

Article

Neural Radiation Fields in a Tidal Flat Environment

Huilin Ge ¹, Zhiyu Zhu ^{1,*}, Haiyang Qiu ² and Youwen Zhang ¹¹ Ocean College, Jiangsu University of Science and Technology, Zhenjiang 212100, China² School of Naval Architecture and Ocean Engineering, Guangzhou Maritime University, Guangzhou 510725, China

* Correspondence: zhuzy@just.edu.cn

Abstract: Tidal flats are critical ecosystems, playing a vital role in biodiversity conservation and ecological balance. Collecting tidal flat environmental information using unmanned aerial vehicles (UAVs) and subsequently utilizing 3D reconstruction techniques for their detection and protection holds significance in providing comprehensive and detailed tidal flat information, including terrain, slope, and other parameters. It also enables scientific decision-making for the preservation of tidal flat ecosystems and the monitoring of factors such as rising sea levels. Moreover, the latest advancements in neural radiance fields (Nerf) have provided valuable insights and novel perspectives for our work. We face the following challenges: (1) the performance of a single network is limited due to the vast area to cover; (2) regions far from the camera center may exhibit suboptimal rendering results; and (3) changes in lighting conditions present challenges for the achievement of precise reconstruction. To tackle these challenges, we partitioned the tidal flat scene into distinct submodules, carefully preserving overlapping regions between each submodule for collaborative optimization. The luminance of each image is quantified by the appearance embedding vector produced by every captured image. Subsequently, this corresponding vector serves as an input to the model, enhancing its performance across varying lighting conditions. We also introduce an ellipsoidal sphere transformation that brings distant image elements into the sphere's interior, enhancing the algorithm's capacity to represent remote image information. Our algorithm is validated using tidal plane images collected from UAVs and compared with traditional Nerf based on two metrics: peak signal-to-noise ratio (PSNR) and learned perceptual image patch similarity (LPIPS). Our method enhances the PSNR value by 2.28 and reduces the LPIPS value by 0.11. The results further demonstrate that our approach significantly enhances Nerf's performance in tidal flat environments. Utilizing Nerf for the 3D reconstruction of tidal flats, we bypass the need for explicit representation and geometric priors. This innovative approach yields superior novel view synthesis and enhances geometric perception, resulting in high-quality reconstructions. Our method not only provides valuable data but also offers profound insights for environmental monitoring and management.

Keywords: tidal flats; neural radiation fields; 3D reconstruction; image processing

Citation: Ge, H.; Zhu, Z.; Qiu, H.; Zhang, Y. Neural Radiation Fields in a Tidal Flat Environment. *Appl. Sci.* **2023**, *13*, 10848. <https://doi.org/10.3390/app131910848>

Academic Editors: Kelvin K.L. Wong, Andrew W.H. Ip, Dhanjoo N. Ghista and Wenjun (Chris) Zhang

Received: 3 August 2023

Revised: 22 September 2023

Accepted: 23 September 2023

Published: 29 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tidal flats [1] are coastal areas located at the interface between land and sea, subject to tidal influences. They form through the accumulation of fine cohesive particles, resulting in bi-directional sedimentation zones where sediment is carried by both land and marine forces. Tidal flats typically comprise various geomorphic units such as flats, vegetation belts, and tidal channels. Within the natural ecosystem, tidal flats hold immense ecological value, playing a pivotal role in preserving biodiversity, mitigating the impact of extreme storm surges, and serving as indicators of sea-level rise. However, the ecological systems of tidal flats are increasingly affected and threatened by global climate change and abnormal fluctuations in sea levels. As a result, monitoring and understanding the developmental trends of tidal flats have become crucially important. This endeavor is essential not only to

ensure the sustainable development of tidal flat resources but also to meet the demands of implementing environmental conservation policies.

Currently, the employment of unmanned systems [2,3] for the acquisition of environmental data in natural settings is a highly promising trend. Unmanned systems offer several advantages: (1) they possess the capability to encompass vast expanses of tidal flat areas; (2) they facilitate considerable time and manpower savings; (3) they enable non-contact data collection, thereby minimizing environmental disruption and degradation; and (4) they are adept at navigating complex terrains and accessing remote locations. Continuous advancements and research in electric motor technology have played a pivotal role in enhancing the performance, efficiency, and sustainability of unmanned systems [4,5]. Presently, numerous endeavors involve the utilization of unmanned aerial vehicles to collect remote sensing imagery of natural environments [6,7], subsequently facilitating the evaluation of ecological well-being. However, relative to merely employing 2D perspectives to detect changes in tidal flat environments, the utilization of image-based 3D reconstruction enables the acquisition of more intricate and comprehensive tidal flat information. Through 3D reconstruction of tidal flats, researchers can obtain detailed and holistic terrain models, encompassing elevation, slope, and topographical characteristics. These models facilitate visualization and simulation of the tidal flats, providing researchers and decision-makers with a more intuitive understanding of the morphological dynamics and enabling more scientifically informed decision-making. Additionally, 3D reconstruction [8,9] empowers more profound quantitative analysis, such as the measurement of tidal flat volume, surface area, slope, and other parameters, which hold immense value for environmental monitoring and management.

The synthesis of novel view images from a collection of 2D pictures has been a long-standing problem in computer vision. The entire process can be described as follows: feature points are detected from the 2D images, and descriptors are added to these feature points [10,11] that serve as unique signatures for subsequent image matching and correlation. Feature point matching is performed among multiple sets of images, establishing correspondences between points in different images. This process allows for the estimation of each camera's position, orientation, and 3D point triangulation within the scene. Ultimately, camera poses and 3D points undergo a combined optimization process known as bundle adjustment, which aims to minimize the differences between projected points and their corresponding image points. Recent works have seen many classical approaches based on structure-from-motion (SfM) [12] or image feature-based 3D reconstruction [13]. However, neural rendering techniques have made significant breakthroughs in this field, offering new insights for the 3D reconstruction of tidal flat environments. Neural radiance fields [14] (Nerf) employ neural networks to represent the radiance field and density of the reconstructed scene, then utilize volume rendering to reconstruct the target scene. Leveraging deep learning technology, Nerf infers the 3D structure and appearance of the scene from a set of observed images, thereby synthesizing new views from previously unobserved viewpoints. This approach has demonstrated remarkable results in many challenging scenarios, significantly enhancing the quality of 3D reconstruction.

