# Neural Radiance Field Methods for Satellite Imagery of Polar Climates

by

Ellemieke Van Kints

(Under the Direction of Deepak Mishra)

Abstract

This research aims to explore modern neural rendering methods for 3D surface reconstruction using remote sensing imagery from Earth observation satellites. Traditional multi-view stereo (MVS) photogrammetric methods rely on rigid surface textures to accurately reconstruct a scene. However, polar regions across Earth exhibit uniform color and texture due to snow and ice coverage, making it difficult for traditional methods to create accurate surface models of the terrain. Neural Radiance Fields (NeRFs) overcome this problem by using neural volumetric rendering techniques to continuously learn the color and geometry of a scene. In this research, NeRF variants, Shadow NeRF (S-NeRF) and Satellite NeRF (Sat-NeRF), are tested using WorldView-2 satellite imagery of a scene in Anchorage, Alaska and Mount Doran, located in Chugach, Alaska. The goal of this research is to assess the feasibility and performance of satellite-based NeRF methods in generating digital surface models (DSMs) of terrain from polar regions on Earth using multi-view satellite imagery. Furthermore, a comparative analysis of S-NeRF and Sat-NeRF using high-density point clouds derived from LiDAR ground-truth is presented to determine which method performs best at surface model generation with sparsely-textured geo-spatial data. The results from this research will ultimately reveal whether NeRF serves as a plausible contender to the current state-of-the-art methods for accurate 3D scene reconstruction from satellite remote sensing data in sparse-textured terrains.

Index words: [neural volumetric rendering, remote sensing, multi-view photogrammetry, 3D deep learning]

# Neural Radiance Field Methods for Satellite Imagery of Polar Climates

by

Ellemieke Van Kints

B.S., University of Georgia, 2023

A Thesis Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Master of Science

Athens, Georgia

2024

©2024
Ellemieke Van Kints
All Rights Reserved

# Neural Radiance Field Methods for Satellite Imagery of Polar Climates

by

Ellemieke Van Kints

Major Professor: Deepak Mishra

Committee: In Kee Kim
Lakshmish Ramaswamy

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2024

# Acknowledgments

I would like to thank Dr. Deepak Mishra, my major advisor and principal investigator of the UGA Small Satellite Research Laboratory, for his continued support, guidance, and encouragement throughout my graduate studies. His insights and expertise were invaluable in shaping the direction and execution of my research. I am deeply grateful for the time he invested in providing feedback, facilitating opportunities, and sharing his knowledge, all of which were instrumental in my development as a researcher.

I am also deeply grateful to Caleb Adams, my project manager at NASA Ames Research Center, whose kindness, mentorship, and unwavering support made a lasting impact on both my professional and personal growth. Caleb was not only a mentor but also a role model who believed in me and encouraged me to reach my full potential. I am endlessly thankful for his guidance, patience, and the confidence he instilled in me. I truly could not have accomplished what I did without him.

I would also like to thank Nancy O'Hare and Batu Ozdener, who were instrumental in the success of this work. Their collaboration and support were critical throughout this project, and I am deeply appreciative of their contributions. I am also grateful to my committee members, Dr. In Kee Kim and Dr. Lakshmish Ramaswamy, for their constructive feedback and support, which significantly improved the quality of this work.

Finally, I would like to thank my family and friends for their unwavering encouragement and support throughout this journey. Their belief in me has been my greatest source of strength. I am especially thankful to my boyfriend, Myles Andrews, for always being by my side and providing me with endless love and support throuhgout this journey.

# Contents

**Acknowledgments** **iv**

**List of Figures** **vi**

**List of Tables** **xi**

**1 Introduction** **1**
1.1 Related Works . . . 2
1.2 Research Objectives . . . 5

**2 Data Collection** **8**
2.1 Data Sources . . . 8
2.2 Study Areas . . . 9
2.3 Data Preparation . . . 10

**3 Reconstructing Polar Regions Using NeRF** **20**
3.1 Experimental Design . . . 21
3.2 Metashape . . . 25
3.3 S-NeRF . . . 28
3.4 Sat-NeRF . . . 33
3.5 Ablation Studies . . . 40
3.6 Discussion & Conclusion . . . 58

