International Journal of Geo-Information

MDPI

Article

Constraining the Geometry of NeRFs for Accurate DSM Generation from Multi-View Satellite Images

Qifeng Wan [1], Yuzheng Guan [1], Qiang Zhao [1], Xiang Wen [1] and Jiangfeng She [1,2,*]

[1] Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Key Laboratory for Land Satellite Remote Sensing Applications of Ministry of Natural Resources, School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China; qfwan@smail.nju.edu.cn (Q.W.); 502022270065@smail.nju.edu.cn (Y.G.); giser.zq@smail.nju.edu.cn (Q.Z.); wenxiang@smail.nju.edu.cn (X.W.)

[2] Jiangsu Center for Collaborative Innovation in Novel Software Technology and Industrialization, Nanjing University, Nanjing 210023, China

* Correspondence: gisjf@nju.edu.cn

Abstract: Neural Radiance Fields (NeRFs) are an emerging approach to 3D reconstruction that use neural networks to reconstruct scenes. However, its applications for multi-view satellite photogrammetry, which aim to reconstruct the Earth's surface, struggle to acquire accurate digital surface models (DSMs). To address this issue, a novel framework, Geometric Constrained Neural Radiance Field (GC-NeRF) tailored for multi-view satellite photogrammetry, is proposed. GC-NeRF achieves higher DSM accuracy from multi-view satellite images. The key point of this approach is a geometric loss term, which constrains the scene geometry by making the scene surface thinner. The geometric loss term alongside *z*-axis scene stretching and multi-view DSM fusion strategies greatly improve the accuracy of generated DSMs. During training, bundle-adjustment-refined satellite camera models are used to cast rays through the scene. To avoid the additional input of altitude bounds described in previous works, the sparse point cloud resulting from the bundle adjustment is converted to an occupancy grid to guide the ray sampling. Experiments on WorldView-3 images indicate GC-NeRF's superiority in accurate DSM generation from multi-view satellite images.

Keywords: neural radiance field; multi-view satellite images; geometric constraint; digital surface model

Citation: Wan, Q.; Guan, Y.; Zhao, Q.; Wen, X.; She, J. Constraining the Geometry of NeRFs for Accurate DSM Generation from Multi-View Satellite Images. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 243. https://doi.org/10.3390/ijgi13070243

Academic Editors: Wolfgang Kainz and Maria Antonia Brovelli

Received: 29 April 2024
Revised: 5 July 2024
Accepted: 5 July 2024
Published: 8 July 2024

1. Introduction

Orbiting high above the Earth, satellites efficiently gather geospatial information, providing valuable resources for numerous human activities [1–6]. Utilizing images captured by satellites, multi-view satellite photogrammetry can generate large-scale 3D models of urban areas represented as digital surface models (DSMs), which provide an essential data foundation for urban design, management, and evaluation [7–12] in the context of rapid worldwide urbanization [13], facilitating the comprehensive analysis of urban areas.

Multi-view satellite photogrammetry pipelines predominantly focus on multi-view stereo (MVS) methods [14–18], while some recent works integrating Neural Radiance Fields (NeRFs) [19] have been proven to achieve better results [20,21]. Using neural network and volume rendering [22] techniques, NeRFs synthesize new photorealistic images from 2D images and are an emerging approach to 3D reconstruction. Due to challenges presented by satellite camera characteristics, its applications for multi-view satellite photogrammetry are limited by its difficulty to acquiring accurate digital surface models (DSMs). For satellite scenes, the closely distributed viewpoints of satellite cameras pose challenges to accurately reconstruct fine geometry, which is the main cause of poor DSM accuracy. Moreover, during training and rendering, NeRFs use camera models to cast rays through the scene and project them onto known image pixels. However, satellites are far from Earth, so the large distance between the cameras and the scene poses challenges for ray sampling. Additionally, common inconsistencies in scene appearance, such as shadow movement

and radiometric variation among satellite images, can easily lead to a lack of convergence during the network training process. Some works [20,21,23] have explored NeRFs for multi-view satellite images, but the accuracy of these methods' DSM outputs must be improved due to a lack of geometric constraint in NeRFs. Moreover, extra input of the altitude bounds of the scene is required in these methods for ray sampling guidance. This extra input reduces their feasibility with the goal of obtaining a DSM that represents the altitude of the scene. Additionally, the network architecture used in these methods for solving inconsistency issues is too complex.

To address these issues, a novel framework, Geometric Constrained Neural Radiance Fields (GC-NeRFs), specifically designed for multi-view satellite photogrammetry, is proposed. The key difference between GC-NeRF and traditional NeRF approaches is a geometric loss term, which constrains the scene geometry by making the scene surface thinner, greatly improving the accuracy of generated DSMs. Moreover, z-axis scene stretching is conducted for finer reconstruction granularity in the z-direction, which plays a constructive role in improving DSM accuracy. In addition, a strategy for fusing multi-view DSMs reduces errors in obstructed areas. During training, bundle-adjustment-refined satellite camera models [20] are used to cast rays through the scene. An occupancy grid converted from sparse point clouds generated by bundle adjustment is used to guide ray sampling, which reduces dependence on altitude bounds. Additionally, GC-NeRF integrates several advanced techniques in NeRF variants [24,25] for a concise network architecture. Experiments with WorldView-3 images indicate that GC-NeRF achieves higher DSM accuracy while minimizing additional input requirements.

2. Related Work

To achieve accurate 3D reconstructions, multi-view satellite photogrammetry faces several challenges, including limited and closely distributed viewpoints leading to ambiguous elevation estimates and inconsistent illumination conditions affecting scene appearance. In this section, relevant research on multi-view stereo and NeRF-based methods is comprehensively reviewed, focusing on approaches aimed at improving reconstruction accuracy and addressing challenges specific to multi-view satellite photogrammetry.

2.1. Multi-View Stereo for Satellite Images

Multi-View Stereo (MVS) was first approached as an extension of stereo pair algorithms by aggregating information from multiple stereo pairs. As a result, true MVS algorithms were mainly developed to reconstruct objects from images photographed at a close distance while considering all the images in the scene [26].

MVS approaches are widely employed for scene reconstruction from aerial and satellite images. However, satellite images characteristic of an extremely small ratio between the depth range and distance from the camera to the scene and inconsistent illumination discourage the use of true MVS methods for satellite images [14]. In the case of satellite images, MVS has traditionally employed pairwise approaches, treating multiple views in pairs using traditional two-view stereo methods and subsequently combining pairwise reconstructions to obtain the final results [15,27]. These methods typically involve pair selection, stereo rectifying, dense stereo matching, triangulation, and depth fusion [28]. Semi-Global Matching (SGM) [29] is a popular choice for the stereo matching step, and its variants such as MGM [30], tSGM [31], and semi-global block matching [32] have been used to enhance efficiency and accuracy. However, these methods usually rely on manually selected stereo pairs and manually designed matching strategies, which may not ensure optimality [33]. Recently, deep learning methods such as GA-Net [34], PSM [35], and HSM [36] have been progressively applied to satellite stereo pipelines [14,37] with some progress. Unlike classical algorithms, deep learning methods are prone to failure or reduced accuracy when encountering unseen scenarios [21]. Moreover, learning-based methods often involve extensive training periods, lasting days or even weeks. Though popular, deep learning methods are still not the preferred option in satellite stereo pipelines, but rather

a step in classic MVS methods [14]. In general, MVS methods mainly focus on pairwise matching and do not fully exploit the benefits of multi-view data, leading to challenges in addressing seasonal and dynamic variations in satellite images [28].

2.2. Neural Radiance Field

Neural Radiance Fields (NeRFs) [19] represent static scenes through a continuous volumetric function F, learned as a fully-connected neural network. This function predicts the emitted RGB color $c_X = (r, g, b)$ and a non-negative scalar volume density σ_X at a 3D point $X = (x, y, z)$ from a given viewing direction $d_{view} = (\theta, \varphi)$:

$$F : (X, d_{view}) \rightarrow (c_X, \sigma_X) \tag{1}$$

Using a collection of input images and their camera poses, rays passing through the scene are projected onto the known pixels. Each ray r is defined by a point of origin o and the viewing direction; o is located at the camera's photography center, in general. To render the ray's color, r is discretized into N 3D points X_i, i.e., ray samples with $X_i = o + d_{view} t_i$, where t_i is the distance between o and X_i. The color $c(r)$ of a ray r is computed using volume rendering [22] as:

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

The rendered color $c(r)$ results from integrating the colors c_i predicted by F at different points of the ray r. The weight w_i of each point X_i in r to the rendered color depends on the opacity α_i and the transmittance T_i:

$$\alpha_i = 1 - \exp(-\delta_i \sigma_i) \tag{3}$$

$$T_i = \prod_{j=1}^{i-1} (1 - \alpha_j) \tag{4}$$

$$w_i = T_i \alpha_i \tag{5}$$

where δ_i is the distance between consecutive points along the ray. Additionally, the depth along a ray is rendered using a similar weighted integration approach:

$$depth(r) = \sum_{i=1}^{N} w_i t_i \tag{6}$$

NeRFs are trained by minimizing the image loss $Loss_c$, which is computed as the mean squared error (MSE) between the rendered color and the observed color of input images along the rays' projections:

$$Loss_c = \sum_{r \in R} \left\| c(r) - c_{gt}(r) \right\|_2^2 \tag{7}$$

where $c_{gt}(r)$ is the observed color, and R is the set of rays in each input batch. During training, F is gradually optimized and predicts accurate colors and densities at different 3D points in the scene.