While Nerf-based techniques have demonstrated exceptional proficiency in representing intricate geometries and smoothly varying appearances concerning viewpoints, Nerf assumes a constant density and radiance across the world, one which remains valid only under static conditions encompassing geometry, material, and lighting. Consequently, when applying Nerf to the 3D reconstruction of tidal environments, we encountered pronounced artifacts, excessive smoothing, and other pseudo-phenomena in terrain units highly sensitive to lighting variations. Notably, water surfaces and smooth rocks exhibited intense specular highlights, while undulating tidal channels resulted in varying shadow and highlight regions across distinct lighting conditions. In recent endeavors to enhance image rendering quality, numerous studies have pursued the decomposition of appearance into scene illumination and materials for re-illumination [15–20]. Some approaches assume fixed lighting conditions [15,18], or fixed reconstructed scene materials [19]. Additionally,

generative image-related techniques have been introduced in the realm of 3D reconstruction, including the utilization of learned latent appearance embeddings as conditioning for neural re-rendering networks [21]. Inspired by these advancements, we address Nerf's limitations in handling lighting variations by introducing appearance embedding vectors for each image, leading to a remarkable enhancement in Nerf's performance across regions that manifests in noticeable lighting changes within tidal environments.

In the pursuit of enhancing Nerf-based 3D reconstructions, various approaches have garnered attention, including scene segmentation into multiple sub-modules. DeRF [22] achieves this by employing spatial Voronoi partitioning to decompose the scene into multiple cells, each independently rendered using smaller MLPs. KiloNerf [23], on the other hand, reconstructs target scenes using thousands of networks, or even more. Similarly, for the 3D reconstruction of tidal flat scenes, we adopt a network structure, as shown in Figure 1, for the task. We retained the repetitive parts between each submodule and optimized the appearance embedding vectors of the submodules during rendering to achieve an overall perspective optimization of the model. Drawing inspiration from Nerf++ [24], we employ an inverted sphere parameterization for image information captured at greater distances from the camera. However, unlike Nerf++, we tailor the shape of the sphere specifically for the tidal flats environment, optimizing it to achieve tighter boundaries around regions of interest and minimizing unnecessary computations.

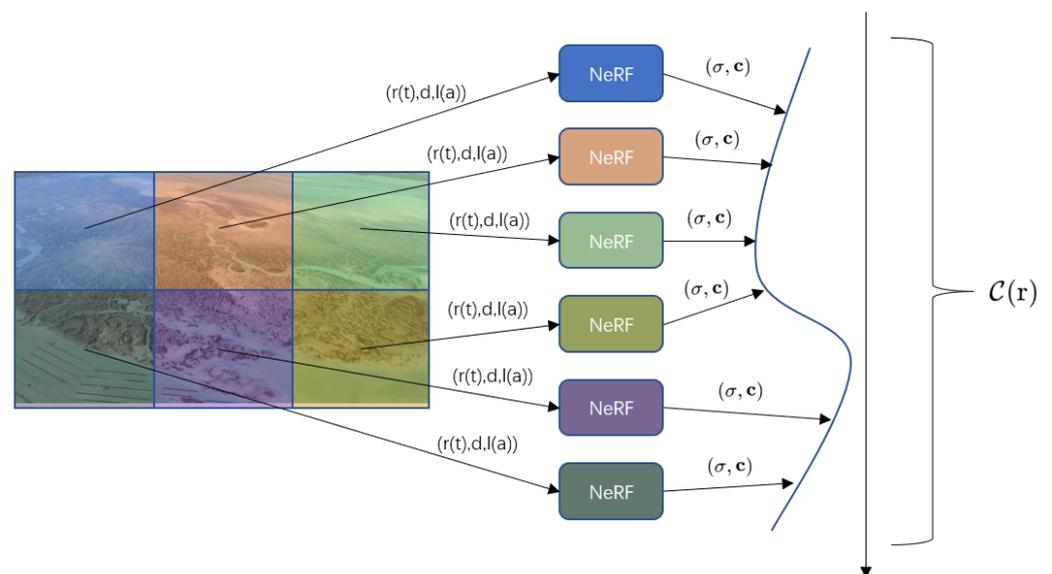


Figure 1. Algorithmic structure.

2. Approach

Firstly, in Section 2.1, we will provide a brief overview of Nerf. Subsequently, in Section 2.2, we will elaborate on the techniques employed to handle the foreground and background in tidal flats. The methods utilized to address variations in tidal flat lighting will be discussed in Section 2.3. Finally, in Section 2.4, we will describe our approach to partitioning the tidal flat environment into sub-modules and the corresponding processing techniques.

2.1. Background

Neural radiance fields represent a neural network model employed for image synthesis and scene reconstruction purposes. This sophisticated neural architecture utilizes multi-layer perceptrons (MLPs) to model the density and color of the scene, thereby encoding the relationship between the 3D coordinates of the scene and the viewing perspective in a functional form. By inputting the scene's positions (x, y, z) and the viewing directions into the MLP network, we obtain functions for volumetric density $\sigma(x)$ and radiance $c(x, d)$

with respect to the positions and viewing directions (θ, ϕ) . Ultimately, the final image is generated through the process of volume rendering. During the rendering stage, stratified sampling is employed to sample along the camera rays $r(t) = o + td$ for each pixel:

$$\hat{C}(r) = \sum_{i=1}^N T_i \cdot (1 - \exp(-\sigma_i \cdot \delta_i)) \cdot c_i \tag{1}$$

To obtain the pixel color, the interval between the near and far rendering boundaries, denoted as t_n and t_f , respectively, is subdivided into N segments. Subsequently, summation is performed for each of these subintervals.

The parameter T_i represents the probability of not being halted by particle interactions from both the near and far rendering boundaries, and it also signifies the radiance from position t to the camera:

$$T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right) \tag{2}$$

Nerf employs stochastic gradient descent to optimize the volumetric density σ and radiance c by reducing the discrepancy between the real image I_i and the rendered predicted image $\hat{I}_i(\sigma, c)$:

$$\min_{\sigma, c} \frac{1}{n} \sum_{i=1}^n \| I_i - \hat{I}_i(\sigma, c) \|_2^2. \tag{3}$$

The disadvantage of neural networks when learning high-frequency information lies in their subpar performance [25]. To counterbalance the biases in synthesizing images, a compensatory measure involves utilizing the γ function to map the positional and viewing information of the input network into a higher-dimensional space:

$$\gamma^k : p \rightarrow (\sin(2^0 p), \cos(2^0 p), \dots, \sin(2^k p), \cos(2^k p)) \tag{4}$$

The network architecture of Nerf is illustrated in the figure, where the input consists of spatial coordinates (x, y, z) and camera view directions (θ, ϕ) . The multi-layer perceptron (MLP) denoted by the green box in the upper part of the image consists of 8 fully connected layers with 256-dimensional features. It connects the output of the upper MLP with the camera view directions (θ, ϕ) . The MLP in the lower part of the image consists of 4 fully connected layers with 256-dimensional features and outputs the color C . The blue box in Figure 2 represents the principle of classic volume rendering, which renders the color of any ray passing through the scene, as indicated by Formula (1). The red box in Figure 2 represents the use of high-frequency functions to map the input to a higher-dimensional space, enabling a better fit for data containing high-frequency variations, as shown in Formula (4).

