

Exploring the Potential of Refractive NeRFs for Photogrammetric Bathymetry - First Application to UAV-based Data from the Pielach River

Anatol Günthner¹, Markus Brezovsky², Frederik Schulte³, Lukas Winiwarter³, Gottfried Mandlbürger², Boris Jutzi¹

¹ Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Karlsruhe, Germany - (anatol.guenther, boris.jutzi)@kit.edu

² Department of Geodesy and Geoinformation, Vienna University of Technology, Vienna, Austria - (markus.brezovsky, gottfried.mandlbuerger)@tuwien.ac.at

³ Unit of Geometry and Surveying, Faculty of Engineering Sciences, University of Innsbruck, Innsbruck, Austria – (frederik.schulte, lukas.winiwarter)@uibk.ac.at

Keywords: Underwater Photogrammetry, Neural Radiance Fields, NeRFrac, Bathymetry, UAV Imaging, Refraction Correction.

Abstract

Accurate 3D reconstruction of underwater environments from above-water photos remains challenging due to refractive distortion at air-water interfaces. This contribution presents the first application of NeRFrac (Zhan et al., 2023) to UAV-based imagery captured in a real-world river area. NeRFrac is a refraction-aware Neural Radiance Field (NeRF) framework that explicitly models the change in direction of light at water surfaces according to Snell's Law. To adapt NeRFrac to complex outdoor scenes, we introduce a mask-based ray selection that selectively applies refractive modeling only to water-covered regions. We systematically evaluate different indices of refraction and compare global versus local training strategies. The results show that masking improves reconstruction quality in submerged areas, with a physically plausible index of refraction (IOR) of 1.333 yielding the best performance in terms of peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). While visual differences between masked and unmasked models remain minor, quantitative metrics confirm the effectiveness of the suggested refraction modeling.

1. Introduction

The original Neural Radiance Field (NeRF) (Mildenhall et al., 2020) is being adapted to many specific cases, such as representing scenes beneath refractive surfaces like glass or water. We apply NeRFrac (Zhan et al., 2023), to UAV data captured at the Pielach river in Lower Austria (Mandlbürger et al., 2025b). So far, NeRFrac has been applied to synthetic data and real data, captured in lab environments. To the best of our knowledge, this is the first application of NeRFrac to UAV-based imagery of a real-world river environment.

For the first time, we show that:

- Using binary water masks during training improves reconstruction quality in outdoor UAV datasets,
- A refinement of the refractive index from $IOR_{\text{water}} = 1.33$ to $IOR_{\text{water}} = 1.333$ results in consistent improvements in PSNR, SSIM, and LPIPS metrics,
- NeRFrac is applicable to real-world UAV imagery and can reconstruct refracted underwater scenes in natural, uncontrolled environments.

2. Related Work

In the following, we briefly summarize prior work on NeRFs, with a particular focus on their ability to handle refraction. This section is structured into four parts: i) general developments in NeRF models, ii) specific challenges posed by refractive media, iii) NeRFrac as an exemplary method for explicit refraction modeling, and iv) the existing research gap in UAV-based bathymetric NeRF applications.

Developments in NeRF Models. NeRFs have emerged as a powerful method for photorealistic 3D scene reconstruction and novel view synthesis from sparse multi-view imagery (Mildenhall et al., 2020). Since their introduction, numerous extensions have addressed limitations in computational efficiency, scalability, and scene complexity. Mip-NeRF proposes a multiscale representation to handle aliasing and resolution-dependent sampling issues (Barron et al., 2021), while TensorRF introduces tensor decomposition to improve rendering speed and memory efficiency (Chen et al., 2022). Other works such as Plenoxels eliminate neural networks entirely in favor of sparse voxel grids (Yu et al., 2021), and IBRNet integrates transformer architectures for improved generalization across scenes (Wang et al., 2021).

Recent research also explores scene-specific uncertainties and occlusion handling in NeRF pipelines, such as density-aware modeling (Jäger et al., 2025) or occlusion-aware comparisons across methods (Petrovska and Jutzi, 2025).

Various overviews (Yao et al., 2024; Xiao et al., 2025) further highlight the evolution of NeRFs towards real-world applicability, including human modeling, robotics, and general scene understanding. Yet, refractive environments remain underrepresented in NeRF literature.

Challenges in Refractive Environments. A major limitation of classical NeRFs is the assumption of straight-line light propagation. This assumption is violated in underwater or refractive scenarios, where light rays are bent at the interface between two media with different refractive indices, such as air and water, causing them to deviate from a straight line. The refractive behavior at the air-water interface depends on both the wavelength of light and the physical properties of the two media, particularly the temperature and density of water

(Schiebener et al., 1990; Bashkatov and Genina, 2003). For accurate 3D reconstruction, these dependencies should be explicitly modeled.

Several approaches, e.g. WaterNeRF (Sethuraman et al., 2023), SeaThru-NeRF (Levy et al., 2023), U2NeRF (Gupta et al., 2024), and AquaNeRF (Gough et al., 2025), focus on scenes that are fully underwater. As a result, they do not account for the refraction that occurs at the air-water interface, which is essential when the camera remains above the waterline.

However, a number of NeRF-based extensions have been proposed to explicitly address the challenges of refraction at media boundaries. Ray deformation approaches like LB-NeRF (Fujitomi et al., 2022) and Ray Deformation Networks (Deng et al., 2024) empirically learn offsets to simulate refraction, but often lack physical grounding.

