CENS: A Coverage-Efficient Pixel Sampling Strategy for Enhancing NeRF-Generated Point Cloud Fidelity

Perpetual Hope Akwensi[1], Frederik Schulte[1], Lukas Winiwarter[1]

[1] Unit of Geometry and Surveying, Faculty of Engineering Sciences, Universität Innsbruck, Innsbruck, Austria
(perpetual.akwensi, frederik.schulte, lukas.winiwarter)@uibk.ac.at

Keywords: 3D Reconstruction, 3D Point Clouds, NeRF, Ortho-photos, Quality Assessment, Simulation

Abstract

Many geospatial workflows critically depend on high-fidelity 3D point clouds for applications such as change detection, orthophoto generation, and modeling. However, NeRF-generated point clouds often suffer from sampling inefficiencies inherent in the predominant random pixel sampling approach. We identify spatial redundancy as one such inefficiency: random sampling has the inevitable consequence of sampling large, low-texture patches more frequently than detailed, high-frequency textured regions. As a result, low-texture areas tend to be oversampled and other pixels remain unsampled – regardless of their importance to the reconstruction task. To overcome this, we propose **CENS** (Coverage-Efficient Non-Redundant Sampling), a deterministic pixel sampling strategy that maximizes spatial coverage, eliminates intra-image sample repetition, and ensures reproducibility via structured initialization. Evaluated on the Jamtal valley dataset, CENS achieves comparable geometric accuracy (cloud-to-mesh (C2M) distances: μ = -0.0027 vs. -0.0011 m; σ = 0.027 vs. 0.028 m) using 50% fewer training steps (11,232 vs. 22,464), while yielding 28.2% more points, higher orthophoto fidelity, and improved point cloud completeness. Beyond CENS, we also explored NeRFs for ALS point cloud simulation, achieving realistic occlusion patterns and accuracy within UAV photogrammetry standards (V_{RMSE} = 24 mm; H_{RMSE} = 17 mm). Crucially, CENS positions NeRFs as a scalable, practical solution for geospatial point cloud and orthophoto generation, advancing them toward real-world mapping workflows, and integrates seamlessly into NeRFStudio.

1. Introduction

Since their introduction by Mildenhall et al. (2020), Neural Radiance Fields (NeRFs) have become a popular approach for photo-realistic rendering, novel-view synthesis and 3D scene reconstruction from 2D images and their camera poses. As such, over the past few years, various aspects of NeRFs such as encoding strategies (Tancik et al., 2020; Barron et al., 2021; Müller et al., 2022; Wang et al., 2024), ray sampling strategies (Neff et al., 2021; Kurz et al., 2022; Bello et al., 2024), rendering acceleration (Liu et al., 2020; Reiser et al., 2021; Müller et al., 2022; Chen et al., 2025b), and scalability (Turki et al., 2022; Tancik et al., 2022; Mi and Xu, 2023) have been investigated and/or improved. Their application scenarios have also been extended to dynamic scenes (Tang et al., 2024; Yu et al., 2025), unbounded scenes (Barron et al., 2022, 2023; Lee et al., 2024), videos, etc. A part of NeRFs that has not seen much exploration is its pixel sampling strategy, although being crucial for the generation of 3D point cloud (PC) at high quality.

The de-facto pixel sampling strategy employed in most existing NeRF models is the equal probability random sampling, hereafter referred to as random sampling (RS). Given a N images of size $H \times W$, RS aims for statistical fairness, thus every pixel has an $\frac{1}{H \times W \times N}$ probability of being sampled. This strategy is simple and easy to implement, but it does not take into account spatial distribution, hence can lead to uneven pixel sampling, multiple pixel repetitions, and unsampled pixels – as illustrated in Fig. 1. There have been some modifications to the original NeRF pixel sampling strategy, such as masking, semantic context, and patch-based sampling, but they all still employed the same RS strategy (Otonari et al., 2022; Tu et al., 2024; Wu et al., 2025).

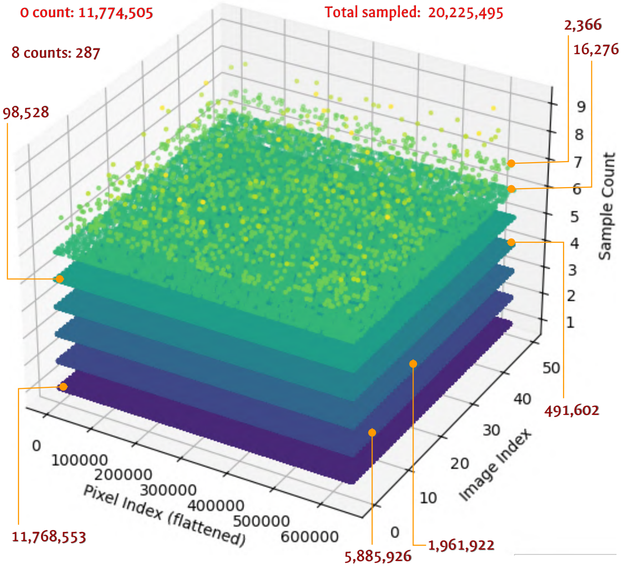


Figure 1. Illustration of the random sampling used in most NeRF models. Given N = 50, H = 800, W = 800, and batch size (BS) of 4096, it takes $\frac{H \times W \times N}{BS}$ iterations to sample all pixels without repetition. This figure shows the distribution of sampled pixels after 7812 iterations. Out of 32 million pixels, only 20.23 million have heen sampled between 1 and 8 times, leaving 11.77 million unsampled. Dark red values show how many pixels were sampled how many times. E.g., 1.96 million pixels were repeatedly sampled 3×.

2D novel view synthesis has predominantly been the primary focus of NeRF-based methods, while 3D point cloud generation is often treated as an auxiliary or secondary outcome. However, in the field of geomatics, point clouds serve as foundational

https://doi.org/10.5194/isprs-annals-XI-2-2026-801-2026

data for a wide range of derived products like terrain models (Lê et al., 2022), tasks like 3D modeling (Akwensi et al., 2024), and analytical applications like change detection (Winiwarter et al., 2021). Consequently, it is essential to evaluate the quality of NeRF-generated 3D point clouds, their corresponding orthophotos, and synthesized ortho-points (i.e., simulated aerial LiDAR point clouds). As such, this paper investigates the influence of random pixel sampling on the quality of NeRF-generated point clouds and introduces **CENS** – a Coverage-Efficient Non-redundant Sampling strategy designed to maximize spatial coverage, ensure reproducibility, and mitigate sample repetition – while being fully integrable into the NeRFStudio (Tancik et al., 2023) framework.