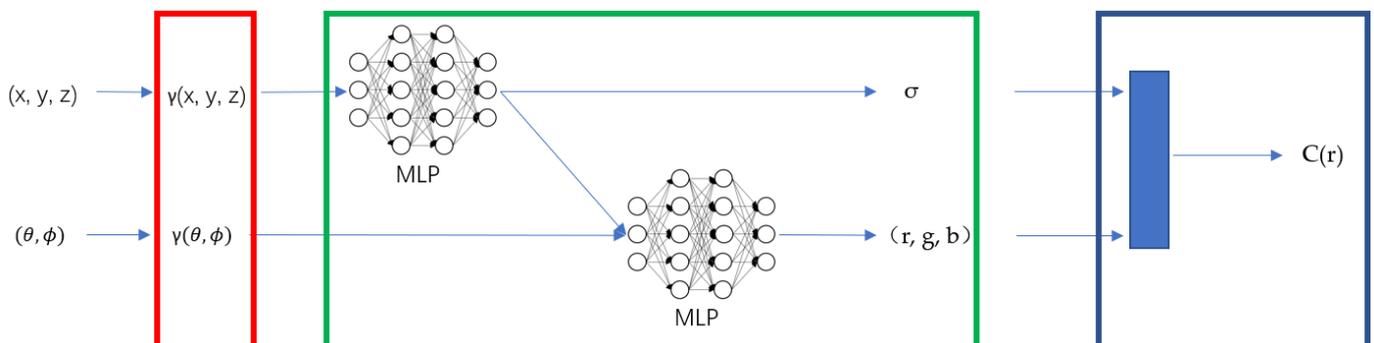


Figure 2. Nerf network structure.

2.2. Foreground and Background Processing

Nerf assumes that the target scene is placed within a fixed-size bounding box before performing 3D reconstruction. However, this assumption is unsuitable for scenes such as

a tidal flat environment. The complexities of such natural environments require a more flexible approach. Before conducting the 3D reconstruction, it is crucial to carefully consider the processing of distant image scenes. If the distant image scenes are not appropriately processed, it could lead to errors in the image matching algorithm. Consequently, incorrect features may be matched together during the generation of the 3D model, resulting in an inaccurate overall reconstruction that fails to faithfully represent the characteristics of the entire area.

In the context of boundaryless scenes, such as tidal flat environments, it becomes essential to process the foreground and background of the scene. Inspired by Nerf++ [24], As shown in Figure 3, we partition the scene space into two parts: an inner unit sphere containing the foreground and all of the cameras, and an outer volume represented by an inverted sphere covering the complement of the inner volume, which includes the remaining parts of the scene. As in formula 5, 3D points (x, y, z) outside the sphere are transformed into quadruples $(x', y', z', 1/r)$, where the direction points outward to the corresponding point outside the sphere.

$$r = \sqrt{x^2 + y^2 + z^2} > 1 \tag{5}$$

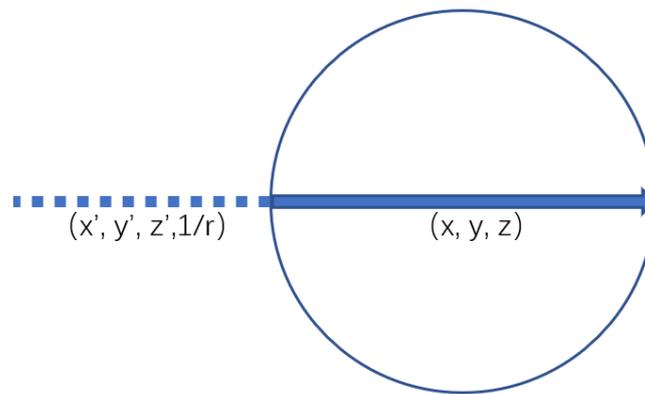


Figure 3. The differences between parameterization of scenes inside and outside a sphere lie in the range and approach used.

The advantage of this approach is that, for the reparameterization of quaternion results for (x, y, z) at infinite distances, the values are all within the range of $[-1, 1]$. This not only conforms to the objective fact of lower resolution for distant objects but also enhances the stability of the data.

Diverging from Nerf++, our approach leverages drone-captured datasets for tidal flat reconstruction. To minimize unnecessary computations, we employ ellipsoids as a more compact representation to enclose the tidal flat environment, replacing the unit sphere. The specifics of this approach are depicted in Figure 4. Nerf++ (left) employs sampling within a unit sphere, centered and enclosing all camera poses, to render its foreground components, while adopting distinct techniques to effectively render the background within the outer volume. In our proposed approach (right), we utilize a similar background parameterization but introduce ellipsoidal modeling for foreground elements, achieving a more tightly bounded region of interest. This adaptation allows for a more efficient and accurate representation of the tidal flat scene, ensuring streamlined rendering processes while preserving the fidelity of reconstruction within the regions of interest. This adaptation allows us to efficiently encapsulate the unique characteristics of the tidal flat scenes, optimizing the rendering process while maintaining accuracy in the reconstruction.

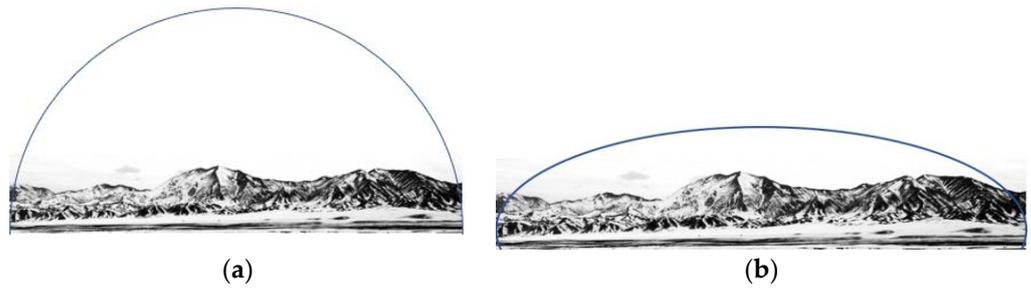


Figure 4. Ray bounds. (a) The unit sphere of Nerf++. (b) The unit sphere of our method.