In contrast, Eikonal Fields (Bemana et al., 2022) optimize a spatially varying index of refraction (IOR) field using the eikonal equation. NeRRF (Chen et al., 2023) combines refraction and reflection modeling using Fresnel equations and a progressive surface extraction scheme. A method to dynamically learn the IOR during training has been proposed to further improve flexibility and accuracy (Wei et al., 2025). A related approach, REF2-NeRF (Kim et al., 2024), jointly models reflection and refraction, particularly in glass-dominated scenes, but is not designed for water-air interfaces.

Explicit Refraction Modeling with NeRFrac. NeRFrac (Zhan et al., 2023), represents one of the first physically grounded NeRF frameworks that explicitly models the interaction of light with a refractive surface. It introduces a dedicated *Refractive Field* to estimate the surface intersection based on Snell’s Law, enabling simultaneous recovery of the water surface and the scene below. Unlike prior deformation-based methods, NeRFrac provides a disentangled representation of the refractive interface and underlying *Radiance Field*.

Quantitative comparisons on synthetic topographic data (Schulte et al., 2025) and real data show that NeRFrac significantly outperforms other baselines in terms of PSNR, SSIM, and LPIPS metrics. Additionally, the extraction of 3D point clouds from trained NeRFrac models is being explored (Brezovsky et al., 2025).

Research Gap: UAV-based Bathymetric NeRF Applications. Despite recent progress, no prior work has applied NeRFrac or comparable refractive NeRF frameworks to UAV-based data captured in outdoor, fluvial environments. All methods for bathymetry such as photogrammetric depth estimation or airborne laser scanning are sensitive to water clarity and surface conditions. Our work represents the first real-world adaptation of NeRFrac on high-resolution UAV imagery from the Pielach River (Mandlbürger et al., 2025b,a). Compared to controlled laboratory setups, the primary challenge in real-world applications is that refraction does not occur uniformly across the entire image. This selective refraction requires the use of per-ray masking to distinguish between submerged and non-refractive regions. Additional complexities include a dynamic water surface, potential turbidity, specular reflections, and pose uncertainties.

3. Methods

To investigate the effect of refractive ray modeling in NeRFs, we use the NeRFrac framework and train it on the UAV data-

set described in Section 4. We compare multiple variants of NeRFrac that differ only in their assumed IOR for water. The IOR values 1.32, 1.33, 1.333 and 1.34 are chosen to reflect realistic conditions for water, where the IOR is typically close to 1.333 for temperatures between 10 and 20 degrees Celsius. This range allows us to analyze the sensitivity of the model to small variations in refractive behavior under realistic environmental conditions.

3.1 Network Architecture

The NeRFrac framework builds upon the standard “Vanilla” NeRF architecture (Mildenhall et al., 2020) and extends it to explicitly model light refraction at transparent surfaces such as water. While Vanilla NeRF assumes that light travels along straight rays in free space, NeRFrac removes this limitation by introducing refractive ray tracing based on Snell’s Law. This enables the generation of accurate novel views of scenes seen through a water surface.

NeRFrac introduces a *Refractive Field* $\Psi_R(\mathbf{o}, \mathbf{v}) = d$ that estimates the distance d from the ray origin \mathbf{o} in direction \mathbf{v} to the intersection point \mathbf{X}_s on the refractive surface. It is implemented as a fully connected multilayer perceptron (MLP) with eight layers and 256 units per layer.

Using the estimated surface point \mathbf{X}_s and the local surface normal \mathbf{N} , computed via least-squares fitting of neighboring surface points, NeRFrac computes the refracted ray direction \mathbf{v}' according to Snell’s Law in vector form:

$$\mathbf{v}' = \eta(\mathbf{I} + c_1\mathbf{N}) - c_2\mathbf{N}, \quad (1)$$

where $\mathbf{I} \in \mathbb{R}^3$ is the incident ray direction and $\mathbf{N} \in \mathbb{R}^3$ is the unit surface normal, both assumed to be normalized. η is the ratio of refractive indices $\frac{n_1}{n_2}$, $c_1 = \mathbf{N} \cdot \mathbf{I}$, and $c_2 = \sqrt{1 - \eta^2(1 - c_1^2)}$.

Along the computed refracted ray, 3D sample points are generated and processed by a second MLP, the *Radiance Field*, which outputs volume density σ and RGB color values conditioned on both the view direction and the spatial location. Volume rendering is then performed by integrating these values along the refracted ray path.

Both the *Refractive Field* and the *Radiance Field* are trained jointly by minimizing the reconstruction loss between rendered and ground-truth images. This implicit supervision allows the *Refractive Field* to learn physically consistent surface geometry without requiring explicit ground-truth depth maps.

To maintain a consistent and compact representation of rays across different views, NeRFrac operates in normalized device coordinates (NDC). In this representation, all rays originate from a common near plane and point along the same axis direction, which simplifies the parameterization of viewing rays. This normalization facilitates stable learning of the *Refractive Field*, as ray origins and directions are expressed in a view-consistent frame. It also allows the surface normals, computed from neighboring refracted points, to be estimated more robustly across varying camera poses. This design choice contributes to improved generalization and stable training behavior in both synthetic and real-world datasets.