2. Related Work

Pixel sampling strategies in NeRFs. Apart from ray sampling strategies, pixel sampling has emerged as vital lever for enhancing the efficiency, stability, and fidelity of NeRF training. Most NeRF implementations rely on random pixel sampling, a naïve approach that often leads to overfitting in (1) low-textured areas – because, compared with high-texture regions, these areas have low feature variation, form large patches, and thus have a higher spatial probability of being sampled frequently; or (2) centered objects within a scene (Mildenhall et al., 2020). To address these limitations, a number of studies have proposed more targeted sampling schemes:

One prominent approach involves importance sampling guided by information-theoretic cues and/or reconstruction metrics. Fukuda et al. (2024) introduced a two-stage strategy: initially sampling high-probability edge pixels, followed by a weighted sampling based on reconstruction error to prioritize pixels that contribute more significantly to scene geometry learning. Active sampling methods such as ActRay (Wu et al., 2023) and Soft Mining (Kheradmand et al., 2024) refine pixel selection using a redesigned upper confidence bound formulation and Langevin Monte-Carlo algorithm, respectively – assigning higher sampling probabilities to hard-to-fit pixels and deprioritizing low-textured regions. Similarly, Otonari et al. (2022) proposed a distortion- and content-aware pixel sampling framework to mitigate sphere-to-plane projection bias and reduce sampling of low-texture pixels in 360-degree imagery. Unlike the above methods, Ge (2025) argued that due to substantial overlap across NeRF input views, it is more beneficial to prioritize hard-to-fit view directions rather than individual pixels. Other studies treat pixel importance as a binary decision, relying on model- or user-defined masks to guide sampling (Tu et al., 2024).

A complementary line of research explores adaptive resolution strategies. That is, instead of sampling all pixels at the finest granularity, dynamic-resolution approaches like MCBlock (Tan et al., 2025) partition images into texture-aware blocks. By leveraging Monte Carlo Tree Search, MCBlock is able to better manage computational resources in sampling high-resolution pixels in complex areas and low-resolution patches in simpler areas. This accelerates training while preserving high-frequency details.

Despite these advances, notable gaps remain. In particular, the interplay between NeRF’s pixel sampling strategies and the task of point cloud generation is underexplored.

Photogrammetric vs NeRF-Based image reconstruction. Classic photogrammetric approaches, while vital for large scale mapping, remain vulnerable to geometric inaccuracies when reconstructing textureless surfaces, managing dynamic illumination, and resolving complex occlusions, which can degrade spatial and perceptual fidelity. Recent extensions of NeRFs to remote sensing applications – most notably orthophoto generation (Wei et al., 2024; Yue et al., 2025; Chen et al., 2025a) and water-related geometry reconstruction (Levy et al., 2023; Schulte et al., 2025) – have mitigated many of these shortcomings. By learning a continuous volumetric representation directly from images, NeRFs obviate the need for a pre-existing 3D surface, thus enabling the robust handling self-occlusion (Karras et al., 2007) and hierarchical occlusion (Wang et al., 2018) while improving the rendering quality of reflective and under-water surfaces. In particular, NeRFOrtho (Yue et al., 2025) successfully eliminated the stitching line effect in orthophoto synthesis, thereby preserving building geometry and texture quality. Consequently, the demonstrated potential of NeRFs for geospatial scene reconstruction warrants further exploration.

3. Methodology

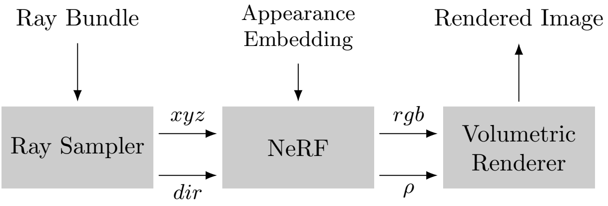


Figure 2. Overview of a basic NeRF pipeline.

Most often, NeRFs and their variants are defined as a model which takes xyz points and their view direction (dir) along a bundle of rays in 3D space given their camera poses (origins and rotations) as illustrated in Fig. 2. Other inputs like appearance embeddings are optional. The model then predicts the rgb color tuple and density ρ of each point, which is then used to render a final image. But one might ask: *how do the ray bundles come about, and how does their determination affect the rendered output?*

In most NeRF pipelines, given N images of size $H \times W$, and T training iterations, one of two strategies is employed in sampling a fixed number of pixels BS across N images per iteration: (1) the naïve approach (Mildenhall et al., 2020), and (2) the list-optimized strategy (Tancik et al., 2023). In each iteration step, the naïve method randomly samples image indices and their corresponding pixels with replacement, often resulting in the oversampling of certain images and pixels, while others receive few or no samples. This leads to uneven image coverage and training dynamics. In contrast, the list-optimized strategy adopts a more structured sampling scheme: although pixel selection within each image remains random and with replacement, it enforces a near-uniform allocation of pixel samples across images – approximately $\frac{BS}{N}$ pixels per image per iteration – ensuring that each image contributes more equitably to the training process.

Given a fixed number of sampled pixels, the rays passing through are computed to form a ray bundle.

3.1 Coverage-Efficient Non-redundant Sampling (CENS)

We propose a new pixel-sampling approach, called coverage-efficient non-redundant sampling (CENS), to mitigate sample repetition, maximize spatial coverage, enhance reproducibility, and better manage the number of training iterations.

To achieve these objectives we first construct a *grid map* – a

2D boolean array of size $H \times W$. The *grid map* indicates the pixel locations that should be sampled across the N images. Each pixel location is determined via Algorithm 1. Note that the terms *grid map* and mask are used interchangeably throughout this paper.

Spatial coverage and reproducibility. To achieve a fairly spaced and well-distributed set of pixels within image bounds, we defined and tested a number of *grid map* (mask) patterns for pixel selection. Examples of these defined patterns are shown in Fig. 3. Some patterns (e.g., *mask-1010* and *mask-0101*) make only $\sim$50% of the total pixels available for sampling, while others (e.g., *mask-1000* and *mask-0010*) make only $\sim$25% available, but with different spatial layouts to choose from. Given a user-chosen mask pattern, the active pixels (i.e., where mask value = 1) are retrieved, shuffled once with a fixed seed for reproducibility, and then sampled in batches of BS pixels across images per iteration. This ensures that all active pixels per image in the course of training get sampled, while also maximizing spatial coverage and ensuring reproducibility.

Additionally, to guide early training toward informative, high-frequency regions, we compute a Canny-based edge-strength map for every input image. Specifically: (1) For each training view, we apply Canny edge detection and retain the raw gradient magnitude image, which serves as a proxy for local texture (edge) strength. (2) During the first 30 percent of training iterations, pixel sampling is performed according to these edge-strength maps. Stronger edges are sampled with higher probability in early iterations, gradually transitioning toward weaker edges as training progresses.