**4 Conclusion** **65**
4.1 Limitations and Future Work . . . 66

**Bibliography** **70**

# List of Figures


2.1 Comparison of features detected in the Jacksonville, Anchorage, and Mount Doran scenes. The original images are shown in the top row, and the corresponding features detected are shown in the bottom row. OpenCV detected 5,090 features in Jacksonville, 4,786 in Anchorage, and 1,467 in Mount Doran. . . . . 10
2.2 The clipping extent for the small Anchorage dataset. For this dataset, the pushbroom images were cropped to the extent of the small AOI ($300 \times 300$ m). . . . . 12
2.3 The clipping extent for the medium Anchorage dataset. For this dataset, the pushbroom images were cropped to the extent of the medium AOI ($600 \times 600$ m). . . . . 13
2.4 The clipping extent for the large Anchorage dataset. For this dataset, the pushbroom images were cropped to the extent of the large AOI ($1,200 \times 1,200$ m). . . . . 14
2.5 The clipping extent for the small Mount Doran dataset. For this dataset, the pushbroom images were cropped to the extent of the small AOI ($600 \times 600$ m). . . . . 15
2.6 The clipping extent for the medium Mount Doran dataset. For this dataset, the pushbroom images were cropped to the extent of the medium AOI ($1,200 \times 1,200$ m). . . . 15
2.7 The clipping extent for the large Mount Doran dataset. For this dataset, the pushbroom images were cropped to the extent of the large AOI ($2,000 \times 2,000$ m). . . . . 16
2.8 The distribution of view and light angles for the Anchorage dataset. . . . . 18
2.9 Sample images from the small, medium, and large Anchorage datasets. . . . . 18
2.10 The distribution of view and light angles for the Mount Doran dataset. . . . . 19
2.11 Sample images from the small, medium, and large Mount Doran datasets. . . . . 19

3.1 Screenshot from the "General Parameters" tab of the Cloud-to-Cloud Distance Computation tool in CloudCompare. . . . . 22
3.2 Screenshot from the "Local Modeling" tab of the Cloud-to-Cloud Distance Computation tool in CloudCompare. . . . . 23
3.3 Results for Metashape using the small Anchorage dataset ($300 \times 300$ m) **without** the RPC models. Metashape achieved a mean distance of 25.374 m for the generated sparse point cloud. . . . . 26
3.4 Results for Metashape using the small Anchorage dataset ($300 \times 300$ m) **with** the RPC models. Metashape achieved a mean distance of 2.652 m for the generated sparse point cloud. . . . . 26

3.5 Results for Metashape using the small Mount Doran dataset (600 × 600 m) **without** the RPC models. Metashape achieved a mean distance of 84.061 m for the generated sparse point cloud. . . . . 27
3.6 Results for Metashape using the small Mount Doran dataset (600 × 600 m) **with** the RPC models. Metashape achieved a mean distance of 15.624 m for the generated sparse point cloud. . . . . 27
3.7 A sample of training images rendered from the small Anchorage dataset (300 × 300 m) using the S-NeRF model. S-NeRF achieved mean PSNR/SSIM scores of 22.13/0.578 on training data. . . . . 29
3.8 Testing images rendered from the small Anchorage dataset (300 × 300 m) using the S-NeRF model. S-NeRF achieved mean PSNR/SSIM scores of 10.07/0.139 on testing data. 30
3.9 Surface reconstruction results for S-NeRF using the small Anchorage dataset (300 × 300 m). S-NeRF achieved a mean distance of 2.883 m for the generated sparse point cloud. . 30
3.10 A sample of training images rendered from the small Mount Doran dataset (600 × 600 m) using the S-NeRF model. S-NeRF achieved mean PSNR/SSIM scores of 22.92/0.699 on training data. . . . . 31
3.11 Testing images rendered from the small Mount Doran dataset (600 × 600 m) using the S-NeRF model. S-NeRF achieved mean PSNR/SSIM scores of 11.541/0.244 on testing data. . . . . 32
3.12 Surface reconstruction results for S-NeRF using the small Mount Doran dataset (600 × 600 m). S-NeRF achieved a mean distance of 61.616 m for the generated sparse point cloud. 32
3.13 A sample of training images rendered from the small Anchorage dataset (300×300 m) using the Sat-NeRF model. Sat-NeRF achieved mean PSNR/SSIM scores of 23.099/0.846 on training data. . . . . 35
3.14 Testing images rendered from the small Anchorage dataset (300 × 300 m) using the Sat-NeRF model. Sat-NeRF achieved mean PSNR/SSIM scores of 12.173/0.249 on testing data. . . . . 36
3.15 Surface reconstruction results for Sat-NeRF using the small Anchorage dataset (300 × 300 m). Sat-NeRF achieved a mean distance of 5.724 m for the generated sparse point cloud. . . . . 36
3.16 A sample of training images rendered from the small Mount Doran dataset (600 × 600 m) using the Sat-NeRF model. Sat-NeRF achieved mean PSNR/SSIM scores of 23.573/0.783 on training data. . . . . 37
3.17 Testing images rendered from the small Mount Doran dataset (600 × 600 m) using the Sat-NeRF model. Sat-NeRF achieved mean PSNR/SSIM scores of 12.797/0.461 on testing data. . . . . 38
3.18 Surface reconstruction results for Sat-NeRF using the small Mount Doran dataset (600×600 m). Sat-NeRF achieved a mean distance of 58.031 m for the generated sparse point cloud. . . . . 38