3.2 Mask-Based Ray Selection

Building on this architecture, we further extend NeRFrac to support heterogeneous scenes with both refractive and non-refractive regions. While the original NeRFrac framework applies refraction uniformly to all rays, our extension introduces mask-based ray selection to enable per-ray control over refractive modeling. These masks are created for each input image and indicate, at pixel level, whether the corresponding ray intersects a water surface. Figure 1 shows the water mask for the test image. During training and rendering, this information is used to conditionally apply refraction only to rays that traverse water, while all other rays follow the standard perspective projection without changing direction. This enables a hybrid ray tracing strategy within a single scene.

Our approach is conceptually related to previous work that uses masks to distinguish between refractive and non-refractive regions in transparent object rendering (Wei et al., 2025). However, that method is based on ray tracing with Fresnel modeling and does not target outdoor hydrographic environments. In contrast, our implementation, based on the NeRFrac architecture, integrates binary masks directly into the forward pass. Unlike prior work, our setup specifically targets shallow-water UAV imagery and enables refractive ray modeling within complex real-world river systems.



Figure 1. Overlay of the manually defined water mask on the test image. The masked region (dark polygonal area) corresponds to the river surface and defines where refractive ray modeling is applied during NeRFrac training and rendering.

3.3 Experimental Setup

To systematically assess the behavior of the proposed model variants, we conduct several controlled experiments using the same UAV dataset. All network variants are trained on the same subset of UAV imagery described in Section 4. The training configuration remains consistent across experiments, including learning rate, number of iterations, ray sampling parameters, and network architecture. We evaluate both the original NeRFrac setup and our mask-based extension, using identical input data to ensure comparability. Variants differ only in their refractive index settings and the presence or absence of per-ray masking.

3.4 2D Evaluation

Following training, we evaluate the rendering quality of each configuration using a fixed test view. Specifically, we select the central image of the 9 captured views, corresponding to the

central viewpoint in the 3×3 camera grid. This image is excluded from training and serves solely for 2D evaluation.

We report the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS) between the rendered and the reference image. Additionally, we compute a composite average score to facilitate global comparison across metrics. This score is defined as the geometric mean of $MSE = 10^{-PSNR/10}$, $\sqrt{1 - SSIM}$, and LPIPS (Barron et al., 2021). This formulation jointly captures pixel-wise accuracy, structural similarity, and perceptual distance in a directionally consistent manner.

Note that LPIPS is computed over the entire image rather than only on water-covered regions defined by the mask. This is because LPIPS relies on perceptual features extracted by convolutional neural networks, which require spatially coherent input regions. Since the water mask defines an irregular, non-rectangular subset of the image, applying LPIPS selectively to only those pixels would break the structural assumptions of the metric and yield unreliable results. In contrast, PSNR and SSIM can be computed on masked areas without violating their formulation, and are therefore used in the water-specific evaluations reported in subsection 5.2.

4. Data

The dataset utilized in this research originates from an inter-institutional research collaboration led by TU Wien (Mandlbauer et al., 2025b). It is publicly available via TU Wien’s research data repository and serves as a benchmark for optical hydrography based on images and laser scans (Mandlbauer et al., 2025a). The set includes nadir and oblique UAV imagery and airborne laser bathymetry captured during a field campaign at the Pielach River in Lower Austria. Data acquisition took place in October 2024 following a major flood event, resulting in clear-water conditions suitable for optical hydrographic analysis.

For this contribution, we use a subset of nine nadir-oriented UAV images taken at an altitude of 80 m above ground level. The imagery is captured using a DJI M350 RTK multicopter. A Zenmuse P1 45 MPix RGB camera is mounted on the UAV, achieving a ground sampling distance of approximately 1 cm. These images are selected based on coverage and image quality for training the NeRFrac network on shallow-water scenes.

To support targeted modeling of light refraction in submerged areas, we create binary masks that distinguish water surfaces from land. These masks are aligned to the images and indicate for each pixel whether it depicts water or land. In the current version, we generate the masks manually, but we aim to automate this process in the future using semantic segmentation techniques.

5. Results

This section presents the experimental results obtained with different refractive settings and masking strategies. The evaluation is structured in three main parts: First, we report quantitative results, starting with a global evaluation over the entire test image (Section 5.1), followed by a more focused regional analysis that considers only the water-covered parts of the scene (Section 5.2). Subsequently, we include qualitative renderings to

visualize differences (Section 5.3). Since the image differences are often subtle, the quantitative metrics provide essential evidence for assessing the impact of refractive modeling. Minor deviations are mainly visible in vegetated areas. Together, these analyses offer a comprehensive view on the impact of refraction modeling and per-ray masking within the NeRFrac framework.

5.1 Global Evaluation (Entire Image)

Table 1 presents quantitative evaluation results based on the entire test image, which includes both submerged and terrestrial areas. In contrast to the regional evaluation in Table 2, where only pixels of submerged areas were considered, this evaluation captures the overall image quality across the full scene and allows comparison using additional perceptual metrics such as LPIPS. To synthesize the results into a single score, we also report a composite average metric as described in Section 3.4.

Figures 2 and 3 visualize key image quality metrics across training iterations, computed on the entire test image. While Figure 2 focuses on PSNR, Figure 3 presents SSIM, LPIPS, and the composite average metric.

In both figures, training with and without water masks is directly compared, both implementations using an IOR of 1.333. Despite the relatively small differences in global metrics which were expected due to the dominance of terrestrial areas in the scene, a consistent improvement can be observed when water masks are used. This indicates that even globally, the selective application of refractive modeling guided by masking has a beneficial effect on reconstruction quality.

