# COMP6411B Final Project
# End-to-end Image Correction for Underwater Images

FRIESKE Rita, YIU Cheuk Tung, BAE Juyoung

The Hong Kong University of Science and Technology
{rmfrieske, ctyiuac, jbaeaa}@connect.ust.hk

**Abstract.** The underwater image correction has commercial and artistic use. Both parametric and end-to-end models are used to enhance the vision. We compare the most successful parametric methods with state-of-the-art end-to-end GAN models, and discover that depth information is crucial for successful denoising strategies. We also explore the performance of the Sea-thru method, SQUID method and a new diffusion model on this task, to our knowledge the first such experiment.

**Keywords:** underwater image correction · computer vision · underwater vision enhancement

## 1 Introduction

In recent years, underwater image perception and analysis technology have developed rapidly and are widely used in marine geological surveys, marine biological detection and protection, marine military and other fields. The acquisition of comprehensive underwater photographs is a prerequisite for using underwater images. However, there are absorption and scattering effects in the underwater light transmission process, which results in issues like low contrast, fuzzy details, and significant color cast in the underwater photographs.

In this work, we explore underwater image correction. Given images taken underwater with blue or green filter coming from water reflection, we explore methods of blue filter removal, that creates an impression of clear images. We perform experiments both on end-to-end and parametric methods and reach a conclusion that the end-to-end models lack depth information that is crucial for parametric models. Furthermore, we explore performance of a novel diffusion model on underwater image correction task, and mark significantly lower performance compared to state-of-the art GAN model. Finally, we introduce our own dataset of underwater videos used in this work for evaluation.[1] The motivation behind using our own dataset is enhancement of underwater videos specifically for the purposes of underwater sport broadcasting. This field is underrepresented in the underwater computer vision research, since most of the

[1] The video result examples are available on https://www.youtube.com/watch?v=jmMjNaARCiE and https://www.youtube.com/watch?v=BBAWEzrvduE

images aim to explore sea waters for commercial and research purposes. Because of that our dataset possesses different set of characteristics i.e. limited depth, varying water luminosity, haze and the presence of humans.

## 2 Related Work

Underwater image correction is an important problem that directly influences marine science and underwater robotics [4,6,7]. The two most common, and parallelly used approaches include parametric image correction and deep learning method. Among parametric methods, [1,29] explore utilizing depth maps to denoise underwater images. Furthermore, Akkaynak et al. explore image correction extending not only to enhancement of the foreground figures but also removal of the blue filter in the background. [1] Apart from the depth information, parametric methods often take into account specific conditions in which the picture or video was taken such as water attenuation, backscatter, luminosity, blurriness etc. [3,13]

End-to-end methods for underwater image correction do not have advantages of the handcrafted features that make certain parametric models perform very well under known conditions, but they tend to generalize better to unknown scenarios. The most popular neural architecture for underwater image enhancement are different types of generative adversarial networks (GANs). [6,23] [5] implement cycle GANs for spherical image correction. [24] propose BA-GAN, block attention GAN model that adds attention mechanism to enhance underwater images. [26] propose Water Fusion GAN architecture equipped with four convolution branches that refine the features of the three prior inputs to then fuse them using multi-scale fusion connections and channel attention decoder to generate results. [27] propose UUGAN for unsupervised learning for underwater image enhancement.

## 3 Methodology

After exploring several types of supervised parametric and non-parametric methods, we noticed two main research directions:

- Approach of image correction that requires depth maps - parametric models
- Approach that does not require depth maps, but rather parallel paired datasets - end-to-end models[2]

[2] All the models and datasets are publicly available at https://github.com/khleeloo/uwh_project

### 3.1 Parametric models

***SQUID*** Based on the 3D structure determined by stereo imaging, this work creates a set of images with various water characteristics obtained in various locations. The dataset consists of images from four distinct locations. The sets are named *Katzaa*, *Michmoret*, *Nachsholim*, and *Satil*, ranging in depth from three to thirty meters.