2.3. Lighting Conditions

Generative image modeling has been a longstanding research focus in computer vision. The application of generative image modeling’s 3D reconstruction techniques has had a profound impact on the representation and replication of tidal flats environments. Tidal flats scenes often face rapid changes in lighting conditions, such as sunrise, sunset, and cloud cover. Leveraging generative image modeling enables us to effectively address the challenges posed by these lighting variations, thereby reducing artifacts, excessive smoothing, and pseudo-phenomena during the reconstruction process. As a result, the accuracy and realism of the reconstructed scenes are significantly enhanced.

In order to address this challenge, we drew inspiration from existing works such as ‘Nerf in the Wild’ [26] and adopted the approach of generative latent optimization (GLO) [27]. For each image x_i in the available image collection $\{x_1, \dots, x_N\}$, we initialized a random m -dimensional vector $L_i = \{l_1, l_2, \dots, l_N\}$ as shown in Figure 5. Recognizing the impact of color variations on the final 3D reconstruction of images, we introduced weather conditions as a conditioning factor, represented by a vector that describes the lighting situation of the corresponding image. This conditioning vector was incorporated into our network to influence the final color generation process.

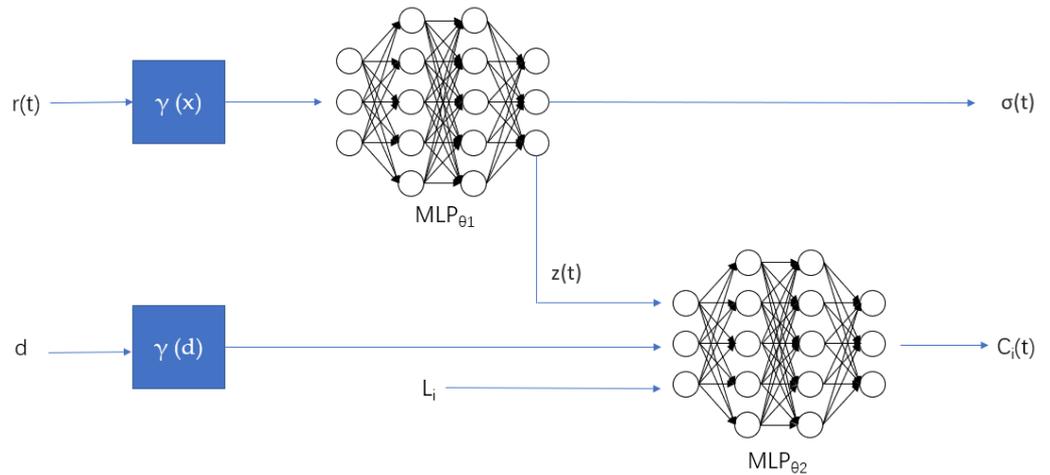


Figure 5. Network structure.

In Equation (6), $\hat{C}_i(\mathbf{r})$ represents the final synthesized color of pixel points for the picture image, and \mathbf{r} denotes the information about the light ray. \mathcal{R} denotes the process of volume rendering, while \mathbf{r} corresponds to the spatial coordinates (x, y, z) and σ denotes the observation direction. These variables are consistent with the original Nerf. The key difference lies in the color inference process.

$$\hat{C}_i(\mathbf{r}) = \mathcal{R}(\mathbf{r}, c_i, \sigma) \tag{6}$$

As shown in Equation (7), where L_i is introduced. $\mathbf{z}(t)$ is the output of MLP_{θ_1} , and $\gamma_{\mathbf{d}}(\mathbf{d})$ represents the angle of observation encoded using positional information, essentially incorporating additional image information.

$$\mathbf{c}_i(t) = \text{MLP}_{\theta_2}(\mathbf{z}(t), \gamma_{\mathbf{d}}(\mathbf{d}), L_i) \quad (7)$$

As shown in Equation (8), MLP_{θ_1} follows the same structure as Nerf, with the input being high-dimensional positional encoding of (x, y, z) , and the output $\mathbf{z}(t)$ representing volume density. This design has the advantage that L_i only affects the color output and does not influence the generated volume density, thus preventing L_i from affecting the geometric shape of the generated 3D model and reducing model errors.

$$[\sigma(t), \mathbf{z}(t)] = \text{MLP}_{\theta_1}(\gamma_{\mathbf{x}}(\mathbf{r}(t))) \quad (8)$$

The main difference between our method and ‘Nerf in the Wild’ lies in the approach. ‘Nerf in the Wild’ divides the scene into dynamic and static parts, employing an additional MLP to identify dynamic components and then combining them with static components to produce the final result. However, for natural landscapes such as tidal flats, having a separate MLP is not meaningful as it has limited impact on the final output and complicates model training, increasing the demand for computational resources.

2.4. Scene Segmentation

The 3D reconstruction of tidal flats using a single Nerf network poses significant challenges due to limitations in the expressive capacity of individual MLP networks. The sheer size and complexity of tidal flats scenes demand immense time and computational resources for a comprehensive reconstruction. Furthermore, the memory requirements escalate as Nerf networks need to store information for every pixel in the scene, particularly daunting for large-scale environments.

Inspired by Mega-nerf [28], we address this challenge by partitioning the scene into manageable blocks and training separate Nerf networks for each block. Mega-nerf emphasizes the parallel training of data, wherein each sub-module operates independently. However, tidal flats environments present a unique issue during data acquisition, as weather conditions and other factors can lead to significant variations in images captured from the same viewpoints, resulting in brightness and contrast differences. In Figure 6, we introduce an approach to facilitate communication among different sub-modules by retaining overlapping image regions between them.

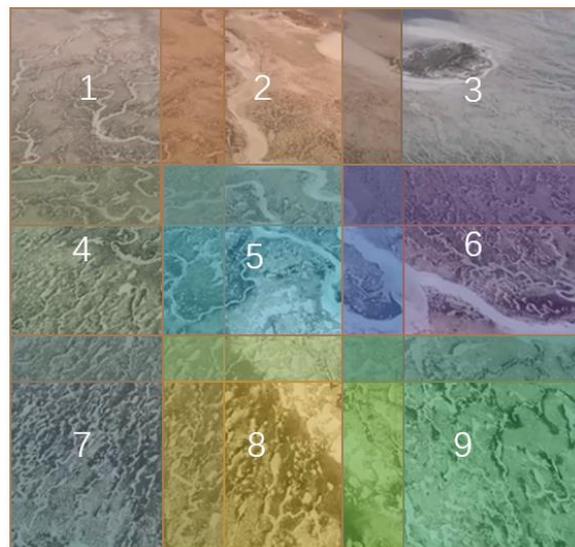


Figure 6. Scene segmentation schematic.

