Wreck-tify – Consistent Diffusion Correction For Shipwreck Mapping

Christian Foreman
cjforema@umich.edu

Ashwin Saxena
ashwinsa@umich.edu

Andrew Scheffer
drewskis@umich.edu

Abstract

Underwater Image Enhancement (UIE) is a difficult and actively studied research problem in computer vision. Underwater images succumb to backscattering and attenuation effects that vary based on the depth from the surface, the distance from the camera, and the frequency of light that is being detected [1]. Current methods for UIE often assume they have access to pairs of ground truth raw-restored image data that can make data collection cumbersome. In this report, our team will first analyze our initial method for training a diffusion model from scratch by implicitly modeling underwater parameters. We will then outline our more consistent method of leveraging the distributional modeling capabilities of pretrained diffusion models to restore underwater images without the need for raw-restored image pairs. The overall goal of this work is to dehaze and color-correct underwater images such that they are conducive to above-water SLAM and scene reconstruction pipelines (i.e. ORBSLAM3 [2], COLMAP [3]). To test the effectiveness of our pipeline, a comparative analysis between our proposed system and simplistic color-distribution shifting is conducted. Fig. 1 shows example UIE output of our method.

1. Introduction

Underwater localization and mapping are challenges in active marine robotics research. Historical deep-water shipwrecks are an important part of history, however, many of them are slowly deteriorating due to human interference, saltwater erosion, and extreme weather conditions [4]. Creating accurate models of these underwater landmarks will be extremely valuable for future historians and scientists. Unfortunately, the underwater domain provides numerous challenges that prevent the adoption of out-of-water algorithms (such as traditional and state-of-the-art localization and mapping systems) from operating effectively.

Vision-based SLAM systems (i.e. ORBSLAM3 [2]) suffer when using images taken deep underwater. Underwater images often suffer from color degradation, distortion, and haze caused by various properties of light. Light refraction

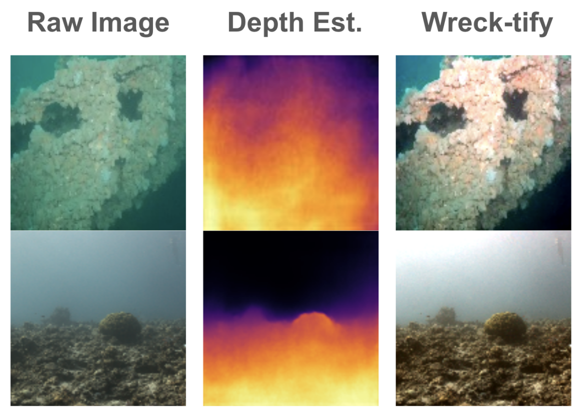

Figure 1. Example of our proposed *Wreck-tify* image enhancement applied to two underwater scenes

mainly contributes to underwater image distortion while attenuation and back-scatter cause color degradation and haze. Suspended particles in the water also make the overall image increasingly blurry. These properties decrease the effectiveness of feature detection and matching algorithms in visual SLAM, which can be seen in Fig. 2 depicting the sparsity of ORB features detected in a shipwreck dataset.

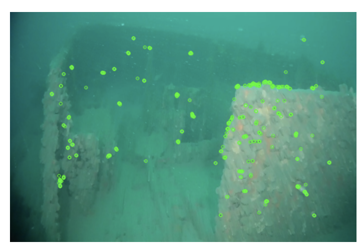

Figure 2. Sparse orb features detected for an underwater shipwreck image

To overcome these challenges, our team proposes to guide the diffusion process of a diffusion model pretrained on only above-water outdoor scenes. By paying attention to globally consistent water parameters and a model of light attenuation and backscattering, our method can consistently dehaze and recolor images across multiple viewpoints of the same scene such that the output is conducive to downstream

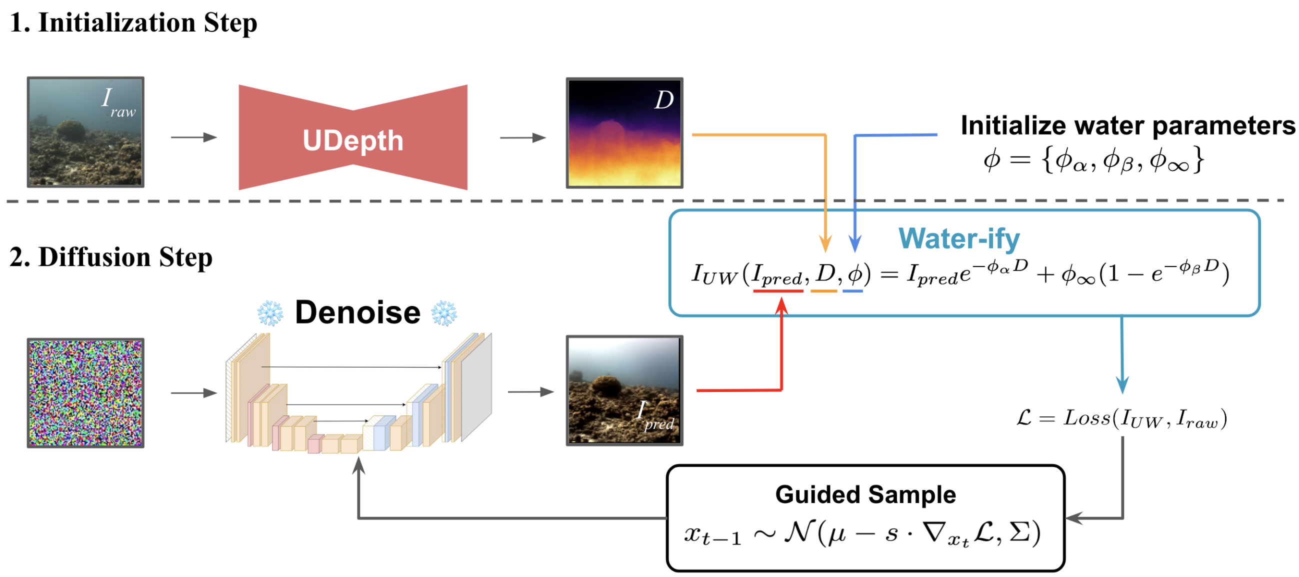

Figure 3. *Wreck-tify* architecture. The initialization step begins by producing a predicted depth-map given the raw underwater image using the off-the-shelf UDepth network [5]. Next, a frozen diffusion model which was pre-trained on outdoor scenes is used to iteratively denoise the input sample. The underwater image formation model [6] and a reconstruction loss are employed to guide the diffusion model to create an above-water counterpart to the underwater input image.