SQUID is not only a dataset, but also provides a method to recover the depth maps and restored image colors in a scene regarding different water types using a single image as input. The attenuation coefficients relate the color-dependent transmission to distance. They presented the haze-lines prior approach which discovered that the color of a haze-free image may be well approximated by hundreds of distinct colors, and that these identically colored pixels form a tight cluster in the RGB space. The essential feature of this algorithm is the non-local distribution of pixels inside a cluster in the haze-free image. In other words, they are dispersed across the entire image plane and located at various camera distances. Because these pixels are part of the same color cluster but are spread across several image regions, their distances from the camera are variable, resulting in different transmittance. As a result, their intensity values will alter under the impact of haze. That is to say, when it is foggy, haze lines will be formed in RGB space by the intensity values of pixels grouped together in a fog-free image. Haze-lines cluster and restore each pixel in the image, restoring the image color more thoroughly than Dark Channel Prior (DCP) [8] models, which are based on the features of image block processing.

***Sea-thru*** The Sea-thru approach can increase analysis efficiency for huge datasets because it is made to reliably eliminate water from underwater photos. Given an RGBD image, it calculates the backscatter image in a way that is similar to the DCP created for haze, but with a known depth map. The optimization approach is then used to predict range-dependent attenuation coefficients using illumination maps produced by locally spatially averaging color. However, for the Sea-thru method, the shot distance must be accurate. The complete distance information has been sorted out by scientists using expert photogrammetry, allowing the method to be successfully applied to perform picture processing.

### 3.2 Deep models

***GAN*** Image to image translation was first explored by [11] who published pix2pix model that became instantaneously popular among the internet users due to its artistic qualities. We have decided to use this particular model for underwater image correction, treating this problem as a style transfer problem between hazy input images and previously cleaned ground truth. Using synthetic data such us the ground truth images in our model is an important alternative in situations where ground truth cannot be obtained naturally, or in cases of data scarcity. [14] support the hypothesis that synthetic sonar image data created by pix2pix can replace ground truth data without significant drop of the performance.

***Diffusion*** Following the widely successful GAN models, Diffusion and score-based generative models have recently gained substantial interests [9, 20–22]. Inspired by non-equilibrium thermodynamics, diffusion models first establish a markov chain of noise diffusion step, where each step iteratively adds small Gaussian random noise to the input image until the input becomes indistinguishable from pure Gaussian noise. Image generation can then be achieved by incrementally reversing the aforementioned noise-adding sequence. Unlike the forward noising step, the reverse denoising step is not trivial, as we do not have the exact parameters that can represent the distribution of denoised predecessors in the chain. In this context, diffusion models achieve the reverse denoising step by employing a deep neural network, thereby acquiring the synthetic output at the end of the reverse chain.

This diffusion-based image generation pipeline can be extended to conditional image generation, where we specify a certain visual constraint along the denoising process. One prominent example of a conditional image generation task is Grayscale-to-RGB colorization, where the diffusion model is expected to generate a fully colored version of the grayscale input. Here, the grayscale input serves as the constraint to which the diffusion model needs to adhere to during the process of reverse denoising, such that the synthetically colored output contains the same semantics as the grayscale input. We hypothesized that same principle can be underwater blue filter removal. Specifically, we experimented with a conditional diffusion model where the input raw underwater image is the constraint and the output is the synthesized version of the input without the blue filter. For the implementation, we made liberal use of Palette [18], a general off-the-shelf framework for image-to-image translation using conditional diffusion models.

## 4 Datasets and Metrics

***Datasets and preprocessing*** For training the deep-learning based models, we utilized public two paired datasets that contain raw underwater images and its color-corrected variants. First is the Enhancing Underwater Visual Perception (EUVP) [10] dataset, which contains 13k pairs of raw underwater images and corresponding color-corrected reference images, collected from real-life photo footprints and expanded with underwater images extracted from the public ImageNet dataset and Flicker$^{TM}$. Sizes of the images were all uniformly aligned to 256x256 resolution. We further augmented our pool of paired samples with the Underwater Image Enhancement Benchmark (UIEB) [15] dataset. UIEB is packed with 890 pairs of raw and color-enhanced high-definition underwater images. Images in the UIEB datasets are of arbitrary sizes, ranging from 640x480 to 2K resolution images. To standardize the scale of the objects in the underwater scenes in respect to the resolution, we resized the every image with 512x512 resolution. We then extracted four 256x256 patches per one image. From both the EUVP and UIEB datasets, we compiled a pool of around 17k pairs of underwater and color-corrected images, all in uniform resolution of 256x256. We used this pool of paired images to train both GAN and Diffusion models.