Using the method introduced above, NeRFs have garnered significant attention for their ability to generate high-quality 3D reconstructions from 2D images. NeRFs' success inspires numerous extensions and adaptations to address various challenges and application domains [38–42]. Many works have improved NeRF rendering effects, such as optical effects [25,43–46], dynamic scenes [47–51], and realistic rendering [52,53]. Other studies focus on its application in multi-view satellite photogrammetry, described in detail in the next subsection.

2.3. NeRF Variants for Multi-View Satellite Photogrammetry

Recent efforts have explored NeRFs' potential for satellite imaging. S-NeRFs [23] pioneer NeRFs' application in multi-view satellite photogrammetry, leveraging sun direction d_{sun} for precise geometry and building shadow rendering. A complex remote sensing irradiance model is adopted in S-NeRF:

$$F: (X, d_{sun}) \rightarrow (a_X, \sigma_X, s_X, sky_X) \tag{8}$$

Physically, a_X represents albedo color, and s_X represents the ratio of incoming solar light with respect to the diffuse sky light sky_X. The color of X is computed as:

$$c_X = a_X s_X + a_X(1 - s_X)sky_X \tag{9}$$

Subsequent works typically follow the irradiance model of S-NeRF [20,21], though it is very complex. Building on S-NeRFs, Sat-NeRFs [20] incorporates transient object modeling similar to NeRF-W [54] and satellite-adapted camera representations, enhancing accuracy by integrating RPC camera models and applying bundle adjustment. EO-NeRFs [21] focus on shadow modeling to align building shadows with scene geometry, resulting in highly accurate and detailed DSM reconstruction. In addition, SatelliteNeRFs [55] directly apply NeRFs to satellite images and can extract mesh, but does not modify special features of satellite images.

Despite progress in simplifying satellite sensors' imaging process, the irradiance model is not entirely accurate. Moreover, these methods require additional input from altitude bounds for sampling guidance, reducing their feasibility. Their training speed is also very low. Various strategies have been adopted to accelerate NeRFs since their proposal in 2020, including optimizing sampling [56,57], scene decomposition [58,59], and combining explicit models [60,61]. Instant-NGP [24] adopts a different approach with a network architecture that uses a multi-resolution hash grid for position encoding and combines it with an occupancy grid to guide sampling, accelerated 3000 times while significantly reducing memory usage. Integrating these methods into multi-view satellite photogrammetry greatly improves efficiency. Table 1 shows the disadvantages of previous NeRF-based methods for multi-view satellite photogrammetry, which are solved with GC-NeRFs.

Table 1. Comparison of NeRF-based multi-view satellite photogrammetry approaches.

	Training Time	Appearance Embedding Module	Complex Irradiance Model	Bounds of Altitude
S-NeRFs [23]	8 h	**no need**	yes	need
Sat-NeRFs [20]	10 h	need	yes	need
EO-NeRFs [21]	15 h	need	yes	need
GC-NeRFs	**6 min**	**no need**	**no**	**no need**

The advantages of these approaches are shown in bold. An NVIDIA GPU with 12 GB RAM was used for the training times.

Overall, NeRFs present a promising solution to address these challenges in multi-view satellite photogrammetry. However, NeRFs are still immature at processing satellite images. Adapting NeRFs to multi-view satellite photogrammetry poses unique challenges posed by satellite camera characteristics. This paper aims to bridge these issues by proposing a novel approach, GC-NeRFs, specifically designed for multi-view satellite photogrammetry to improve reconstruction accuracy.

3. Methods

GC-NeRFs aim to enhance geometric reconstruction accuracy for satellite scenes and generate accurate DSMs. The overview of GC-NeRFs is shown in Figure 1. Its key contributions include z-axis scene stretching, an occupancy grid converted from sparse

point clouds, DSM fusion, and a geometric loss term for network training, which are shown in bold. After applying a bundle adjustment to the satellite images, refined camera models and a point cloud are obtained (Figure 1a,b). Then, the scene is stretched in the z-axis to enlarge the scale in the vertical direction (Figure 1c). Afterward, satellite camera models are used to cast rays through the scene, and an occupancy grid converted from the sparse point cloud generated by the bundle adjustment is used to guide sampling along the rays (Figure 1d). These sample points are input into the GC-NeRF network to query colors and densities. GC-NeRF is trained with the image loss and geometric loss terms proposed in Section 3.3 (Figure 1e). After optimization, GC-NeRF renders multi-view DSMs (Figure 1g) using volume rendering (Figure 1f). Additionally, the fusion of multi-view DSMs also contributes to accuracy improvement (Figure 1h).

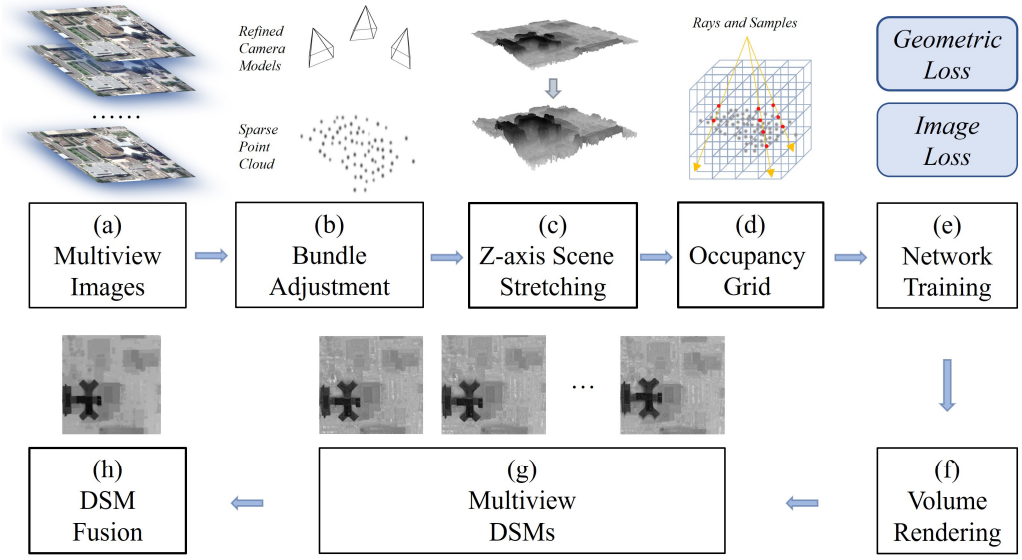


Figure 1. Overview of GC-NeRFs. GC-NeRFs use satellite images and corresponding satellite camera models to reconstruct scenes and generate accurate DSMs. Their key contributions include z-axis scene stretching, an occupancy grid converted from sparse point clouds, DSM fusion, and a geometric loss term for network training, which are shown in bold.

3.1. Z-axis Stretched Radiance Model

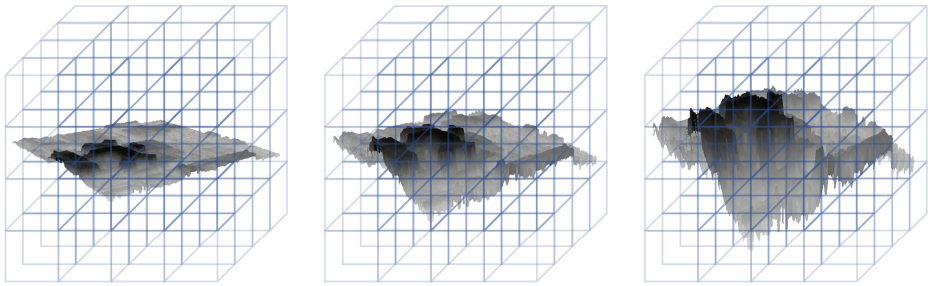
GC-NeRF represents the scene as a static surface. However, minimizing $Loss_c$ in Formula (7) can lead to a lack of convergence due to inconsistent appearances between satellite images. Therefore, the sun direction d_{sun} was used for appearance encoding, turning Formula (1) into:

$$F: (X, d_{view}, d_{sun}) \rightarrow (c_X, \sigma_X) \tag{10}$$

For appearance encoding, d_{sun} efficiently handles different weather conditions or seasonal variations in satellite images, whereas without d_{sun}, the experiments would fail. Compared to previous works [20], which used an extra appearance module, GC-NeRF's network architecture is very concise. Furthermore, GC-NeRF avoids using the complex remote sensing optical model of Formulas (8) and (9) proposed by S-NeRF [23] to ensure network simplicity.