3.19 Qualitative view synthesis results for each method, with WorldView-2 satellite imagery of Anchorage (top row) and Mount Doran (bottowm row). Sat-NeRF produced the most visually accurate novel views, achieving a mean PNSR/SSIM of 12.173/0.249 for the Anchorage dataset and 12.797/0.461 for the Mount Doran dataset. . . . . 39
3.20 Qualitative surface reconstruction results for each method, with WorldView-2 satellite imagery of Anchorage (top row) and Mount Doran (bottowm row). S-NeRF produced the most geometrically accurate surface model for the Anchorage scene, achieving a mean distance error of 2.883 m. Sat-NeRF produced the most geometrically accurate surface model the for the Mount Doran scene, achieving a mean distance error of 58.031 m. . . 39
3.21 View synthesis results on training data with Anchorage scenes of varying AOI sizes. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 23.099/0.846 for the small Anchorage dataset. . . . . 42
3.22 View synthesis results on testing data with Anchorage scenes of varying AOI sizes. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 12.173/0.249 for the small Anchorage dataset. . . . . 43
3.23 View synthesis results on training data with Mount Doran scenes of varying AOI sizes. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 23.573/0.783 for the small Mount Doran dataset. . . . . 44
3.24 View synthesis results on testing data with Mount Doran scenes of varying AOI sizes. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 12.797/0.462 for the small Mount Doran dataset. . . . . 45
3.25 Surface reconstruction results for Sat-NeRF with Anchorage scenes of varying AOI sizes. The model achieved a mean distance error of 3.687 m for the medium Anchorage scene. 46
3.26 Surface reconstruction results for Sat-NeRF with Mount Doran scenes of varying AOI sizes. The model achieved a mean distance error of 58.031 m for the small Mount Doran scene. . . . . 46
3.27 Surface reconstruction results for Metashape with Anchorage scenes of varying AOI sizes. The method achieved a mean distance error of 2.652 m for the sparse point cloud of the small Anchorage scene. . . . . 47
3.28 Surface reconstruction results for Metashape with Mount Doran scenes of varying AOI sizes. The method achieved a mean distance error of 13.979 m for the sparse point cloud of the large Mount Doran scene. . . . . 48
3.29 View synthesis results on training data with various positional encoding (PE) frequencies, for the small Anchorage scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 23.75/0.888 with a positional encoding frequency of 8. . . . . 51
3.30 View synthesis results on testing data with various positional encoding (PE) frequencies, for the small Anchorage scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 15.573/0.657 with a positional encoding frequency of 8. . . . . 51

3.31 View synthesis results on training data with various positional encoding (PE) frequencies, for the small Mount Doran scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 24.014/0.773 with a positional encoding frequency of 8. . . . . 51
3.32 View synthesis results on testing data with various positional encoding (PE) frequencies, for the small Mount Doran scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 12.986/0.483 with a positional encoding frequency of 8. . . . . 52
3.33 Surface reconstruction results for Sat-NeRF with various positional encoding (PE) frequencies, for the small Anchorage scene. The model achieved a mean distance error of 4.553 m with a positional encoding frequency of 16. . . . . 52
3.34 Surface reconstruction results for Sat-NeRF with various positional encoding (PE) frequencies, for the small Mount Doran scene. The model achieved a mean distance error of 52.934 m with a positional encoding frequency of 8. . . . . 52
3.35 View synthesis results on training data with each activation function, for the small Anchorage scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 23.099/0.846 with the SIREN activation function. . . . . 55
3.36 View synthesis results on testing data with each activation function, for the small Anchorage scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 12.173/0.249 with the SIREN activation function. . . . . 55
3.37 View synthesis results on training data with each activation function, for the small Mount Doran scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 23.573/0.783 with SIREN activations. . . . . 56
3.38 View synthesis results on testing data with each activation function, for the small Mount Doran scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 12.797/0.461 with SIREN activation. . . . . 56
3.39 Surface reconstruction results for Sat-NeRF with each activation function, for the small Anchorage scene. The model achieved a mean distance error of 2.136 m with the ReLU activation function. . . . . 57
3.40 Surface reconstruction results for Sat-NeRF with each activation function, for the small Mount Doran scene. The model achieved a mean distance error of 51.854 m with the ReLU activation function. . . . . 57
3.41 Metashape error distributions for the sparse point cloud of the small Anchorage scene ($300 \times 300$ m). The images and RPC models were used as input. The sparse point cloud (top) and histogram (bottom) are colorized by the absolute distance error along the Z-dimension, with saturation bounds set to (-10.0, 10.0) m. Metashape achieved a mean distance error of 2.652 m for the sparse point cloud of the small Anchorage scene. 59

3.42 Metashape error distributions for the dense point cloud of the small Anchorage scene ($300 \times 300$ m). The images and RPC models were used as input. The dense point cloud (top) and histogram (bottom) are colorized by the absolute distance error along the Z-dimension, with saturation bounds set to ($-10.0$, $10.0$ m). Metashape achieved a mean distance error of 2.024 m for the dense point cloud of the small Anchorage scene. . . . 60
3.43 Metashape error distributions for the sparse point cloud of the small Mount Doran scene ($600 \times 600$ m). The images and RPC models were used as input. The sparse point cloud (top) and histogram (bottom) are colorized by the absolute distance error along the Z-dimension, with saturation bounds set to ($-10.0$, $10.0$ m). Metashape achieved a mean distance error of 15.624 m for the sparse point cloud of the small Mount Doran scene. . 61
3.44 Metashape error distributions for the dense point cloud of the small Mount Doran scene ($600 \times 600$ m). The images and RPC models were used as input. The dense point cloud (top) and histogram (bottom) are colorized by the absolute distance error along the Z-dimension, with saturation bounds set to ($-10.0$, $10.0$ m). Metashape achieved a mean distance error of 13.034 m for the sparse point cloud of the small Mount Doran scene. . 62