This curriculum-style sampling ensures that the network first focuses on reconstructing high-frequency structures (object boundaries, fine details, and textured surfaces) before allocating capacity to smoother regions. This should reduce early-stage ambiguity and improve convergence.

After the initial 30 percent of the training iterations, we replace the edge-guided sampling with a *grid map* sampling strategy according to the user-chosen mask pattern. This ensures (1) low-textured areas are eventually incorporated without dominating the sampling distribution, and (2) uniform spatial coverage in later training stages.

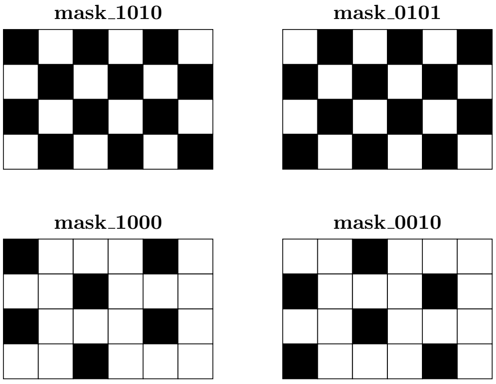


Figure 3. Examples of defined mask patterns.

Non-repetitive sampling. Given the overlapping nature of multi-view images generally used for NeRF reconstructions, pixel repetition can not be fully avoided. However, the *grid map* derived from mask patterns ensures that, within each image, no pixel is sampled more than once during T training iterations. While randomization of selected pixels per image in each batch is crucial for achieving accurate rendering, we argue that this should not come at the expense of reproducibility or spatial non-redundancy. Guided by the $grid\_map$, we thus in Algorithm 1 enable random yet non-repetitive sampling of pixels per image in each batch/iteration. In this way, all sample locations in the $grid\_map$ across N images are visited over T iterations (albeit in a random order), thereby eliminating intra-image repetition while preserving spatial coverage and reproducibility.

Training iteration management. The total number of training iterations T required to cover all selected pixels depends on the image size, the number of images, the batch size, and the sparsity of the $grid\_map$ (denoted as gf). Thus, we define T as: $T = \frac{(N \cdot H \cdot W)}{BS} \cdot gf$.

Algorithm 1: Coverage-Efficient Non-redundant Sampling (CENS) with optional edge-guided initialization

```
Input: $idx$ ;                       /* image index */
    $batch$ ;          /* dict of lists of tensors */
    $H, W$ ;              /* image height, width */
    $BS$ ;               /* batch size per image */
    $ts$ ;               /* train iteration step */
    $N$ ;                    /* number of images */
    $pname$ ;              /* mask pattern name */
    $canny$ ;        /* whether to use edge-guided
sampling */
Output: pixel_batch $\in \mathbb{R}^{[BS \times 3]}$;
cond ← $canny$ and $ts < 30\% \cdot T$
if not cond then
    1. mask ← GetMaskPattern ($H, W, pname$)
    2. mask_idxes ← mask.nonzero().shuffle(seed)
    3. curr_idxes ← mask_idxes [$BS \cdot ts : BS \cdot (ts + 1)$]
    if $ts$ != $T$ and curr_idxes < $BS$ then
        Cycle to a new $pname$
        Re-run steps 1 - 3
for $i \leftarrow 0$ to $N - 1$ do
    if cond then
        edge_map ← $batch$[‘edge_map’][idx]
        emap_idxes ← edge_map.argsort()
        curr_idxes ← edge_map [$BS \cdot ts : BS \cdot (ts + 1)$]
        pixel_batch ← concat($idx$, curr_idxes)
    else
        pixel_batch ← concat($idx$, curr_idxes)
return pixel_batch
```

3.2 Simulated ALS Point Cloud

Given technological advancements in UAV flight planning and ‘smart’ oblique capture, we show that NeRFs can be a means of generating simulated point clouds of various laser-scanning platforms – in this case, airborne laser scanning (ALS) PCs. Having trained a NeRF model with images from a well planned smart oblique mission, ALS PCs can be approximated from this trained model using camera intrinsics and simulated camera poses. To not stray from the initial flight and mission parameters: (1) the camera origins used during training, hereon refered to as initial origins (O_I), are used to generate a mesh surface via delaunay triangulation. (2) Based on the faces and vertices (O_I) of this surface, the centroid (new camera origins (O_N)) and normal of each face is computed. (3) The normals – which essentially give the camera a nadir view – are transformed into rotation matrices (camera orientation), then concatenated with O_N to form the simulated camera poses. (4) Using these simulated poses, image dimensions, and the camera intrinsics, we compute a ray bundle for NeRF ingestion and simulated ALS point cloud generation.

3.3 Ortho-photo Reconstruction via NeRF

In addition to the generation of high-fidelity point clouds via NeRFs, we explored NeRF-based orthophoto reconstruction. Motivated by the methodology of Yue et al. (2025), orthophotos were generated by rendering the trained NeRF model from a nadir perspective over a defined area of interest. Virtual cameras were configured as described in steps (2) and (3) of Section 3.2, ensuring that true scale and spatial resolution is preserved. Using these virtual cameras and the trained NeRF model, we rendered orthophotos that faithfully represent terrain and scene features. This procedure demonstrates that NeRF-derived volumetric representations can produce orthophotos without requiring dense point clouds or explicit 3D mesh generation. It also highlights the versatility of NeRFs in delivering both 3D (point clouds) and 2D (orthophotos) products for a range of geospatial applications.

4. Experiments, Results, and Discussions

Data. Since this work examines NeRFs through the lens of geospatial applications, we utilize the simulated dataset from Schulte et al. (2025), which was constructed using real-world LiDAR data captured from the Jamtal glacial valley combined with a mesh model to generate synthetic views. It encompasses 130 images of size 768 × 1024 at a ground sampling distance (GSD) of 3 cm.

Training. We adopt the NeRFacto model as it incorporates advances from multiple research efforts and is widely used within the NeRFStudio community. Our primary modification involves replacing the default pixel sampler with our proposed CENS strategy. Notably, Winiwarter et al. (2025) and Schulte et al. (2025) reported a vertical shift in NeRFacto-generated point clouds, which was attributed to the model’s camera optimization module. Accordingly, we conducted training both with and without camera optimization, obtaining a mean (μ), standard deviation (σ), and vertical shift (vs) values of [-0.103, 0.052, 0.111] (with camera optimization) and [-0.003, 0.027, 0.0] meters (without camera optimization), respectively. Hence, unless otherwise stated, all experiments in this section are performed without camera optimization.

