A Appendix

A.1 Underwater View Synthesis Dataset

Due to the absence of multi-view underwater scene datasets suitable to evaluate novel-view rendering, we establish a new benchmark Underwater View Synthesis (UVS) Dataset containing 8 scenes, equally split into *easy* (real-world high quality) and *hard* (real-world low quality). For real-world data, we hand-pick 4 scenes from high quality youtube videos to form the easy split, while the hard split is composed of low-quality noisy real-world captures obtained during a diving expedition. For each scene from the easy and hard splits, we select roughly 100-150 images and calibrate them using Structure-from-Motion (SfM) algorithm in an open-source software package COLMAP Schönberger & Frahm (2016); Schönberger et al. (2016). For COLMAP, we use a "simple radial" camera model with a single radial distortion coefficient and a shared intrinsic for all images. We use a "sift feature guided matching" option in the exhaustive matcher step of SfM and also refine principle points of the intrinsic during the bundle adjustment.

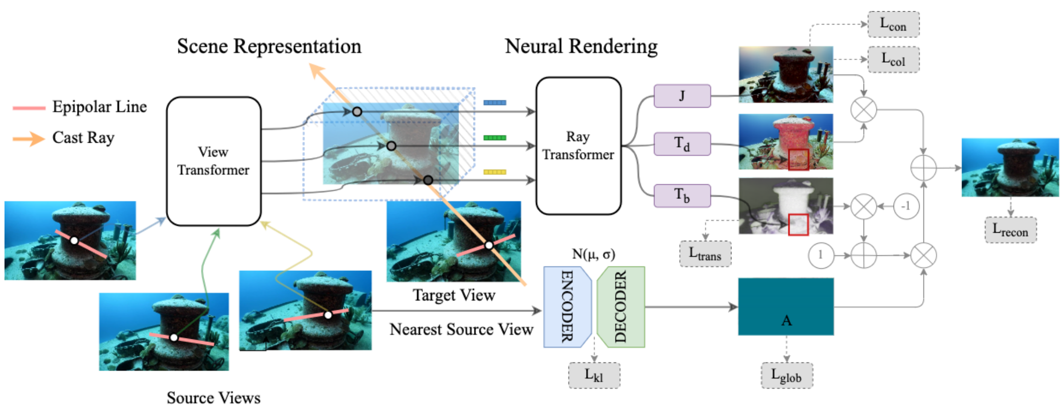


Figure 2: Overview of U2NeRF: 1) Identify source views for a given target view, 2) Extract features for epipolar points using a trainable U-Net like model, 3) For each ray in the target view, sample points and directly predict a target patch disentangled into scene radiance, direct and backscatter transmission maps, and global background light. 4) The individual components are combined based on the image formation model to reconstruct the underwater image which is used as a self-supervision loss.