4.1 Plotted view and light angles for the Jacksonville dataset. The minimum and maximum viewing elevation angles are plotted in red. . . . . 67
4.2 Plotted view and light angles for the Anchorage dataset. The minimum and maximum viewing elevation angles are plotted in red. . . . . 68
4.3 Plotted view and light angles for the Mount Doran dataset. The minimum and maximum viewing elevation angles are plotted in red. . . . . 68
4.4 The images with the lowest and highest camera elevation angles from the Jacksonville, Anchorage, and Mount Doran datasets. . . . . 69

# List of Tables


3.1 Quantitative results for Metashape. The best metrics for each scene are in **bold**. Metashape achieved the highest reconstruction accuracies for both scenes with the use of the RPC models. The generated point cloud of Anchorage had an average distance error of 2.652 m, while the generated point cloud of Mount Doran had an average distance error of 15.624 m. . . . . 27

3.2 Quantitative results for each method. The best metrics for each scene are in **bold**. Metashape performed best at surface reconstruction, achieving a mean distance error of 2.652 m for Anchorage and 15.624 m for Mount Doran. Sat-NeRF performed best at novel view synthesis, achieving mean PSNR/SSIM scores of 12.173/0.249 for Anchorage and 12.797/0.461 for Mount Doran. . . . . 40

3.3 Quantitative results for Sat-NeRF with datasets covering various scene sizes. The best metrics for each scene are in **bold**. For the task of novel view synthesis, the highest accuracy was achieved with the small Anchorage scene and small Mount Doran scene. For the task of 3D surface reconstruction, the highest accuracy was achieved with the medium Anchorage scene and small Mount Doran scene. . . . . 41

3.4 Quantitative results for Metashape with datasets covering various scene sizes. The best metrics for each scene are in **bold**. Metashape achieved the highest surface reconstruction accuracy with the small Anchorage scene and large Mount Doran scene. . . . . 49

3.5 Quantitative results for Sat-NeRF with various positional encoding (PE) frequencies. The best metrics for each scene are in **bold**. For the task of novel view synthesis, the highest accuracy was achieved using a frequency 8 for both the Anchorage and Mount Doran scenes. For the task of 3D surface reconstruction, the highest accuracy was achieved using a frequency of 16 for the Anchorage scene and 8 for the Mount Doran scene. . . . . 53

3.6 Quantitative results for Sat-NeRF with ReLU and SIREN activation functions. The best metrics for each scene are in **bold**. For the task of novel view synthesis, the highest accuracy was achieved using SIREN activations for both the Anchorage and Mount Doran scenes. For the task of 3D surface reconstruction, the highest accuracy was achieved using ReLU activations for both scenes. . . . . 58

# Chapter 1

# Introduction

Remote sensing imagery offers a passive way to monitor various regions on Earth. Photogrammetric methods transform 2D images into 3D visual models, which can provide a more comprehensive representation of a particular target compared to the 2D images alone. The resulting 3D models can be used for quantifying landscape parameters, such as relief or surface texture [31], [32]. Traditional multi-view stereo (MVS) photogrammetry performs 3D reconstruction by identifying and matching feature points across multiple images [6]. However, these feature points are difficult to detect in areas with complicated lighting conditions and homogeneous physical characteristics, resulting in sparse and discontinuous reconstructions. Polar regions on Earth, which are critical targets for climate monitoring and environmental studies, exhibit high specularity due to the significant amount of light being reflected by snow and ice. Thus, traditional MVS methods often struggle to create accurate 3D models of the terrain [26]. Today, critical algorithmic gaps exist in traditional MVS methods that hinder their use for observing polar regions.

Thus, the scientific rationale driving this work is to advance the capabilities of Earth observation through the assessment of Neural Radiance Field (NeRF) methods for MVS photogrammetry of polar landscapes from satellite remote sensing data. NeRF offers a promising alternative to traditional methods by enabling continuous and volumetric scene representations. NeRFs break away from traditional photogrammetric assumptions of rigid surfaces, and instead, use multi-layer perceptron (MLP) networks to predict the radiance and density of points in 3D space [25]. Due to their ability at modeling dynamic conditions such as changing lighting and evolving terrain, these methods could potentially overcome the challenges posed by traditional MVS methods in reconstructing polar regions [23], [24]. By building on recent advancements in NeRF technology, this study aims to evaluate the effectiveness of NeRFs at generating accurate 3D models of polar landscapes from satellite remote sensing data. The results from this research will ultimately reveal whether NeRF serves as a plausible contender to the current state-of-the-art for scene reconstruction from satellite data in sparse-textured terrains.