Metrics. Following the evaluation protocol of Schulte et al. (2025), we employ the Cloud-to-Mesh (C2M) distance and completeness (Seitz et al., 2006) metrics to assess geometric accuracy of reconstructed PCs. For perceptual accuracy, we evaluated generated digital orthophotos on common measures for NeRFs. Additionally, we also analyze point count as a proxy for sample repetition.

4.1 Effects of training steps on NeRF-gen PC accuracy

We trained the NeRFacto model using both random and CENS sampling strategies for 44928, 22464, and 11232 steps, corresponding to gf values of 2.0, 1.0, and 0.5, respectively. A $gf \leq 1$ indicates that $(gf \times 100)\%$ of all pixels have been visited during training, while $gf > 1$ implies all pixels have been visited at least once. For each training configuration, we generated 100 million points, cropped the reconstructed scene to its core area (Fig. 4), and performed cleaning (i.e., statistical outliers removal, artifacts removal, and elimination of duplicate points using a minimum spacing threshold of 1 mm.) PC geometric accuracy was then evaluated using the C2M metric. The results, summarized in Fig. 5, show that CENS generally outperforms or matches random sampling across all training step configurations. Notably, even with only half the $grid\_map$ $(gf = 0.5)$, CENS achieved superior point coverage and a significant accuracy boost compared to random sampling. This is evidenced by poorly reconstructed areas (highlighted by red ellipses in the image), as well as improved μ and σ values in Fig. 5. As discussed in Section 3.1, pixel repetition cannot be entirely avoided due to image overlap. Nevertheless, CENS on average retains ∼24 million more points (a 28.2% increase) compared to random sampling, owing to its spatial non-redundancy. Overall, Fig. 5 demonstrates that CENS can achieve comparable or better accuracy with fewer training steps, underscoring its sampling efficiency.

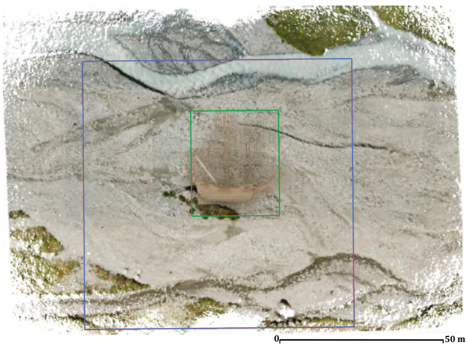


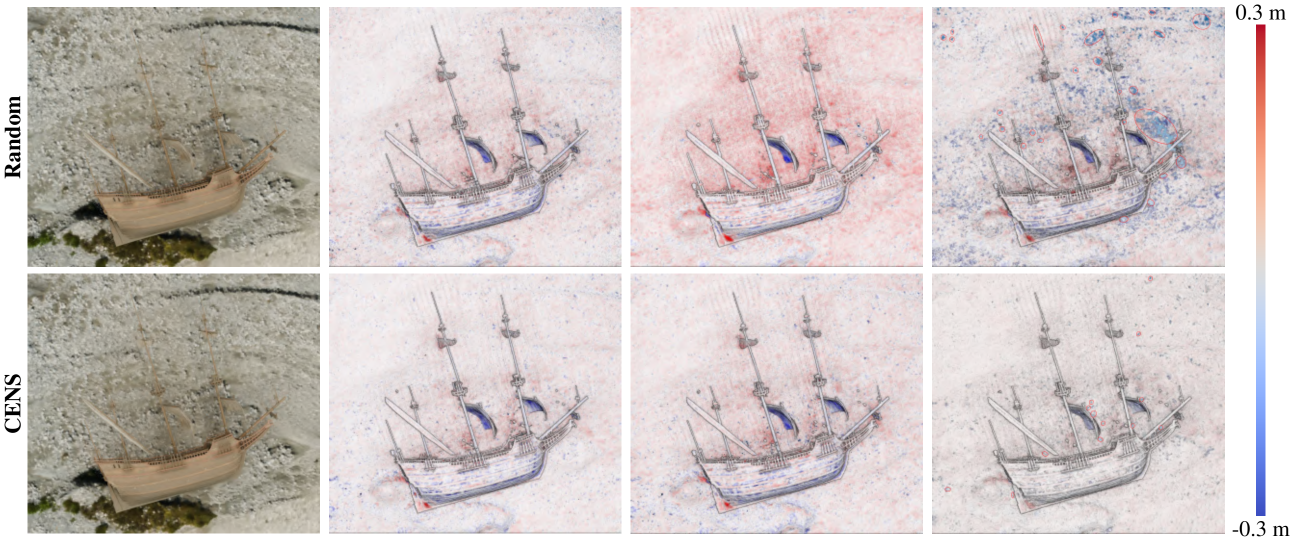
Figure 4. Jamtal valley study area. Blue box: core area of interest (75 × 75 m) centered on the ship. Green box: area (25 × 30 m) used for completeness quality assessment.

4.2 Spatial completeness of reconstructed PCs

To evaluate completeness of reconstructed PCs as a function of geometric fidelity, we used the best performing trained models (random and CENS) from Section 4.1 (i.e., $44928 \Leftrightarrow (gf = 2)$) to generate 50 million points (pre-cleaning). Considering in the core area the ship holds most of the details and complexity in the scene, we cropped the scene (both reference and reconstructed) to a 25 × 30 meter area centered on the ship. Given the reference PC $P_{gt} \in \mathbb{R}^{n\times 3}$ and generated PC $P_{rec} \in \mathbb{R}^{m\times 3}$, we estimate completeness in Eq. 1 as:

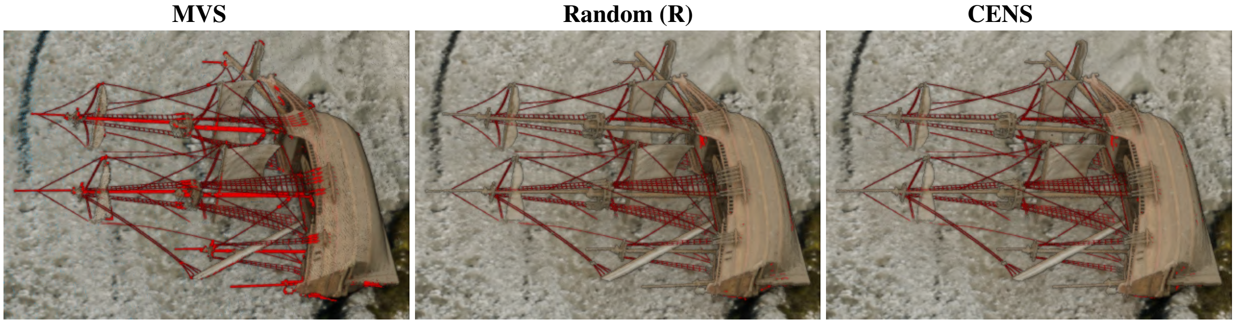
$$\forall i \in \{1,\dots,|P_{gt}|\},$$
$$\texttt{Comp} = \frac{1}{|P_{gt}|} \cdot \|\{P_{rec}|d(P^i_{gt},P_{rec}) \leq \tau\}\| \times 100, \qquad (1)$$