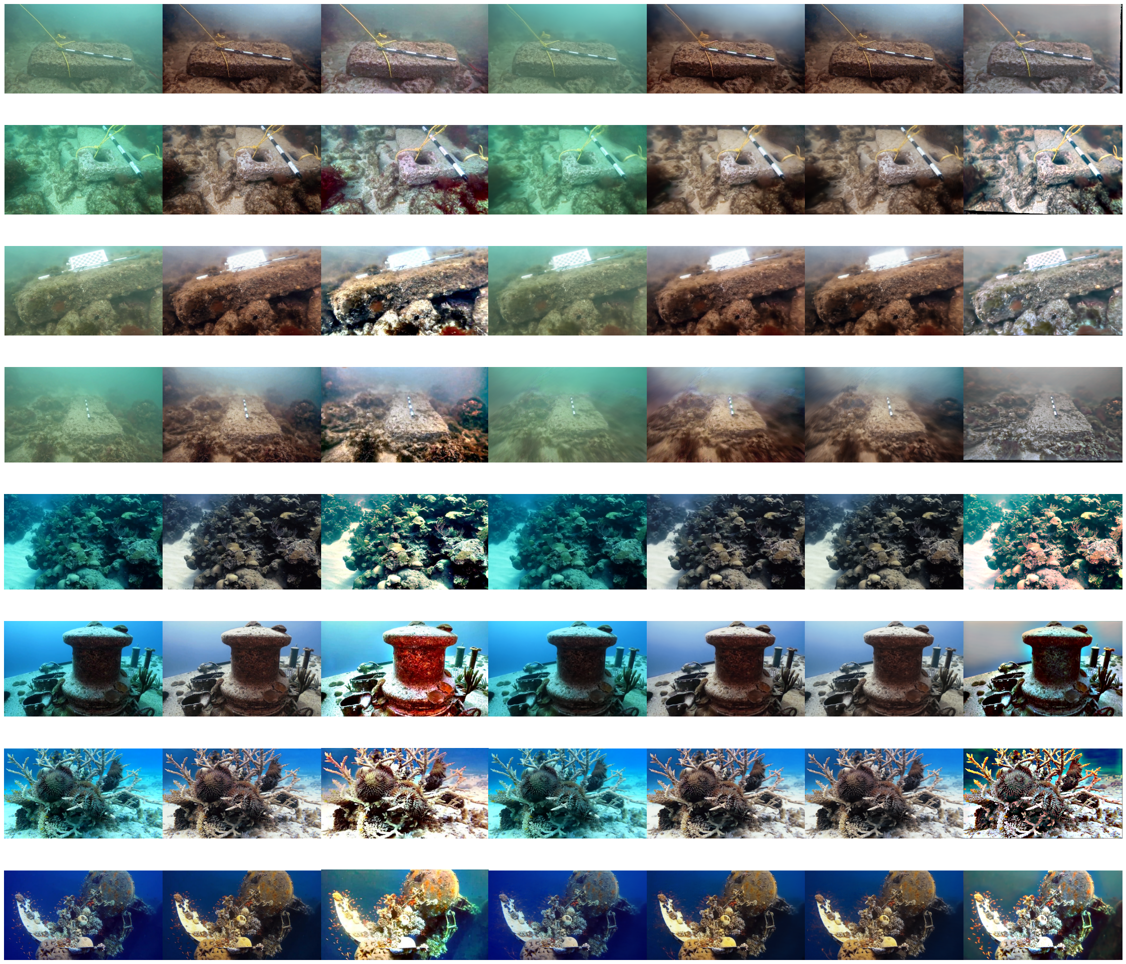
Models	UIQM↑	UCIQE↑	LPIPS (gray)↓
UPIFM	1.424	32.940	-
UIESS	1.136	30.534	-
NeRF	0.501	31.622	0.208
NeRF + UIESS	0.865	31.054	0.223
UIESS + NeRF	0.858	30.336	0.198
U2NeRF	**1.570**	**32.556**	**0.174**

(a) Easy Split

Models	UIQM↑	UCIQE↑	LPIPS (gray)↓
UPIFM	1.182	28.537	-
UIESS	0.649	27.161	-
NeRF	0.463	18.370	0.334
NeRF + UIESS	0.486	**27.453**	0.328
UIESS + NeRF	0.456	26.530	0.292
U2NeRF	**1.100**	26.788	**0.260**

(b) Hard Split

Table 1: Comparison of U2NeRF against baseline methods for single-scene rendering on the UVS dataset



(a) Degraded Image (b) UIESS (c) UPIFM (d) NeRF (e) NeRF + Clean (f) Clean + NeRF (g) U2NeRF

Figure 3: Qualitative results on single-scene rendering for Easy and Hard Scenes. The top 4 rows depict the scenes from the hard split(Scene 1, Scene 2, Scene 3 and Scene 4 respectively) and the last 4 rows depict the scene from the easy split(coral, shipwreck, starfish and debris respectively). (a) represents the actual underwater image from the scene, (b) & (c) represents the no rendering baseline methods (Chen & Pei (2022) Chai et al. (2022)), (d), (e) & (f) refers to the renderings from NeRF on raw underwater image, restored view after NeRF rendering and NeRF rendering on restored input underwater images respectively, and (g) refers to results from our method: U2NeRF. U2NeRF is able to render better high-quality images when compared to other rendering+restoring methods.

REFERENCES

Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5855–5864, 2021.

Shu Chai, Zhenqi Fu, Yue Huang, Xiaotong Tu, and Xinghao Ding. Unsupervised and untrained underwater image restoration based on physical image formation model. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2774–2778. IEEE, 2022.

Tianlong Chen, Peihao Wang, Zhiwen Fan, and Zhangyang Wang. Aug-nerf: Training stronger neural radiance fields with triple-level physically-grounded augmentations. In *Proceedings of the*

IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15191–15202, 2022.

Yu-Wei Chen and Soo-Chang Pei. Domain adaptation for underwater image enhancement via content and style separation. *IEEE Access*, 10:90523–90534, 2022.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I*, pp. 405–421, 2020.

Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5480–5490, 2022.

Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.

Mukund Varma T, Peihao Wang, Xuxi Chen, Tianlong Chen, Subhashini Venugopalan, and Zhangyang Wang. Is attention all nerf needs? *arXiv preprint arXiv:2207.13298*, 2022.