## 1.1 Related Works

In recent years, the field of digital surface modeling has seen significant advancements, particularly with the introduction of neural rendering methods such as NeRF. Traditional MVS photogrammetric methods have long been the standard for reconstructing 3D scenes from multiple 2D images, relying heavily on the presence of distinct textures and colors in the environment to accurately map the geometry [6], [9]. However, these methods face challenges in polar regions, where uniform surfaces covered by snow and ice provide little to no texture for reliable reconstruction [26]. To address these limitations, researchers have explored neural-based approaches, including NeRF and its variants, which leverage volumetric rendering and deep learning to capture more detailed and accurate models of complex scenes. This chapter will delve into the existing body of work on neural rendering and 3D scene reconstruction, focusing on the evolution of techniques from traditional photogrammetry to modern NeRF-based methods, and highlighting studies that have applied these techniques to satellite remote sensing data.

### 1.1.1 Traditional Methods for 3D Surface Reconstruction

Traditional MVS photogrammetric methods, such as Structure-from-Motion (SfM), enable the creation of detailed digital surface models (DSMs) from sets of overlapping images. SfM aims to reconstruct the camera pose and scene geometry simultaneously through automatic identification and matching of features across multiple images. The overall SfM pipeline takes in 2 or more images, extracts key features from the images, matches these feature points across images, triangulates them into 3D space, and finally, performs bundle adjustment to minimize the reprojection error of the 3D points back onto the 2D image planes [36]. This process results in a sparse 3D point cloud, which can be used to render new views from previously unseen viewing angles.

Agisoft Metashape is an off-the-shelf, industry-standard software for generating 3D reconstructions from 2D remote sensing images. It employs SfM under-the-hood to produce accurate and high-resolution models of terrain. When using satellite images to reconstruct a scene in Metashape, the Rational Polynomial Coefficient (RPC) model for each image can be used as additional input to enhance the precision and spatial accuracy of the 3D data product. RPCs are native satellite camera models that establish correspondences between the 2D pixel space and the 3D world space using projection and localization functions. Compared to approximate pinhole camera models, the use of RPC camera models has been found to improve the performance of photogrammetric methods with satellite remote sensing images [7], [21]. They are commonly provided as auxiliary data with satellite imagery.

However, a major limitation of MVS methods are their reliance on stereo matching. Surface reconstruction based on feature matching between stereo images requires varying texture on the object surface [6]. Scenes that contain homogeneous physical characteristics often result in sparse and discontinuous reconstructions, as stereo matching techniques have difficulty differentiating between the extracted feature points in a set of images [26]. For example, polar regions on Earth appear relatively uniform in color and texture when viewed from a distance. The lack of distinct visual features can make it challenging for MVS methods like SfM to identify and match feature points for accurate 3D reconstruction.

Another limitation of these methods is that they do not model transmission nor reflection, which are crucial components when wanting to reconstruct scenes with high specularity. Polar regions on Earth exhibit high specularity due to the significant amount of light being reflected by snow and ice. This leads to overexposure in images, preventing SfM from accurately modeling how light interacts with the surface. Image overexposure from specular reflections can also contribute to inaccurate depth estimation of the surface due to the lack of distinctive features. While MVS photogrammetric methods like SfM have proven effective for satellite imagery of urban scenes, they face significant challenges when applied to polar remote sensing data. This necessitates the exploration of surface reconstruction algorithms that can accurately and comprehensively capture both the shape and reflectance properties of polar terrain.

### 1.1.2 Neural Radiance Fields

Neural Radiance Fields, or NeRFs, first introduced by Mildenhall et al., are the latest methods in 3D computer graphics [25]. NeRFs represent a static scene as a continuous volumetric function, or field, parameterized as basic MLP networks. The input to the field is a 5D coordinate $(x, y, z, \theta, \phi)$, containing the spatial location $(x, y, z)$ and viewing direction $(\theta, \phi)$ of points along a viewing ray, and the output is the volume density and view-dependent emitted radiance at that spatial location [25]. Novel views of the scene are then rendered by accumulating the color and density values along the viewing ray and projecting their output onto an image. The NeRF method optimizes the MLP network using a differentiable rendering procedure based on classical volumetric rendering techniques [25]. Specifically, NeRF uses gradient decent to minimize the error between each observed image and the corresponding views rendered from the estimated scene representation. By doing this, NeRF succeeds in creating photorealistic and volumetrically accurate 3D scenes.

A clear advantage of NeRF is that it breaks away from traditional MVS methods that rely on extracting sufficient feature descriptors from a set of images. Because the key technique of MVS photogrammetry is simply matching feature descriptors across images, these methods fail to model complex scene effects, such as transmission and reflection. NeRF overcomes this limitation and achieves a superior representation of a scene by simultaneously learning the color and geometry of points in 3D space, using only a set of images and their known camera poses. NeRF also employs techniques such as positional encoding to capture high-frequency scene details [25]. The accuracy of NeRF has been found to be comparable with MVS standards, specifically for satellite imagery of urban scenes [33].