where $d(P^i_{gt},P_{rec}) = \min_{j\in P_{rec}} \|P^i_{gt} - P^j_{rec}\|$ is the cloud-to-cloud distance, and τ is the distance threshold (set to 0.1 meters) below which points contribute to completeness evaluation. We compare the NeRF-based methods to the classic MVS dense PC generation approach in Agisoft Metashape (Version 2.2.1) – a popular commercial software. Considering the limited magnitude of MVS’ reconstructed PC (2.44 million), the NeRF-based PCs were subsampled with a 5 cm spacing to produce comparable point counts (∼2.5 million) across methods to ensure a fair comparison. From Fig. 6 we observe that while the NeRF-based approaches were able to reconstruct fine details, with the exception of the thin ropes of the ship (which are subpixel structures), MVS was unable to reconstruct all of the ship’s masts and ropes resulting in 10% lower completeness. For the NeRF-based methods, CENS achieved a slightly higher completeness (∼0.6% i.e., about 87,878 points) compared to random sampling. This suggests that if given a bigger, more complex detailed scene, CENS’ non-redundancy and better spatial coverage may significantly contribute to superior completeness.



μ (m) `(R`$\mid$`CENS):`	$-0.0011 \mid -0.0027$	$-0.0014 \mid -0.0009$	$-0.0138 \mid 0.0030$
σ (m) `(R`$\mid$`CENS):`	$0.028 \mid 0.027$	$0.036 \mid 0.032$	$0.062 \mid 0.045$
`pre DP cnt(R`$\mid$`CENS):`	$85.346 \mid 85.739$	$85.419 \mid 85.839$	$85.159 \mid 85.729$
`post DP cnt(R`$\mid$`CENS):`	$50.654 \mid 74.388$	$50.704 \mid 74.489$	$50.300 \mid 74.452$
`steps = gf:`	$44928 = 2\times$	$22464 = 1\times$	$11232 = 0.5\times$

Figure 5. Comparison of random (R) (top row) and CENS (bottom row) strategies on geometric fidelity. From left to right: reconstructed rgb images and C2M (distance) values for different choices of grid sampling factors. For C2M distance, upper-bound (red) and lower-bound (blue) are $+0.3$ m and -0.3 m, respectively. `DP`: duplicate points removal. `cnt`: point count (in millions). `steps`: training steps. `gf`: grid sampling factor.



`Completeness(%):` 72.73 82.89 83.46

Figure 6. Completeness of reconstructed PCs across various methods. From left to right: Multi-view Stereo (MVS), NeRF with random sampling (R), and NeRF with CENS. Red points highlight areas where the reference scene was not reconstructed.

4.3 Perceptual accuracy on generated orthophotos

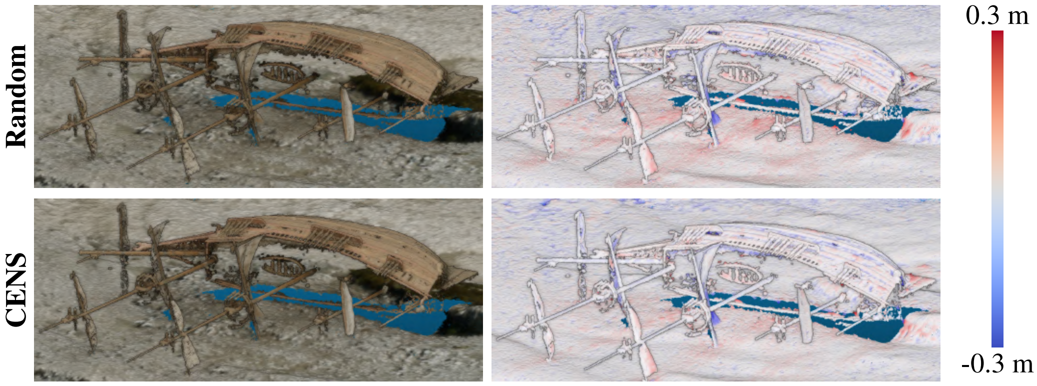
In Sections 4.1 and 4.2 we demonstrated the superior geometric fidelity of CENS-based NeRF PCs compared to classic methods. In geospatial applications, the perceptual quality of derived PC products – here, orthophotos – is equally as important. Considering the relatively poor geometric accuracy of MVS (Fig. 6), we evaluate perceptual quality only on orthophotos derived from NeRF-based PCs. Using the core 75×75 meter area of the scene, we generated orthophotos following section 3.3. These orthophotos were then compared against reference orthoimages rendered from the ground-truth mesh. The results, summarized in Table 1, indicate that CENS consistently outperforms random sampling across all metrics, achieving a notable 1.37 dB increase in PSNR, a 0.04 improvement in SSIM, and a 0.05 reduction in LPIPS. This suggests that CENS' enhanced spatial coverage and non-redundancy contribute to the generation of higher-quality orthophotos, which are crucial for various geospatial applications, e.g., urban planning and mapping, infrastructure inspection and asset management, archaeology and cultural heritage documentation, etc (Lv et al., 2024). Statistically, the lower standard deviations in CENS' results indicate consistency, reliability, and a capacity for better generalisation.

	PSNR(dB) (↑)	SSIM (↑)	LPIPS (↓)
Random	20.60 ± 2.56	0.67 ± 0.02	0.31 ± 0.07
CENS	21.97 ± 1.45	0.71 ± 0.02	0.26 ± 0.03

Table 1. Perceptual quality of orthophotos generated via random sampling and CENS based strategies. The ($\pm$) values indicate standard deviation (1σ).

4.4 NeRF-based simulated airborne laser scanning PCs

From Section 3.2, new nadir-only camera poses were computed outside the NeRF pipeline to emulate ALS acquisition geometry. For inference with both trained models, we used all pixels from all simulated camera poses thus obviating the need



	Random	CENS
μ (m)	-0.0056	-0.0079
σ (m)	0.0250	0.0230
V_{RMSE} (m)	0.0256	0.0240
H_{RMSE} (m)	0.0170	0.0160

Figure 7. Illustration of simulated ALS PCs from random (R) and CENS models. From left to right: simulated ALS colored PCs and their corresponding C2M distance maps. V_{RMSE} and H_{RMSE}: vertical and horizontal root mean square errors of simulated ALS PCs. All reported values are in meters.