above-water SLAM feature extractors. This technique of employing image enhancement algorithms for underwater images has been explored before [7], however, the use of effective diffusion models for image preprocessing as input to SLAM algorithms has been unexplored and is the focus of our work.

In this paper, our team will first analyze and assess the limitations of our initial method: training a diffusion model from scratch using pairs of underwater and restored images. We will then outline our proposed method, *Wreck-tify*, that leverages the distributional modeling capabilities of pre-trained diffusion models to restore underwater images without ground-truth restored image pairs. The high-level architecture for *Wreck-tify* is presented in Fig. 3.

2. Related Work

In this section, related work in the area of Underwater Image Enhancement (UIE) will be overviewed. Additionally, some recent advances and background in diffusion model techniques will be discussed as our methods build on these ideas.

2.1. Underwater Image Enhancement

We identify two main categories of UIE in recent work: simple color compensation algorithms and methods that model water effects.

2.1.1 Naive Color Compensation and the Gray World

The first class of UIE methods are simple color compensation algorithms like Color Channel Compensation (3C) [8] and Contrast Limited Adaptive Histogram Equalization (CLAHE) [9]. In more traditional image color correction pipelines, oftentimes the complete loss of color channels can lead to adverse effects like image artifacts and color distribution shifts in enhanced images. The 3C method works by reconstructing lost color channels by using an "opponent color space" like LAB. This works by blurring the image and subtracting the local mean from each opponent color pixel to reshift the mean color back to an assumed "neutral" color across most natural scenes. Similarly, CLAHE operates by shifting color distributions of local image patches and limits the values such that it amplifies the contrast of different sections in an image.

Figure 4. Example of 3C color compensation algorithm for underwater scenes

These methods are simple yet effective solutions to colorize images that have severely non-uniform color spectrum distributions captured in low-light, hazy, or underwater conditions. Fig. 4 shows an example of a colorized image from the aforementioned 3C algorithm in low-light underwater conditions.

2.1.2 Modeling Water Effects

In recent years, there has been a growing interest in developing methods that explicitly model the physical properties of underwater environments. This class of algorithms can further be divided into two based on the process of modeling those water effects.

Coefficient Estimation Techniques: Berman et al. introduced a method for estimating water transmission in their paper [10]. By incorporating multiple spectral profiles of different water types and estimating just two additional global parameters—the attenuation ratios of blue-red and blue-green color channels—the method simplifies the problem to single-image dehazing. Importantly, when the water type is unknown, the approach evaluates various parameters from a library of water types, automatically selecting the best result based on color distribution. Fig. 5 shows how the overall pipeline works. The authors also created a dataset, SQUID, using this methodology that is discussed later in our paper.

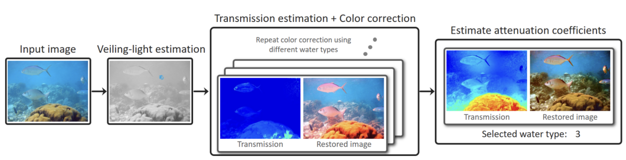

Figure 5. The proposed color restoration and transmission estimation method. First, the veiling-light is estimated. Then, the transmission estimation and color restoration are repeated for multiple water types that have different optical characteristics. Finally, the best result is selected automatically based on the gray-world assumption.

GAN Techniques: Newer UIE techniques have relied on using generative adversarial networks (GANs) to train their algorithms. One notable advancement in this direction is the WaterGAN framework proposed by Li et al. [11]. As shown in Fig. 6, WaterGAN utilizes a GAN to learn the mapping between underwater images and their corresponding enhanced versions, simulating the process of underwater image degradation and restoration. By training on large datasets of paired GAN-generated underwater images and their above water counterparts, an image restoration network can effectively generate enhancements that mitigate the effects of scattering, absorption, and turbulence commonly observed in underwater scenes.

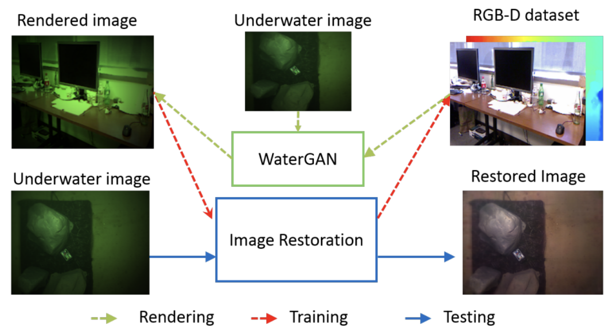

Figure 6. Flowchart displaying both the WaterGAN and image restoration network. WaterGAN takes input in-air RGB-D and a sample set of underwater images and outputs synthetic underwater images aligned with the in-air RGB-D. The image restoration network uses this aligned data for training.

2.2. Diffusion Models

In the past few years, the image-generation, editing, and restoration capabilities of Denoising Diffusion Probabilistic Models (DDPMs) [12] have been very inspiring. We believe that the advances in this area can be readily applied to the underwater image enhancement problem. Recently, image restoration tasks in above-water domains have enjoyed drastic quality improvements through the use of DDPMs. These "diffusion models" have been shown to significantly outperform GAN based counterparts at image restoration and conditional generation tasks [13, 14].