### 1.1.3 NeRF for Satellite Photogrammetry

NeRF methods for satellite-based MVS photogrammetry have recently emerged, namely Shadow-NeRF (S-NeRF) [15], Satellite-NeRF (Sat-NeRF) [21], and Earth Observation NeRF (EO-NeRF) [22]. These methods propose a number of improvements to the original NeRF method, including an explicit light transport model that takes into account occlusion and varying illumination and an altitude-based point sampling strategy attuned to the characteristics of satellite imagery [15]. As a result, the color and geometry of the 3D scene is improved, particularly in shadowed areas that lack distinct visual features.

### Shadow NeRF

Shadow NeRF (S-NeRF), developed by Derksen et al. at the European Space Agency (ESA), is the first application of NeRF for multi-view satellite imagery [15]. It extends the basic NeRF method in two ways. First, it introduces a shadow-aware irradiance model, which predicts the intensity of incoming light at a 3D location. Instead of directly modeling the radiance, like in the original NeRF method, S-NeRF learns the total incoming light, or irradiance, and the albedo. A simple Lambertian reflectance model is then used to calculate the radiance of a point in 3D space. The S-NeRF irradiance model is dependent on the solar direction, which is added as a network input. By incorporating the learned irradiance and albedo values into the radiance calculation, as opposed to learning the color directly as a network output, S-NeRF produces more visually realistic scene representations. Additionally, the explicit prediction of albedo enables an independent synthesis of the scene with only that property. Thus, S-NeRF can also produce shadow-free renderings.

S-NeRF introduces an altitude-based sampling technique better attuned to Earth observation images captured from satellites. Because the distance from camera sensor to surface is much larger for satellite imagery, S-NeRF samples points along the absolute Z-axis, within a minimum and maximum altitude bound, instead of sampling points along the sensor axis. This technique greatly improves the depth estimation of the reconstructed scene, compared to the original NeRF method. The altitude bounds for a given scene must be explicitly defined, as they will vary from location to location. The experiments in the S-NeRF paper obtain the altitude bounds for a scene from airborne LiDAR data [15], but they can be obtained by other means as well. Overall, the combination of these features enable S-NeRF to work well with multi-view satellite imagery. Specifically, S-NeRF is able to accurately reconstruct the color and geometry of a scene using relatively few training images, images with an off-nadir angle smaller than 30°, and images with varying lighting conditions. By taking into account the solar direction, S-NeRF also succeeds in modeling the light and reflectance interactions of the learned scene.

### Satellite NeRF

Satellite NeRF (Sat-NeRF), developed by Marí et al. from the European Space Agency (ESA), incorporates well-known techniques for satellite image processing into a NeRF framework. In particular, Sat-NeRF represents each input image as a Rational Polynomial Camera (RPC) model, characteristic of satellite images, instead of a pinhole camera model commonly used in NeRF for close-range imagery [21]. RPC models consist of a set of polynomial functions that can translate between 2D pixel space and 3D scene space via projection and localization functions. Sat-NeRF directly uses these RPC models to cast rays into the scene space. An advantage of representing the input images as RPC models is that they provide independence from the satellite system, thus allowing them to describe complex acquisition systems without relying on the satellite-specific physical modeling [21]. A study conducted by the authors of Sat-NeRF found that this more general approach leads to better results, compared to S-NeRF, which uses an approximate pinhole camera model.

Sat-NeRF further refines the RPC model for each image by means of bundle adjustment, which minimizes the reprojection error of points in the 3D scene. Bundle adjustment is commonly done in MVS photogrammetry to correct RPC inconsistencies. Before training a Sat-NeRF model, bundle adjustment can be applied to correct the RPC models of all input images. The authors of Sat-NeRF found that the use of bundle adjusted RPC models directly increased the accuracy of the point sampling strategy [21]. Lastly, like S-NeRF, Sat-NeRF assumes a Lambertian surface to simulate realistic lighting by accounting for the angle-dependent reflectance of light rays. The combination of these factors enables Sat-NeRF to render new views and infer surface models of similar quality to those obtained with traditional state-of-the-art MVS pipelines.

**Earth Observation NeRF**

Earth-Observation NeRF (EO-NeRF) improves upon S-NeRF and Sat-NeRF by using a geometrically consistent shadow model [22]. Previous satellite-based NeRF variants predict the shadows in a scene as a color property depending on the solar direction. However, EO-NeRF renders them according to the geometry and the position of the Sun at each optimization step [22]. In addition, EO-NeRF uses Universal Transverse Mercator (UTM) coordinates to handle geo-referenced data in a more appropriate way, as opposed to using Earth-centered Earth-fixed (ECEF) coordinates as in S-NeRF and Sat-NeRF. The use of UTM coordinates offers the advantage of having the altitude of the scene be aligned with the Z-axis [22]. These key features enable EO-NeRF to outperform the altitude accuracy of state-of-the-art MVS methods and previous satellite-based NeRF approaches.

