ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XI-2-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission II, 4–11 July 2026, Toronto, Canada

SatGeo-NeRF: Geometrically Regularized NeRF for Satellite Imagery

Valentin Wagner[1], Sebastian Bullinger[1], Michael Arens[1], Rainer Stiefelhagen[2]

[1] Fraunhofer IOSB, Ettlingen, Germany - {*firstname.lastname*}@iosb.fraunhofer.de
[2] Karlsruhe Institute of Technology, Karlsruhe, Germany - rainer.stiefelhagen@kit.edu

Keywords: Neural Radiance Fields, Satellite Imagery, Geometrical Regularization, Gravity Alignment, Granularity Regularization

Abstract

We present *SatGeo-NeRF*, a geometrically regularized *NeRF* for satellite imagery that mitigates overfitting-induced geometric artifacts observed in current state-of-the-art models using three model-agnostic regularizers. *Gravity-Aligned Planarity Regularization* aligns depth-inferred, approximated surface normals with the gravity axis to promote local planarity, coupling adjacent rays via a corresponding surface approximation to facilitate cross-ray gradient flow. *Granularity Regularization* enforces a progressive coarse-to-fine geometry-learning scheme, and *Depth-Supervised Regularization* stabilizes early training using sparse depth cues for improved geometric accuracy. On the DFC2019 satellite reconstruction benchmark, *SatGeo-NeRF* improves the Mean Altitude Error by 14.0% and 11.4% relative to state-of-the-art baselines such as *EO-NeRF* and *EO-GS*.

1. Introduction

In recent years, the number of earth observation satellites featuring high-resolution camera systems has increased drastically. While 3D information from satellite data is highly impactful for urban, environmental, and disaster domains, reliable reconstruction remains an open research question because the inverse problem is fundamentally ambiguous, hindering a direct mapping from 2D projections to 3D structure.

Satellite images pose domain-specific challenges including a) specialized camera models due to the vast distances involved, b) images captured over multiple satellite passes, resulting in variable shadow and lighting conditions, and c) temporary objects such as vehicles moving in between image captures.

Popular photogrammetry approaches such as de Franchis et al. (2014), Beyer et al. (2018) and Zhang et al. (2019) focus on image-based feature matching to extract explicit geometry representations such as point clouds. Recent works (Derksen and Izzo, 2021; Marí et al., 2022, 2023; Behari et al., 2024) reapproach the problem as a novel-view synthesis task using adaptations of *Neural Radiance Fields* (*NeRF*) (Mildenhall et al., 2020). *NeRF* reconstruct images on a per-pixel basis from multiple viewpoints, learning an implicit geometrical scene understanding as byproduct. The learned geometry is extracted into a *Digital Surface Model* (*DSM*) to quantitative evaluate the altitude error of the derived results.

NeRFs build upon the multi-view consistency assumption, effectively expecting the appearance of the geometry to be consistent across views. Thus, *NeRFs* optimize a pure photometric consistency at pixel level and impose no explicit geometric constraints on the scene. This view-consistency assumption breaks under the high variability of multi-temporal satellite data. Therefore, previous works model variable elements such as lighting as part of the rendering process (Derksen and Izzo, 2021; Marí et al., 2023; Behari et al., 2024) and introduce uncertainty to handle transient objects (Marí et al., 2022). The geometry is hereby still unconstrained.

We observe that *NeRFs* tackle cross-view inconsistencies by subtly warping geometry of nominally flat regions, producing wave-like geometric artifacts. Such effects are evident in fig. 1; the state-of-the-art *EO-NeRF* depth map shows artificial structures on walkways and flat roof surfaces. In this work, we therefore focus on geometrical regularization based loss functions to mitigate the artifacts caused by geometrical overfitting.

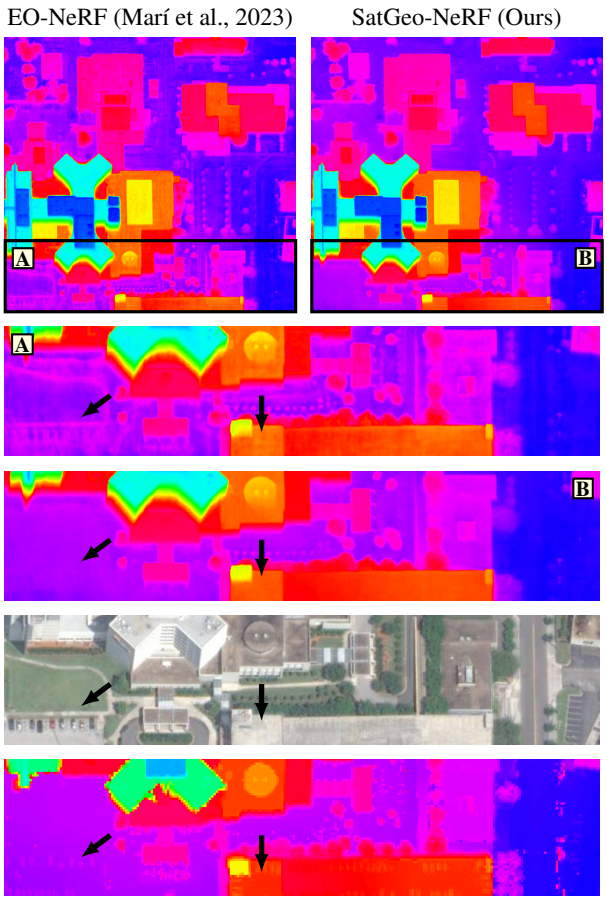


Figure 1. Rendered depth estimations (top), closeups (middle), RGB and Lidar DSM ground truth (bottom). Our geometry regularization technique reduces high-frequency artifacts caused by overfitting to individual training images.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-367-2026 | © Author(s) 2026. CC BY 4.0 License.

Urban scenes naturally contain large nominally flat areas such as streets, parking lots or roof structures. Because these surfaces are expected to be perpendicular to gravity, we propose aligning approximated local surface planes with the axis of gravity to provide a physically grounded prior that suppresses wave-like geometric artifacts. We apply the regularizer scene-wide and rely on the dominant photometric objective to preserve detail and avoid over-smoothing.