IOR _{water}	PSNR ↑	SSIM ↑	LPIPS ↓	Average ↓
Training without masks				
1.32	23.692	0.486	0.571	0.120
1.33	23.711	0.487	0.565	0.120
1.333	23.802	0.496	0.561	0.118
1.34	23.782	0.493	0.564	0.119
Training with masks				
1.32	23.814	0.501	0.556	0.118
1.325	23.839	0.500	0.561	0.118
1.33	23.870	0.504	0.554	0.117
1.333	23.899	0.503	0.554	0.117
1.335	23.836	0.502	0.557	0.118
1.34	23.835	0.501	0.555	0.117

Table 1. Quantitative evaluation results for the full test image, including both water and terrestrial areas, at training iteration 200 000. Reported metrics are PSNR, SSIM, and LPIPS, as well as the average score combining all three. The average is computed as the geometric mean of $MSE = 10^{-PSNR/10}$, $\sqrt{1 - SSIM}$, and LPIPS (Barron et al., 2021). The 200 raw metric values, obtained during rendering of test images every 1000 iterations, were smoothed using a moving average filter (size 15 points). The best-performing refractive index configuration and the corresponding metrics are highlighted in bold.

5.2 Regional Evaluation (Water-Only Pixels)

To better assess the impact of refraction modeling in water-covered areas, we compute PSNR and SSIM only for pixels marked as water in the binary mask. Table 2 summarizes the results. Across all IOR values, masking consistently improves

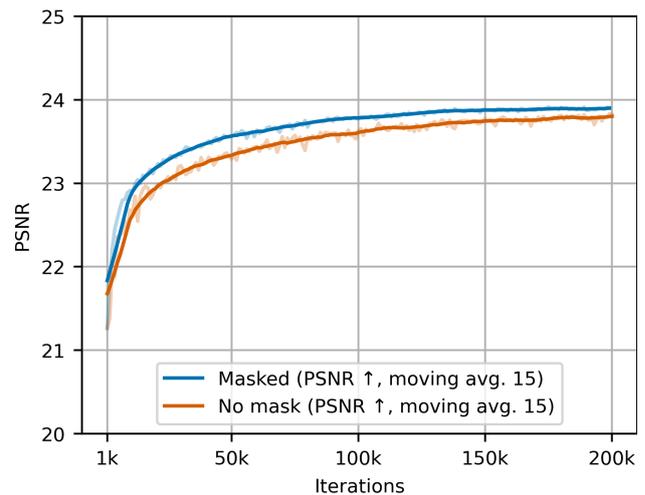


Figure 2. PSNR across iterations for two training variants. The blue curve shows the PSNR computed on the entire image when training was performed using water masks and an IOR of 1.333 was applied only in the masked (i.e., water-covered) regions. The orange curve represents the PSNR when no masks were used and an IOR of 1.333 was applied globally across the entire image. Training was performed for 200 000 iterations. The 200 raw metric values, obtained during rendering of test images every 1000 iterations, were smoothed using a moving average filter (size 15 points).

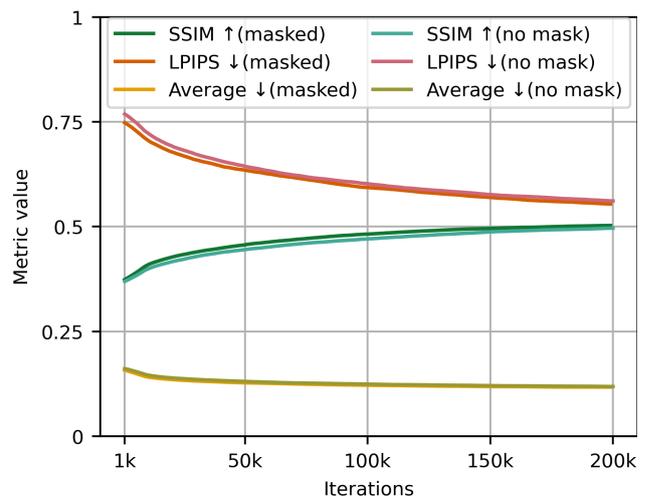


Figure 3. SSIM, LPIPS and average across iterations for two training variants. The green, red, and orange curves show the SSIM (↑), LPIPS (↓) and average (↓) respectively for training with masks, where an IOR of 1.333 was applied only in the water-covered (masked) regions. The corresponding cyan, pink, and olive curves represent the same metrics for training without masks, where an IOR of 1.333 was applied globally across the entire image. Training was performed for 200 000 iterations. The 200 raw metric values, obtained during rendering of test images every 1000 iterations, were smoothed using a moving average filter (size 15 points).

results. This evaluation is based on the previously described binary masks that indicate the presence of water at the pixel level, allowing a targeted assessment of the network’s capacity to model light behavior below the water surface.

IOR_{water}	PSNR \uparrow	SSIM \uparrow
Training without masks		
1.32	36.009	0.841
1.33	36.063	0.843
1.333	36.243	0.843
1.34	36.134	0.842
Training with masks		
1.32	36.288	0.843
1.325	36.264	0.843
1.33	36.435	0.844
1.333	36.447	0.844
1.335	36.262	0.844
1.34	36.235	0.843

Table 2. Quantitative evaluation of different refractive index values (IOR_{water}) in NeRFrac with and without per-ray masking. Metrics are computed exclusively for pixels corresponding to water-covered regions in the test image, as defined by the binary water masks. The reported values represent the mean PSNR and SSIM at training iteration 200 000. The 200 raw metric values, obtained during rendering of test images every 1000 iterations, were smoothed using a moving average filter (size 15 points). Note that LPIPS and the composite average metric are not reported, as they require full-image context and cannot be meaningfully computed for masked regions only. The best-performing refractive index configuration and the corresponding metrics are highlighted in bold.