Fast convergence is achieved by integrating multi-resolution hash encoding (MH) proposed in Instant-NGP [24]. MH divides the scene into multi-resolution voxel grids and uses hash tables to store optimizable space features at each grid cell vertex (Figure 2a). Each input 3D coordinate X is encoded into a 32-length vector.

Unlike typical scenes, satellite scenes are usually flat, with a significant difference in horizontal and vertical scales as their horizontal range is large and vertical range is small (Figure 2a).



(a) Original scene (b) Suitable stretched scene (c) Over stretched scene

Figure 2. Z-axis scene stretching. (**a**) The original satellite scene is flat. (**b**) The suitably stretched scene makes full use of multi-resolution hash encodings. (**c**) The over-stretched scene presents excessive hash conflicts.

This study aimed to improve the accuracy of output DSMs in the z-direction. To make the reconstruction granularity in the z-direction finer, z-axis scene stretching was conducted so that $X = (x, y, z_{sretched})$:

$$z_{sretched} = \frac{1}{s_z} z \tag{11}$$

where s_z denotes a stretching scale factor between 0 and 1. Stretching only on the z-axis makes the scene scale more coordinated and increases the sampling density in the z-direction, thus improving the utilization rate of the hash grid and DSM accuracy. However, over-stretching (Figure 2c) leads to excessive hash conflicts in the multi-resolution hash encodings, reducing model performance. Therefore, s_z should not be too small. Setting s_z to 0.80 optimizes model performance, as discussed in Section 5. Stretching the z-axis of the scene will deform the 3D model, resulting in corresponding squeezing when outputting DSMs.

The architecture of the GC-NeRF network is shown in Figure 3. The network receives 3D spatial coordinate X, sun direction d_{sun}, and viewing direction d_{view} as inputs to predict the volume density σ_X and color c_X at X. F_σ and F_c are small fully-connected networks. The former has only one hidden layer, whereas the latter has two hidden layers, with each hidden layer containing 64 neurons and activated by the ReLU function. Furthermore, the output σ_X is activated by an exponential function and the output c_X is activated by a sigmoid function. Spherical harmonic encodings [61] were used to encode the viewing direction and sun direction to a 16-length vector, respectively.

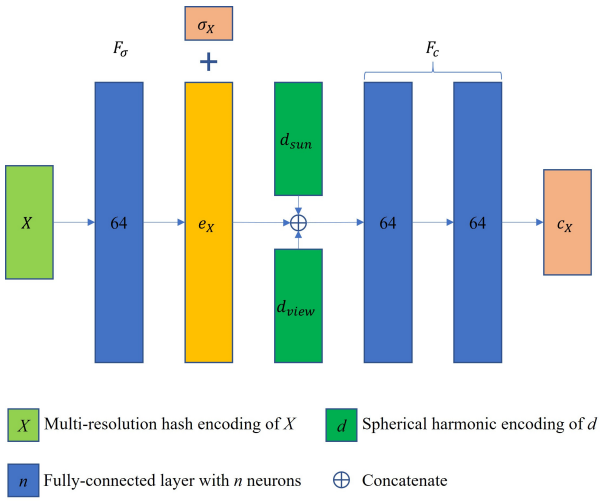


Figure 3. The architecture of the GC-NeRF network. The model receives 3D spatial coordinate X, sun direction d_{sun}, and viewing direction d_{view} as inputs to predict the volume density σ_X and color c_X at X.

3.2. Occupancy Grid Converted from Sparse Point Cloud

In this paper, satellite camera models are parameterized with rational polynomial coefficients. GC-NeRF is trained by casting rays from bundle-adjustment-refined satellite camera models [20,62] into known image pixels. To render a ray's color, NeRFs typically require multiple samples, which can be challenging for satellite images due to their vast distance from Earth. Using the same 128-point sampling approach as NeRFs would yield poor results. Previous methods (such as Sat-NeRFs [20], EO-NeRFs [21], S-NeRFs [23], and RS-NeRFs [63]) addressed this problem by sampling only within the scene's altitude bounds, which are difficult to acquire. In fact, obtaining the altitude of the scene is the goal of multi-view satellite photogrammetry.

When applying bundle adjustment, an extra point cloud indicates scene geometry generation. Therefore, this paper proposes a method of converting the sparse point cloud to an occupancy grid (Figure 4a,b) for ray sampling guidance, eliminating the need to input altitude bounds.

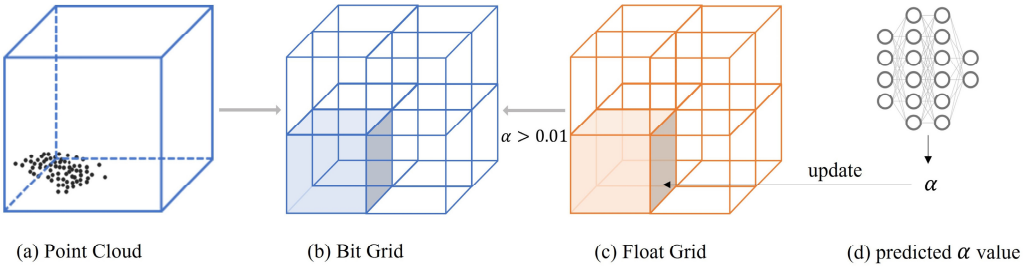


Figure 4. The converting and updating of the occupancy grid. (**a**) The point cloud is freely obtained from the bundle adjustment. (**a**,**b**) The point cloud is converted into a bit occupancy grid. (**b**,**c**) The bit grid cells are classified by the float grid cells. (**c**,**d**) The float grid cells are updated by the α value predicted from the network.

The occupancy grid divides the entire scene into 128^3 cells, with each cell storing a bit to indicate whether an object occupies the area (Figure 4b). Empty areas are skipped during ray sampling, reducing the number of samples to increase training and rendering efficiency. When converting point clouds to occupancy grids, the occupancy tensor is set to positive if a point falls within its cell (Figure 4a,b). However, due to sparsity and errors in the point cloud, training cannot solely depend on the initial occupancy grid. To mitigate this, volume density is re-evaluated every 16 training iterations to update the occupancy grid following previous works [24,64]. Cells in the occupancy grid are classified based on their counterparts, float grid cells, which store cell density as float values (Figure 4b,c). During updates, float grids decay old density by 0.95 and sample an alternative density value from a random point in the cell. In this way, the maximum of old and alternative densities is retained (Figure 4c,d). A classification threshold, opacity α, derived from density and step size using Formula (3), determines non-empty cells ($\alpha > 0.01$).

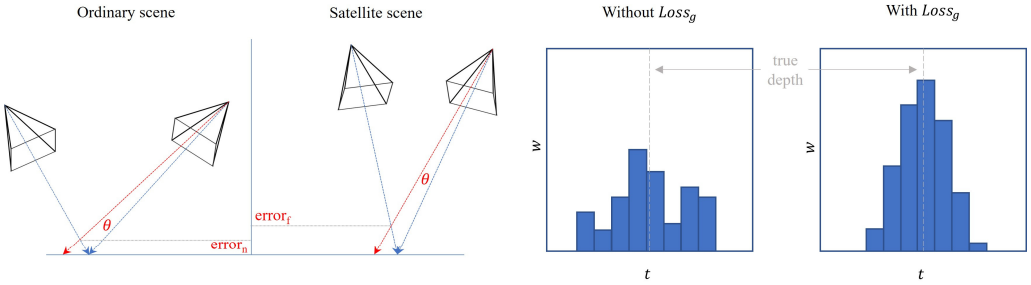
3.3. Geometric Loss Term

A geometric loss term is designed to constrain scene geometry. Multi-view satellite photogrammetry mainly focuses on geometric accuracy, while original NeRFs only constrain the appearance without explicitly constraining the geometry. Geometric consistency is implicitly constrained by the intersection of rays. However, camera parameters produce inevitable errors, creating inconsistent depths of different rays intersecting at the same point. Satellite cameras observe the scene from closely distributed viewpoints with the same camera parameter errors, i.e., at angle θ, the geometric error is greater in the satellite scene (Figure 5a).

The large geometric error leads to a wide range of depths and sample weights along rays dispersed around the true depth (left graph in Figure 5b). Obtaining true depth from a

large depth range is difficult and decreases DSM accuracy. To reduce depth estimation error, a geometric loss term is introduced for GC-NeRFs to constrain scene geometry as follows:

$$Loss_g = \sum_{r \in R} \sum_{i \in r} (t_i - depth(r))^2 w_i \tag{12}$$



(a) The error of elevation caused by camera errors (b) The distribution of weights along a ray

Figure 5. (**a**) At the same camera parameter error angle θ, the geometric error is greater in the satellite scene. (**b**) Without $Loss_g$, the sample weight distribution is scattered around the true depth, resulting in significant errors in depth estimation. By contrast, the weight distribution is compactly around the true depth.