Model	CENS		CENS+Canny	
Priority scheme	50% AP	25% AP	canny+50% AP	canny+25% AP
Total iterations (T)	11232	5612	14600	7300
PSNR(dB) (↑)	$\mathbf{34.30 \pm 1.18}$	31.8 ± 1.44	33.07 ± 4.13	32.45 ± 2.29
SSIM (↑)	0.94 ± 0.01	0.91 ± 0.01	$\mathbf{0.95 \pm 0.01}$	0.92 ± 0.01
LPIPS (↓)	0.06 ± 0.02	0.10 ± 0.02	$\mathbf{0.05 \pm 0.02}$	0.08 ± 0.02

Table 2. Evaluation of high-frequency (HF) area-guided CENS (CENS+canny), CENS without prioritized HF areas (CENS), and random sampling based on synthesized novel view images. The (±) values indicate standard deviation calculated over 13 test images. **Bold** values indicate the best performance in the 50% AP case, while gray values indicate the best results in the 25% AP case.

for any pixel sampling strategy and ensuring consistent inference samples across both models. Consequently, the inferred point clouds exhibit a common characteristic of ALS data, depending on the sensor: ground-only coverage with vertical structure/feature occlusions – a direct result of ALS's mostly nadir-pointing laser pulses, which have difficulty capturing side views or resolving hierarchical occlusions. Fig. 7 shows the results of the generated simulated PCs. Visual inspection indicates that these simulated point clouds successfully emulate the occlusion signatures usually observed in ALS PCs. Specifically, only the camera-facing surface (which in our case is top half of the hull) of the ship was reconstructed, while the bottom half of the hull, the ground beneath, and the inner structures of the ship were occluded. Additionally, parts of the sails were also poorly reconstructed.

Quantitatively, we also assessed the accuracy of the simulated PCs. According to UAV photogrammetry standards (Cho et al., 2025) (i.e., H_{RMSE} threshold: $0.5 - 2\times$ GSD; V_{RMSE} threshold: $1 - 3\times$ GSD. Note: our GSD = 3 cm), both NeRF outputs yielded RMSE values (Fig. 7) that fall within the thresholds established by ASPRS, Japan, the UK, Turkey, and South Korea. Simulating an ALS PC entails generating a PC that meets the accuracy standard of a captured ALS PC – which for example, according to Abdullah et al. (2024) requires $V_{RMSE} \leq 10$ mm. However, our NeRF-simulated PCs' V_{RMSE} values of $24.0 - 25.6$ mm do not meet this accuracy criterion, and this deviation cannot be compensated for by visual (qualitative) accuracy.

In addition, CENS' lower spread ($\sigma = 0.023$ m vs. 0.025 m) and coefficient of variation (2.92 vs. 4.05) indicate superior consistency; but this is insufficient to meet ALS simulation accuracy requirements. That said, we remain optimistic that with further research, NeRFs can simulate ALS PC that fall within standard error bounds.

4.5 Reproducibility of CENS vs. RS NeRF model output

In this section, we investigate the reproducibility of the proposed CENS-based NeRF pipeline relative to its random sampling counterpart. To this end, we performed five independent training and point-generation runs for each pipeline, using identical model configurations and training data. For every run, we computed the C2M distance errors between the ground-truth mesh and the generated point cloud. The results show that both RS and CENS produce near-unbiased point clouds (mean bias < 0.5 mm), indicating that neither model introduces a systematic elevation bias. However, CENS consistently achieves a smaller spread of C2M distances – approximately 11% lower than RS – demonstrating superior precision in geometry reconstruction. The very small variability of this spread across runs ($\sigma = 0.07$ mm) further confirms that CENS is more reproducible than random sampling with respect to the internal noise of the reconstruction process.

The bias variability is almost identical for both models, confirming that training-run randomness does not differentially affect systematic error. These quantitative metrics substantiate the claim that CENS is a more stable and consistent method for generating photogrammetric point clouds, even though both models produce satisfactory results.

4.6 Effect of prioritizing high-frequency (HF) pixels

High-textured regions typically have a lower spatial probability of being sampled under uniform random pixel selection,

Model	Random	CENS			CENS+Canny	
Priority scheme	None	100% AP	50% AP	25% AP	canny+50% AP	canny+25% AP
Total iterations (T)	Any	22464	11232	5612	14600	7300
Time per iteration (ms)	55.39 ± 1.23	61.46	60.22	60.42	128.25	124.32
Peak GPU memory (GB)	4.189	4.189	4.189	4.189	4.961	4.961

Table 3. Comparison of various model priority schemes on computational complexity using the experimental dataset.

as discussed in Section 2. In practice, however, these regions are often more informative for 3D reconstruction because they contain stronger geometric cues than low-textured areas. In this section, we examine how prioritizing high-frequency (HF) pixels during training affects reconstruction quality.

Given the 80/60 front/side overlap ratio used during data acquisition, a 50% active-pixel (AP) mask (e.g., a ‘1010’ or ‘0101’ pattern) provides nearly complete spatial coverage. Under such conditions, the need to explicitly prioritize HF regions is reduced. This expectation is confirmed by experiments as shown in Table 2, where the performance of 50% AP CENS and Canny+50% AP CENS is nearly identical.

Furthermore, as the percentage of active pixels (*gf*) decreases, so does reconstruction quality – relatively more in the case of random sampling, as shown in Fig. 5. To investigate whether HF prioritization becomes more beneficial under sparse sampling, we focus on the 25% AP case. Here, we train a NeRF model that uses Canny-derived edge-strength maps to guide pixel sampling during the first 30% of training iterations, followed by sampling guided by a 25% AP *grid map* (e.g., a ‘1000’ or ‘0010’ pattern).

As shown in Table 2, this enhanced version of the CENS strategy outperforms the CENS baseline for 2D novel view synthesis of UAV imagery. While PSNR differences are statistically inconclusive, both SSIM and LPIPS exhibit small but perceptually meaningful improvements in favour of the CENS+Canny model.

4.7 Computational complexity and scalability

We evaluate computational efficiency in terms of the average time per training iteration. Across all random sampling models trained in this study (irrespective of T), we observe an average iteration time of 55.39 ± 1.23 milliseconds. The time and memory costs of the CENS and random sampling methods are broadly comparable, whereas the CENS+Canny variant is more than twice as slow and exhibits slightly higher memory usage due to the required edge-guided computation.