In general, diffusion models work in two stages. First, the *forward process* iteratively adds noise to images drawn from particular distribution. Next, the *reverse process* is learned to iteratively "de-noise" these arbitrary images, allowing one to sample from the learned image distribution. Recent techniques have proposed guiding or conditioning the output of these image samples in different ways [15, 16]. The following sections describe these processes more in detail.

2.2.1 Forward Diffusion Process

Following [12] one can define a forward Markov chain which adds increasing Gaussian noise over T timesteps to an image in the desired distribution, x_0, until it becomes a completely noisy image x_T. The following distribution, $q(x_t|x_{t-1})$ depicts this process where t is the iteration step and $\beta_t \in [0, 1]$ corresponds to the variance of the Gaussian noise at each timestep. This variance typically decreases linearly according to the timestep.

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t|\sqrt{1-\beta}x_{t-1}, \beta_t I) \tag{1}$$

This formulation can actually be simplified such that the value of x_t can be sampled in closed form at any timestep given a linearly decreasing β_t shown in the below equation where $\gamma_t = \prod_{n=1}^{t}(1-\beta_t)$ and $\epsilon \sim \mathcal{N}(0, I)$ follows a standard normal distribution [12]. This is due to the fact that a multiplication of 2 Gaussians is still Gaussian. This simplication allows sampling times to be greatly improved.

$$x_t = \sqrt{\gamma_t}x_0 + \sqrt{1-\gamma_t}\epsilon \tag{2}$$

2.2.2 Reverse Diffusion Process

In the diffusion reverse process, the model is defined as the Markov process that is opposite of the forward process. This is modeled by the following distribution $p_\theta(x_{t-1}|x_t, c)$ where $\mu_\theta(x_t, c, t)$ represents the mean of the distribution, σ_t^2 is the variance, θ represents the model parameters, and c represents the optional conditioning image.

$$p_\theta(x_{t-1}|x_t, c) = \mathcal{N}(x_{t-1}|\mu_\theta(x_t, c, t), \sigma_t^2 I) \tag{3}$$

The reverse process takes T iterative refinement steps to progressively "denoise" a noisy image until it eventually resembles images similar to the training distribution. This works by using the conditional distribution $p_\theta(x_{t-1}|x_t, c)$, which is learned by the neural network model f_θ, to estimate the added forward process noise at the previous timestep. During training, one can use the denoising model f_θ to estimate the values of ϵ. The inputs to f_θ during training are the noisy image x_t (generated from the forward process), a conditioning image c, and the current timestep t. Given this definition, one can train f_θ to produce quality estimates of ϵ using the following loss function.

$$Loss = \mathbb{E}_{(x,y)}\mathbb{E}_{(\epsilon,t)}\left\|f_\theta(\underbrace{\sqrt{\gamma_t}x_0 + \sqrt{1-\gamma_t}\epsilon}_{x_t}, c, t) - \epsilon\right\|_1^1 \tag{4}$$

2.2.3 Conditional and Guided Diffusion

Recent research in generative diffusion models has established ways of guiding or conditioning the generated output on various inputs. For example, the Super Resolution 3 (SR3) model established a simple image conditioning scheme for diffusion models [14]. The authors found that in the reverse-diffusion process, by modeling f_θ as a simple U-Net, one can condition the diffusion by simply concatenating the conditioning image to the input "noisy" image at each timestep.

Another option for conditionally generating in-distribution observations from a diffusion model is through *guidance*. Guidance techniques such as those presented in [17, 18] offer the benefit of not modifying the training procedure of the diffusion model and only affecting the inference procedure. Oftentimes, these works utilize some conditioning factor (such as an image reconstruction loss for example) to guide the next sample for the diffusion model to de-noise. Overall, we found that guidance-based techniques applied to large pre-trained diffusion models are the most effective way to achieve conditional image generation without costly training.

3. Technical Approach

In this section, we will discuss both of our attempts at creating a diffusion model for underwater image correction.

3.1. Attempt 1: Training a Model from Scratch

Our first attempt at performing this UIE task involved training a diffusion model on a dataset containing underwater images paired with reference "corrected" images. For this task, our team used the LSUI dataset [19] depicted in Fig. 7. This dataset contains roughly 5000 image pairs with varying depths, lighting conditions, and water types. This dataset was created with a transformer-based method. Therefore the "corrected" images should not be interpreted as true-color ground truth but instead as "enhanced reference images." Given this dataset, we can define our model's task as taking an underwater image as input and producing an image similar to the LSUI "reference" images as output.

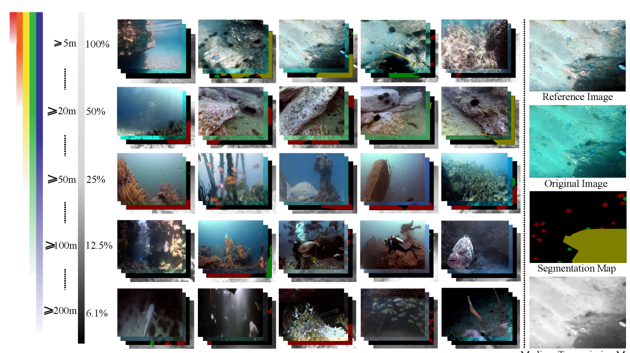

Figure 7. LSUI dataset example images

In determining how to formulate our diffusion model for this task, we drew inspiration from the super-resolution diffusion model (SR3) [14] that shows how diffusion can be conditioned on input images by simply concatenating the conditioned image onto the partially denoised image x_t at timestep t of the denoising process. Our team used the preprocessed underwater image after running 3C color compensation to provide a condition with a more above-water color distribution. This high-level structure is depicted in Fig. 8.