Minimizing $Loss_g$ impacts sample weights far away from depth on a ray r by making them smaller, creating compact weight distribution (right graph in Figure 5b) and a thinner surface. These changes result in increased geometric precision, surface position, and DSM output accuracy. Meanwhile, $Loss_g$ is applied to a batch of rays, R, reducing depth inconsistency between different rays. using $Loss_g$ directly makes the density σ of the entire scene tend toward zero as training progresses, indicating that the reconstructed scene is empty, and the reconstruction has failed. To avoid this, $Loss_g$ is rewritten as:

$$Loss_g = \sum_{r \in R} \left(\sum_{i \in r} (t_i - depth(r))^2 w_i + \exp\left(-\sum_{i \in r} \sigma_i \right) \right) \tag{13}$$

The additional item for the geometric loss term in Formula (13) ensures that the scene does not become empty during training. The main term of the GC-NeRF loss function is the $Loss_c$ defined in Formula (7), which is complemented by $Loss_g$. The complete loss function can be expressed as:

$$Loss = Loss_c + \lambda_g Loss_g \tag{14}$$

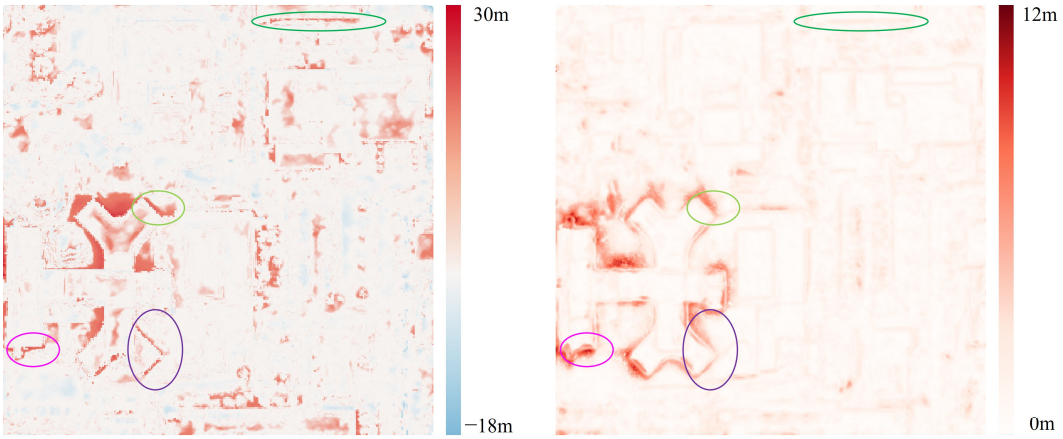
where λ_g is a weight given to $Loss_g$ and is empirically set to 0.02. The model is trained following the ray casting strategy for NeRFs, and $Loss_g$ significantly contributes to the improved accuracy of output DSMs.

3.4. Multi-View DSMs Fusion

To generate a DSM, a depth map was rendered according to Formula (6). Subsequently, the corresponding camera parameters were used to convert the depth map into a 3D point cloud, which was flattened into a DSM [20]. Due to the obstruction caused by tall buildings and trees, the surface depth in some areas was inconsistent with the rendered depth. In other words, the generated point cloud did not cover this area, resulting in inaccurate DSM information at the edges of these tall objects.

Therefore, a multi-view DSM fusion strategy was proposed to improve accuracy. Multi-view DSMs were generated using all viewpoints in each dataset. DSMs from a single viewpoint may be hindered by tall objects, but DSMs from multiple viewpoints can complete each other. A simple approach is to merge point clouds from multiple viewpoints. However, flattening the merged point cloud does not improve DSM accuracy significantly.

By analyzing DSMs from multiple viewpoints, we found large root mean square error (RMSE) values for the elevation between different viewpoints in occluded areas (Figure 6b). Therefore, merging these point clouds directly cannot distinguish noisy points. Furthermore, errors in these areas were mostly positive (Figure 6a), indicating that the predicted elevation was greater than the actual elevation in most areas with a large RMSE value.



(a) DSM errors in merged point clouds (b) The RMSE of multi-view DSMs

Figure 6. The relativity between positive DSM errors in merged point clouds and the root mean square error (RMSE) of multi-view DSMs. The predicted elevation is greater than the actual elevation in most areas with a large elevation standard deviation.

Therefore, $RMSE$ can be used to improve DSM accuracy:

$$elevation = \begin{cases} avg, \ RMSE < Thres_{RMSE} \\ min, \ RMSE \geq Thres_{RMSE} \end{cases} \quad (15)$$

In this formula, $elevation$ denotes an elevation value in a DSM pixel and $RMSE$ denotes the standard deviation of elevations at one certain pixel under different viewpoints. avg denotes areas where the $RMSE$ is small and outliers are removed using the 3-sigma rule. The remaining average the elevation values are considered the final elevation at the corresponding pixel. min denotes areas with a large $RMSE$, where the minimum elevation is taken. $Thres_{RMSE}$ is set to 0.1 times the maximum RMSE in the entire scene as a threshold for determining whether the RMSE is large or small. After multi-view DSM fusion, the elevation accuracy of some obstructed areas improved.

4. Experiments and Results

GC-NeRFs were assessed in four areas of interest (AOI), each spanning 256 × 256 m, using approximately 10–20 crops from the WorldView-3 optical sensor with a pixel resolution of 0.3 m. The four AOI are the same as those of S-NeRFs and Sat-NeRFs for comparison purposes. The four AOI are from different locations, with one from rural areas and three from urban areas. The images were sourced from publicly available data from the 2019 IEEE GRSS Data Fusion Contest (DFC2019) [65,66]. Detailed information for each AOI is shown in Table 2.

Table 2. Detailed information for each AOI.

Area Index	004	068	214	260
# train images	9	17	22	15
# test images	2	2	2	2
Altitude Bounds [m]	[−24, 0]	[−27, 30]	[−29, 73]	[−30, 13]
Type	Rural areas	Urban Areas	Urban Areas	Urban Areas

4.1. Implementation Details

GC-NeRFs were trained with an Adam optimizer starting with a learning rate of 0.01, which was decreased every 4k iterations by a factor 0.33 according to a step scheduler. A size of 2^{18} samples in each batch of rays was used, and the convergence required about 15 k iterations and 6 min to converge on a NVIDIA GeForce RTX 3080 GPU with 12 GB RAM. DSMs were accessed using lidar DSMs (from DFC2019 datasets) with a resolution of 0.5 m per pixel. The peak signal-to-noise ratio (PSNR) [67] of the image renderings and elevation mean absolute error (MAE) [68], with respect to lidar data, were used as evaluation metrics.

4.2. Result Analysis

The experimental results indicated that GC-NeRF is superior to previous methods including S-NeRFs, Sat-NeRFs, and SatelliteRFs [69], both qualitatively and quantitatively. Table 3 displays detailed experimental data. The 3D models in Figure 7 show that GC-NeRFs are can acquire fine geometry of the scene.

Table 3. Numerical results of reconstruction.

Area Index	004		068		214		260		Mean	
	PSNR	MAE	PSNR	MAE	PSNR	MAE	PSNR	MAE	PSNR	MAE
S-NeRFs	26.1	4.42 m	24.2	3.64 m	24.9	4.83 m	21.5	7.71 m	24.2	5.15 m
Sat-NeRFs	26.6	1.37 m	25.0	1.28 m	25.7	1.68 m	21.7	1.64 m	24.7	1.49 m
SatelliteRFs	26.6	-	25.3	-	25.5	-	22.0	-	24.9	-
GC-NeRFs	**26.8**	**1.33 m**	**27.3**	**1.15 m**	**26.9**	**1.47 m**	**23.6**	**1.63 m**	**26.2**	**1.40 m**

The best PSNR and MAE values are shown in bold. Overall, GC-NeRFs provide the best image quality measured by PSNR and DSM accuracy measured by MAE.