1.1 Contributions

- We propose *SatGeo-NeRF*, a geometrically regularized *NeRF* for satellite imagery, featuring three model-agnostic regularization techniques to mitigate overfitting-induced geometric artifacts commonly observed in current state-of-the-art models.
- The first model-agnostic regularization technique, *Gravity-Aligned Planarity Regularization*, provides a physically grounded prior by aligning depth-inferred, approximated surface normals with the axis of gravity, suppressing wave-like geometric artifacts. This regularization approach connects adjacent rays via a corresponding surface approximation, facilitating gradient flow across the participating rays.
- The second model-agnostic regularization technique, *Granularity Regularization* for satellite-domain *NeRFs*, enforces a coarse-to-fine geometry-learning scheme. We also demonstrate the continued benefit of the third model-agnostic regularizer, *Depth-Supervised Regularization*, which regularizes depth during the initial training stage using sparse 3D points.
- *SatGeo-NeRF* achieves state-of-the-art results on the DFC-2019 benchmark scenes, improving the MAE by 14.0% percent and 11.4% relative to the previous state-of-the-art.

2. Related Works

The flexibility of the learned, differentiable rendering *NeRFs* provide has shown to be beneficial in handling many of the challenges of multi-date satellite data. Notable contributions include *Shadow-NeRF* (Derksen and Izzo, 2021) as one of the first adaptations of *NeRF* to the satellite domain, mainly proposing to render shadows as their own lighting component based on the solar position. *Sat-NeRF* (Marí et al., 2022) introduces transient uncertainty and utilizing the satellite-domain-specific *Rational-Polynomial-Camera* (*RPC*) models. *EO-NeRF* (Marí et al., 2023) and *SUNDIAL* (Behari et al., 2024) both improve upon shadow rendering through simulating solar shading by casting additional rays. *Season-NeRF* (Gableman and Kak, 2023) allow rendering images across seasonal appearance changes.

Other works propose expanding the adaptable *NeRF* rendering mechanism to other modalities. Pic et al. (2024) combine high resolution panchromatic data with lower resolution color information. *Semantic-Sat-NeRF* (Wagner et al., 2025) integrate semantic information into the model, decreasing rendering artifacts caused by moving objects such as vehicles.

EO-GS (Aira et al., 2024) adapt *Gaussian Splatting* (Kerbl et al., 2023) to the satellite domain by approximating the satellite-domain specific cameras as affine projections and encode the scene using a sparse set of 3D-Gaussians.

Geometric regularization within *NeRFs* has so far been explored predominantly in few-shot scenarios. Niemeyer et al. (2021) enforce depth consistency for image patches from novel viewpoints, and Ehret et al. (2024) extend this idea to pixel-wise depth constraints. Seo et al. (2023) regularize the geometry along orthogonal rays to remove floating artifacts. Yang et al. (2023) propose a coarse-to-fine strategy by gradually increasing the number of frequencies used in the input encoding during training. Guo et al. (2022) and Zhou et al. (2024) leverage semantic priors to impose localized planar constraints, whereas Popovic et al. (2023) propose semantic-free planar constraints tailored to indoor scenes by clustering explicit surface normals.

3. Foundations for Satellite-Specific NeRF

3.1 General NeRF Principles

NeRFs (Mildenhall et al., 2020) represent a static three-dimensional scene as a continuous volumetric function $\mathcal{F}_{NeRF}$ encoded with an *Multi-Layer Perceptron* (*MLP*) network.

$$\mathcal{F}_{NeRF} : (\mathbf{x}, \mathbf{d}) \mapsto (\sigma, \mathbf{c}) \tag{1}$$

For a given 3D scene coordinate $\mathbf{x} = (x, y, z)$ and viewing direction $\mathbf{d} = (d_x, d_y, d_z)$, the network predicts a color $\mathbf{c}$ and density σ.

To render images, a ray $\mathbf{r}(t) = \mathbf{o} + t \cdot \mathbf{d}$ is created for each image pixel. Here, $\mathbf{o}$ and $\mathbf{d}$ denote the origin and direction vectors. Each ray $\mathbf{r}$ is discretized into N 3D points $\mathbf{x}_i$ and used as input for $\mathcal{F}_{NeRF}$. The pixel color $\mathbf{c}(\mathbf{r})$ is calculated by aggregating the color values $\mathbf{c}_i$ predicted for each sampled position $\mathbf{x}_i$ along the ray $\mathbf{r}$.

$$\mathbf{c}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i \mathbf{c}_i \tag{2}$$

The contribution of each predicted color value $\mathbf{c}_i$ to the overall ray color $\mathbf{c}(\mathbf{r})$ is based on its opacity α_i and transmittance T_i.

$$\alpha_i = 1 - \exp(-\sigma_i \delta_i) \text{ and } T_i = \prod_{j=1}^{i-1} (1 - \alpha_j) \tag{3}$$

Both attributes depend on the predicted volume density σ_i representing the scene geometry. The opacity α_i hereby defines the visibility of the current sample $\mathbf{x}_i$ based on its density and the distance $\delta_i = t_i - t_{i-1}$ to the previous sample. The transmittance T_i is based on previous samples visibility and is used to handle occlusions.

Analog to aggregating the color values in eq. (2) the sample position t_i is accumulated along the ray to determine a depth value $\mathbf{d}(\mathbf{r})$ representing the distance of the pixel to the scene.

$$\mathbf{d}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i t_i \tag{4}$$

NeRFs minimize the $L2$-Loss between the rendered ray color $\mathbf{c}(\mathbf{r})$ and the ground-truth pixel color $\mathbf{c}_{GT}(\mathbf{r})$ for rays $\mathbf{r} \in \mathcal{R}$ randomly sampled from all training views, thereby enforcing multi-view consistency.

$$L_{color}(\mathcal{R}) = \sum_{\mathbf{r} \in \mathcal{R}} ||\mathbf{c}(\mathbf{r}) - \mathbf{c}_{GT}(\mathbf{r})||_2^2 \tag{5}$$

3.2 Satellite-Domain Adapted NeRF

To handle the domain-specific requirements related to lighting, transient objects and camera model accuracy *EO-NeRF* (Marí

et al., 2023) proposes an extended volumetric *NeRF* function:

$$\mathcal{F}_{EONeRF} : (\mathbf{x}, \boldsymbol{\omega}, \mathbf{t}_j) \mapsto (\sigma, \mathbf{c}_a, \mathbf{a}, \beta, \tau, \mathbf{A}_j, \mathbf{b}_j, \mathbf{q}_j) \quad (6)$$

The inputs are the 3D scene coordinate $\mathbf{x}$, the sun direction $\boldsymbol{\omega}$ and an image-specific embedding vector $\mathbf{t}_j$ (where j is the image index). Whereas the geometrical density σ is unchanged, the color prediction is decomposed into multiple components: the albedo color $\mathbf{c}_a$ and the ambient color $\mathbf{a}$. Additionally, $\mathbf{A}_j$ and $\mathbf{b}_j$ describe an affine color transformation between the predicted albedo color and the current image j. To handle transient, non-stationary objects such as vehicles an uncertainty β and transient scalar τ are predicted using the image-specific embedding $\mathbf{t}_j$. Finally, to increase camera accuracy the network predicts a 2D coordinate adjustment component $\mathbf{q}_j$ for each image j.