However, a notable limitation of S-NeRF, Sat-NeRF, and EO-NeRF is that they were tested and evaluated using Earth observation scenes of urban environments, which have defined edges from buildings and considerable color variation and surface texture. Such urban scenes are considered to be idealized data, as their visual characteristics make it easier for traditional photogrammetry methods to differentiate between points in a scene and create geometrically accurate surface models. Although this data is sufficient for initial performance evaluation, it does not capture the full extent of the abilities of these methods. Urban scenes account for only a fraction of all Earth observation scenes. Thus, to better assess the feasibility of satellite-based NeRF methods in a broader remote sensing context, they must be tested with more diverse and challenging data.

## 1.2 Research Objectives

The objective of this research is to assess the feasibility of satellite-based NeRF methods at generating digital surface models of polar landscapes on Earth, from multi-view satellite imagery. These areas appear relatively uniform in color and texture when viewed from a distance, thus making it difficult for traditional MVS methods to accurately reconstruct the terrain. Due to this limitation, the current research explores alternative methods that use neural volumetric rendering techniques, as opposed to standard feature-based matching, for 3D surface reconstruction. The objectives of this study are listed below. The following sections will discusses these objectives in greater detail.

1. Compile high-resolution datasets of WorldView-2 ($\leq$ 2 m/pxl) satellite images of polar regions.
2. Test a traditional MVS method using high-resolution polar datasets.
3. Test S-NeRF and Sat-NeRF using high-resolution polar datasets.

Overall, this research aims to answer the question: How do satellite-based NeRF methods S-NeRF and Sat-NeRF perform at novel view synthesis and surface reconstruction using sparsely-textured satellite remote sensing data of polar landscapes, compared to a traditional MVS method?

**Research Objective 1: Compile High-Resolution Satellite Datasets of Polar Regions**

The first objective of this research is to compile datasets of high resolution ($\leq$ 2 m/pxl) WorldView-2 satellite imagery of polar regions across Alaska. The scenes of interest for this objective are Mount Doran located in Chugach, Alaska and the University of Alaska Anchorage located in Anchorage, Alaska. These scenes were chosen because they exhibit minimal color variation and surface texture due to snow on the ground. The Anchorage scene presents an easier reconstruction problem, as it is semi-urban and contains the sharp edges from buildings that traditional MVS methods rely on. The Mount Doran scene, on the other hand, appears much more uniform in color and texture, presenting a more challenging reconstruction problem. Testing both datasets will reveal whether satellite-based NeRF methods posses the same limitations as traditional MVS methods with reconstructing feature-sparse scenes.

**Research Objective 2: Test a Traditional MVS Photogrammetry Method**

The second research objective is to test Agisoft Metashape with both polar datasets. Metahsape is one of the leading commercial software tools for scene reconstruction from remote sensing data. It employs SfM under-the-hood, a traditional MVS photogrammetric method, to generate high-quality 3D models from aerial or satellite imagery. Metashape will serve as the baseline method in this research. Due to its ability at leveraging the RPC camera models of satellite images, experiments will be run with and without the RPC data. This will reveal the influence of RPC data in constructing the scene from an SfM approach, and further, whether or not Metashape can achieve accurate 3D reconstructions from images alone. It is important to note that Metahshape only reconstructs the scene geometry. The color of each point in 3D space is merely averaged from the projected pixel values used as input. Thus, at the current stage in this research, evaluation is done only on the surface reconstruction results of the generated sparse point clouds.

**Research Objective 3: Test S-NeRF and Sat-NeRF**

The third research objective is to test satellite-based NeRF methods S-NeRF and Sat-NeRF with both polar datasets. Doing so will enable a quantitative assessment of NeRF performance at novel view synthesis and 3D surface reconstruction using data that has been historically challenging for traditional MVS photogrammetric methods. This research will provide valuable insight into the potential advantages

and limitations of these approaches in a broader geo-spatial context and ultimately reveal whether NeRFs are feasible in such an application. For novel view synthesis, the performance of S-NeRF and Sat-NeRF will be evaluated on the image quality scores of rendered novel views, as measured by the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). For surface reconstruction, the performance of S-NeRF and Sat-NeRF will be evaluated on the quality of the shape of the learned scene at its maximal altitude, computed as the average distance in meters between the predicted and ground-truth point clouds. Both the predicted and ground-truth point clouds will be derived from DSMs and used to evaluate the overall learned scene geometry.

# Chapter 2

# Data Collection

Structure-from-Motion (SfM), Shadow NeRF (S-NeRF), and Satellite NeRF (Sat-NeRF) will be tested using high-resolution satellite imagery of snow-covered urban and mountainous polar terrain. The University of Alaska Anchorage was chosen as the study area for the urban scene, while Mount Doran in the Chugach Range of Alaska was selected for the mountainous polar scene. Both locations were specifically chosen due to their consistent snow coverage, which minimizes the color variation and surface texture in the satellite images. This creates a controlled yet challenging environment for evaluating 3D reconstruction methods. The uniform surface characteristics of the snow-covered terrain will allow for a focused assessment of how well these methods model the geometry of feature-sparse scenes.