Considering the results in Table 3, together with the superior qualitative and quantitative performance of CENS as shown in Fig. 5, we argue that CENS is the preferred strategy for scalable training—whether scaling to more images, higher resolutions, or both. The marginal gains offered by CENS+Canny do not justify its substantially higher computational overhead in our use-case.

This outcome can be attributed to the effective spatial-coverage property of CENS, which ensures that a sufficient number of high-frequency pixels are sampled to support high-quality reconstruction. As a result, the additional high-frequency emphasis introduced by CENS+Canny yields only limited improvements.

5. Conclusions

Several factors influence the quality of NeRF-generated point clouds, and consequently, the fidelity of derived geospatial data products and downstream analyses. Building on our initial argument – that the de-facto pixel sampler in NeRFs is repetitive and spatially unaware, often resulting in uneven and redundant PCs – we introduced a coverage-efficient, non-redundant pixel sampling strategy called CENS. CENS was designed to maximize spatial coverage, reduce point repetition, and promote reproducibility. Accordingly, this study examined key influencing factors such as pixel sample distribution, training iteration count, high-frequency pixels prioritization, and camera pose optimization.

Experimental results demonstrate that CENS meets its design objectives and achieves superior performance compared to the de-facto standard. Moreover, CENS is fully compatible with the widely adopted NeRFStudio framework.

We also explored the potential of NeRFs to simulate airborne laser scanning point clouds. While the generated PCs successfully replicated characteristic ALS occlusion patterns and met UAV photogrammetry accuracy thresholds, they did not satisfy the stricter ALS accuracy standards. Nonetheless, we are optimistic that with further research – such as incorporating depth supervision into the NeRF pipeline – accuracies closer to ALS-level could be achieved. Additionally, extending NeRFs to simulate point clouds for other platforms (e.g., mobile terrestrial systems) presents a promising future direction for comprehensive exploration.

References

Abdullah, Q., Munjy, R., Nimetz, J., Zoltek, M., Lee, C., 2024. Asprs positional accuracy standards for digital geospatial data, edition 2, version 2. Technical report, American Society for Photogrammetry and Remote Sensing (ASPRS), Bethesda, MD.

Akwensi, P. H., Bharadwaj, A., Wang, R., 2024. APC2Mesh: Bridging the gap from occluded building façades to full 3D models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 438-451. https://doi.org/10.1016/j.isprsjprs.2024.04.009.

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. *International Conference on Computer Vision (ICCV)*, 5835–5844. https://doi.org/10.1109/ICCV48922.2021.00580.

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., Hedman, P., 2022. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 5460–5469. https://doi.org/10.1109/CVPR52688.2022.00539.

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., Hedman, P., 2023. Zip-NeRF: Anti-aliased grid-based neural radiance fields. *International Conference on Computer Vision (ICCV)*, 19640–19648. https://doi.org/10.1109/ICCV51070.2023.01804.

Bello, J. L. G., Bui, M. V., Kim, M., 2024. ProNeRF: Learning Efficient Projection-Aware Ray Sampling for Fine-Grained Implicit Neural Radiance Fields. *IEEE Access*, 12, 56799–56814. https://doi.org/10.1109/ACCESS.2024.3390753.

Chen, S., Yan, Q., Qu, Y., Gao, W., Yang, J., Deng, F., 2025a. Ortho-NeRF: Generating a true digital orthophoto map using the neural radiance field from unmanned aerial vehicle images. *Geo spatial Inf. Sci.*, 28(2), 741–760. https://doi.org/10.1080/10095020.2023.2296014.

Chen, Y., Zhang, L., Zhao, S., Zhou, Y., 2025b. ATM-NeRF: Accelerating Training for NeRF Rendering on Mobile Devices via Geometric Regularization. *IEEE Trans. Multim.*, 27, 3279–3293. https://doi.org/10.1109/TMM.2025.3535288.

Cho, J., Jeong, S., Lee, B., 2025. Optimal ground control point layout for UAV photogrammetry in high precision 3D mapping. *Measurement*, 257, 118343. https://doi.org/10.1016/j.measurement.2025.118343.

Fukuda, K., Kurita, T., Aizawa, H., 2024. Important pixels sampling for NeRF training based on edge values and squared errors between the ground truth and the estimated colors. *Int. Joint Conf. Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*, 102–111. https://doi.org/10.5220/0012346200003660.

Ge, A., 2025. Importance sampling of views in NeRF. https://www.albertge.com/projects/nerf.pdf (Accessed 26 Sept. 2025).

Karras, G. E., Grammatikopoulos, L., Kalisperakis, I., Petsa, E., 2007. Generation of orthoimages and perspective views with automatic visibility checking and texture blending. *Photogrammetric Engineering & Remote Sensing*, 73(4), 403–411. https://doi.org/10.14358/PERS.73.4.403.

Kheradmand, S., Rebain, D., Sharma, G., Isack, H., Kar, A., Tagliasacchi, A., Yi, K. M., 2024. Accelerating neural field training via soft mining. *Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 20071–20080. https://doi.org/10.1109/CVPR52733.2024.01897.

Kurz, A., Neff, T., Lv, Z., Zollhöfer, M., Steinberger, M., 2022. AdaNeRF: Adaptive sampling for real-time rendering of neural radiance fields. *European Conference on Computer Vision (ECCV)*, 254–270. https://doi.org/10.1007/978–3–031–19790–1_16.

Lê, H.-Â., Guiotte, F., Pham, M.-T., Lefèvre, S., Corpetti, T., 2022. Learning Digital Terrain Models From Point Clouds: ALS2DTM Dataset and Rasterization-Based GAN. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 4980-4989. https://doi.org/10.1109/JSTARS.2022.3182030.

Lee, J., Jung, H., Park, J., Bae, I., Jeon, H., 2024. Geometry-aware projective mapping for unbounded neural radiance fields. *International Conference on Learning Representations (ICLR)*, 1–25. https://openreview.net/forum?id=w7BwaDHppp.

Levy, D., Peleg, A., Pearl, N., Rosenbaum, D., Akkaynak, D., Korman, S., Treibitz, T., 2023. SeaThru-NeRF: Neural radiance fields in scattering media. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 56–65. https://doi.org/10.1109/CVPR52729.2023.00014.

Liu, L., Gu, J., Lin, K. Z., Chua, T., Theobalt, C., 2020. Neural sparse voxel fields. *Advances in Neural Information Processing Systems (NeurIPS)*, 33, 15651–15663. https://proceedings.neurips.cc/paper_files/paper/2020/file/b4b75–8962f17808746e9bb832a6fa4b8–Paper.pdf.