The upper part of Table 2 shows the results for models trained without water masks, treating all rays equally, regardless of whether they intersect water or not. The lower section presents the same evaluation for models trained with masking, allowing selective application of refractive modeling.

Across all configurations, the differences in PSNR and SSIM values are generally small, indicating that NeRFrac performs robustly under slight variations in the refractive index. Nonetheless, the results consistently show that combining masking with a physically plausible refractive index of $IOR_{\text{water}} = 1.333$ yields the best reconstruction quality for submerged areas. This confirms the effectiveness of mask-based hybrid ray modeling and highlights the relevance of accurate IOR selection for shallow-water environments.

Since water covers only about 20 % of the image, its influence on the global score is diluted, and the observed differences are smaller than in the regional evaluation. Still, the use of masks consistently leads to the best performance across all reported metrics.

Figure 4 shows the PSNR across training iterations, but in contrast to Figure 2, the evaluation here is restricted to submerged areas only. A clear difference between masked and unmasked training is visible: when using masks, the PSNR converges faster and exhibits less noise throughout the training process. This stability makes it more suitable for loss computation based on PSNR and indicates that the masking strategy effectively guides the network in learning refractive behavior in submerged areas.

5.3 Qualitative Results

Beyond quantitative evaluation, visual inspection remains an essential complement for identifying subtle structural deviations. In the following, we present qualitative renderings to illustrate the differences between masked and unmasked training

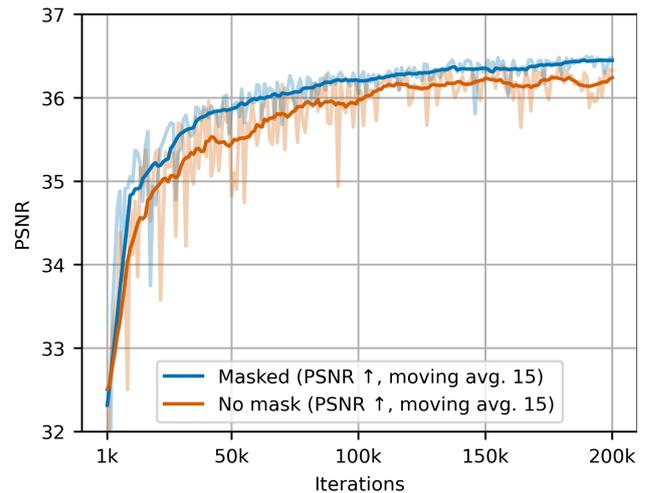


Figure 4. PSNR across iterations for two training variants, evaluated only on the water-covered regions of the image. The blue curve shows the PSNR when training was performed using water masks and an index of refraction (IOR) of 1.333 was applied only in the masked (i.e., water-covered) regions. The orange curve represents the PSNR when no masks were used and an IOR of 1.333 was applied globally across the entire image. Training was performed for 200 000 iterations. The 200 raw metric values, obtained during rendering of test images every 1000 iterations, were smoothed using a moving average filter (size 15 points).

configurations. These visualizations help contextualize the numerical results by showing where and how the observed differences manifest in the image domain. However, the synthesized images appear almost identical, regardless of whether masking was used during training, which further emphasizes the importance of the preceding quantitative evaluations.

Figures 5 and 6 show visual comparisons of the reference image and the corresponding reconstructions obtained from masked and unmasked training, respectively. In both figures, the top row includes the full reference image (a), the synthesized image from the corresponding camera pose (b), and a pixel-wise difference map (c) visualized as a false-color composite. The bottom row shows cropped regions (d–f) that correspond to the areas marked in the full images above, allowing a more focused inspection of localized errors.

In Figure 5, which shows results for training with per-ray masks and $IOR = 1.333$, only minor differences are visible between the synthesized image and the reference. These small deviations become more evident in the difference map (Figure 5 c), particularly in the vegetated regions.

Figure 6 presents the same comparison setup, but for a model trained without any masking. The differences between the synthesized image and the reference image remain small.

When directly comparing the two synthesized images, with and without masking, only minor visual differences can be discerned. Figure 7 illustrates this comparison explicitly: Figure 7a shows the pixel-wise difference between the two renderings, while Figure 7b shows a zoomed-in crop of the same region previously highlighted in Figures 5 and 6. Despite the near-identical appearance of the outputs, the difference map still reveals small localized deviations. This underlines the importance of quantitative metrics for systematic evaluation.