In practice, we implement the reverse diffusion step, f_θ, using a simple U-Net by concatenating x_t and c_t as seen in

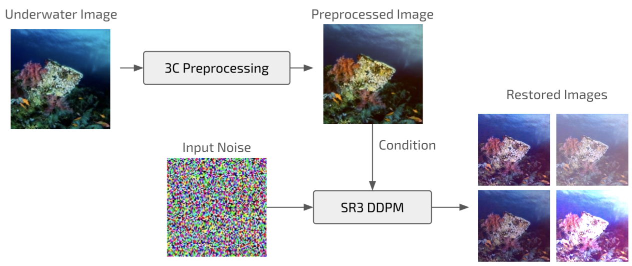

Figure 8. Our SR3-based diffusion model including 3C preprocessing and conditional diffusion model generation.

Fig. 9. As seen, the value of c_t is determined by performing the procedure defined in 3C [8] on a raw underwater image. Our intuition was that this mean-color-shifted image would guide the diffusion model to produce reasonable colors as well as align the output image at each timestep with the contours, shapes, and gradients defined in the raw image. The model was also given access to the specific timestep, t, allowing it to learn the scheduling routine for γ_t and therefore better predict the noise. At the end of training, our function f_θ should accurately predict noise in a noisy input image. Then, using a deterministic function, this predicted noise can be iteratively removed from the image over T denoising iterations following the procedure in SR3 [14].

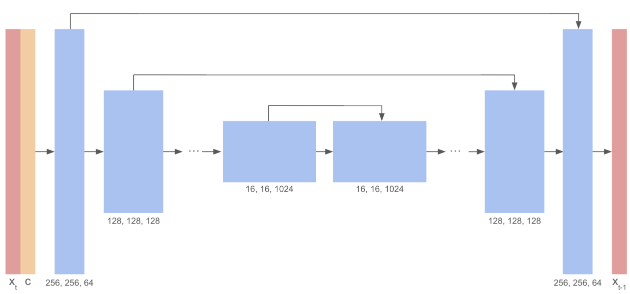

Figure 9. Description of the U-Net reverse diffusion model. 3C is applied to the raw underwater image to create c which is concatenated to a noisy reference image, x_t, to guide the denoising of the underwater reference image.

After getting some preliminary results from this method our team noticed a few large issues when applying this method to downstream SLAM systems. First, as seen in Fig. 8, the output of our diffusion model actually varies for a given input underwater image. We believe that this is largely due to the fact that the act of performing 3C color compensation looses valuable information about the true distribution of color in the image. We noticed that this simple fact actually destroys the usefulness of this pipeline as an input to SLAM systems because different viewpoints of the same scene (at potentially variable depths) will produce drastically different color profiles – making feature matching quite difficult. Additionally, another limitation of this method is the amount of time that it takes to train. Our team found that in order to produce visually pleasing images using this method we had to train the diffusion model for over 30 hours. Finally, another major limitation stems from the severe shortage of raw-restored image pairs in readily available datasets. These limitations inspired our team to design the proposed *Wreck-tify* model, described in the next section.

3.2. Attempt 2: The Wreck-tify Pipeline

Building off of the ideas and limitations from the previous SR3-based diffusion method for UIE, we propose the *Wreck-tify pipeline*. The model diagram for this new architecture is shown in Fig. 3. This architecture contains a few important design principles. First, we chose to utilize a large (2GB) pre-trained image diffusion network trained unconditionally on a large number of outdoor scene datasets [20]. By training the diffusion model on these datasets the model is effectively learning a large prior distribution of what in-air outdoor scenes look like (i.e. their color distributions, the fact that color is not dependent on distance, etc). The main idea of the *Wreck-tify* method is to guide the test-time diffusion of this model so that it produces an image that is in the "in-air" distribution but still contains the same information contained in the original underwater image.

As seen in Fig. 3, this guidance is achieved by applying attenuation and back-scattering effects to the noisy "corrected" image at every step of the denoising process. This degradation process is defined using the underwater image formation model per channel for each pixel [6] using Eq. 5). Here, I_{pred} is the in-air image, D is the depth from the camera, ϕ_∞ is the color of the water at infinity, and ϕ_α and ϕ_β are the attenuation and back-scatter coefficients respectively.

$$I_{UW}(I_{pred}, D, \phi) = I_{pred}e^{-\phi_\alpha D} + \phi_\infty(1 - e^{-\phi_\beta D}) \quad (5)$$

This underwater-degraded prediction, I_{UW}, is directly compared to ground truth underwater image I_{raw} that was the initial input to the model. Following [20], we use a diffusion model library to compute the reconstruction loss of the underwater image and use this information to guide the next denoising step.

Importantly, the above water degradation step only works if there exists an accurate depth estimation for each pixel in the image. For this task, we use an off-the-shelf underwater depth prediction network named UDepth [5]. As seen in Fig. 9, this model directly operates on the input monocular underwater image and outputs a normalized depth map. Another important set of parameters that need to be correctly initialized in this framework are the water parameters $\phi = \{\phi_\alpha, \phi_\beta, \phi_\infty\}$ defined uniquely for each channel. As a simplification of our method, we assume that these parameters are known a-priori. In practice, one can iteratively refine these parameters by hand until reasonable diffusion results are attained.

When applying this system to multiple frames of a video we have the assumption that the water parameters do not change across the different images. We have found that this simple physics-based assumption enforces temporal UIE consistency such that downstream SLAM systems can accurately retrieve and match ORB features from images with varying viewpoints. Importantly, we also find that even if the water parameters are incorrect for a particular scene, the diffusion outputs are still temporally consistent in their color distribution. One obvious limitation of this assumption, however, is that if the path of the underwater vehicle involves drastic changes in the depth of the water column, the assumption that the water parameters are constant will not hold. For more discussion and results regarding the temporal consistency of this algorithm and the importance of the water parameters please see Section 4.