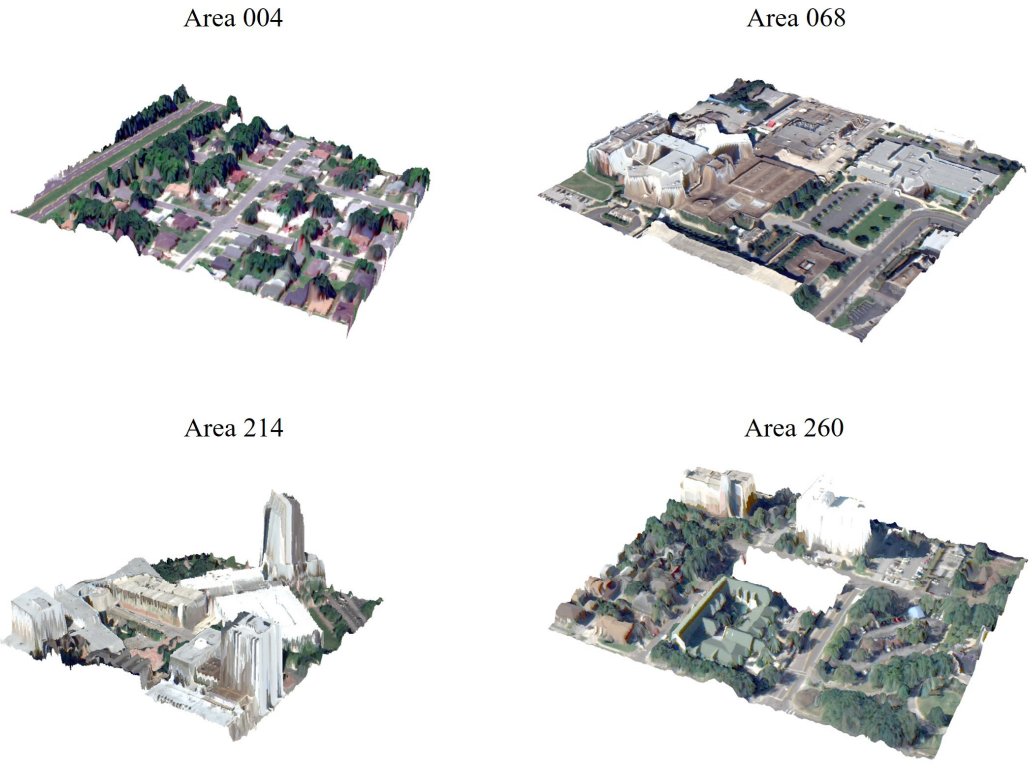


Figure 7. Visualization of 3D models derived by superimposing DSMs onto images. The DSMs and images are generated by GC-NeRFs.

NeRFs were originally used for novel view synthesis, and reconstruction quality is mainly measured by the quality of the synthesized views. Therefore, PSNR was first used to evaluate GC-NeRF reconstruction quality. Figure 8 compares the quality of rendered images

between S-NeRFs, Sat-NeRFs, and GC-NeRFs. Table 3 shows that GC-NeRFs produce statistically superior image quality scores in appearance, as measured by PSNR. However, Sat-NeRFs remove transient objects such as cars (Figure 8a,b), which may decrease PSNR. In fact, quantitative appearance comparisons are not meaningful because satellite images have inconsistent lighting conditions. Qualitative comparison found that GC-NeRF rendered images were clearer, as shown in Figure 8c,d.

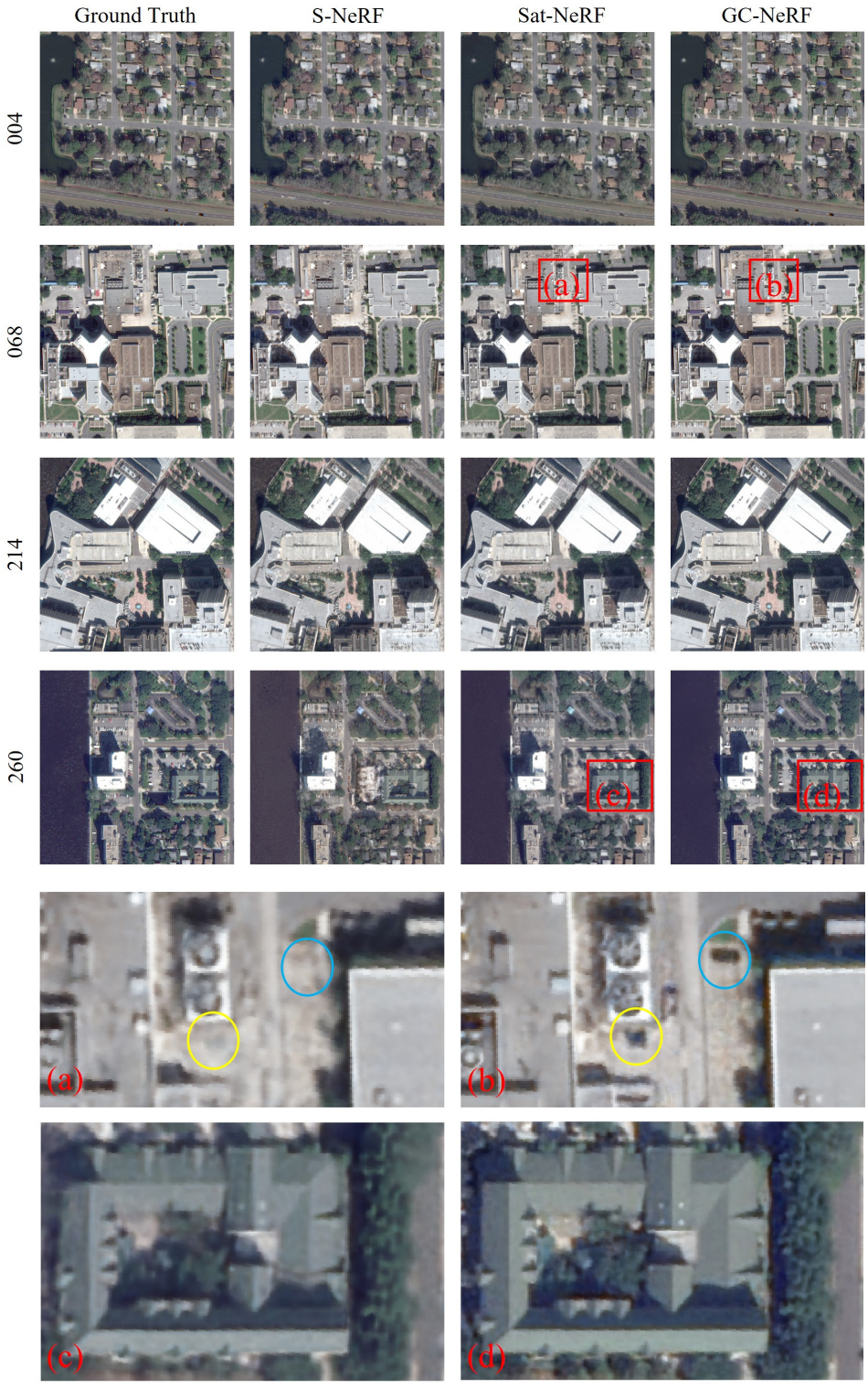


Figure 8. Images rendered by S-NeRFs, Sat-NeRFs, and GC-NeRFs. (**a**,**b**) Sat-NeRFs are robust to transient phenomenon such as cars. (**c**,**d**) GC-NeRFs render clearer images.

Figure 9 shows the visualization of DSMs generated by lidar, S-NeRFs, Sat-NeRFs, and GC-NeRFs. Compared to lidar DSM, GC-NeRFs generate unexpectedly uneven surfaces, such as road areas (Figure 9a,b). Sharper building edges were obtained by the multi-view DSM fusion strategy (Figure 9c,d). Thanks to the geometric loss term proposed in Section 3.3, the quality of DSMs generated by GC-NeRFs is superior to that of S-NeRFs (Figure 9e,f). Table 3 shows that GC-NeRFs produce statistically superior DSM quality scores, as measured by the MAE.