Additionally, we use Underwater Hockey Video Dataset for evaluation. For the Underwater Hockey Video footprints, we simply decomposed the video to multiple frames of images, and performed similar operations that we performed to the UIEB datasets. We resized to 512x768 resolution and extracted a set of 256x256 patches and performed inference on our trained Pix2Pix and Diffusion models.

***Metrics*** The qualitative comparison of two frames/videos includes two frames/videos with and without the blue filter aligned side-by-side. The metrics used for quantitative evaluation include Peak Signal to Noise Ratio (PSNR) [10, 12, 15], and the second structural similarity index measure (SSIM) [10, 12, 15, 25], underwater image quality measure (UIQM) [10, 15, 16], underwater color image quality evaluation metric (UCIQE) [12, 15, 17, 28] and accuracy.

**Structural similarity index measure (SSIM)** is used to calculate a comparison between the images.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1}$$

$\mu_x$and $\mu_y$ describe average of row and column data; $\sigma_x$ and $\sigma_y$ are variance of row and column data and $C_1, C_2$ stabilize variables with weak denominator.

**Peak Signal to Noise Ratio (PSNR)** describes the ratio between the most realizable power of signal and corrupting noise that influence the consistency of its representation.

$$PSNR(x,y) = \frac{10\log_{10}(\max(\max(x), \max(y))^2}{|x-y|^2} \tag{2}$$

**Underwater Color Image Quality Evaluation UCIQE** is a metric that combines chroma, saturation, and contrast, and it is used to quantify the non-uniform color cast, blurring, and low-contrast that characterize underwater engineering and monitoring images. [28]

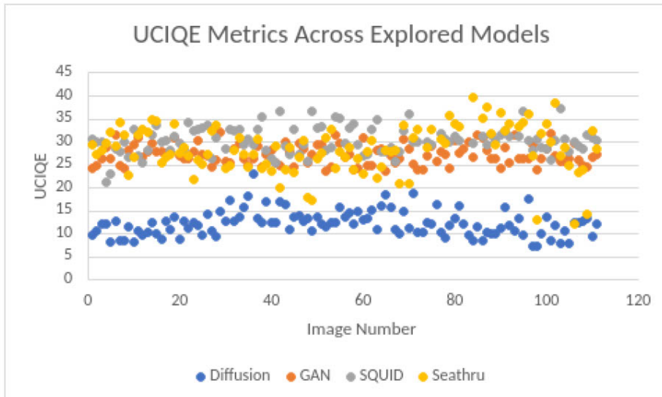
$$UCIQE = c_1 \times \sigma_c + c_2 \times con_l + c_3 \times \mu_s \tag{3}$$

Where $\sigma_c$ stands for the deviation of chroma, $con_l$ is a contrast of luminance $\mu_s$ is an average of saturation and $c1 = 0.4680, c2 = 0.2745, c3 = 0.2576$. [2]

**Underwater Image Quality Measure (UIQM)** is a measure proposed by [19] that utilizes colorfulness, sharpness and contrast.

$$UIQM = c_1 \times UICM + c_2 \times UISM + c_3 \times UIConM \tag{4}$$

**Fig. 1.** UCIQE Metrics for all the compared images. Diffusion method shows decreased performance and Sea-thru huge variance, presumably from the lack of stability of the generated depth maps.




Where UICM stands for underwater image colorfulness measure, UISM is underwater image sharpness measure and UIConM is underwater image contrast measure. A higher UIQM score implies that the output is more congruent with human visual perception, whereas a higher UCIQE score shows that the chroma, saturation, and contrast are better balanced.