3.3. Downstream Visual SLAM

Traditionally, visual SLAM systems do not work very well underwater [21]. This is likely due to poor image quality resulting in a lack of features and inaccurate depth measurements. By first performing underwater image enhancement, we aim to be able to extract more visual features underwater. To compare the effectiveness of our method, our team wrote a simple ORB-feature matcher that can be used to establish a sense of the quality of our processed images. Due to the slow inference of our methods, running the entire SLAM pipeline was deemed infeasible for this project.

4. Evaluation and Results

Our team evaluated our method at two distinct stages in the *Wreck-tify* pipeline. First, we evaluate the single-image quality both qualitatively and quantitatively using image quality metrics. Second, groups of underwater images collected from distinct scenes are analyzed to determine both our method's effectiveness at being color-consistent for different camera viewpoints and our method's ability to generate images conducive to visual SLAM feature detectors like ORB.

All evaluation was performed using a single Nvidia RTX 3070 Ti GPU, requiring roughly two minutes of processing time per image. This long inference time for our method disqualifies it from being run in real-time as a preprocessing stage for SLAM systems; however, our method still has the potential to be applied as a preprocessing stage before running offline localization and mapping algorithms such as COLMAP.

4.1. Datasets

The **Stereo Quantitative Underwater Image Dataset (SQUID)** [10] is a small-scale underwater image dataset. The creators collected a dataset of images taken in different locations with varying water properties, showing color charts in the scenes. Moreover, to obtain ground truth, the 3D structure of the scene was calculated based on stereo imaging enabling a quantitative evaluation of restoration algorithms on natural images.

The **Shipwreck Dataset** [22], which we use in our evaluation, consists of continuous camera feeds and multi-beam imaging sonar on four different shipwrecks in the Mediterranean Sea (50-68 meters in depth).

4.2. Qualitative Single-Image Analysis

The image quality of the enhanced images after running through our *Wreck-tify* pipeline can be analyzed using Fig. 10 above. In this figure, the raw underwater images were captured in drastically different underwater environments. This demonstrates our model's versatility and adaptability when applied to scenes with varying water parameters.

Quite immediately, one can see that the color distribution of the processed input images in Fig. 10 are similar to those of the above-air images. The color of distant objects seems to not fade as quickly as it did in the raw image. This provides evidence that the pre-trained above-water diffusion model of *Wreck-tify* is providing a useful above-water image prior.

One interesting aspect of Fig. 10 to note is that the extracted depth from the RAW underwater images seem to be quite blurry on some of the input images. This may be because the distribution of the underwater images that we are evaluating are not in the distribution of images that UDepth was trained on. It's also interesting to see that at obvious "cutoff" points in the raw images (i.e. at the edges of three pillars in the third column) the predicted depth value appears to have a "smooth dropoff" that does not seem to be consistent with the input image. This low-resolution and potentially unreliable depth prediction step has definite areas for improvement as a component of our underwater image enhancement pipeline.

We also compare the single-image output of our method to the CLAHE color compensation baseline introduced in Section 2. As seen in Fig. 11, CLAHE appears to have purple and red artifacts that likely do not reflect the actual in-air color distribution of the shipwreck. In contrast, our method seems to produce a more reasonable color distribution for the image.

4.3. Quantitative Single-Image Analysis

To quantitatively evaluate our method compared to the raw image and the CLAHE baseline, we evaluate corresponding Underwater Image Quality Metric (UIQM) values

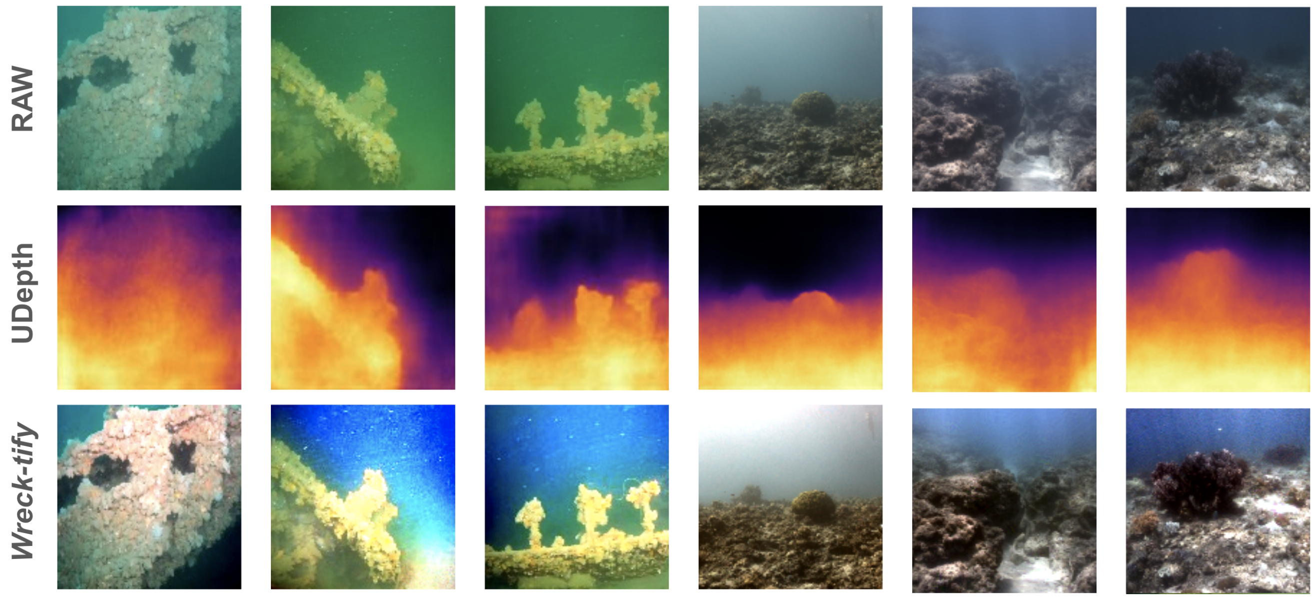

Figure 10. *Wreck-tify* results on the Shipwreck Dataset [22] (three leftmost images) and the SQUID Dataset [10] (three rightmost images).