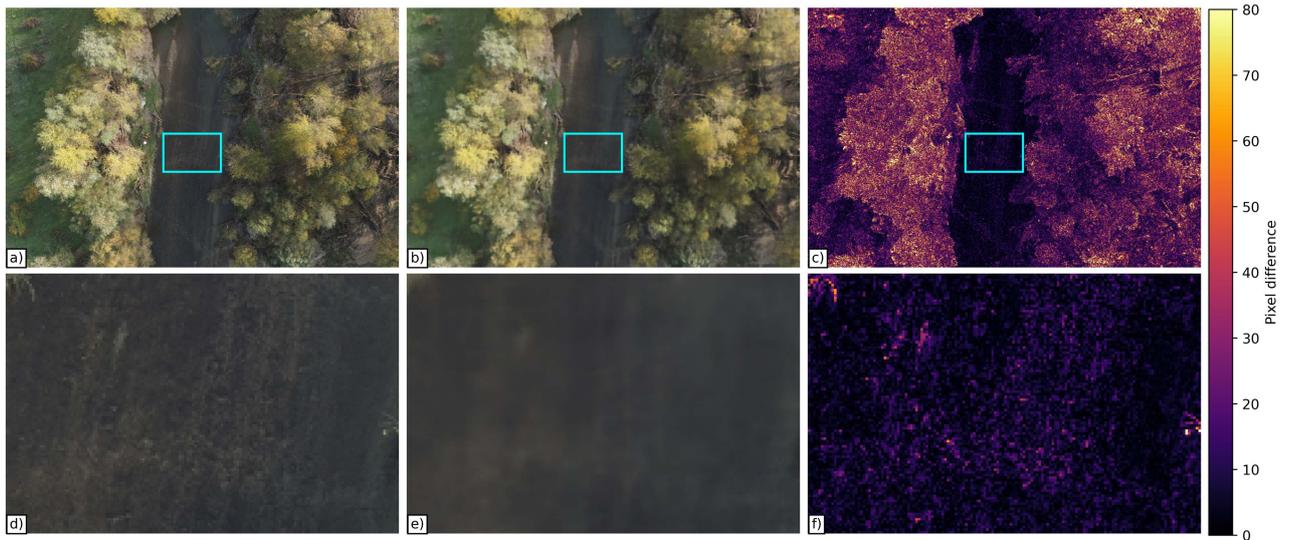


Figure 5. Results trained with masks: a) reference image, b) synthesized image trained with masks and IOR = 1.333 for the water covered part of the scene, c) difference map, d) crop of reference image, e) crop of synthesized image and f) crop of difference map. The difference maps are shown as false color composites representing the Euclidean distance in 3x8-bit RGB space (dimensionless).

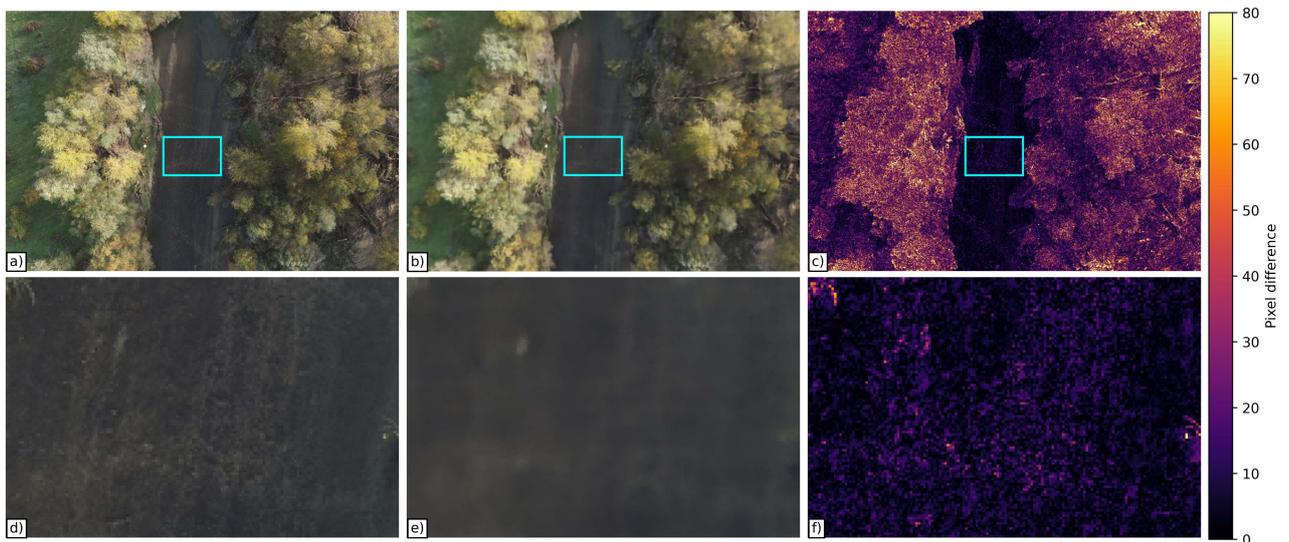


Figure 6. Results trained without masks: a) reference image, b) synthesized image trained without masks and IOR = 1.333 for the entire scene, c) difference map, d) crop of reference image, e) crop of synthesized image and f) crop of difference map. Again, the difference maps are shown as false color composites representing the Euclidean distance in 3x8-bit RGB space (dimensionless).

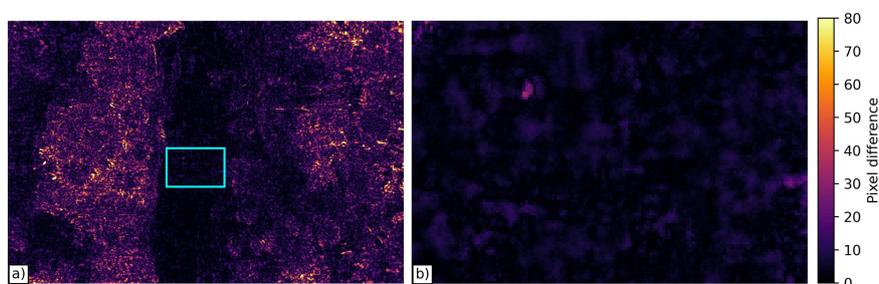


Figure 7. Difference between synthesized images trained with and without masks: a) difference for the entire image (i.e. difference between Figure 5b and Figure 6b), b) difference for the crop in the water covered area (i.e. difference between Figure 5e and Figure 6e). Shown as false color composites representing the Euclidean distance in 3x8-bit RGB space (dimensionless).