## 5 Experiments and Results

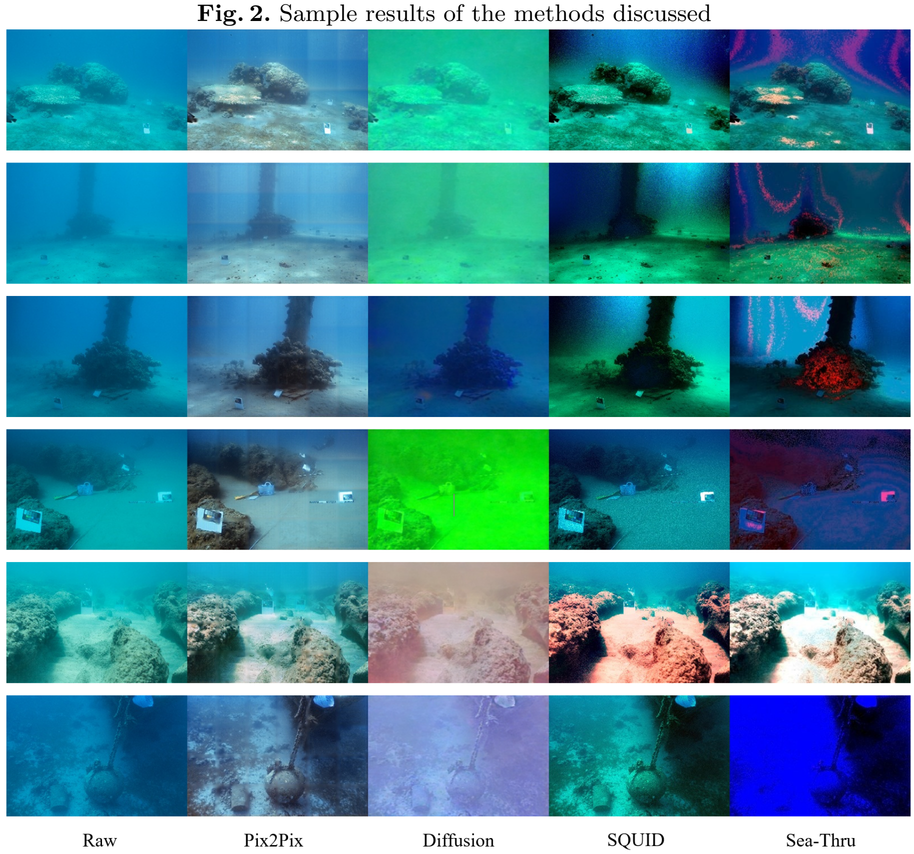
In this section, we describe the implementation of different models including *SQUID*, *Sea-thru*, *GAN* and *Diffusion*. Afterwards, the datasets and metrics employed during our experimental evaluation are described.

### Parametric baselines

***SQUID and Sea-thru*** We would like to get the output from the Sea-thru method using the SQUID dataset and compare the results with other models using the same dataset.

For Sea-thru, it requires a raw image and a depth map as input to remove water. We would like to use the SQUID method to generate a depth map while using a raw image as input. The method is running on a MatLab environment. We first extracted all the **.tif** files from the SQUID dataset and downsized them

**Fig. 2.** Sample results of the methods discussed




to **.jpg** files. The resolution of the images is 5474 × 3653. To ensure comparability to the models, the resolution remains unchanged. In addition, we also use the SQUID method to generate the color-corrected image to compare with other models. It first estimates every possible water type (I, IA, IB, II, III for open ocean waters, and 1 through 9 for coastal waters), and then determine the attenuation value to generate the restored color image.

Because the image size is relatively large (5474 × 3653), so the generation time of SQUID in MatLab is very time-consuming. About the computational setting, the speed for generating one image using GPU NVIDIA GeForce RTX 2060 Super is about seventeen minutes; and using GPU NVIDIA GeForce GTX 1080 Ti is about ten minutes. There are 110 images from SQUID and they take around 30 hours to finish generating all the depth maps and the restored images.

After obtaining all the depth maps of the SQUID image. We put the corresponding original input image and the depth map into the Sea-thru model. The output images are resized to 320 × 214.

| Model | UCIQE | UIQM | PSNR | SSIM |
|---|---|---|---|---|
| Sea-Thru | 28.21 | 0.26 | 11.54 | 0.21 |
| SQUID | 22.63 | 0.27 | 12.15 | 0.27 |
| Diffusion | 12.15 | 0.30 | 11.01 | 0.19 |
| Pix2Pix | 27.1 | 0.55 | N/A | N/A |

**Table 1.** Results of color enhancement experiments with various models. The higher values of UIQM and UCIQE metrics indicate better results. For reference, we also calculated the PSNR and SSIM values of Sea-Thru, SQUID, and Diffusion outputs in respect to the Pix2Pix output.

The UCIQE metric is designed to evaluate the nonuniform color cast and low-contrast that define underwater images. And the UIQM addresses three underwater picture quality criteria: colourfulness, sharpness, and low-contrast. According to Table 1, the result of SQUID is 25% better than Sea-thru. The colourfulness and sharpness of SQUID and Sea-thru results are similar. For PSNR and SSIM, both SQUID and Sea-thru values are lower. That means the SQUID retains more comprehensive information and image details.