Since multiple sub-modules often cover a single scene, during the rendering step, we identify and filter out sub-modules with poorly collected image data. To achieve this, we train an additional MLP to output the corresponding transmittance value (T_i) for each scene, as shown in Figure 7. The transmittance values are bounded within the range $[-1, 1]$. When sampling rays intersect the objects for the first time, the corresponding T_i value is close to 1, whereas it becomes close to 0 after passing through the object's interior or surface. For cases with multiple views of the same object, we calculate the average T_i value. Sub-modules with T_i values below a predefined threshold are discarded.

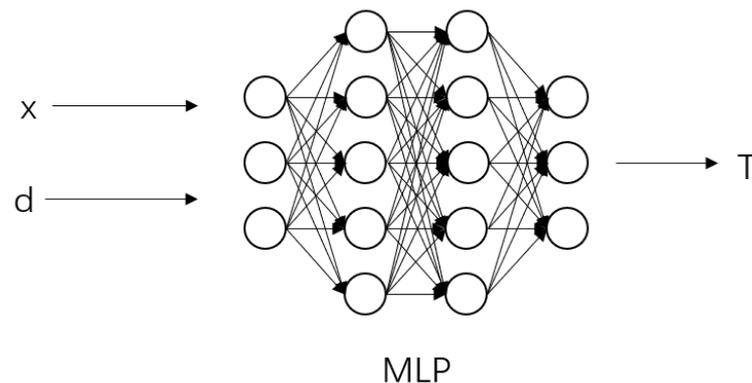


Figure 7. Transmittance Value Estimation Process.

We perform 2D interpolation on the rendered results to achieve smooth transitions between different sub-modules. By calculating the spatial distance between pixel points and their initial poses, we determine the corresponding weights for each point. The weight coefficients are inversely proportional to the distance from the initial pose, ensuring a smooth blending effect across sub-modules.

Addressing the varying lighting conditions among different sub-modules, we employ Formula (6) to determine the appearance embedding vector L_i . We select a reference sub-module, determine the appearance embedding vector L_1 for this module, and identify the 3D points visible in both the reference sub-module and that requiring lighting adjustments (L_2). By fixing the MLPs of the two sub-modules and rendering the RGB for this point, we efficiently calculate and update L_2 , achieving convergence in approximately 100 iterations.

3. Experiments

3.1. Datasets

We employed internet-sourced, drone-captured images of the Smithton–Wool north tidal flats, acquired on 31 December 2018. These images were carefully categorized into distinct groups, including “Tidal Trees” formed during low tide, river mouths, ground textures, vegetation, and deep-water areas.

3.2. Training

We have previously classified aerial photographs into different scenes. Similar to Nerf, we used COLMAP [29], which incorporates two camera parameters for radial distortion and tangential distortion, to estimate the camera poses. We employed the PyTorch framework and followed the same steps to load the data. Each batch consisted of 2048 ray samples, and we utilized the Adam optimizer [30] with an initial learning rate of 4×10^{-4} , which gradually decayed to 4×10^{-5} .

3.3. Evaluation

We employed two different approaches to obtain the results and compared the rendered outcomes with real-world ground truth. Additionally, we generated quantitative reports based on PSNR (the bigger, the better) and LPIPS [31] (the smaller, the better) metrics.

We present our results in Table 1, where our approach demonstrates better overall performance in terms of PSNR and LPIPS compared with Nerf, PSNR improved by an average of 2.28 and LPIPS decreased by an average of 0.11, indicating smaller differences between our reconstructed results and the original images. However, it should be noted that our LPIPS performance is not as good as Nerf in deep water regions.

Table 1. Comparison of our method with nerf in a tidal flats environment.

	PSNR \uparrow		LPIPS \downarrow	
	nerf	ours	nerf	ours
"Tidal Trees"	24.24	28.99	0.4985	0.3013
River mouths	31.31	29.28	0.3591	0.2889
Ground textures	24.91	29.16	0.5078	0.3069
Vegetation	25.06	28.63	0.5243	0.4065
Deep-water areas	27.76	28.62	0.3741	0.4176

Figure 8 illustrates the comparison between our method and the Nerf approach in rendering results and ground truth in a tidal flats environment. Our method exhibits stronger expressiveness in capturing important image details such as ground texture, vegetation, and estuary, which are crucial for the evaluation of a tidal flats environment.