3.3 Satellite-Domain-Specific Ray Generation

Analog to recent works such as Marí et al. (2022), Marí et al. (2023), Behari et al. (2024) and Wagner et al. (2025) the rays for each pixel are created using the satellite-domain specific *Rational-Polynomial-Camera* (*RPC*) (Tao and Hu, 2001) model. As it is typically used for georegistration of satellite images, the *RPC* model is directly extracted from the satellite image metadata. Each pixel is projected into the scene using the *RPC* model based on two predefined scene height boundaries $[h_{max}, h_{min}]$. These points are used as starting and end point of the ray, respectively. This limits the space covered by rays to closely align with the actual scene content and minimizes the sampling of empty areas.

Based on Marí et al. (2023) the projected points are converted into the *Universal Transverse Mercator* (*UTM*) coordinate system. *UTM* divides the earth into zones, projecting each zone onto a planar surface. This approximates the earths ellipsoid shape locally using a flat ground plane. The scene content is therefore aligned with the ground plane, which is defined by the up-vector, i.e. the unit vector along the z-axis

To increase accuracy, a *Bundle Adjustment* preprocessing step is performed across the *RPC* models following the approach in Marí et al. (2021). Additionally, a learned camera adjustment $\mathbf{q}_j$ embedding is used. This parameter defines a shift on the XY-axis learned per-image j, allowing the network to adjust camera-accuracy during training.

4. Geometrical Regularization for Satellite Data

In this section, we propose *SatGeo-NeRF*, a *NeRF* model for satellite images based on three geometric regularization principles: *Gravity-Aligned Planarity*, *Geometrical Granularity*, and *Depth-Supervised*. The regularizers are model-agnostic and are transferable to any satellite-domain–adapted *NeRF*. We realize them on *EO-NeRF* as described in section 3.2, with the interaction shown in fig. 3.

4.1 Gravity-Aligned Planarity Regularization

Although photometric supervision alone suffices to train *NeRFs*, it can induce geometry overfitting in single views, producing wave-like artifacts on nominally flat surfaces. We address this with *Gravity-Aligned Planarity Regularization*, which enforces planarity on locally estimated surface planes using the axis of gravity as a prior. The term counterbalances the photometric loss by penalizing deviations from local planarity. Local surface orientation is estimated by computing normals from *NeRF*-predicted depths along adjacent rays, as illustrated in fig. 2. We apply the regularizer scene-wide, but optimization is dominated by the photometric term, preserving detail and preventing oversmoothing.

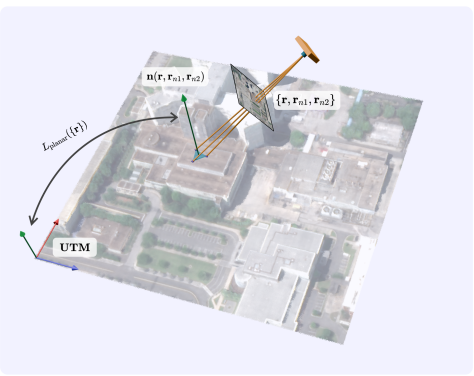


Figure 2. *Gravity-Aligned Planarity Regularization*: A local plane is estimated from three surface points on adjacent rays; its normal is constrained to align with the axis of gravity (approximated as *UTM*-up).

4.1.1 Explicit Surface Normals

For each ray $\mathbf{r}$ we determine two random adjacent rays $\mathbf{r}_{n_1}, \mathbf{r}_{n_2}$ from the four immediate neighboring pixels. To prevent unstable normal calculation, we make sure that we sample one neighbor from each of the two adjacent pixels in X- and Y-axis. All three rays are evaluated to determine their estimated depth values $d(\mathbf{r})$ as seen in eq. (4). The depth is converted into the three-dimensional surface point $\mathbf{p}_s(\mathbf{r})$ by following each ray along its direction: $\mathbf{p}_s(\mathbf{r}) = \mathbf{r}(d(\mathbf{r})) = \mathbf{o} + d(\mathbf{r})\mathbf{d}$. We approximate a local surface plane by determining the surface normal $\hat{\mathbf{n}}(\mathbf{r}, \mathbf{r}_{n_1}, \mathbf{r}_{n_2})$ based on the three surface points.

$$\hat{\mathbf{n}}(\mathbf{r}, \mathbf{r}_{n_1}, \mathbf{r}_{n_2}) = (\mathbf{p}_s(\mathbf{r}_{n_1}) - \mathbf{p}_s(\mathbf{r})) \times (\mathbf{p}_s(\mathbf{r}_{n_2}) - \mathbf{p}_s(\mathbf{r})) \quad (7)$$

To ensure correct length, we additionally normalize each calculated surface normal, denoted as $\mathbf{n}(\mathbf{r}, \mathbf{r}_{n_1}, \mathbf{r}_{n_2})$. To summarize, $\mathbf{n}$ describes a unit vector orthogonal to the plane defined by the three surface points using the set of rays $\mathbf{r}$, $\mathbf{r}_{n_1}$ and $\mathbf{r}_{n_2}$. With this process, we approximate local surface planes directly from the learned geometry, closely following the foundational *NeRF* rendering equations in eqs. (3) and (4).

