Neural Underwater Scene Representation

Supplementary Material

Yunkai Tang[1,2†] Chengxuan Zhu[3†] Renjie Wan[4*] Chao Xu[3] Boxin Shi[1,2*]

[1]National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

[2]National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University

[3]National Key Lab of General AI, School of Intelligence Science and Technology, Peking University

[4]Department of Computer Science, Hong Kong Baptist University

tangyunkai@stu.pku.edu.cn, peterzhu@pku.edu.cn, renjiewan@hkbu.edu.hk,

xuchao@cis.pku.edu.cn, shiboxin@pku.edu.cn

https://freebutuselesssoul.github.io/uwnerf

A. More on underwater dynamic factors

Underwater environments present unique challenges for scene modeling, primarily due to the three dynamic factors introduced in Sec. 1. In this part, we explain more about the three factors.

Firstly, ① **distance-dependent visibility** plays a crucial role. As an object moves further away from the camera, it becomes increasingly difficult to observe. This phenomenon results from the inherent scattering and absorbing properties of water, which significantly diminish the clarity and intensity of light as it travels through the medium. The scattering leads to light rays deviating from their original paths, while absorption reduces their overall strength. As a result, distant objects appear more blurry and less discernible, presenting a major challenge for accurate visual representation in underwater scenes.

Secondly, the aquatic environment is characterized by ② **unstable illumination**. This variability stems from the scattering effect of water particles and the fluctuating lighting conditions, often influenced by factors such as the time of day, weather, wave turbulence, and water turbidity. Light rays bend, scatter, and get absorbed differently at various points, creating non-uniform lighting conditions. These changes in illumination are not only spatially diverse but also temporally varying, making the lighting observed from different viewpoints inconsistent over time. Such instability in illumination complicates the task of modeling underwater scenes accurately.

Lastly, the underwater realm is a dynamic ecosystem, bustling with marine life, which introduces ③ **moving objects** as the third dynamic factor. This includes a myriad of marine plants and animals, each contributing to the ever-changing visual landscape. These moving entities defy the static assumptions commonly held in vanilla NeRF models, which are typically designed for static scenes. The continuous movement of these elements not only alters the visual scene but also interferes with the light paths, adding another layer of complexity.

The confluence of these factors creates a highly complex and constantly evolving environment. This complexity poses significant challenges for current NeRF models, which struggle to comprehend and represent such dynamic underwater scenes accurately.