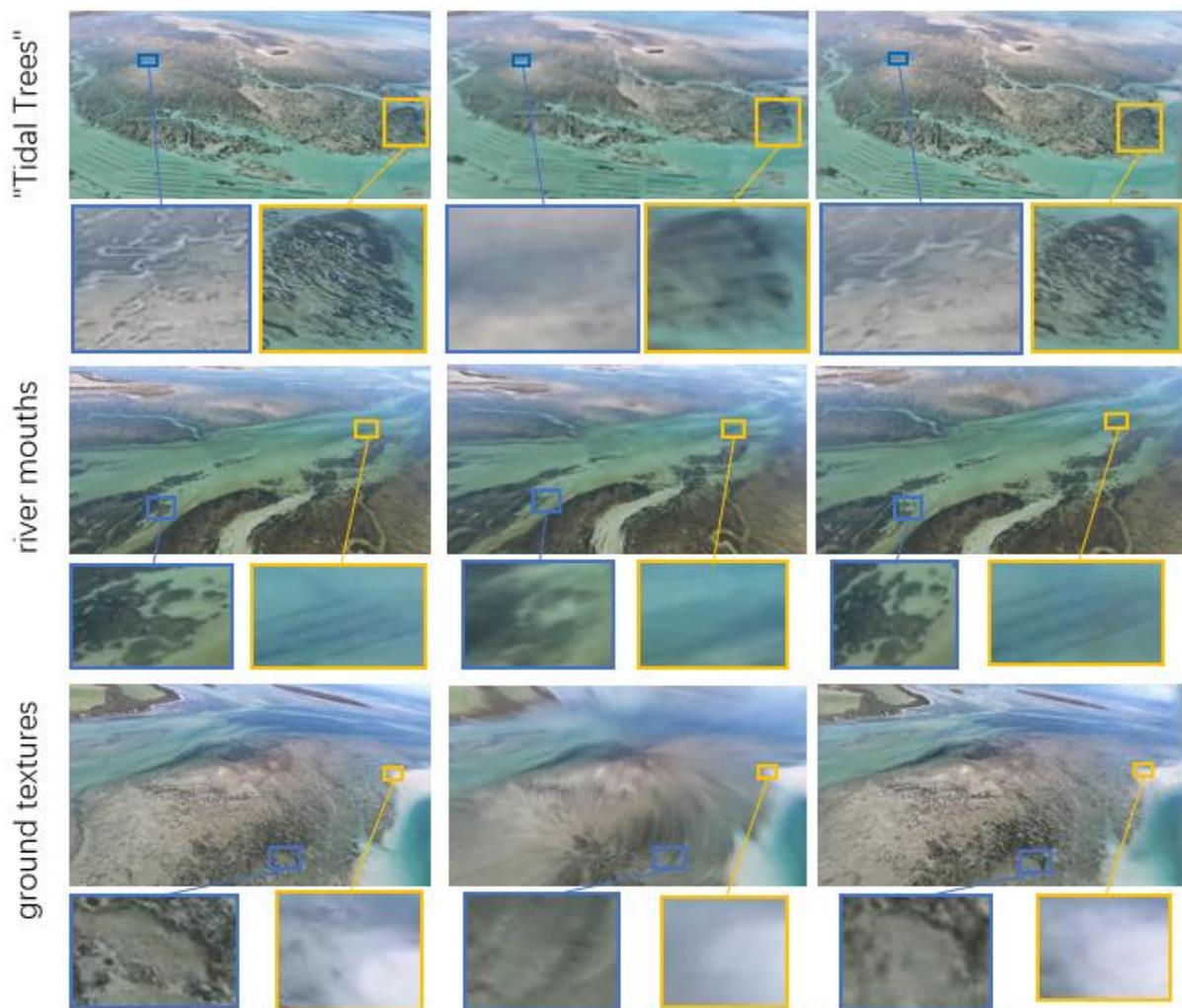


Figure 8. Cont.

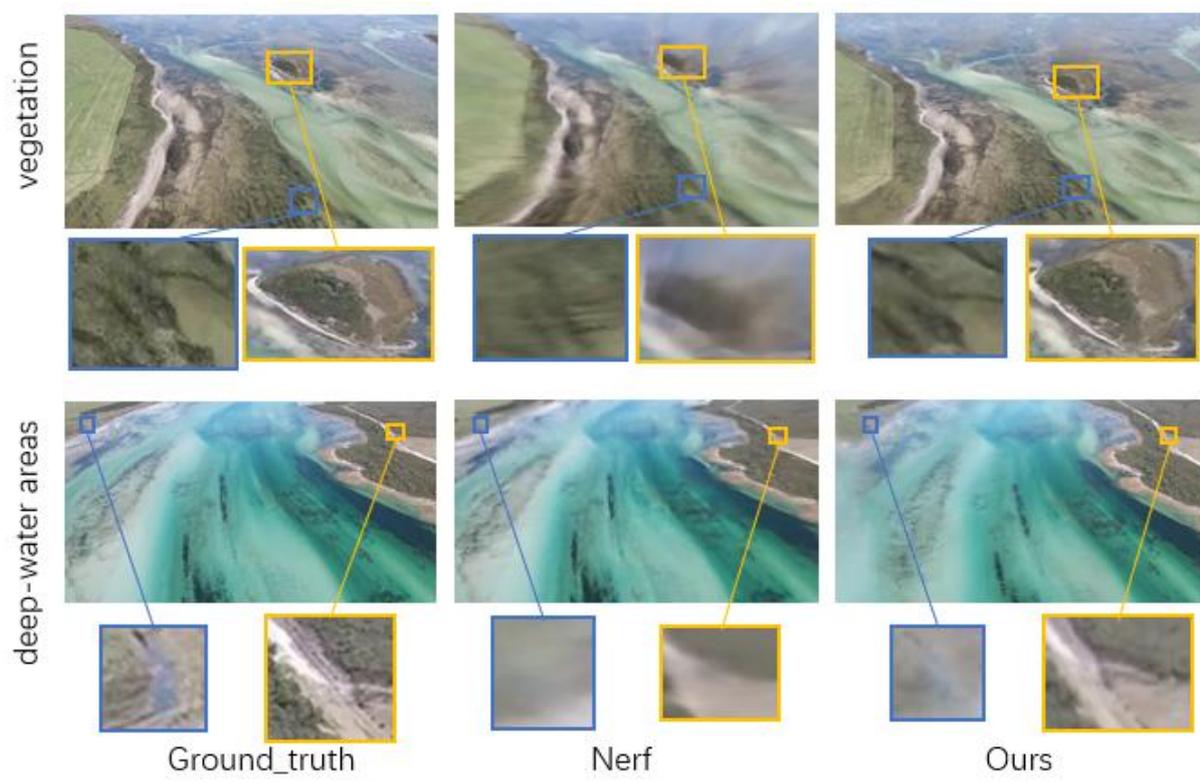


Figure 8. Comparison of results.

3.4. Ablation Experiment

While our method has achieved satisfactory results in the 3D reconstruction of tidal flats, we conducted further investigations to explore the specific effects of our improvement. In addition to Nerf, we conducted two separate control experiments to investigate: (1) the impact of using an ellipsoidal sphere to transform distant image elements into the interior of the sphere, in relation to distant tidal flats image information; and (2) the introduction of an image embedding vector for each submodule, optimizing them under the same lighting conditions.

As shown in Figure 9, we solely employed method (1) and performed a detailed comparison of reconstruction results within a submodule. It is evident that our algorithm enhances the reconstruction of distant image information.

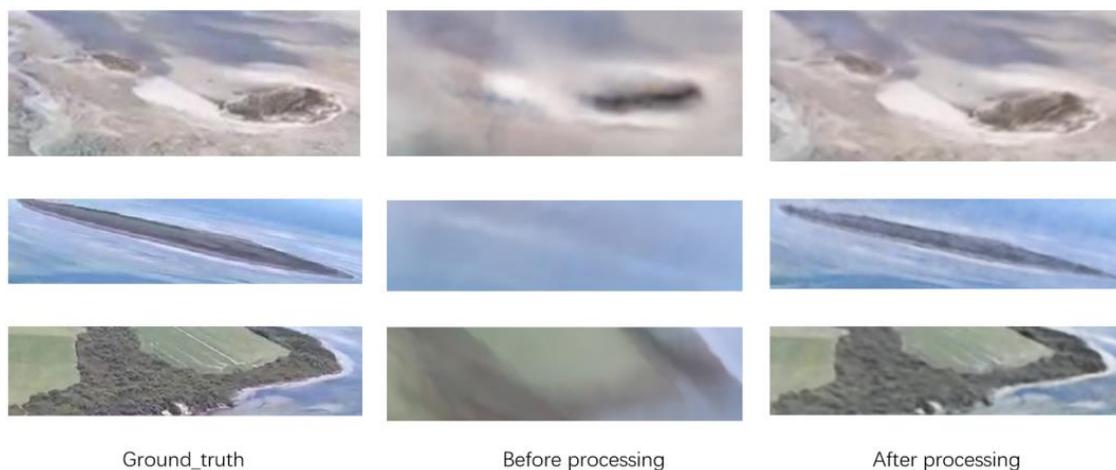


Figure 9. Comparison of remote image information of tidal flat.