4.1.2 Gravity Alignment

We align the calculated explicit surface normals along the approximated axis of gravity $\mathbf{g}$ with the regularization loss term L_{planar}.

$$L_{\text{planar}}(\mathcal{R}) = \sum_{\mathbf{r} \in \mathcal{R}} ||\mathbf{g} - \mathbf{n}(\mathbf{r}, \mathbf{r}_{n_1}, \mathbf{r}_{n_2})||_2 \quad (8)$$

To estimate the gravity axis $\mathbf{g}$, we leverage the known alignment of scene content in the *UTM* coordinate system. *UTM* places the ground in the XY-plane, so the Z-axis approximates the inverse gravity vector $\mathbf{g} = (0, 0, 1)^T$.

By explicitly computing surface normals with two auxiliary adjacent rays $\mathbf{r}_{n_1}, \mathbf{r}_{n_2}$, the *Gravity-Aligned Planarity* loss L_{planar} backpropagates through all samples along the three rays. Unlike standard *NeRF* training that treats rays independently, this design couples adjacent rays and imposes geometric regularization over an area.

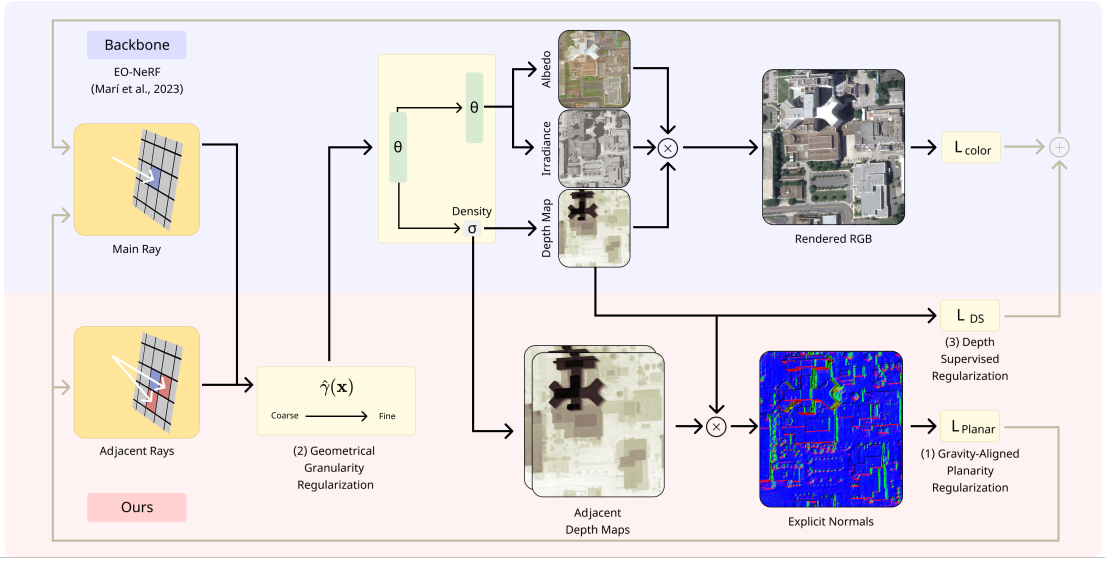


Figure 3. We augment a satellite-domain–adapted backbone (*EO-NeRF*, Marí et al. (2023)) with three model-agnostic regularizers: (1) *Gravity-Aligned Planarity Regularization*: estimate explicit normals by sampling two adjacent rays per training ray and evaluating depth only; align normals with the axis of gravity to mitigate overfitting and enable cross-ray smoothing. (2) *Geometrical Granularity Regularization*: coarse-to-fine masking of the input frequency encoding. (3) *Depth-Supervised Regularization*: depth constraints from a sparse 3D point cloud obtained from a *Bundle Adjustment* preprocessing step.

4.1.3 Minimizing Computational Cost Calculating explicit normals by inferencing two additional rays for each ray $\mathbf{r}$ increases computational cost by a factor of three if computed naively. We propose additional measures to reduce the computational cost required by the adjacent rays. A major reduction in computational cost is achieved by using only the partial network $\mathcal{F}_{Depth}$, predicting only the geometrical density σ based on the 3D coordinate input $\mathbf{x}$. Output heads related to color, uncertainty or transient prediction are skipped.

$$\mathcal{F}_{Depth} : (\mathbf{x}) \mapsto (\sigma) \tag{9}$$

We calculate the explicit normals based solely on the depth values, making any outputs related to color unnecessary. We therefore also do not inference the additional shadow rays $\mathbf{r}_{\text{sun}}$ required for the shadow component $\mathbf{s}(\mathbf{r})$. Additionally, we decrease the number of samples and reduce the length of the ray to minimize sampling empty space. Similar to the normal calculation, we assume that the selected points define a local (flat) approximation of the surface. We therefore derive the length of the adjacent rays to cover the immediate range surrounding the predicted depth $\mathbf{d}(\mathbf{r})$ of the main ray $\mathbf{r}$.

With the main ray $\mathbf{r}$ covering the length $t_l = t_{far} - t_{near}$, we define the partial ray length as $t_n = t_l \cdot p_n$ with $p_n \in [0, 1]$ as scaling factor. The partial adjacent rays $\mathbf{r}_n$ are then defined by shifting the near and far points based on the partial ray length t_n.

$$\mathbf{r}_n(t) = \mathbf{o}_n + t\mathbf{d} \text{ with } t \in [t_{near} + \tfrac{t_l - t_n}{2}, t_{far} - \tfrac{t_l - t_n}{2}] \tag{10}$$

To keep a similar spacing of the samples along the rays, the number of samples N_n is reduced to $N_n = N \cdot p_n$.

4.2 Geometrical Granularity Regularization

The general concept of *NeRFs* (Mildenhall et al., 2020) struggles to capture high-frequency color and geometric detail when operating directly on low-frequency sample coordinates $\mathbf{x}$. To address this, Mildenhall et al. (2020) apply a high-frequency positional encoding $\gamma(\mathbf{x})$ composed of sine and cosine functions at L increasing frequencies.

$$\begin{aligned}\boldsymbol{\gamma}(\mathbf{x}) = (&\sin(2^0\pi\mathbf{x}), \cos(2^0\pi\mathbf{x}), \dots, \\ &\sin(2^{L-1}\pi\mathbf{x}), \cos(2^{L-1}\pi\mathbf{x}))\end{aligned} \tag{11}$$