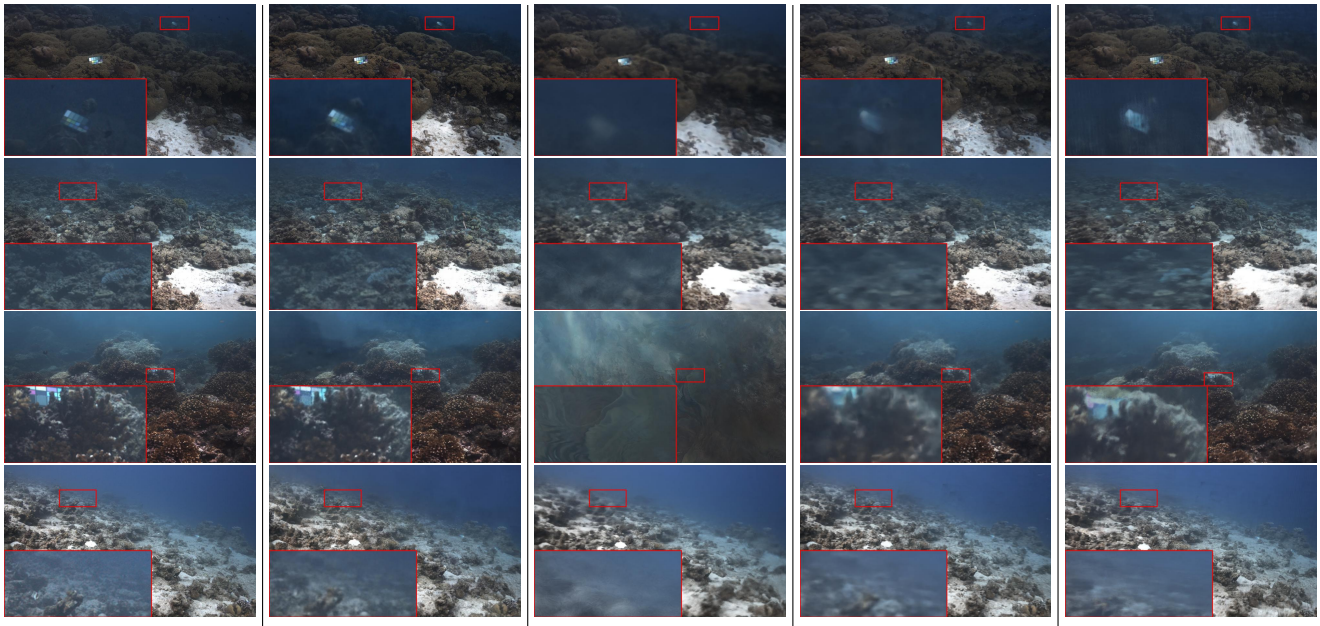
B. More explanations on equations

B.1. Explanation on Eq. (3)

Consistent with assumption in SeaThru-NeRF [4], we assume objects are opaque, thus σ_{obj} becomes very high near object surfaces and almost zero elsewhere. Water is semi-transparent and owns an empirically low non-zero density. This results in $\sigma_{\text{obj}} \gg \boldsymbol{\sigma}_{\text{w}}$ near object surfaces and $\sigma_{\text{obj}} \ll \boldsymbol{\sigma}_{\text{w}}$ in media. In this unified rendering equation, the minor items of σ_{obj} and $\boldsymbol{\sigma}_{\text{w}}$ can always be ignored, simplifying Eqs. (3) and (4) to similar equations in works such as DehazeNeRF [2].

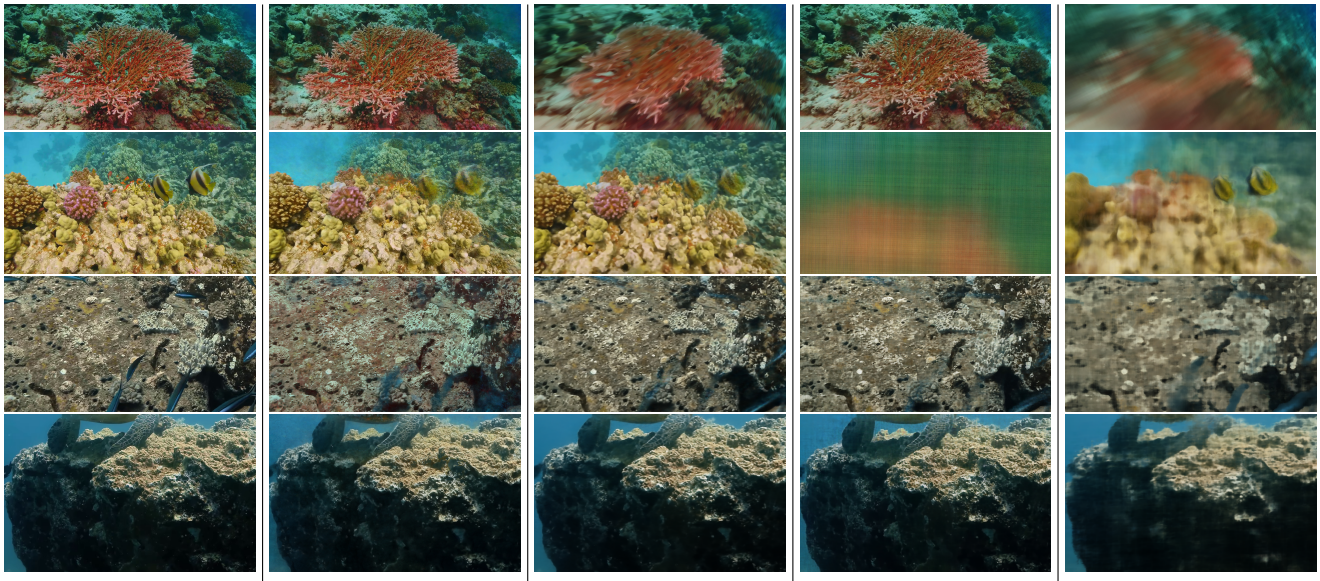
B.2. Explanation on Eq. (8)

Instead of using a linear function in Eq. (8), we designed a sinusoidal function which maps the original proportions of density $\frac{\{\sigma_{\text{sta},i},\sigma_{\text{dyn},i},\boldsymbol{\sigma}_{\text{w}}\}}{\sigma_{\text{sta},i}+\sigma_{\text{dyn},i}+\boldsymbol{\sigma}_{\text{w}}} \in [0,1]$ into weight factors $\boldsymbol{\beta}_{\{\text{sta,dyn,w}\},i} \in [0,1]$. This function results in a more obvious separation between density of different scene components (static object, moving object and media) in the same



(a) Ground Truth | (b) Proposed-T | (c) MIP-360 [1] | (d) Instant-NGP [5] | (e) SeaThru-NeRF [4]

Figure S1. More qualitative comparisons on the SeaThru dataset [4]. Images are in high resolution, and please zoom-in for details.