Furthermore, in Figure 10, we selected scenes with evident variations in the collected tidal flats image data, including changes caused by water surface reflections and smooth geological environments. We validated method (2) in two submodules, demonstrating an improvement in the representation capability of tidal flats environments.

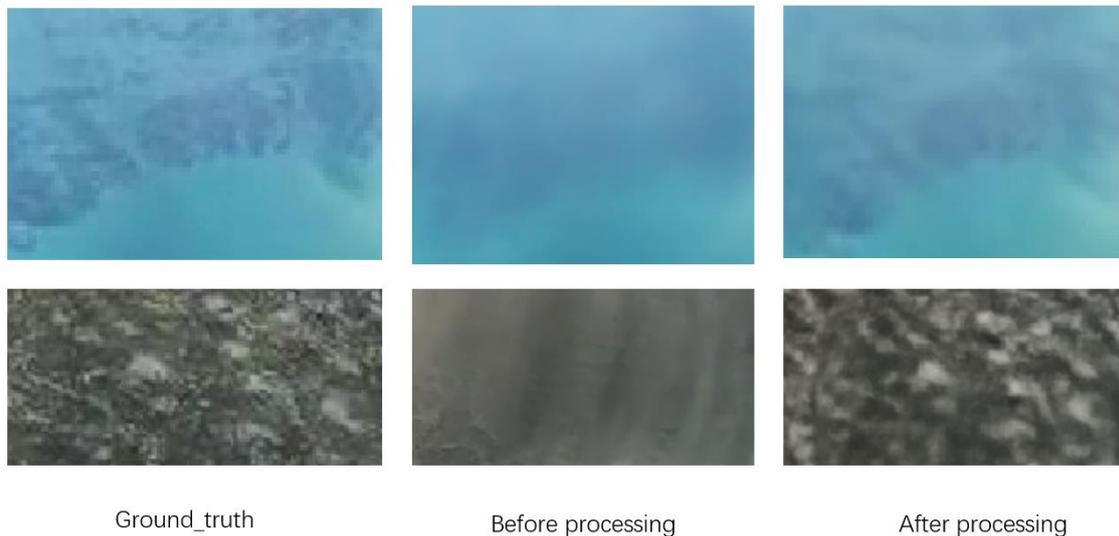


Figure 10. Comparison of the results of tidal flats under varying lighting conditions.

4. Discussion

In regard to the “ground textures” dataset presented in Figure 8, it is evident that there are excessively bright highlights on the rocks in the ground texture. We conducted further investigation into this phenomenon. Upon analyzing the image data of the tidal flat scene, we observed that the area in question constitutes a protruding section of the ground, surrounded by bright water bodies. We hypothesize that this result is influenced by the brightness of neighboring submodules during joint optimization, leading to an enhanced representation of brightness in this specific submodule. To mitigate this issue and enhance our reconstruction quality, we intend to incorporate a weight parameter in our future work. This parameter will help control the impact of neighboring submodules on our module, particularly in specific cases.

In Table 1, our method exhibits suboptimal performance in terms of LPIPS values within the “deep water” dataset. The primary challenge in capturing glossy surfaces lies in the generation of sporadic glossy artifacts that intermittently appear and vanish between rendered views, rather than smoothly traversing surfaces in a physically plausible manner. We hypothesize that factors such as illumination, reflection, perspective, and transparency within the deep-water region significantly influence image generation. These factors can lead to a more intricate and irregular radiation field in the deep-water region, potentially necessitating an increased number of sampling points and greater network capacity for accurate representation.

It is worth noting that methods employing remotely operated vehicles (ROVs) or autonomous underwater vehicles (AUVs) are better equipped to capture information in deep water areas [32]. However, adopting such methods inevitably escalates data acquisition costs and may compromise result reliability due to potential data quality issues. In future research, we intend to address this limitation by utilizing the reflection of the observation vector on the local normal vector as input, as opposed to using the observation vector itself. Alternatively, we may introduce a function that characterizes the outgoing radiation within our model for materials with varying degrees of roughness. This approach aims to distinguish between diffuse reflection and specular reflection in smooth materials, further enhancing the expressiveness of our model.

Overall, in order to facilitate the application of neural radiance fields for 3D reconstruction in tidal flat environments, we introduced an embedded vector for global optimization. This vector takes into account specific crucial lighting conditions unique to tidal flats. To incorporate information from distant regions of the tidal flats, we employed an inversion transformation that effectively transfers external scenery onto the inner surface of a sphere, thereby enhancing the algorithm's representational capabilities.

In the context of tidal flat 3D reconstruction, we employ two crucial parameters, PSNR and LPIPS, to validate the superiority of our model over the conventional Nerf model. When compared with the traditional Nerf model, our approach demonstrates an average increase of 2.28 in PSNR and a corresponding average decrease of 0.11 in LPIPS.

Our method optimizes the utilization of images acquired through drone-based surveys, resulting in a significantly enhanced capacity for capturing the intricacies of the tidal flat environment. This improvement empowers researchers to conduct more comprehensive assessments of tidal flat ecosystems. Furthermore, it offers invaluable support for a diverse range of scientific investigations, environmental preservation initiatives, and sustainable development objectives. This pioneering approach introduces a fresh perspective to the realm of 3D reconstruction in natural environments by leveraging neural radiance fields. Its particular efficacy in challenging tidal flat environments sets the stage for innovative advancements in the field.