However, we observe that jointly learning coarse and fine scales induces geometric artifacts in satellite imagery We therefore introduce a *Geometrical Granularity Regularization* that masks encoding frequencies during training, based on Yang et al. (2023), thereby enforcing a coarse-to-fine schedule at negligible extra computational cost. Based on the current training iteration t, final regularization iteration T and maximum encoding frequency L, a bitmask $\boldsymbol{\alpha}(t, T, L)$ is calculated. We apply the bitmask $\boldsymbol{\alpha}$ to the input encoding $\boldsymbol{\gamma}$ to form the regularized input encoding $\boldsymbol{\gamma}'$:

$$\boldsymbol{\alpha}_i(t, T, L) = \begin{cases} 1, & i < 2(\frac{t \cdot L}{T} + b), \\ 0, & i \geq 2(\frac{t \cdot L}{T} + b). \end{cases} \tag{12}$$

$$\boldsymbol{\gamma}'(\mathbf{x}, t) = \boldsymbol{\gamma}(\mathbf{x}) \odot \boldsymbol{\alpha}(t, T, L) \tag{13}$$

In practice, this mask increases the visibility by one *sine* or *cosine* frequency each time training advances by a fixed amount, defined by the final regularization iteration T. The bias term b specifies the initial number of active frequency bands, providing sufficient low-frequency signals at the start.

Originally introduced by Yang et al. (2023) for few-shot scenarios, we employ a short regularization window T with a high initial bias b as a geometrical granularity regularization, reducing geometric error in satellite scenes with negligible overhead. Because the schedule is brief, we substitute gradual unmasking of input frequencies with a binary on/off threshold.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-367-2026 | © Author(s) 2026. CC BY 4.0 License.

4.3 Depth Supervised Regularization

To increase accuracy of the used camera model, *Bundle Adjustment* is performed across all *RPC*-cameras (Marí et al., 2021) of a given scene by minimizing multi-view reprojection error over image-based correspondences. As byproduct a set of sparse points is derived from image features. In contrast to recent work, we observe that guiding the network during the initial stage of training with this known sparse point cloud positively impacts performance.

We construct an additional set of rays $\mathcal{R}_{\mathrm{DS}}$ for the derived 3D points and its corresponding image pixel. The *Depth Supervised Regularization* loss $L_{\mathrm{DS}}(\mathcal{R}_{\mathrm{DS}})$ compares the predicted depth $d(\mathbf{r})$ with the depth based on the known 3D position $\mathbf{X}(\mathbf{r})$ and ray origin $\mathbf{o}(\mathbf{r})$, weighted with the reprojection error $w(\mathbf{r})$.

$$L_{\mathrm{DS}}(\mathcal{R}_{\mathrm{DS}}) = \sum_{\mathbf{r}\in\mathcal{R}_{\mathrm{DS}}} w(\mathbf{r}) \left(d(\mathbf{r}) - \|\mathbf{X}(\mathbf{r}) - \mathbf{o}(\mathbf{r})\|_2\right)^2 \quad (14)$$

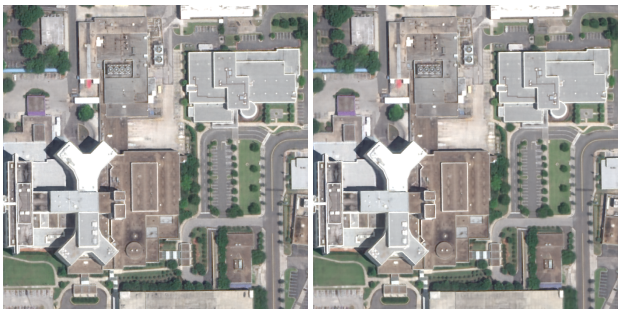
Originally proposed by Marí et al. (2022), expanding works such as Marí et al. (2023) and Behari et al. (2024) chose to forfeit this additional regularization, arguing that image-based loss terms are sufficient. However, similar to the *Gravity-Aligned Planarity Regularization* described in section 4.1.2, we instead argue that any known prior should be utilized to guide the network and increase the quality of the learned geometry. The positive effects of *Depth Supervised Regularization* are shown in table 2.

5. Implementation

We use *EO-NeRF* (Marí et al., 2023) as backbone, expanding the architecture with three additional regularization terms. At the time of our experiments, the official implementation of *EO-NeRF* was not publicly available, and we therefore reimplement the method. As Marí et al. (2023) describe only the high-level network structure, we base our concrete architecture on the publicly available model of Marí et al. (2022). We split the network into a main feature backbone and several specialized heads. The backbone additionally outputs the density σ, while the heads are small one-layer MLPs fed by an additional projection layer. The backbone consists of 8 layers with a width of 256 and heads use one layer with a width of 128. We employ the ReLU activation function between hidden layers, sigmoid for all color outputs, and softplus for β and σ prediction.

Following established satellite specific training strategies (Marí et al. (2022), Marí et al. (2023), Behari et al. (2024) and Wagner et al. (2025)), we train with the standard *NeRF* color loss from eq. (5) for the first two epochs and afterwards switch to the uncertainty-aware loss from Martin-Brualla et al. (2020). The *Gravity-Aligned Planar Regularization* L_{planar} is turned on at epoch three once a coarse scene representation is learned, and its weight is empirically chosen as $\lambda_{\mathrm{planar}} = 0.1$. The *Depth Supervised Regularization* L_{DS} is enabled for the first 25% of iterations, guiding the network in the initial learning stage with a weight of $\lambda_{\mathrm{DS}} = 1000$.

We set the positional and directional encoding frequency ranges to $L_p = 10$ and $L_d = 4$, following Mildenhall et al. (2020). We apply *Geometrical Granularity Regularization* to the positional encoding frequencies, using a starting bias $b = L_p/2$, which enables half of the available frequencies initially. We use a short regularization window $T = \lfloor 0.10 \cdot \textit{total\_iterations} \rfloor$, as proposed by Yang et al. (2023) for non–few-shot settings. Each ray is sampled with $N = 128$ samples.



(a) *EO-NeRF*: 29.20* (b) *SatGeo-NeRF* (Ours): 27.26

Figure 4. Rendered RGB images with their respective PSNR values. To human observers, the 6.6% relative PSNR difference is not perceptually noticeable under normal viewing.

All experiments are optimized over 300.000 iterations using an Adam optimizer with an initial learning rate of $5e-4$ and a batch size of 1024. A single training takes approximately 10 hours using an NVIDIA RTX 4090.