6. Discussion

Our results demonstrate that training with binary water masks outperforms training with a global approach to refraction modeling (Figure 2). This advantage is especially evident in the PSNR and SSIM metrics computed exclusively for submerged areas (Table 2). Here, per-ray refractive modeling enables a more physically consistent rendering process, selectively applying Snell's Law only where image rays are expected to change direction due to the air–water interface. In contrast, models without masks apply the same refractive assumptions to all rays, which can lead to reduced fidelity in terrestrial areas.

In global evaluations that include both land and water areas, however, the improvements introduced by masking are less pronounced. This can be attributed to the spatial distribution of the test image: approximately 20 % of the pixels belong to water-covered regions, while the remaining 80 % represent terrestrial features. As a result, the contribution of refractive ray modeling is diluted in the full-frame metrics, despite its targeted benefits.

Among all tested refractive indices, the value $IOR_{\text{water}} = 1.333$ consistently yielded the best performance, particularly when combined with masking. While the absolute differences compared to neighboring IOR values (e.g., 1.32 or 1.34) are small, the consistency of this result across metrics underscores the model's sensitivity to physically accurate refraction parameters.

Visually, the differences between all model outputs are remarkably subtle. Even in cropped close-ups of the water region, deviations between masked and unmasked renderings are difficult to discern. This emphasizes the importance of robust, quantitative evaluation metrics such as PSNR and SSIM in assessing the effectiveness of refraction-aware NeRFs. In practice, perceptual metrics like LPIPS cannot be meaningfully applied in masked regions due to their reliance on full-frame context, and are therefore omitted from the regional evaluation.

The introduced mask-based extension to NeRFrac currently relies on manually created water masks. This manual effort is time-consuming and does not scale well to larger or more complex datasets. Future work should therefore explore automated segmentation approaches to streamline and scale the mask creation process.

Another limitation is the lack of explicit geometric validation. While qualitative renderings and image-based metrics indicate improved fidelity, no 3D reconstruction quality was evaluated against external ground truth. A promising avenue for future research is the analysis of point cloud accuracy, following the approach of Brezovsky et al. (2025), who assess refractive NeRFs in terms of geometric reconstruction performance.

7. Conclusion

This contribution presents the first application of a refraction aware NeRF variant, specifically NeRFrac, to real-world UAV data captured in an uncontrolled outdoor riverine setting. Our findings demonstrate that refractive NeRFs can be successfully trained on high-resolution aerial imagery to model complex light interactions at the air–water interface. In particular, applying a physically plausible index of refraction ($IOR = 1.333$) exclusively to rays intersecting water, using per-pixel masks, yields the highest reconstruction accuracy in submerged regions. While the global image quality does not improve uniformly through refraction modeling, due to its localized benefit

in water and potential interference in terrestrial areas, targeted refractive ray modeling proves highly effective for bathymetric regions.

This study demonstrates the feasibility of integrating physically grounded refraction modeling into NeRFs for aerial photogrammetry. It contributes to the field of optical bathymetry by opening up new possibilities for 3D scene reconstruction beneath refractive surfaces. Moreover, it extends the scope of hybrid NeRF architectures, where rays are treated heterogeneously depending on scene semantics, an approach recently explored (Wei et al., 2025), but not yet applied to natural water surfaces.

In addition, future studies should include direct comparisons with methods that explicitly model refractive effects at air–water interfaces, such as Eikonal Fields (Bemana et al., 2022), NeRRF (Chen et al., 2023), and REF2-NeRF (Kim et al., 2024). Integrating such comparisons would help to more precisely situate NeRFrac within the broader landscape of refraction-aware neural rendering.

To enhance the scalability and automation of the presented approach, future work should focus on integrating semantic segmentation for automatic water mask generation and incorporating refractive NeRFs into broader photogrammetric workflows, such as dense point cloud extraction and georeferencing. In addition, combining the method with LiDAR or depth-based measurements will enable full 3D validation and extend its applicability to full-scale hydrographic mapping.

Acknowledgments

The research presented was carried out within the transnational WEAVE project BathyNeRF funded by the Austrian Science Fund (FWF) [10.55776/PIN1353223] and the German Research Foundation (DFG) [538522540].

References

- Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P. P., 2021. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 5835–5844. ISSN: 2380-7504.
- Bashkatov, A. N., Genina, E. A., 2003. Water refractive index in dependence on temperature and wavelength: a simple approximation. *Saratov Fall Meeting 2002: Optical Technologies in Biophysics and Medicine IV*, 5068, SPIE, 393–395.
- Bemana, M., Myszkowski, K., Revall Frisvad, J., Seidel, H.-P., Ritschel, T., 2022. Eikonal Fields for Refractive Novel-View Synthesis. *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings*, ACM, Vancouver BC Canada, 1–9.
- Brezovsky, M., Guenther, A., Schulte, F., Winiwarter, L., Jutzi, B., Mandlbürger, G., 2025. Analysis of refraction-aware neural radiance fields for 3D reconstruction of underwater

scenes. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H., 2022. TensorRF: Tensorial Radiance Fields. *Computer Vision – ECCV 2022*, Springer Nature Switzerland, Cham, 333–350.