### Deep models

***Pix2Pix*** We train pix2pix model on the aligned and filtered splits of EUVP and UIEB datasets. With the Unet 256 generator and 70 x70 PatchGAN discriminator. The learning rate used 0.0002 with the linear learning rate policy. The model was trained for 200 epochs with warm up and weight decay policy on NVIDIA 3090 GPU. The obtained results are most natural compared to the other models, which often display red or green filters. The naturalness of the output creates the impression that the change is often not visible enough, which is visible, especially in the video output. While the generated video is smoother and free of colored artifacts generated by parametric methods, the output is only slightly enhanced, the more saturated colors are visible in silhouettes of people, but the blue water filter is still there unlike in sea-thru outputs.

***Diffusion*** To train the diffusion-based palette model, we used the same paired and processed EUVP and UIEB datasets used to train the Pix2Pix model. We used a learning rate of 0.00002 with a batch size of 4 with on-the-fly augmentations of random flips and rotations. Despite the Palette's state-of-the-art performance in other conditional generation tasks, we observed that its performance was subpar for our blue filter removal task. For diffusion models, the image generation process starts with a random Gaussian input, and the trained model iterative generates the output of desired style and semantics. The output therefore highly depends on the random starting input and has extremely high variability. On the other hand, the expected output of our task - cleared underwater image - is far from requiring high variability, as the objective is to obtain the image with as little blue filter as possible with clear visual semantics. However, as can be seen from Fig. 2, the high variability led to an inconsistent color scheme of the output of the Diffusion model.

## 6 Conclusion

In this work, we explored several different methods for the underwater blue-filter removal task. Overall, we suggest the superiority of deep based model, especially the GAN based image generation model, over parametric counterparts. Given the versatility of GAN-based model and its benefit of not requiring a dedicated depth map serves as a huge merit for underwater blue filter removal task, let alone the superior qualitative and quantitative outcomes of the GAN-based model. In the future we would like to combine the input of the depth maps since they have proven to be successful in parametric models and the strengths of GAN model, by combining both methods and seek even better outcome.

## References


1. Akkaynak, D., Treibitz, T.: Sea-thru: A method for removing water from underwater images. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1682–1691 (2019). https://doi.org/10.1109/CVPR.2019.00178
2. Barbosa, W.V., Amaral, H.G., Rocha, T.L., Nascimento, E.R.: Visual-quality-driven learning for underwater vision enhancement. In: 2018 25th IEEE International Conference on Image Processing (ICIP). pp. 3933–3937 (2018). https://doi.org/10.1109/ICIP.2018.8451356
3. Berman, D., Avidan, S.: Diving into haze-lines: Color restoration of underwater images www.eng.tau.ac.il/ bermanTaliTreibitz2marsci.haifa.ac.il/profiles/tTreibitzwww.eng.tau.ac.il/ avidan
4. Bryson, M., Johnson-Roberson, M., Pizarro, O., Williams, S.B.: True color correction of autonomous underwater vehicle imagery. Journal of Field Robotics **33**, 853–874 (9 2016). https://doi.org/10.1002/ROB.21638
5. Cui, J., Jin, L., Kuang, H., Xu, Q., Schwertfeger, S.: Underwater depth estimation for spherical images. Journal of Robotics **2021** (2021). https://doi.org/10.1155/2021/6644986
6. Fabbri, C., Islam, M.J., Sattar, J.: Enhancing underwater imagery using generative adversarial networks. Proceedings - IEEE International Conference on Robotics and Automation pp. 7159–7165 (9 2018). https://doi.org/10.1109/ICRA.2018.8460552
7. Garcia, R., Gracias, N., Nicosevici, T., Prados, R., Hurtos, N., Campos, R., Escartin, J., Elibol, A., Hegedus, R., Neumann, L.: Exploring the seafloor with underwater robots. Computer Vision in Vehicle Technology pp. 75–99 (2 2017). https://doi.org/10.1002/9781118868065.CH4

8. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1956–1963 (2009). https://doi.org/10.1109/CVPR.2009.5206515
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. CoRR **abs/2006.11239** (2020), https://arxiv.org/abs/2006.11239
10. Islam, M.J., Xia, Y., Sattar, J.: Fast underwater image enhancement for improved visual perception. IEEE Robotics and Automation Letters **5**(2), 3227–3234 (2020). https://doi.org/10.1109/LRA.2020.2974710
11. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 **2017-January**, 5967–5976 (11 2017). https://doi.org/10.1109/CVPR.2017.632
12. Jamadandi, A., Mudenagudi, U.: Exemplar-based underwater image enhancement augmented by wavelet corrected transforms. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 11–17 (2019)
13. Jarina, R.A., Abas, P.E., Silva, L.C.D.: Depth estimation for underwater images from single view image. IET Image Processing **14**, 4188–4197 (2020). https://doi.org/10.1049/IET-IPR.2019.1533
14. Lee, E.H., Jeon, M.H., Jang, H., Park, B., Kim, A., Lee, S.: Study on the training effectiveness of deep learning with synthesized underwater sonar image using pix2pix and fcn. 2020 IEEE/OES Autonomous Underwater Vehicles Symposium, AUV 2020 (9 2020). https://doi.org/10.1109/AUV50043.2020.9267923
15. Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D.: An underwater image enhancement benchmark dataset and beyond. IEEE Transactions on Image Processing **29**, 4376–4389 (2020). https://doi.org/10.1109/TIP.2019.2955241
16. Panetta, K., Gao, C., Agaian, S.: Human-visual-system-inspired underwater image quality measures. IEEE Journal of Oceanic Engineering **41**(3), 541–551 (2016). https://doi.org/10.1109/JOE.2015.2469915
17. Park, J., Han, D.K., Ko, H.: Adaptive weighted multi-discriminator cyclegan for underwater image enhancement. Journal of Marine Science and Engineering (2019)
18. Saharia, C., Chan, W., Chang, H., Lee, C.A., Ho, J., Salimans, T., Fleet, D.J., Norouzi, M.: Palette: Image-to-image diffusion models (2022), https://openreview.net/forum?id=FPGs276lUeq
19. Sahotra, P., Singh, S.: Quality measures technique for underwater images using artificial neural network. Journal of Engineering and Applied Sciences **13**, 3916–3921 (01 2018). https://doi.org/10.3923/jeasci.2018.3916.3921
20. Sohl-Dickstein, J., Weiss, E.A., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37. p. 2256–2265. ICML'15, JMLR.org (2015)
21. Song, Y., Ermon, S.: Improved techniques for training score-based generative models. CoRR **abs/2006.09011** (2020), https://arxiv.org/abs/2006.09011
22. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. CoRR **abs/2011.13456** (2020), https://arxiv.org/abs/2011.13456
23. Wang, J., Li, P., Deng, J., Du, Y., Zhuang, J., Liang, P., Liu, P.: Ca-gan: Class-condition attention gan for underwater image enhancement. IEEE Access **8**, 130719–130728 (2020). https://doi.org/10.1109/ACCESS.2020.3003351

24. Wang, Y., Fan, H., Liu, S., Liu, J., Li, S., Tang, Y.: Ba-gan: Block attention gan model for underwater image enhancement. Proceedings of 2021 IEEE International Conference on Unmanned Systems, ICUS 2021 pp. 809–813 (2021). https://doi.org/10.1109/ICUS52573.2021.9641455
25. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
26. Wu, J., Liu, X., Lu, Q., Lin, Z., Qin, N., Shi, Q.: Fw-gan: Underwater image enhancement using generative adversarial network with multi-scale fusion. Signal Processing: Image Communication **109**, 116855 (11 2022). https://doi.org/10.1016/J.IMAGE.2022.116855
27. Xu, H., Long, X., Wang, M.: Uugan: a gan-based approach towards underwater image enhancement using non-pairwise supervision. International Journal of Machine Learning and Cybernetics pp. 1–14 (9 2022). https://doi.org/10.1007/S13042-022-01659-8/TABLES/5, https://link.springer.com/article/10.1007/s13042-022-01659-8
28. Yang, M., Sowmya, A.: An underwater color image quality evaluation metric. IEEE Transactions on Image Processing **24**(12), 6062–6071 (2015). https://doi.org/10.1109/TIP.2015.2491020
29. Yu, H., Li, X., Lou, Q., Lei, C., Liu, Z.: Underwater image enhancement based on dcp and depth transmission map. Multimedia Tools and Applications **79**, 20373–20390 (7 2020). https://doi.org/10.1007/S11042-020-08701-3/TABLES/3, https://link.springer.com/article/10.1007/s11042-020-08701-3