[23]. This metric comprises of three distinct components: UISM which measures image sharpness, UIConM which measures image contrast, and UICM which measures image colorfulness. Each of these metrics are averaged and shown in Table 1 for raw underwater images, CLAHE-processed images, and *Wreck-tify* processed images.

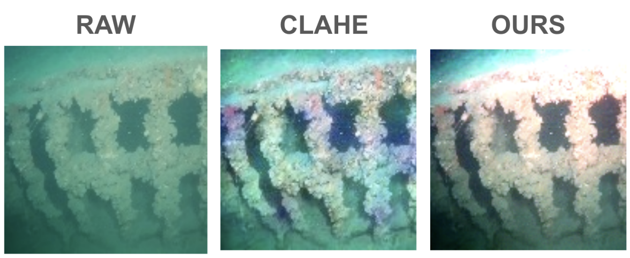

Figure 11. Comparison of our method compared to the color compensation baseline, CLAHE, on an underwater shipwreck image.

Table 1. Comparison of Underwater Image Quality Metrics (UIQM) for underwater images, CLAHE preprocessed images and *Wreck-tify* preprocessed images.

	UISM	UIConM	UICM	UIQM
Raw	4.78	1.27	1.09	1.19
CLAHE	**5.68**	**2.10**	4.40	**1.34**
Ours	4.64	1.87	**5.62**	1.12

The only metric where our method consistently outperforms the baseline is UICM (measuring colorfulness). Although CLAHE outperforms our method in terms of image contrast, our method directly improves contrast from the raw image. Also, as seen in Fig. 11, the increased contrast of CLAHE preprocessing seems to come at the cost of image artifacts that are likely not consistent across viewpoint changes.

4.4. Investigating Image Consistency

One of the core challenges we were trying to solve with our *Wreck-tify* model was with keeping the color correction consistent across consecutive frames in the same scene. To qualitatively evaluate the performance of our method, we ran our model on several frames of the same scene. For this case, we set the water parameters to be the same across each generated frame. We obtained the 9 water parameters via an estimation technique from [20].

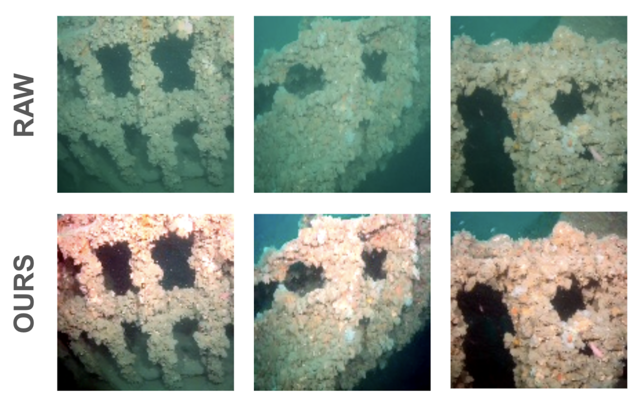

Figure 12. Example *Wreck-tify* output on the same scene.

As we can see in Fig. 12, our method seems to output images that are consistent with each other. However, for this case, we noticed that our method consistently over corrects in the red channel as the shipwreck is over saturated with red color. We believe this is due to the water parameters of our input being slightly off. For more examples of color consistency and the effects of different water parameters, refer to the appendix.

4.5. SLAM Evaluation

Since the image enhancement for each image takes approximately two minutes, we were unable to run complete ORBSLAM3 on the shipwreck dataset. To evaluate the effectiveness of our algorithm we carried out ORB feature matching and compared it to a CLAHE enhanced image as shown in Fig. 13 and 14. Qualitatively, our results show that our method has comparable ORB matches in the examples shown. Since these images were taken from successive frames of a video, CLAHE is able to perform relatively well. We believe, however, that due to the color inconsistency issues of CLAHE, our method would be better when running a complete SLAM algorithm, and identifying loop closures from varying viewpoints.

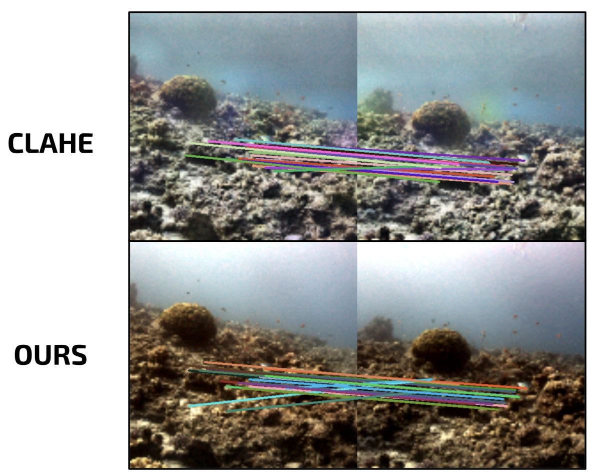

Figure 13. Comparison of ORB features matching between CLAHE and our proposed method Example 1

5. Conclusion

Our paper addresses the significant challenges of Underwater Image Enhancement (UIE) for applications in underwater localization and mapping. To overcome these challenges, we introduced the *Wreck-tify* pipeline, which leverages diffusion models pre-trained on above-water scenes. By guiding the diffusion process with an understanding of water parameters and light effects, our method aims to produce images conducive to downstream visual SLAM pipelines. Our approach shows improvements in image