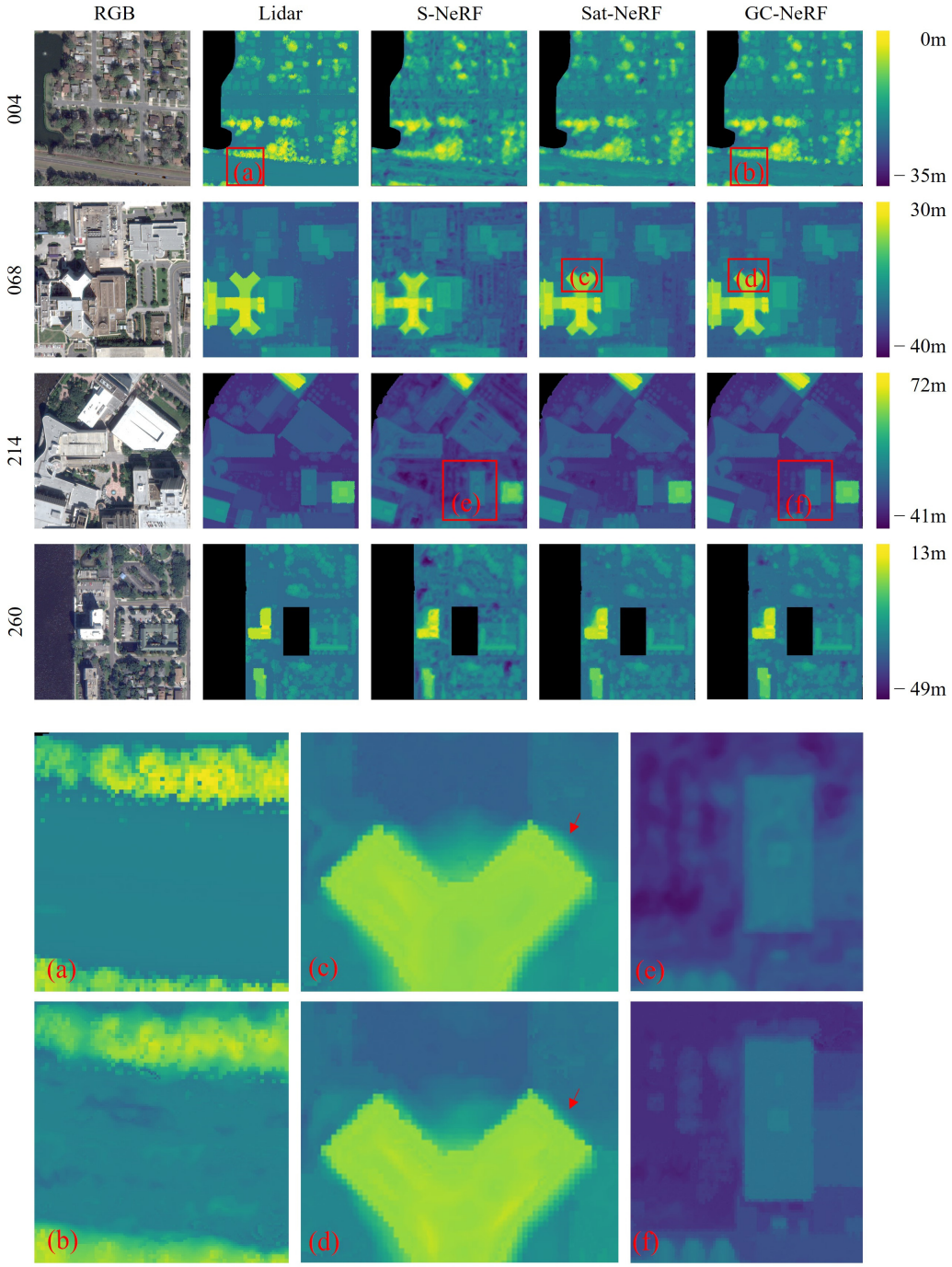


Figure 9. Visualization of lidar, S-NeRFs, Sat-NeRFs, and GC-NeRF DSMs. Areas marked by water and building changes are masked. (**a**,**b**) The DSM rendered by a GC-NeRF shows that the road is uneven compared to lidar DSM. (**c**,**d**) The GC-NeRF DSM displays sharper building edges than the Sat-NeRF DSM. (**e**,**f**) The DSM quality rendered by the GC-NeRF is superior to that of the S-NeRF.

4.3. Ablation Analysis

To verify the effectiveness of the proposed method, ablation experiments were conducted. This paper focuses on geometric accuracy; therefore, the ablation experiment mainly analyzes the impact of the proposed method on DSM accuracy.

GC-NeRFs with all or without one of three measures, namely scene stretching, geometric constraint, and DSM fusion, were evaluated. The results are shown in Table 4. Without geometric constraint, the accuracy of DSMs generated by GC-NeRFs was almost the same as without adding any improvement strategies, emphasizing the importance of geometric constraint. Additionally, scene stretching and DSM fusion can also improve accuracy. Comparing different scenes, the accuracy improvement of scene stretching and DSM fusion for area 004 was minimal. In rural areas, the small altitude range and low terrain complexity may result in better reconstruction. Overall, the various methods proposed by the GC-NeRF framework contributed to accuracy improvement in the final DSM.

Table 4. The MAE of DSMs rendered by GC-NeRFs with all or without one of three measures, namely scene stretching, geometric constraint, and DSM fusion.

Area Index	004	068	214	260
GC-NeRFs without scene stretching	1.59 m	2.11 m	3.70 m	2.65 m
GC-NeRFs without geometric constraint	9.03 m	10.31 m	14.19 m	8.14 m
GC-NeRFs without DSM fusion	1.72 m	2.01 m	3.16 m	2.49 m
GC-NeRFs without all the above	10.21 m	13.44 m	16.88 m	9.47 m
GC-NeRFs	1.33 m	1.15 m	1.47 m	1.63 m

The geometric constraint contributes most to improving DSM accuracy.

5. Discussion

Despite improvements in DSM accuracy, some limitations to the proposed method must be acknowledged. Firstly, the hyperparameter s_z must be adjusted manually. Appropriate values of s_z were obtained through extensive experiments so that GC-NeRF universality could be reduced. Figure 10 shows the MAE of DSMs under different s_z. The best s_z value was around 0.80.

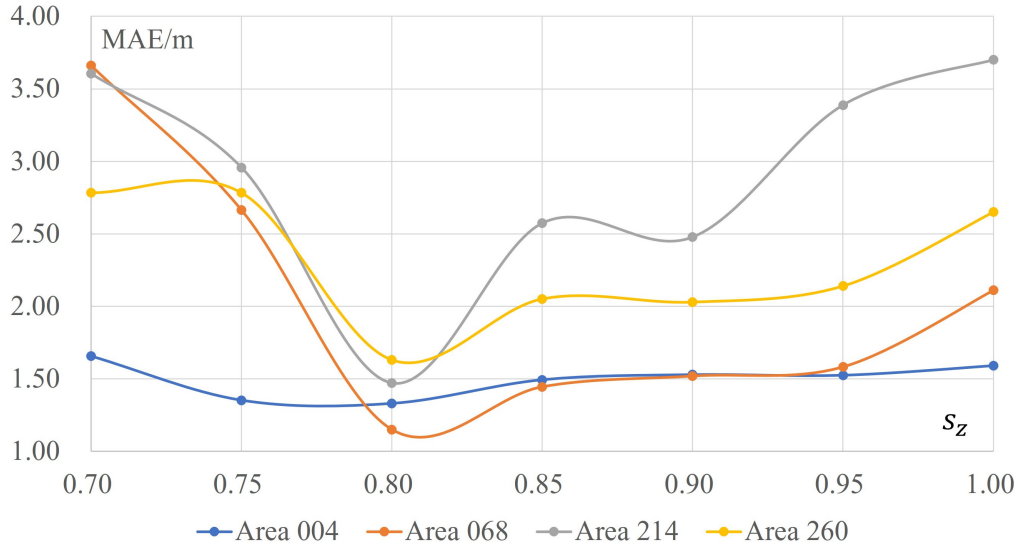


Figure 10. The MAE of DSMs under different s_z. Low MAE indicates high DSM accuracy, and the best s_z value is around 0.8.

Secondly, using the sun direction as appearance encoding, GC-NeRF implicitly addresses shadow issues in a concise manner. However, the limited number of input images

may result in some areas always being shadowed. The reconstruction accuracy of these areas is low, so more effort is needed to eliminate the influence of shadows on geometry.

Thirdly, GC-NeRF is only applicable to high-resolution satellite images, and the results are poor when using low-resolution satellite images as input. Moreover, in areas with excessive terrain undulations, some hyperparameters must be adjusted. Overall, the robustness of GC NeRFs to different types of satellite images and terrain still needs improvement.

Additionally, a common phenomenon with GC-NeRFs and other NeRF-based multi-view satellite photogrammetry approaches is that they produce uneven road and building surfaces, which is far from expected. Other geometric constraints could be designed to overcome this weakness. RegNeRFs [70] and CL-NeRFs [71] add geometric constraints of local plane or slanted plane regularization to make the surface more uniform. However, there are some areas that should not be regular planes in the real world, such as vegetation. Satellite images encompass multiple bands beyond RGB, which can be used to calculate vegetation indices; therefore, subsequent work involves applying plane regularization to areas with low vegetation index values, which may help generate a more uniform surface and improve DSM accuracy.

6. Conclusions

The main contribution of this paper is the proposal of a GC-NeRF framework for multi-view satellite photogrammetry, which generates highly accurate DSMs without extra input. Through a combination of geometric constraint, scene stretching, and multi-view DSM fusion techniques, GC-NeRFs achieve notable improvements in output DSM accuracy. Occupancy grids converted from sparse point clouds avoid the extra input of altitude bounds for training. Additionally, the integration of advanced techniques in NeRF variants greatly enhances efficiency and conciseness in processing satellite images. Experimental results demonstrated the effectiveness of GC-NeRFs. Overall, GC-NeRFs offer a promising solution for reconstructing accurate 3D scenes from multi-view satellite images. Scaling GC-NeRFs to larger datasets and addressing computational efficiency issues may greatly benefit from their practical deployment. Future research can focus on handling building changes. GC-NeRFs treat the scene as static, but the construction and demolition of buildings can lead to significant geometric changes within the scene, which is difficult to manage. Currently, there are some NeRF extensions for dynamic scenes, but these methods deal with continuously changing scenes, where the input data is continuous video. Satellite images are shot very far apart, and handling the modeling problems of such discrete dynamic scenes is challenging. However, it is still very beneficial for change detection.