(a) Ground Truth | (b) Proposed | (c) Instant-NGP [5] | (d) SeaThru-NeRF [4] | (e) DynamicNeRF [3]

Figure S2. Qualitative comparisons on the proposed dataset. Images are in high resolution, and please zoom-in for details.

position. Scene components with density proportion larger than 0.5 will be given an weight factor larger than its original proportion, while components with density proportion less than 0.5 will obtain a smaller factor. Utilizing this sinusoidal function allows a more quickly convergence for our model.

C. More results

Our proposed method is capable of representing underwater scenes consistently in both spacial and temporal dimensions. We conduct several experiments to illustrate the consistency of our model.

In addition to those already displayed in Fig. 4 of the main paper, we show more results obtained by our method under "Proposed-T" setting, Instant-NGP [5], SeaThru-NeRF [4], and MIP-360 [1] on the SeaThru dataset [4].

"Proposed-T" refers to the architecture obtained by removing all time-related components in our proposed method. Since SeaThru dataset [4] is composed of sparsely captured photos and does not include temporal information, we train with "Proposed-T" on this dataset. We demonstrate the synthesized images on the validation view of the SeaThru dataset [4] in Fig. 8. A patch is selected for closer observation on details. The proposed method can render vivid water medium effects and also reconstruct the intricate structures. MIP-360 [1] and Instant-NGP [5], as state-of-the-art NeRF methods for scenes above sea level, are ignorant of the distance-dependent visibility. Therefore both of them fail to reconstruct the correct appearance of the scene. In especial, MIP-360 [1] fails to predict the depth of the scene and results in even worse degenerated effects. SeaThru-NeRF [4] manages to model the static scenes under water, but struggles in high-frequency details due to its network design, such as frequency-based encoding.

We also show qualitative results on the proposed dataset in Fig. 9, comparing results obtained by the proposed method, Instant-NGP [5], SeaThru-NeRF [4], and DynamicNeRF [3]. Due to a lack of consideration of time dimension, Instant-NGP [5] and Seathru-NeRF [4] fail to model the time-dependent changes in the scene. Blurry artifacts are observed around the moving objects. By zooming in, it is clear that Instant-NGP [5] cannot reconstruct as much details as the proposed method. It also hinders the ray termination estimation in MIP-NeRF [1], causing significant degeneration in the rendered test view. Though DynamicNeRF [3] is able to represent moving objects in the scene, it neglects the absorbing and scattering effects of the water medium, resulting in distortion in scene reconstruction. It also preserves fewer high-frequency details, since the scene flow supervision in DynamicNeRF [3] leads to a over-smoothness. Our proposed method distinctively models different underwater dynamics, achieving the most realistic results of underwater scene representation.

For a more intuitive comparison of the different methods, we render a video with a fixed time step and changing camera poses in CORAL, and another with a fixed camera pose and changing time step in COMPOSITE. The latter shows significant time-varying illumination and moving fish in the scene, validating the proposed method. **Please refer to the project page for video.**

References

[1] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2, 3

[2] Wei-Ting Chen, Wang Yifan, Sy-Yen Kuo, and Gordon Wetzstein. DehazeNeRF: Multiple image haze removal and 3D shape reconstruction using neural radiance fields. *arXiv preprint arXiv:2303.11364*, 2023. 1

[3] Chen Gao, Ayush Saraf, Johannes Kopf, and Jia-Bin Huang. Dynamic view synthesis from dynamic monocular video. In *Proc. of International Conference on Computer Vision*, 2021. 2, 3

[4] Deborah Levy, Amit Peleg, Naama Pearl, Dan Rosenbaum, Derya Akkaynak, Simon Korman, and Tali Treibitz. SeaThru-NeRF: Neural radiance fields in scattering media. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 1, 2, 3

[5] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 2022. 2, 3