6. Experiments

In this section, we evaluate the impact of our proposed *SatGeo-NeRF* on reconstruction quality with respect to both learned geometry and image reproduction. Additionally, we provide an ablation study covering all proposed regularization terms. Finally, we demonstrate the reduction of computational costs for explicit surface estimation by evaluating only partial segments of adjacent rays.

6.1 Dataset

Analog to previous works we evaluate our model on a widely used subset of four scenes from the 2019 IEEE GRSS Data Fusion Contest (DFC2019) (Le Saux et al., 2019), comprising multi-date imagery (2014–2016) over Jacksonville, Florida. Each site includes 10–20 captures at 0.3 m ground sampling distance with varying off-nadir angles. The imagery is pan-sharpened from multispectral and panchromatic sources and pre-processed to 8-bit true-color RGB. Experiments are conducted on a $256 \times 256\,\mathrm{m}^2$ area aligned to the provided LiDAR Digital Surface Model (DSM). We perform an initial bundle adjustment of the *RPC* models of each scene; the optimized RPCs are used for all methods, including baselines.

6.2 Comparison with Baseline

We conduct a comparative evaluation against state-of-the-art models such as *EO-NeRF* (Marí et al., 2023) and *EO-GS* (Aira et al., 2024). At the time of experiments, the official *EO-NeRF* implementation was not publicly available. We therefore rely on our reimplementation (*EO-NeRF**). Our reproduced results do not reach the performance reported by Marí et al. (2023). This discrepancy in performance is in line with other reimplementations of *EO-NeRF* such as Behari et al. (2024).

To assess the quality of the reconstructed geometry, we report the *Mean Altitude Error* (MAE). For each training image, we convert predicted depth to a *Digital Surface Model* (DSM), align the model to a ground-truth LiDAR DSM, and compute the

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-367-2026 | © Author(s) 2026. CC BY 4.0 License.

Foliage Mask	Model	MAE [m] ↓ 004	068	214	260	Mean	PSNR ↑ 004	068	214	260	Mean
×	SatNeRF *(Marí et al., 2022)*	1.39	1.45	2.57	1.77	1.80	32.11	29.60	28.22	29.20	29.78
	EO-NeRF* *(Marí et al., 2023)*	1.57	1.23	2.43	1.63	1.72	29.20	27.33	26.51	27.28	27.58
	EO-GS *(Aira et al., 2024)*	1.45	1.25	2.30	1.69	1.67	33.55	28.71	26.16	28.78	29.30
	SatGeo-NeRF (Ours)	**1.30**	**1.08**	**2.13**	**1.39**	**1.48**	27.26	26.68	26.02	26.02	26.50
	Compared to EO-NeRF*	17.2% ↓	12.2% ↓	12.4% ↓	14.7% ↓	14.0% ↓	6.6% ↓	2.4% ↓	1.8% ↓	4.6% ↓	3.9% ↓
✓	SatNeRF *(Marí et al., 2022)*	1.02	1.43	2.62	1.73	1.70	31.76	29.42	28.12	28.94	29.56
	EO-NeRF* *(Marí et al., 2023)*	0.91	1.21	2.42	1.40	1.49	29.60	27.08	26.37	27.12	27.54
	EO-GS *(Aira et al., 2024)*	0.92	1.22	2.25	1.55	1.49	35.88	28.92	26.35	29.32	30.12
	SatGeo-NeRF (Ours)	**0.78**	**1.05**	**2.10**	**1.16**	**1.27**	27.28	26.42	25.88	25.70	26.32
	Compared to EO-NeRF*	14.3% ↓	13.2% ↓	13.2% ↓	17.1% ↓	14.8% ↓	7.8% ↓	2.4% ↓	1.9% ↓	5.2% ↓	4.4% ↓

Table 1. Evaluation of our proposed *SatGeo-NeRF* with state-of-the-art baselines. Our proposed regularization are able to improve the MAE by a mean of 14.0% on the DFC2019 dataset (Le Saux et al., 2019) with only minimal impact on image render quality.

Regularization: Granularity	Planarity	Depth	MAE [m] ↓ 004	068	214	260	Mean	PSNR ↑ 004	068	214	260	Mean
×	×	×	1.57	1.23	2.43	1.63	1.72	29.20	27.33	26.51	27.28	27.58
✓	×	×	1.39	1.16	2.20	1.44	1.55	28.60	26.98	26.22	26.54	27.09
✓	✓	×	1.72	**1.06**	**2.10**	1.56	1.61	28.72	27.18	26.37	26.93	27.30
✓	✓	✓	**1.30**	1.08	2.13	**1.39**	**1.48**	27.26	26.68	26.02	26.02	26.50

Table 2. Ablation study demonstrating improvements in reconstruction quality from our proposed regularization terms.

per-view altitude error. The final scene score represents to the mean across all training views. In combination with the wide range of off-nadir angles in the training data this yields a robust geometric evaluation. While *EO-GS* originally reports an altitude error for a single nadir view, we extend their evaluation to all training views for a fair comparison. Because PSNR is unreliable for novel views due to ambient illumination changes and transient effects, we report PSNR only on the input (training) views.

The quantitative results are presented in table 1. Across all four scenes our proposed regularization are able to decrease the MAE by a mean of 14.0% compared to the state-of-the-art *EO-NeRF* (Marí et al., 2023) method. As the network is not able to overoptimize the geometry as freely, the PSNR value for the training views drops slightly by 3.9%. As shown in fig. 4, this difference is not perceptually noticeable to human observers.

Whereas urban scene content remains relatively static across images, vegetation such as trees feature increased variance due to seasonal changes. We apply a foliage mask during evaluation to filter out all vegetation to evaluate the impact of our proposed regularization on urban scene content specifically. The semantic masks for the Lidar Ground-Truth are provided by *EO-GS* (Aira et al., 2024) and for the RGB images by *Semantic-SatNeRF* (Wagner et al., 2025). We are able to decrease the MAE by 14.8% for urban scene content compared to *EO-NeRF* (Marí et al., 2023).

Figure 5 shows that *NeRF*s systematically warp geometry to fit each view, as evidenced by misaligned surface normals. Our proposed *Planarity Regularization* improves surface normal quality by enforcing gravity alignment. Figure 6 demonstrates that our method produces more accurate scene geometry, effectively mitigating local geometrical artifacts introduced by overfitting.