Author Contributions: Conceptualization, Qifeng Wan; methodology, Qifeng Wan and Yuzheng Guan; software, Qifeng Wan and Qiang Zhao; validation, Qifeng Wan and Qiang Zhao; visualization, Qifeng Wan and Yuzheng Guan; writing—original draft, Qifeng Wan; writing—review and editing, Qifeng Wan, Xiang Wen and Jiangfeng She; project administration, Jiangfeng She. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 41871293.

Data Availability Statement: The data presented in this paper are openly available at https://ieee-dataport.org/open-access/data-fusion-contest-2019-dfc2019 (accessed on 4 December 2023).

Acknowledgments: The authors would like to thank the website https://github.com/NVlabs/tiny-cuda-nn (accessed on 20 October 2023) for providing the source code for multi-resolution hash encoding.

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

References

1. Peter, B.G.; Messina, J.P.; Carroll, J.W.; Zhi, J.; Chimonyo, V.; Lin, S.; Snapp, S.S. Multi-Spatial Resolution Satellite and sUAS Imagery for Precision Agriculture on Smallholder Farms in Malawi. *Photogramm. Eng. Remote Sens.* **2020**, *86*, 107–119. [CrossRef]

2. Barrile, V.; Simonetti, S.; Citroni, R.; Fotia, A.; Bilotta, G. Experimenting Agriculture 4.0 with Sensors: A Data Fusion Approach between Remote Sensing, UAVs and Self-Driving Tractors. *Sensors* **2022**, *22*, 7910. [CrossRef]
3. Pearse, G.D.; Dash, J.P.; Persson, H.J.; Watt, M.S. Comparison of High-Density LiDAR and Satellite Photogrammetry for Forest Inventory. *ISPRS J. Photogramm. Remote Sens.* **2018**, *142*, 257–267. [CrossRef]
4. Njimi, H.; Chehata, N.; Revers, F. Fusion of Dense Airborne LiDAR and Multispectral Sentinel-2 and Pleiades Satellite Imagery for Mapping Riparian Forest Species Biodiversity at Tree Level. *Sensors* **2024**, *24*, 1753. [CrossRef]
5. Jaud, M.; Geoffroy, L.; Chauvet, F.; Durand, E.; Civet, F. Potential of a Virtual Reality Environment Based on Very-High-Resolution Satellite Imagery for Structural Geology Measurements of Lava Flows. *J. Struct. Geol.* **2022**, *158*, 104569. [CrossRef]
6. Demarez, V.; Helen, F.; Marais-Sicre, C.; Baup, F. In-Season Mapping of Irrigated Crops Using Landsat 8 and Sentinel-1 Time Series. *Remote Sens.* **2019**, *11*, 118. [CrossRef]
7. Bhattacharya, S.; Braun, C.; Leopold, U. An Efficient 2.5D Shadow Detection Algorithm for Urban Planning and Design Using a Tensor Based Approach. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 583. [CrossRef]
8. Chen, C.; Ye, S.; Bai, Z.; Wang, J.; Nedzved, A.; Ablameyko, S. Intelligent Mining of Urban Ventilated Corridor Based on Digital Surface Model under the Guidance of K-Means. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 216. [CrossRef]
9. Zhang, J.; Xu, W.; Qin, L.; Tian, Y. Spatial Distribution Estimates of the Urban Population Using DSM and DEM Data in China. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 435. [CrossRef]
10. Zhu, L.; Shen, S.; Gao, X.; Hu, Z. Urban Scene Vectorized Modeling Based on Contour Deformation. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 162. [CrossRef]
11. McClean, F.; Dawson, R.; Kilsby, C. Implications of Using Global Digital Elevation Models for Flood Risk Analysis in Cities. *Water Resour. Res.* **2020**, *56*, e2020WR028241. [CrossRef]
12. Qin, R.; Tian, J.; Reinartz, P. 3D Change Detection—Approaches and Applications. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 41–56. [CrossRef]
13. Zhang, L.; Yang, L.; Zohner, C.M.; Crowther, T.W.; Li, M.; Shen, F.; Guo, M.; Qin, J.; Yao, L.; Zhou, C. Direct and Indirect Impacts of Urbanization on Vegetation Growth across the World's Cities. *Sci. Adv.* **2022**, *8*, eabo0095. [CrossRef] [PubMed]
14. Gómez, A.; Randall, G.; Facciolo, G.; von Gioi, R.G. An Experimental Comparison of Multi-View Stereo Approaches on Satellite Images. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 844–853.
15. de Franchis, C.; Meinhardt-Llopis, E.; Michel, J.; Morel, J.-M.; Facciolo, G. An Automatic and Modular Stereo Pipeline for Pushbroom Images. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *II–3*, 49–56. [CrossRef]
16. d'Angelo, P.; Kuschk, G. Dense Multi-View Stereo from Satellite Imagery. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 6944–6947.
17. Gong, K.; Fritsch, D. DSM Generation from High Resolution Multi-View Stereo Satellite Imagery. *Photogramm. Eng. Remote Sens.* **2019**, *85*, 379–387. [CrossRef]
18. Facciolo, G.; de Franchis, C.; Meinhardt-Llopis, E. Automatic 3D Reconstruction from Multi-Date Satellite Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 57–66.
19. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Computer Vision—ECCV 2020*; Springer International Publishing: Cham, Switzerland, 2020; pp. 405–421.
20. Marí, R.; Facciolo, G.; Ehret, T. Sat-NeRF: Learning Multi-View Satellite Photogrammetry with Transient Objects and Shadow Modeling Using RPC Cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022; pp. 1311–1321.
21. Marí, R.; Facciolo, G.; Ehret, T. Multi-Date Earth Observation NeRF: The Detail Is in the Shadows. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Vancouver, BC, Canada, 17–24 June 2023; pp. 2034–2044.
22. Kajiya, J.T.; Von Herzen, B.P. Ray Tracing Volume Densities. *SIGGRAPH Comput. Graph.* **1984**, *18*, 165–174. [CrossRef]
23. Derksen, D.; Izzo, D. Shadow Neural Radiance Fields for Multi-View Satellite Photogrammetry. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; pp. 1152–1161.
24. Müller, T.; Evans, A.; Schied, C.; Keller, A. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.* **2022**, *41*, 1–15. [CrossRef]
25. Srinivasan, P.P.; Deng, B.; Zhang, X.; Tancik, M.; Mildenhall, B.; Barron, J.T. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 7495–7504.
26. Schönberger, J.L.; Zheng, E.; Frahm, J.-M.; Pollefeys, M. Pixelwise View Selection for Unstructured Multi-View Stereo. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 501–518.
27. Krauß, T.; d'Angelo, P.; Schneider, M.; Gstaiger, V. THE FULLY AUTOMATIC OPTICAL PROCESSING SYSTEM CATENA AT DLR. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *XL-1-W1*, 177–183. [CrossRef]