Author Contributions: Conceptualization, H.G. and Z.Z.; methodology, H.G.; software, H.G.; validation, H.G., Z.Z. and H.Q.; formal analysis, H.G.; investigation, Z.Z.; resources, H.Q.; data curation, Y.Z.; writing—original draft preparation, H.G.; writing—review and editing, Y.Z.; visualization, Y.Z.; supervision, Z.Z.; project administration, Z.Z.; funding acquisition, H.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Zhenjiang key research and development plan—social development project (SH2022013). Additionally, this research was supported by the Jiangsu Province key research and development plan—Social development project (BE2022783).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Murray, N.J.; Phinn, S.R.; DeWitt, M.; Ferrari, R.; Johnston, R.; Lyons, M.B.; Clinton, N.; Thau, D.; Fuller, R.A. The global distribution and trajectory of tidal flats. *Nature* **2019**, *565*, 222–225. [\[CrossRef\]](#)
- Manfreda, S.; McCabe, M.F.; Miller, P.E.; Lucas, R.; Madrigal, V.P.; Mallinis, G.; Ben Dor, E.; Helman, D.; Estes, L.; Ciraolo, G.; et al. On the use of unmanned aerial systems for environmental monitoring. *Remote Sens.* **2018**, *10*, 641. [\[CrossRef\]](#)
- Hardin, P.J.; Jensen, R.R. Small-scale unmanned aerial vehicles in environmental remote sensing: Challenges and opportunities. *GIScience Remote Sens.* **2011**, *48*, 99–111. [\[CrossRef\]](#)
- Soudagar, M.E.M.; Mujtaba, M.A.; Safaei, M.R.; Afzal, A.; Raju, V.D. Effect of Sr@ZnO nanoparticles and Ricinus communis biodiesel-diesel fuel blends on modified CRDI diesel engine characteristics. *Energy* **2021**, *215*, 119094. [\[CrossRef\]](#)
- Soudagar, M.E.M.; Nik-Ghazali, N.-N.; Kalam, M.; Badruddin, I.A.; Banapurmath, N.; Bin Ali, M.A.; Kamangar, S.; Cho, H.M.; Akram, N. An investigation on the influence of aluminium oxide nano-additive and honge oil methyl ester on engine performance, combustion and emission characteristics. *Renew. Energy* **2020**, *146*, 2291–2307. [\[CrossRef\]](#)
- Tang, L.; Shao, G. Drone remote sensing for forestry research and practices. *J. For. Res.* **2015**, *26*, 791–797. [\[CrossRef\]](#)
- Syifa, M.; Park, S.J.; Lee, C.W. Detection of the pine wilt disease tree candidates for drone remote sensing using artificial intelligence techniques. *Engineering* **2020**, *6*, 919–926. [\[CrossRef\]](#)
- Zhao, S.; Kang, F.; Li, J.; Ma, C. Structural health monitoring and inspection of dams based on UAV photogrammetry with image 3D reconstruction. *Autom. Constr.* **2021**, *130*, 103832. [\[CrossRef\]](#)
- Eltner, A.; Schneider, D. Analysis of different methods for 3D reconstruction of natural surfaces from parallel-axes UAV images. *Photogramm. Rec.* **2015**, *30*, 279–299. [\[CrossRef\]](#)
- Aguilera, C.; Barrera, F.; Lumbreras, F.; Sappa, A.D.; Toledo, R. Multispectral image feature points. *Sensors* **2012**, *12*, 12661–12672. [\[CrossRef\]](#)
- Rodehorst, V.; Koschan, A. Comparison and evaluation of feature point detectors. In Proceedings of the 5th International Symposium Turkish-German Joint Geodetic Days, Berlin, Germany, 28–31 March 2006.
- Schonberger, J.L.; Frahm, J.M. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2016; pp. 4104–4113.

13. Pei, J.F.; Huang, Y.L.; Huo, W.B.; Zhang, Y.; Yang, J.Y.; Yeo, T.S. SAR automatic target recognition based on multiview deep learning framework. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2196–2210. [[CrossRef](#)]
14. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106. [[CrossRef](#)]
15. Bi, S.; Xu, Z.; Srinivasan, P.; Mildenhall, B.; Sunkavalli, K.; Hašan, M.; Hold-Geoffroy, Y.; Kriegman, D.; Ramamoorthi, R. Neural reflectance fields for appearance acquisition. *arXiv* **2020**, arXiv:2008.03824.
16. Boss, M.; Braun, R.; Jampani, V.; Barron, J.T.; Liu, C.; Lensch, H.P. NeRD: Neural reflectance decomposition from image collections. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021.
17. Boss, M.; Jampani, V.; Braun, R.; Barron, J.T.; Liu, C.; Lensch, H.P. Neural-PIL: Neural pre-integrated lighting for reflectance decomposition. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–14 December 2021.
18. Srinivasan, P.P.; Deng, B.; Zhang, X.; Tancik, M.; Mildenhall, B.; Barron, J.T. NeRV: Neural reflectance and visibility fields for relighting and view synthesis. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
19. Zhang, K.; Luan, F.; Wang, Q.; Bala, K.; Snavely, N. PhysSG: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
20. Zhang, X.; Srinivasan, P.P.; Deng, B.; Debevec, P.; Freeman, W.T.; Barron, J.T. NeR Factor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Trans. Graph. (SIGGRAPH Asia)* **2021**, *40*, 1–18.
21. Meshry, M.; Goldman, D.B.; Khamis, S.; Hoppe, H.; Pandey, R.; Snavely, N.; Martin-Brualla, R. Neural rerendering in the wild. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
22. Rebain, D.; Jiang, W.; Yazdani, S.; Li, K.; Yi, K.; Tagliasacchi, A. Derf: Decomposed radiance fields. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; IEEE Computer Society: Washington, DC, USA, 2021; pp. 14148–14156.
23. Reiser, C.; Peng, S.; Liao, Y.; Geiger, A. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 14335–14345.
24. Zhang, K.; Riegler, G.; Snavely, N.; Koltun, V. Nerf++: Analyzing and improving neural radiance fields. *arXiv* **2020**, arXiv:2010.07492.
25. Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.A.; Bengio, Y.; Courville, A.C. On the spectral bias of neural networks. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018.
26. Martin-Brualla, R.; Radwan, N.; Sajjadi, M.S.M.; Barron, J.T.; Dosovitskiy, A.; Duckworth, D. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 7210–7219.
27. Bojanowski, P.; Joulin, A.; Lopez-Paz, D.; Szlám, A. Optimizing the latent space of generative networks. *arXiv* **2017**, arXiv:1707.05776.
28. Turki, H.; Ramanan, D.; Satyanarayanan, M. Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12922–12931.
29. Fisher, A.; Cannizzaro, R.; Cochrane, M.; Nagahawatte, C.; Palmer, J.L. ColMap: A memory-efficient occupancy grid mapping framework. *Robot. Auton. Syst.* **2021**, *142*, 103755. [[CrossRef](#)]
30. Kingma, D.P.; Ba, J.; Bengio, Y.; LeCun, Y. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; p. 6.
31. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; p. 6.
32. Wong, K.K.L. *Cybernetical Intelligence: Engineering Cybernetics with Machine Intelligence*, 1st ed.; The Institute of Electrical and Electronics Engineers, Inc.: Piscataway, NJ, USA; John Wiley & Sons, Inc.: Chichester, UK, 2024; ISBN 9781394217489.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.