6.3 Ablation Study

We present an ablation study evaluating the impact of our proposed regularization terms (*Geometrical Granularity*, *Gravity-Aligned Planarity* and *Depth Supervised*) on reconstruction quality in table 2. The results demonstrate that introducing individual geometric regularizations improve the MAE compared to the state-of-the-art baselines. Notably, the combination of all three regularization terms achieves the lowest mean MAE, indicating improved geometric accuracy overall. However, we observe that for certain scenes, such as scene 068, the combination of frequency and planar regularization yields the lowest MAE, while the full combination achieves the best result for other scenes such as 260. This suggests that while the regularizations are complementary and their combination provides the most robust performance across all scenes, specific combinations may be more effective for individual cases. PSNR values remain relatively stable across different configurations, indicating that improvements in geometry do not come at the expense of image reconstruction quality.

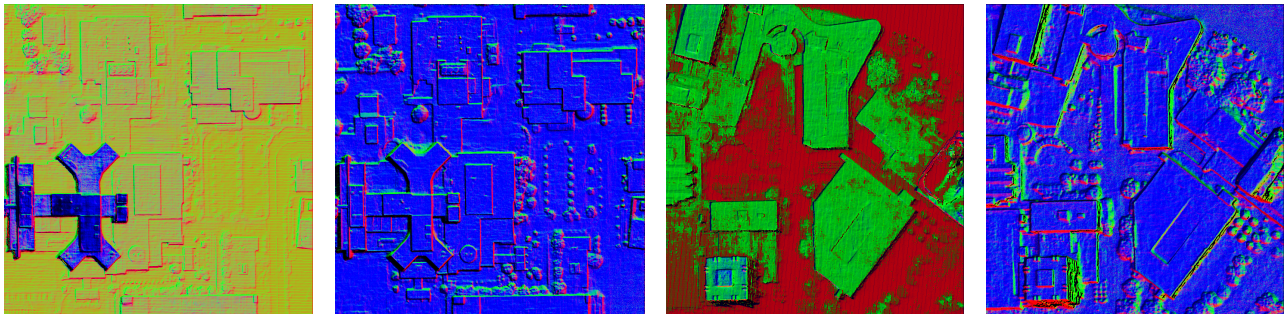
6.4 Minimizing Computational Cost

To reduce the computational impact of determining explicit surface normals, we propose reducing the evaluated length of the adjacent rays in section 4.1.3, therefore reducing the required samples which in turn lowers the computational cost. In table 3 we show that centering adjacent rays around the main surface depth with a length of 50% only has marginal impacts on reconstruction quality for most scenes, even decreasing the MAE for some such as 068.

Length	Samples	MAE [m] ↓ 004	068	214	260	Mean
100%	2×128	1.30	1.08	2.13	1.39	1.48
50%	2×64	1.37	1.05	2.11	1.47	1.50
25%	2×32	1.91	1.02	2.12	1.45	1.63

Table 3. Ablation study on adjacent ray length and sample density for explicit normal calculation. Centering adjacent rays around main surface point with a length of 50 % reduces computation with minimal impact on MAE for most scenes.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-367-2026 | © Author(s) 2026. CC BY 4.0 License.



(a) W/o Planarity Regularization (b) With Planarity Regularization (c) W/o Planarity Regularization (d) With Planarity Regularization

Figure 5. Explicit surface normal vectors determined based on predicted depth values of three adjacent rays; colors show normalized orientation (e.g., blue = upright). *Planarity Regularization* increases quality through alignment with the axis of gravity.