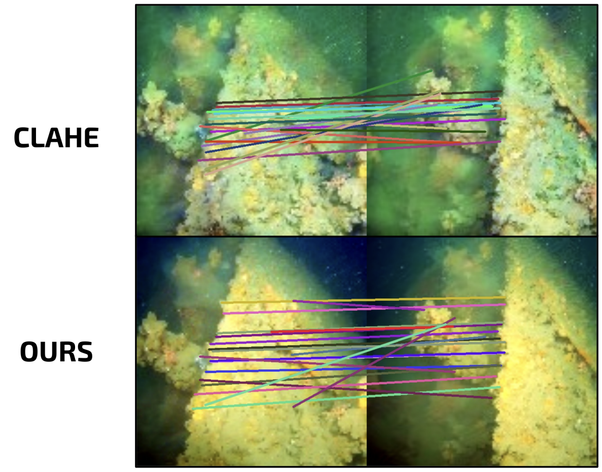

Figure 14. Comparison of ORB features matching between CLAHE and our proposed method Example 2

quality metrics, such as contrast and colorfulness, compared to the raw image. Moreover, our method maintains color consistency across consecutive frames for applications like SLAM. While there are many opportunities for improvements, such as decreasing inference time and improving depth estimation, our work provides a step forward in addressing UIE challenges for underwater robotics.

References

[1] Tomasz Łuczyński and Andreas Birk. *Underwater Image Haze Removal and Color Correction with an Underwater-ready Dark Channel Prior*. 2018. arXiv: 1807.04169 [cs.CV].

[2] Carlos Campos et al. "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM". In: *IEEE Transactions on Robotics* 37.6 (Dec. 2021), pp. 1874–1890. ISSN: 1941-0468.

[3] Johannes Lutz Schönberger and Jan-Michael Frahm. "Structure-from-Motion Revisited". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

[4] Charlotte Gjelstrup Björdal and Thomas Nilsson. "Reburial of shipwrecks in marine sediments: a long-term study on wood degradation". In: *Journal of Archaeological Science* 35.4 (2008), pp. 862–872. ISSN: 0305-4403.

[5] Boxiao Yu, Jiayi Wu, and Md Jahidul Islam. *UDepth: Fast Monocular Depth Estimation for Visually-guided Underwater Robots*. 2023.

[6] Derya Akkaynak and Tali Treibitz. "A Revised Underwater Image Formation Model". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 6723–6732. DOI: 10.1109/CVPR.2018.00703.

[7] A. Quattrini Li et al. "On Understanding the Challenges in Vision-Based Shipwreck Mapping". In: *ICRA 2016 Workshop on Marine Robot Localization and Navigation*. Stockholm, Sweden, May 2016, pp. 1–3.

[8] Codruta O. Ancuti et al. "Color Channel Compensation (3C): A Fundamental Pre-Processing Step for Image Enhancement". In: *IEEE Transactions on Image Processing* 29 (2020), pp. 2653–2665.

[9] Garima Yadav, Saurabh Maheshwari, and Anjali Agarwal. "Contrast limited adaptive histogram equalization based enhancement for real time video system". In: *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. 2014, pp. 2392–2397. DOI: 10.1109/ICACCI.2014.6968381.

[10] Dana Berman et al. "Underwater Single Image Color Restoration Using Haze-Lines and a New Quantitative Dataset". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[11] Jie Li et al. "WaterGAN: Unsupervised Generative Network to Enable Real-time Color Correction of Monocular Underwater Images". In: *IEEE Robotics and Automation Letters (RA-L)* (2017). accepted.

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. *Denoising Diffusion Probabilistic Models*. 2020.

[13] Xin Li et al. *Diffusion Models for Image Restoration and Enhancement – A Comprehensive Survey*. 2023.

[14] Chitwan Saharia et al. *Image Super-Resolution via Iterative Refinement*. 2021.

[15] Kangfu Mei et al. *CoDi: Conditional Diffusion Distillation for Higher-Fidelity and Faster Image Generation*. 2024. arXiv: 2310.01407 [cs.CV].

[16] Jiwen Yu et al. *FreeDoM: Training-Free Energy-Guided Conditional Diffusion Model*. 2023. arXiv: 2303.09833 [cs.CV].

[17] Jonathan Ho and Tim Salimans. *Classifier-Free Diffusion Guidance*. 2022. arXiv: 2207.12598 [cs.LG].

[18] Mingqiang Wei et al. *RainDiffusion: When Unsupervised Learning Meets Diffusion Models for Real-world Image Deraining*. 2023. arXiv: 2301.09430 [cs.CV].

[19] Lintao Peng, Chunli Zhu, and Liheng Bian. *U-shape Transformer for Underwater Image Enhancement*. 2021.

[20] Opher Bar Nathan et al. "Osmosis: RGBD Diffusion Prior for Underwater Image Restoration". In: *arXiv preprint arXiv:2403.14837* (2024).

[21] Felipe Guth et al. "Underwater SLAM: Challenges, state of the art, algorithms and a new biologically-inspired approach". In: *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*. 2014, pp. 981–986.

[22] Nicolas Pecheux et al. "Self Calibration of a Sonarndash;Vision System for Underwater Vehicles: A New Method and a Dataset". In: *Sensors* 23.3 (2023).

[23] Karen Panetta, Chen Gao, and Sos Agaian. "Human-Visual-System-Inspired Underwater Image Quality Measures". In: *IEEE Journal of Oceanic Engineering* 41.3 (2016), pp. 541–551.