Chen, X., Liu, J., Zhao, H., Zhou, G., Zhang, Y.-Q., 2023. NeRRF: 3D Reconstruction and View Synthesis for Transparent and Specular Objects with Neural Refractive-Reflective Fields. arXiv:2309.13039.

Deng, W., Campbell, D., Sun, C., Kanitkar, S., Shaffer, M., Gould, S., 2024. Ray Deformation Networks for Novel View Synthesis of Refractive Objects. 3118–3128.

Fujitomi, T., Sakurada, K., Hamaguchi, R., Shishido, H., Onishi, M., Kameda, Y., 2022. LB-NeRF: Light Bending Neural Radiance Fields for Transparent Medium. *2022 IEEE International Conference on Image Processing (ICIP)*, 2142–2146. ISSN: 2381-8549.

Gough, L., Azzarelli, A., Zhang, F., Anantrasirichai, N., 2025. AquaNeRF: Neural Radiance Fields in Underwater Media with Distractor Removal. arXiv:2502.16351.

Gupta, V., S, M., T, M. V., Mitra, K., 2024. U2NeRF: Unsupervised Underwater Image Restoration and Neural Radiance Fields. arXiv:2411.16172 [cs].

Jäger, M., Landgraf, S., Jutzi, B., 2025. Density uncertainty quantification with NeRF-Ensembles: Impact of data and scene constraints. 137, 104406. <https://publikationen.bibliothek.kit.edu/1000179539>.

Kim, W., Fukiage, T., Oishi, T., 2024. REF2-NeRF: Reflection and Refraction aware Neural Radiance Field. *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7196–7203. ISSN: 2153-0866.

Levy, D., Peleg, A., Pearl, N., Rosenbaum, D., Akkaynak, D., Korman, S., Treibitz, T., 2023. SeaThru-NeRF: Neural Radiance Fields in Scattering Media. arXiv:2304.07743.

Mandlbürger, G., Glas, S., Amon, P., Rhomberg-Kauert, J., Gueguen, L.-A., Mulsow, C., Brezovsky, M., Dammert, L., Haines, J., Schulte, F., Winiwarter, L., Jutzi, B., Maas, H.-G., 2025a. Pielach River research dataset October 2024 - Mapping shallow inland running waters with UAV-borne photo and laser bathymetry.

Mandlbürger, G., Rhomberg-Kauert, J., Gueguen, L.-A., Mulsow, C., Brezovsky, M., Dammert, L., Haines, J., Glas, S., Himmelsbach, T., Schulte, F., Amon, P., Winiwarter, L., Jutzi, B., Maas, H.-G., 2025b. Mapping shallow inland running waters with UAV-borne photo and laser bathymetry : The Pielach River showcase. *Journal of applied hydrography*, 130(3), 42–54.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. arXiv:2003.08934.

Petrovska, I., Jutzi, B., 2025. Seeing beyond vegetation: A comparative occlusion analysis between Multi-View Stereo, Neural Radiance Fields and Gaussian Splatting for 3D reconstruction. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 100089. <https://publikationen.bibliothek.kit.edu/1000181231>.

Schiebener, P., Straub, J., Levelt Sengers, J. M. H., Gallagher, J. S., 1990. Refractive index of water and steam as function of wavelength, temperature and density. *Journal of Physical and Chemical Reference Data*, 19(3), 677–717. <https://pubs.aip.org/jpr/article/19/3/677/457268/Refractive-index-of-water-and-steam-as-function-of>.

Schulte, F., Guenther, A., Brezovsky, M., Mandlbürger, G., Jutzi, B., Winiwarter, L., 2025. Simulation and validation of underwater scenes for two-media optical 3D reconstruction. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

Sethuraman, A. V., Ramanagopal, M. S., Skinner, K. A., 2023. WaterNeRF: Neural Radiance Fields for Underwater Scenes. *OCEANS 2023 - MTS/IEEE U.S. Gulf Coast*, 1–7. ISSN: 0197-7385.

Wang, Q., Wang, Z., Genova, K., Srinivasan, P. P., Zhou, H., Barron, J. T., Martin-Brualla, R., Snavely, N., Funkhouser, T., 2021. IBRNet: Learning Multi-View Image-Based Rendering. 4690–4699.

Wei, J., Yue, Z., Li, S., Cheng, Z., Lian, Z., Song, M., Cheng, Y., Zhao, H., 2025. An Index of Refraction Adaptive Neural Refractive Radiance Field for Transparent Scenes. *Electronics*, 14(6), 1073. <https://www.mdpi.com/2079-9292/14/6/1073>.

Xiao, W., Chierchia, R., Cruz, R. S., Li, X., Ahmedt-Aristizabal, D., Salvado, O., Fookes, C., Lebrat, L., 2025. Neural Radiance Fields for the Real World: A Survey. arXiv:2501.13104 [cs].

Yao, M., Huo, Y., Ran, Y., Tian, Q., Wang, R., Wang, H., 2024. Neural Radiance Field-based Visual Rendering: A Comprehensive Review. arXiv:2404.00714 [cs].

Yu, A., Fridovich-Keil, S., Tancik, M., Chen, Q., Recht, B., Kanazawa, A., 2021. Plenoxels: Radiance fields without neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zhan, Y., Nobuhara, S., Nishino, K., Zheng, Y., 2023. NeRFrac: Neural Radiance Fields through Refractive Surface. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 18356–18366. ISSN: 2380-7504.