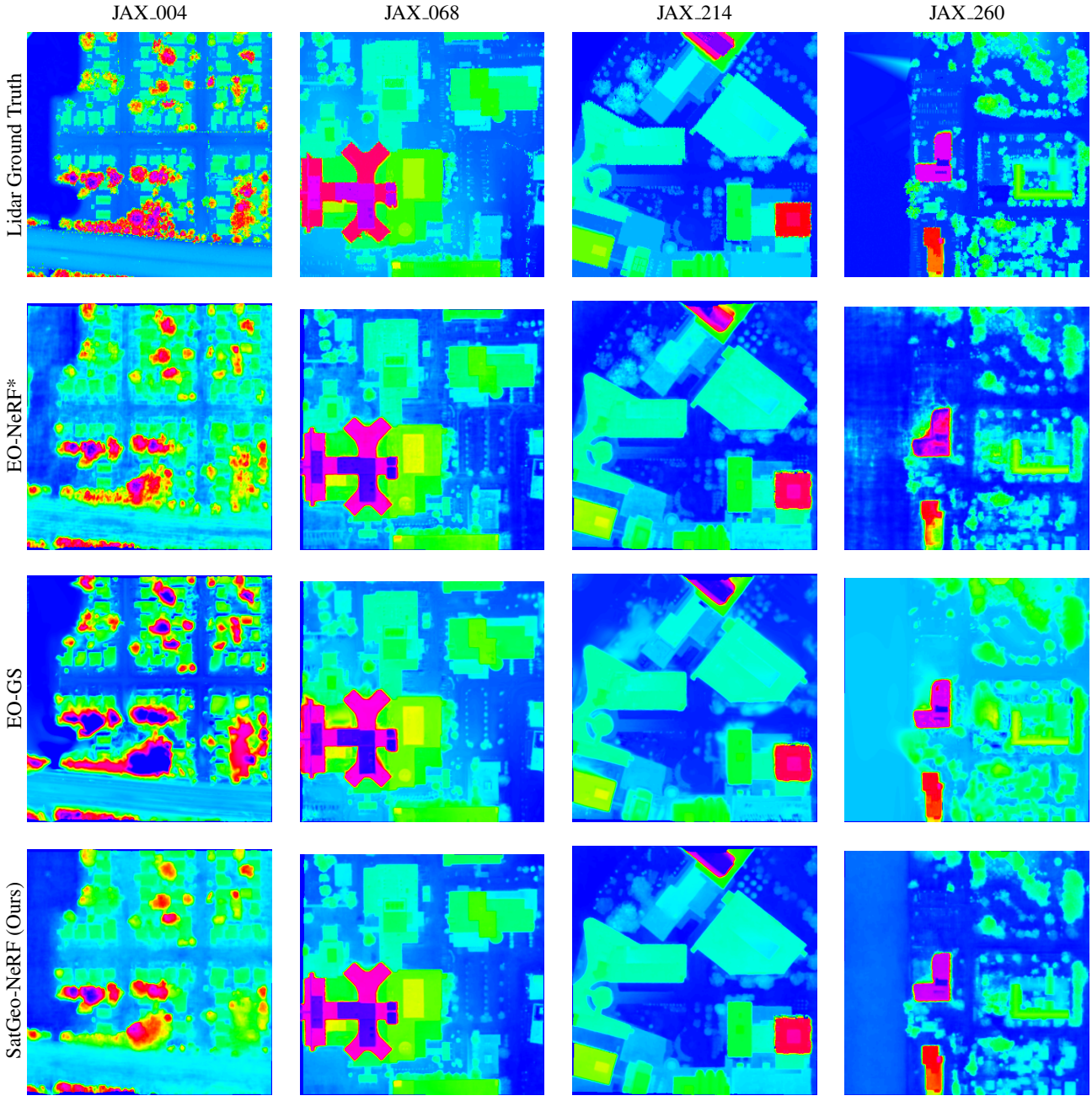


Figure 6. Extracted DSM of our proposed regularization in comparison to the Lidar Ground Truth, *EO-NeRF** (Marí et al., 2023) and *EO-GS* (Aira et al., 2024). For each scene the view with minimal MAE is shown. Colors map height from blue (low) to red (high).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-367-2026 | © Author(s) 2026. CC BY 4.0 License.

7. Conclusion

This paper introduces *SatGeo-NeRF* featuring three model-agnostic geometry regularization techniques: *Gravity-Aligned Planarity*, *Geometrical Granularity* and *Depth Supervised Regularization*. By aligning local surface approximations with the axis of gravity, we are able to provide geometrical guidance to the *NeRF*. The *Granularity Regularization* limits the available frequencies during training, forcing the network to learn the geometry in a coarse-to-fine manner. Lastly we reintroduce *Depth Supervised Regularization*, guiding the network during the initial training stage using a coarse 3D point cloud of the scene.

Our proposed regularization techniques are able to remove high-frequency geometric artifacts caused by overfitting, improving the MAE by a mean of 14.0% and 11.4% for the DFC2019 dataset compared to state-of-the-art baselines such as *EO-NeRF* and *EO-GS*.

References

Aira, L. S., Facciolo, G., Ehret, T., 2024. Gaussian Splatting for Efficient Satellite Image Photogrammetry. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5959-5969.

Behari, N., Dave, A., Tiwary, K., Yang, W., Raskar, R., 2024. Sundial: 3d satellite understanding through direct ambient and complex lighting decomposition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 522–532.

Beyer, R. A., Alexandrov, O., McMichael, S., 2018. The Ames Stereo Pipeline: NASA’s Open Source Software for Deriving and Processing Terrain Data. *Earth and Space Science*, 5(9), 537-548.

de Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., 2014. An automatic and modular stereo pipeline for pushbroom images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3, 49–56.

Derksen, D., Izzo, D., 2021. Shadow Neural Radiance Fields for Multi-view Satellite Photogrammetry. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1152-1161.

Ehret, T., Marí, R., Facciolo, G., 2024. A generic and flexible regularization framework for nerfs. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 3088–3097.

Gableman, M., Kak, A. C., 2023. Incorporating Season and Solar Specificity Into Renderings Made by a NeRF Architecture Using Satellite Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46, 4348-4365.

Guo, H., Peng, S., Lin, H., Wang, Q., Zhang, G., Bao, H., Zhou, X., 2022. Neural 3d scene reconstruction with the manhattan-world assumption. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 5501–5510.

Kerbl, B., Kopanas, G., Leimkuehler, T., Drettakis, G., 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics (TOG)*, 42, 1 - 14.

Le Saux, B., Yokoya, N., Hänsch, R., Brown, M., 2019. Data fusion contest 2019 (dfc2019).

Marí, R., Facciolo, G., Ehret, T., 2022. Sat-NeRF: Learning Multi-View Satellite Photogrammetry With Transient Objects and Shadow Modeling Using RPC Cameras. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1310-1320.

Marí, R., Facciolo, G., Ehret, T., 2023. Multi-date earth observation nerf: The detail is in the shadows. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2034–2044.

Martin-Brualla, R., Radwan, N., Sajjadi, M. S. M., Barron, J. T., Dosovitskiy, A., Duckworth, D., 2020. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7206-7215.

Marí, R., de Franchis, C., Meinhardt-Llopis, E., Anger, J., Facciolo, G., 2021. A Generic Bundle Adjustment Methodology for Indirect RPC Model Refinement of Satellite Imagery. *Image Processing On Line*, 11, 344–373.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2020. NeRF. *Communications of the ACM*, 65, 99 - 106.

Niemeyer, M., Barron, J. T., Mildenhall, B., Sajjadi, M. S. M., Geiger, A., Radwan, N., 2021. RegNeRF: Regularizing Neural Radiance Fields for View Synthesis from Sparse Inputs. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5470-5480.

Pic, E., Ehret, T., Facciolo, G., Marí, R., 2024. Pseudo pan-sharpening nerf for satellite image collections. *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 2650–2655.

Popovic, N., Paudel, D. P., Van Gool, L., 2023. Surface normal clustering for implicit representation of manhattan scenes. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, 17814–17824.

Seo, S., Chang, Y., Kwak, N. J., 2023. FlipNeRF: Flipped Reflection Rays for Few-shot Novel View Synthesis. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 22826-22836.

Tao, C. V., Hu, Y., 2001. A Comprehensive Study of the Rational Function Model for Photogrammetric Processing. *Photogrammetric Engineering and Remote Sensing*, 67, 1347-1357.

Wagner, V., Bullinger, S., Bodensteiner, C., Arens, M., 2025. Semantic neural radiance fields for multi-date satellite data. *Proceedings of the Winter Conference on Applications of Computer Vision (WACV) Workshops*, 1238–1246.

Yang, J., Pavone, M., Wang, Y., 2023. Freenerf: Improving few-shot neural rendering with free frequency regularization. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

Zhang, K., Sun, J., Snavely, N., 2019. Leveraging Vision Reconstruction Pipelines for Satellite Imagery. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2139-2148.

Zhou, X., Guo, H., Peng, S., Xiao, Y., Lin, H., Wang, Q., Zhang, G., Bao, H., 2024. Neural 3D Scene Reconstruction With Indoor Planar Priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46, 6355-6366.