Appendix A: Pictures

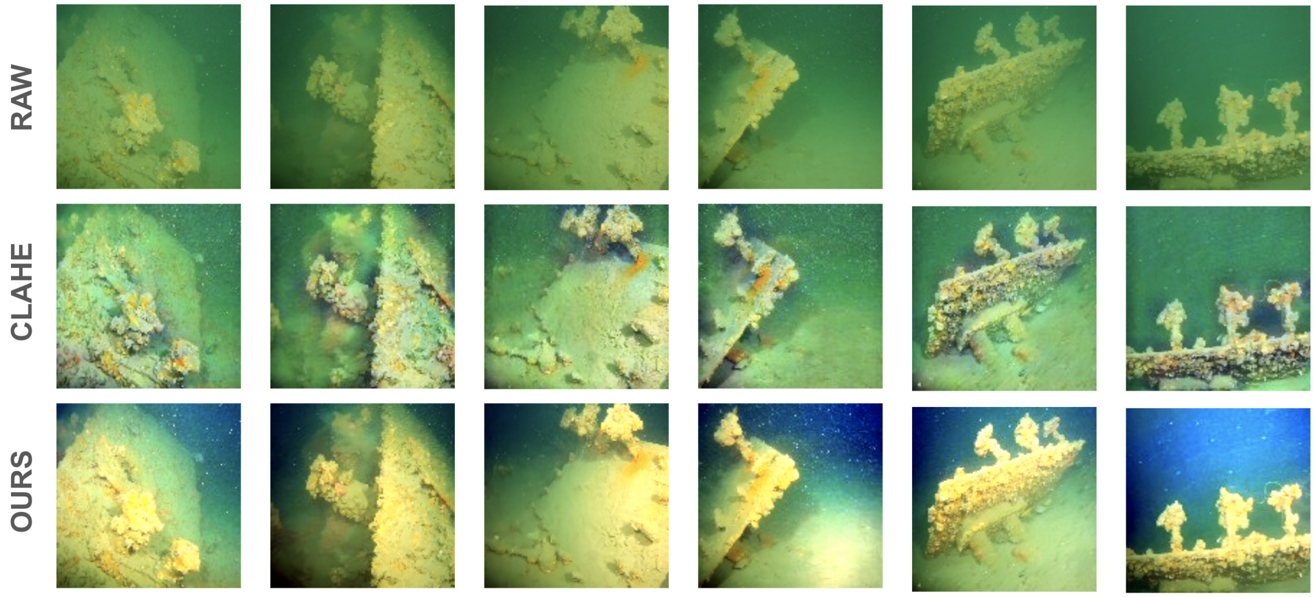

Figure 15. Color consistency comparison between CLAHE and *Wreck-tify* on images of the same shipwreck from the shipwreck dataset.

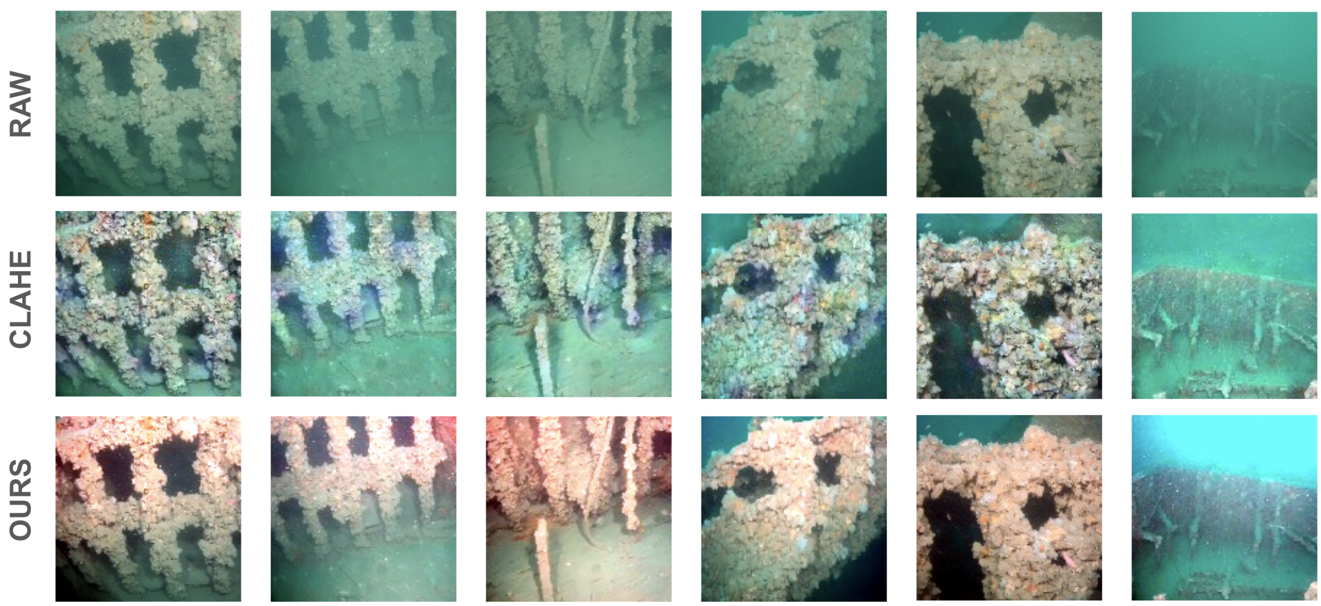

Figure 16. Color consistency comparison between CLAHE and *Wreck-tify* on images of the same shipwreck from the shipwreck dataset. Notice how CLAHE tends to have these artifacts of blue patches and tends to not be as consistent as our method.

Appendix B: Different Water Parameters

Figure 17. The effects of different water parameters on the same raw source images. The water parameters are defined by 9 constants where there are 3 constants for each RGB channel. As seen in the figure, the initialized water parameters greatly affect the quality of the image enhancement.

Appendix C: Code Implementation

To view the code of our method and our evaluation, please refer to our Google Drive: https://drive.google.com/drive/folders/1eYD1rIn3VRbFfLm-DRhclyFVmLX9FI7T?usp=drive_link