28. Qu, Y.; Deng, F. Sat-Mesh: Learning Neural Implicit Surfaces for Multi-View Satellite Reconstruction. *Remote Sens.* **2023**, *15*, 4297. [CrossRef]
29. Beyer, R.A.; Alexandrov, O.; McMichael, S. The Ames Stereo Pipeline: NASA's Open Source Software for Deriving and Processing Terrain Data. *Earth Space Sci.* **2018**, *5*, 537–548. [CrossRef]
30. Facciolo, G.; Franchis, C.D.; Meinhardt, E. MGM: A Significantly More Global Matching for Stereovision. In Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 7–10 September 2015.
31. Rothermel, M.; Wenzel, K.; Fritsch, D.; Haala, N. SURE: Photogrammetric Surface Reconstruction from Imagery. In Proceedings of the LC3D Workshop, Berlin, Germany, 4–5 December 2012; Volume 8.
32. Lastilla, L.; Ravanelli, R.; Fratarcangeli, F.; Di Rita, M.; Nascetti, A.; Crespi, M. FOSS4G DATE FOR DSM GENERATION: SENSITIVITY ANALYSIS OF THE SEMI-GLOBAL BLOCK MATCHING PARAMETERS. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2-W13*, 67–72. [CrossRef]
33. Han, Y.; Wang, S.; Gong, D.; Wang, Y.; Ma, X. STATE OF THE ART IN DIGITAL SURFACE MODELLING FROM MULTI-VIEW HIGH-RESOLUTION SATELLITE IMAGES. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *V-2–2020*, 351–356. [CrossRef]
34. Zhang, F.; Prisacariu, V.; Yang, R.; Torr, P.H.S. GA-Net: Guided Aggregation Net for End-To-End Stereo Matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 185–194.
35. Chang, J.-R.; Chen, Y.-S. Pyramid Stereo Matching Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5410–5418.
36. Yang, G.; Manela, J.; Happold, M.; Ramanan, D. Hierarchical Deep Stereo Matching on High-Resolution Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 5515–5524.
37. Marí, R.; Ehret, T.; Facciolo, G. Disparity Estimation Networks for Aerial and High-Resolution Satellite Images: A Review. *Image Process. Line* **2022**, *12*, 501–526. [CrossRef]
38. Xiangli, Y.; Xu, L.; Pan, X.; Zhao, N.; Rao, A.; Theobalt, C.; Dai, B.; Lin, D. BungeeNeRF: Progressive Neural Radiance Field for Extreme Multi-Scale Scene Rendering. In *Computer Vision—ECCV 2022*; Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T., Eds.; Springer Nature Switzerland: Cham, Switzerland, 2022; pp. 106–122.
39. Tancik, M.; Casser, V.; Yan, X.; Pradhan, S.; Mildenhall, B.; Srinivasan, P.P.; Barron, J.T.; Kretzschmar, H. Block-NeRF: Scalable Large Scene Neural View Synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 8248–8258.
40. Moreau, A.; Piasco, N.; Tsishkou, D.; Stanciulescu, B.; de La Fortelle, A. LENS: Localization Enhanced by NeRF Synthesis. In Proceedings of the 5th Conference on Robot Learning, London, UK, 11 January 2022; pp. 1347–1356.
41. Turki, H.; Ramanan, D.; Satyanarayanan, M. Mega-NeRF: Scalable Construction of Large-Scale NeRFs for Virtual Fly-Throughs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 12922–12931.
42. Chen, A.; Xu, Z.; Zhao, F.; Zhang, X.; Xiang, F.; Yu, J.; Su, H. MVSNeRF: Fast Generalizable Radiance Field Reconstruction from Multi-View Stereo. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 14124–14133.
43. Boss, M.; Braun, R.; Jampani, V.; Barron, J.T.; Liu, C.; Lensch, H.P.A. NeRD: Neural Reflectance Decomposition from Image Collections. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 12684–12694.
44. Guo, Y.-C.; Kang, D.; Bao, L.; He, Y.; Zhang, S.-H. NeRFReN: Neural Radiance Fields with Reflections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 18409–18418.
45. Verbin, D.; Hedman, P.; Mildenhall, B.; Zickler, T.; Barron, J.T.; Srinivasan, P.P. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5481–5490.
46. Mildenhall, B.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P.P.; Barron, J.T. NeRF in the Dark: High Dynamic Range View Synthesis from Noisy Raw Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 16190–16199.
47. Pumarola, A.; Corona, E.; Pons-Moll, G.; Moreno-Noguer, F. D-NeRF: Neural Radiance Fields for Dynamic Scenes. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 10313–10322.
48. Fang, J.; Yi, T.; Wang, X.; Xie, L.; Zhang, X.; Liu, W.; Nießner, M.; Tian, Q. Fast Dynamic Radiance Fields with Time-Aware Neural Voxels. In *SIGGRAPH Asia 2022 Conference Papers*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 1–9.
49. Attal, B.; Laidlaw, E.; Gokaslan, A.; Kim, C.; Richardt, C.; Tompkin, J.; O' Toole, M. TöRF: Time-of-Flight Radiance Fields for Dynamic Scene View Synthesis. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 26289–26301.

50. Liu, J.-W.; Cao, Y.-P.; Mao, W.; Zhang, W.; Zhang, D.J.; Keppo, J.; Shan, Y.; Qie, X.; Shou, M.Z. DeVRF: Fast Deformable Voxel Radiance Fields for Dynamic Scenes. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 36762–36775.
51. Athar, S.; Shu, Z.; Samaras, D. FLAME-in-NeRF: Neural Control of Radiance Fields for Free View Face Animation. In Proceedings of the 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), Waikoloa Beach, HI, USA, 5–8 January 2023; pp. 1–8.
52. Barron, J.T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P.P. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 5835–5844.
53. Barron, J.T.; Mildenhall, B.; Verbin, D.; Srinivasan, P.P.; Hedman, P. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5470–5479.
54. Martin-Brualla, R.; Radwan, N.; Sajjadi, M.S.M.; Barron, J.T.; Dosovitskiy, A.; Duckworth, D. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 7206–7215.
55. Zhang, K.; Snavely, N.; Sun, J. Leveraging Vision Reconstruction Pipelines for Satellite Imagery. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019.
56. Piala, M.; Clark, R. TermiNeRF: Ray Termination Prediction for Efficient Neural Rendering. In Proceedings of the 2021 International Conference on 3D Vision (3DV), London, UK, 1–3 December 2021; pp. 1106–1114.
57. Neff, T.; Stadlbauer, P.; Parger, M.; Kurz, A.; Mueller, J.H.; Chaitanya, C.R.A.; Kaplanyan, A.; Steinberger, M. DONeRF: Towards Real-Time Rendering of Compact Neural Radiance Fields Using Depth Oracle Networks. *Comput. Graph. Forum* **2021**, *40*, 45–59. [CrossRef]
58. Rebain, D.; Jiang, W.; Yazdani, S.; Li, K.; Yi, K.M.; Tagliasacchi, A. DeRF: Decomposed Radiance Fields. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 14148–14156.
59. Reiser, C.; Peng, S.; Liao, Y.; Geiger, A. KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 14315–14325.
60. Xu, Q.; Xu, Z.; Philip, J.; Bi, S.; Shu, Z.; Sunkavalli, K.; Neumann, U. Point-NeRF: Point-Based Neural Radiance Fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5438–5448.
61. Fridovich-Keil, S.; Yu, A.; Tancik, M.; Chen, Q.; Recht, B.; Kanazawa, A. Plenoxels: Radiance Fields Without Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5501–5510.
62. Marí, R.; de Franchis, C.; Meinhardt-Llopis, E.; Anger, J.; Facciolo, G. A Generic Bundle Adjustment Methodology for Indirect RPC Model Refinement of Satellite Imagery. *Image Process. Line* **2021**, *11*, 344–373. [CrossRef]
63. Xie, S.; Zhang, L.; Jeon, G.; Yang, X. Remote Sensing Neural Radiance Fields for Multi-View Satellite Photogrammetry. *Remote Sens.* **2023**, *15*, 3808. [CrossRef]
64. Li, C.; Wu, B.; Pumarola, A.; Zhang, P.; Lin, Y.; Vajda, P. INGeo: Accelerating Instant Neural Scene Reconstruction with Noisy Geometry Priors. In *Computer Vision—ECCV 2022 Workshops*; Karlinsky, L., Michaeli, T., Nishino, K., Eds.; Springer Nature Switzerland: Cham, Switzerland, 2023; pp. 686–694.
65. Le Saux, B.; Yokoya, N.; Haensch, R.; Brown, M. 2019 IEEE GRSS Data Fusion Contest: Large-Scale Semantic 3D Reconstruction. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 33–36. [CrossRef]
66. Bosch, M.; Foster, K.; Christie, G.; Wang, S.; Hager, G.D.; Brown, M. Semantic Stereo for Incidental Satellite Images. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 7–11 January 2019; pp. 1524–1532.
67. Hu, M.; Luo, X.; Chen, J.; Lee, Y.C.; Zhou, Y.; Wu, D. Virtual Reality: A Survey of Enabling Technologies and Its Applications in IoT. *J. Netw. Comput. Appl.* **2021**, *178*, 102970. [CrossRef]
68. Xia, G.; Bi, Y.; Wang, C. Optimization Design of Passive Residual Heat Removal System Based on Improved Genetic Algorithm. *Ann. Nucl. Energy* **2023**, *189*, 109859. [CrossRef]
69. Zhou, X.; Wang, Y.; Lin, D.; Cao, Z.; Li, B.; Liu, J. SatelliteRF: Accelerating 3D Reconstruction in Multi-View Satellite Images with Efficient Neural Radiance Fields. *Appl. Sci.* **2024**, *14*, 2729. [CrossRef]
70. Niemeyer, M.; Barron, J.T.; Mildenhall, B.; Sajjadi, M.S.M.; Geiger, A.; Radwan, N. RegNeRF: Regularizing Neural Radiance Fields for View Synthesis from Sparse Inputs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5480–5490.
71. Chen, X.; Song, Z.; Zhou, J.; Xie, D.; Lu, J. Camera and LiDAR Fusion for Urban Scene Reconstruction and Novel View Synthesis via Voxel-Based Neural Radiance Fields. *Remote Sens.* **2023**, *15*, 4628. [CrossRef]