQUID- Stereo Quantitative Underwater Image Dataset [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5744037

[29] Roboflow, "Aquarium Object Detection Dataset," Roboflow, Link (accessed Jul. 21, 2023).

[30] Berman, Dana, Deborah Levy, Shai Avidan, and Tali Treibitz. "Underwater Single Image Color Restoration Using Haze-Lines and a New Quantitative Dataset." IEEE Transactions on Pattern Analysis and Machine Intelligence (2020).

[31] Ancuti, Codruta O., et al. "Color balance and fusion for underwater image enhancement." IEEE Transactions on image processing 27.1 (2017): 379-393.

[32] Peng, Yan-Tsung, Xiangyun Zhao, and Pamela C. Cosman. "Single underwater image enhancement using depth estimation based on blurriness." 2015 IEEE International Conference on Image Processing (ICIP). IEEE, 2015.

[33] Ancuti, Codruta O., et al. "Color transfer for underwater dehazing and depth estimation." 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017.

[34] Drews, Paul, et al. "Transmission estimation in underwater single images." Proceedings of the IEEE international conference on computer vision workshops. 2013.

[35] Peng, Yan-Tsung, and Pamela C. Cosman. "Underwater image restoration based on image blurriness and light absorption." IEEE transactions on image processing 26.4 (2017): 1579-1594.

[36] Li, Chongyi, Saeed Anwar, and Fatih Porikli. "Underwater scene prior inspired deep underwater image and video enhancement." Pattern Recognition 98 (2020): 107038.

[37] "Platform for underwater image quality evaluation," PUIQE, https://puiqe.eecs.qmul.ac.uk/ (accessed Jul. 22, 2023).

[38] Li, Zhaoshuo, et al. "Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers." Proceedings of the IEEE/CVF international conference on computer vision. 2021.