Lv, J., Jiang, G., Ding, W., Zhao, Z., 2024. Fast Digital Orthophoto Generation: A Comparative Study of Explicit and Implicit Methods. *Remote Sensing*, 16(5), 786. https://doi.org/10.3390/rs16050786.

Mi, Z., Xu, D., 2023. Switch-NeRF: Learning scene decomposition with mixture of experts for large-scale neural radiance fields. *International Conference on Learning Representations (ICLR)*, 1–15. https://openreview.net/forum?id=PQ2zoIZqvm.

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. *European Conference on Computer Vision (ECCV)*, 12346, 405–421. https://doi.org/10.1007/978–3–030–58452–8_24.

Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4), 102:1–102:15. https://doi.org/10.1145/3528223.3530127.

Neff, T., Stadlbauer, P., Parger, M., Kurz, A., Mueller, J. H., Chaitanya, C. R. A., Kaplanyan, A., Steinberger, M., 2021. DONeRF: Towards Real-Time Rendering of Compact Neural Radiance Fields using Depth Oracle Networks. *Comput. Graph. Forum*, 40(4), 45–59. https://doi.org/10.1111/cgf.14340.

Otonari, T., Ikehata, S., Aizawa, K., 2022. Non-uniform sampling strategies for NeRF on 360° images. *British Machine Vision Conference 2022 (BMVC)*, 1–12. https://bmvc2022.mpi–inf.mpg.de/344 (Accessed 20 Sept. 2025).

Reiser, C., Peng, S., Liao, Y., Geiger, A., 2021. KiloNeRF: Speeding up neural radiance fields with thousands of tiny MLPs. *International Conference on Computer Vision (ICCV)*, 14315–14325. https://doi.org/10.1109/ICCV48922.2021.01407.

Schulte, F., Brezovsky, M., Günthner, A., Jutzi, B., Mandlburger, G., Winiwarter, L., 2025. Simulation and validation of underwater scenes for two-media optical 3D reconstruction. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, XLVIII-2/W10-2025, 271–278. https://doi.org/10.5194/isprs-archives-XLVIII-2-W10-2025-271-2025.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 519–528. https://doi.org/10.1109/CVPR.2006.19.

Tan, Y., Hao, J., Wu, J., Liu, L., Li, Q., Zhang, X., 2025. MCBlock: Boosting Neural Radiance Field Training Speed by MCTS-based Dynamic-Resolution Ray Sampling. *CoRR*, abs/2504.09878, 1-11. https://doi.org/10.48550/arXiv.2504.09878.

Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B., Srinivasan, P. P., Barron, J. T., Kretzschmar, H., 2022. Block-NeRF: Scalable large scene neural view synthesis. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 8238–8248. https://doi.org/10.1109/CVPR52688.2022.00807.

Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., Ng, R., 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems (NeurIPS)*, 7537 – 7547.

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., McAllister, D., Kerr, J., Kanazawa, A., 2023. NeRFStudio: A modular framework for neural radiance field development. *ACM SIGGRAPH 2023 Conference Proceedings*, 72:1–72:12. https://doi.org/10.1145/3588432.3591516.

Tang, X., Yang, M., Sun, P., Li, H., Dai, Y., Zhu, F., Lee, H., 2024. PaReNeRF: Toward fast large-scale dynamic NeRF with patch-based reference. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 5428–5438. https://doi.org/10.1109/CVPR52733.2024.00519.

Tu, P., Zhou, X., Wang, M., Yang, X., Peng, B., Chen, P., Su, X., Huang, Y., Zheng, Y., Xu, C., 2024. NeRF2Points: Large-Scale Point Cloud Generation From Street Views' Radiance Field Optimization. *CoRR*, abs/2404.04875, 1-18. https://doi.org/10.48550/arXiv.2404.04875.

Turki, H., Ramanan, D., Satyanarayanan, M., 2022. Mega-NeRF: Scalable construction of large-scale NeRFs for virtual fly-throughs. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 12912–12921. https://doi.org/10.1109/CVPR52688.2022.01258.

Wang, Q., Yan, L., Sun, Y., Cui, X., Mortimer, H., Li, Y., 2018. True orthophoto generation using line segment matches. *The Photogrammetric Record*, 33(161), 113-130. https://doi.org/10.1111/phor.12229.

Wang, Y., Gong, Y., Zeng, Y., 2024. Hyb-NeRF: A multiresolution hybrid encoding for neural radiance fields. *Winter Conference on Applications of Computer Vision (WACV)*, 3677–3686. https://doi.org/10.1109/WACV57701.2024.00365.

Wei, J., Zhu, G., Chen, X., 2024. NeRF-Based Large-Scale Urban True Digital Orthophoto Map Generation Method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18, 1070-1084. https://doi.org/10.1109/JSTARS.2024.3491869.

Winiwarter, L., Anders, K., Höfle, B., 2021. M3C2-EP: Pushing the limits of 3D topographic point cloud change detection by error propagation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 178, 240-258. https://doi.org/10.1016/j.isprsjprs.2021.06.011.

Winiwarter, L., Schulte, F., Wang, J., Zhang, Q., Anders, K., Jutzi, B., 2025. Assessing the potential of neural radiance fields and Gaussian splatting for change detection and change quantification. *6th Joint International Symposium on Deformation Monitoring (JISDM 2025)*, 1–8. https://doi.org/10.5445/IR/1000180533.

Wu, J., Liu, L., Tan, Y., Jia, Q., Zhang, H., Zhang, X., 2023. ActRay: Online active ray sampling for radiance fields. *SIGGRAPH Asia 2023*, 97:1–97:10. https://doi.org/10.1145/3610548.3618254.

Wu, Y., Lee, J. Y., Zou, C., Wang, S., Hoiem, D., 2025. MonoPatchNeRF: Improving neural radiance fields with patch-based monocular guidance. *International Conference on 3D Vision (3DV)*, 480–489. https://doi.org/10.1109/3DV66043.2025.00049.

Yu, S., Sun, X., Li, W., Xu, Q., Yuan, Z., Wang, S., She, R., Wang, C., 2025. STGC-NeRF: Spatial-temporal geometric consistency for lidar neural radiance fields in dynamic scenes. *AAAI Conference on Artificial Intelligence*, 9644–9652. https://doi.org/10.1609/aaai.v39i9.33045.

Yue, D., Liu, X., Wan, Y., Zhang, Y., Zheng, M., Fan, W., Zhong, J., 2025. NeRFOrtho: Orthographic Projection Images Generation based on Neural Radiance Fields. *International Journal of Applied Earth Observation and Geoinformation*, 136, 104378. https://doi.org/10.1016/j.jag.2025.104378.