U2NERF: UNSUPERVISED UNDERWATER IMAGE RESTORATION AND NEURAL RADIANCE FIELDS

Vinayak Gupta[1]*, Manoj S[1]*, Mukund Varma T[1]*, & Kaushik Mitra[1]
[1]Indian Institute of Technology Madras
{vinayakguptapokal, mukundvarmat, manoj.s.2908}@gmail.com,
kmitra@ee.iitm.ac.in

ABSTRACT

Underwater images suffer from colour shifts, low contrast, and haziness due to light scattering and refraction. Restoration of these images has received significant attention, particularly in a multi-view setup. In this work, we present (**U2NeRF**), a transformer-based architecture that learns to render and restore novel views simultaneously while conditioned on multi-view geometry. We attempt to implicitly bake the restoring capabilities onto the NeRF pipeline and disentangle the predicted colors into several components, which are then combined to reconstruct the underwater image in a self-supervised manner. In addition, we release an Underwater View Synthesis (**UVS**) dataset consisting of 8 real underwater scenes. Our experiments demonstrate that when optimized on a single scene, U2NeRF outperforms several baselines and showcases improved rendering and restoration capabilities. (Link to Source code)

1 INTRODUCTION

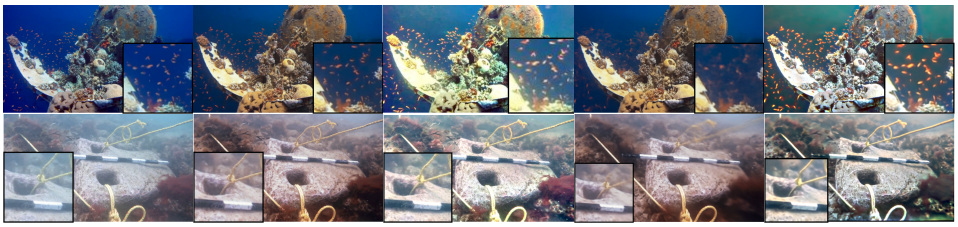
Neural Radiance Fields (Mildenhall et al., 2020) and its follow-up works (Barron et al., 2021; Niemeyer et al., 2022; Chen et al., 2022) have achieved remarkable success in novel view synthesis, by generating high-quality photo-realistic scenes. However, rendering and restoring target views still pose a challenge in complex scenarios such as underwater scenes. In this paper, we propose a method that combines NeRF's rendering capabilities with the restoration abilities of Chai et al. (2022) in a fully self-supervised manner. This is particularly significant as obtaining accurate ground truth labels for underwater images can be difficult, especially in a multi-view scenario. Chai et al. (2022) presents a physics-based self-supervised method that disentangles the prediction into four separate light transmission maps, which are then combined to reconstruct the original image.

Most NeRF methods operate at a pixel level, limiting its color restoration capacity. We demonstrate that by predicting image patches (rather than pixels), we provide sufficient spatial context for image restoration. Additionally, we adapt the recently proposed Generalizable NeRF Transformer (GNT) (T et al., 2022) into our work, utilizing its multi-view information aggregation to further enhance underwater restoration.

2 METHOD

NeRF represents the 3D scene as a radiance field $\mathcal{F} : (x, \theta) \mapsto (c, \sigma)$, where each spatial coordinate $x \in \mathbb{R}^3$ together with the viewing direction $\theta \in [-\pi, \pi]^2$ is mapped to a color $c \in \mathbb{R}^3$ plus density $\sigma \in \mathbb{R}_+$ tuple. However, a single pixel does not provide sufficient context for automatic restoration. In our work, we first adapt GNT to render an image patch of size p. The final ray feature obtained from the ray transformer block is passed on to a sequence of convolution and upsampling layers. Motivated by Chai et al. (2022), we disentangle the underwater image into several components - scene radiance (J), global background light (A) and degradation components - direct and back scatter transmission maps (T_D, T_B) that account for attenuation and light reflection respectively.

*Equal contribution.



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

Figure 1: Qualitative results for single-scene rendering on debris scene from easy split and on scene 2 from hard split of the UVS Dataset

These individual components can be combined to reconstruct the original image I at pixel i as:

$$I(i) = J(i)T_D(i) + (1 - T_B(i))A \tag{1}$$

This allows our network to be trained in a fully self-supervised manner, without the need for ground truth images. Additionally, the individual maps are regularized with appropriate image-level losses, which was made possible due to patch-based predictions by our model. To predict J, T_D, and T_B, we initialize separate output heads to project the final ray feature to the desired patch size. Since A is independent of the input image content, we pass the nearest source image from the target view direction onto a Variational AutoEncoder (VAE) to estimate global background light. Please refer appendix for a detail description given on the model architecture in Figure 2.

In addition to the photometric loss ($\mathcal{L}_{\text{rec}}$), we (1) minimize the difference between encoded feature z and latent code sampled from Gaussian $\hat{z}$ in the vae ($\mathcal{L}_{\text{kl}}$), (2) minimize the difference between the saturation and brightness of the predicted scene radiance to reduce haze ($\mathcal{L}_{\text{con}}$), (3) minimize the potential color deviations in the scene radiance ($\mathcal{L}_{\text{col}}$), (4) ensure constant back-scatter coefficients ($\mathcal{L}_{\text{trans}}$) across channels, and (5) enforce constant global background light by minimizing variance within each local neighbourhood ($\mathcal{L}_{\text{glob}}$) as proposed in the original paper (Chai et al., 2022). Together, the network is trained to optimize the loss:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{rec}, + \lambda_2 \mathcal{L}_{con} + \lambda_3 \mathcal{L}_{col} + \lambda_4 \mathcal{L}_{kl} + \lambda_5 \mathcal{L}_{trans} + \lambda_6 \mathcal{L}_{glob}$$

where λ indicates the weight for each loss term.

3 Experiments

We compare our results with restoration methods like UIESS (Chen & Pei, 2022) and UPIFM (Chai et al., 2022). In the case of more complex scenes present in the easy and hard data splits, we can clearly see the superiority of U2NeRF both in terms of rendering, and color restoration quality. More interestingly, we find that U2NeRF even outperforms no rendering baselines, that is, those algorithms that assume direct access to the target view and perform only restoration. Although our method extends upon UPIFM, we still manage to outperform the 'only restoration' baseline with sufficient margin. This signifies the relevance of multi-view geometry to automatically restore a target view. We show qualitative results in Fig. 1, and can clearly see that U2NeRF renders and restores images with greater visual quality when compared to other methods. In the case of debris, U2NeRF successfully recovers the fishes and enhances its visibility to improve restoration quality, while in the case of scene 2 from hard split, U2NeRF is able to render complex, moving structures like ropes while still maintaining higher detail along the surface of the rock. Please check the appendix for more results on the other scenes of the UVS Dataset given in Figure 3.

4 Conclusion

We extend radiance fields to simultaneously render and restore novel views, in the context of underwater images. By augmenting existing radiance fields with spatial awareness, and when combined with a physics-informed underwater image formation model, our model can successfully restore underwater images in a multi-view setup. These results demonstrate that transformers can be successfully used to model the underlying physics in 3D vision.

URM STATEMENT

We acknowledge that all authors of this work meet the URM criteria of ICLR 2024 Tiny Papers Track

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.

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.