The data utilized in this study consists of high-resolution ($\leq$ 2 m/pxl) WorldView-2 satellite imagery and high-density point clouds derived from ground-truth LiDAR digital surface maps (DSMs). The Geospatial Data Abstraction Library (GDAL) software was used to crop the raw multi-spectral 8-band WorldView-2 satellite images and extract their RGB bands [17]. The associated RPC models for each image were updated to represent the extent of the cropped images. This processed data serves as the primary input for the SfM, S-NeRF, and Sat-NeRF methods. The high-density ground-truth point clouds were used to quantitatively evaluate the performance of the explored methods at 3D surface reconstruction.

## 2.1 Data Sources

The methods in this research were tested with high-resolution imagery from the WorldView-2 satellite. WorldView-2 collects panchromatic images at 0.41 m/pxl and multi-spectral images at 1.64 m/pxl. Along with the panchromatic (450-800 nm) and the four typical multi-spectral bands (blue, green, red, NIR), WorldView-2 has four additional bands for enhanced multi-spectral analysis (coastal blue, yellow, red edge, NIR2) [5]. For this research, WorldView-2 Basic All Bands (BAB) data products were downloaded from the Maxar Global-Enhanced GEOINT Delivery (G-EGD) service. This specific data product includes all the spectral bands captured by the satellite sensor, without any additional processing or enhancements. The associated image metadata is also provided in the BAB data product, including the camera azimuth and elevation, Sun azimuth and elevation, and RPC camera models.

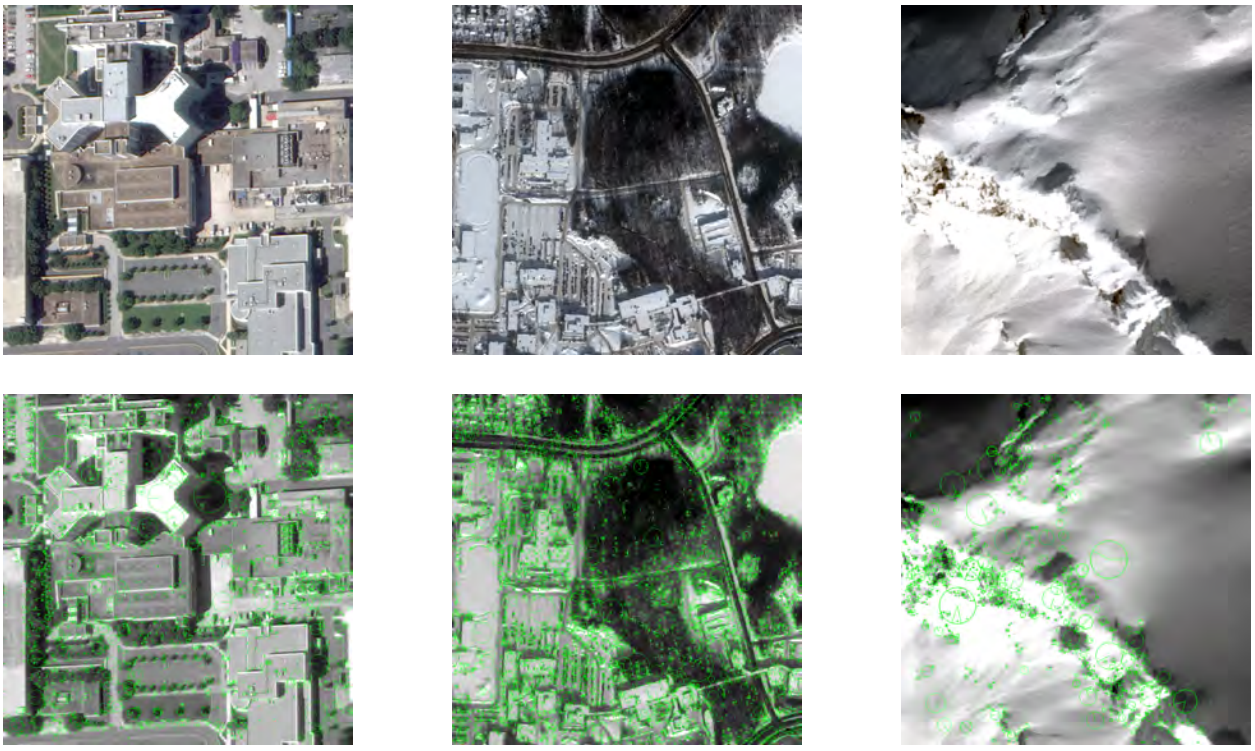
The high-resolution ground-truth LiDAR DSMs were provided through the Geospatial-Intelligence Access and Information Sharing (GEOAxIS/Gx) services. The LiDAR DSMs have $\leq$ 2 m/pxl ground sampling distance (GSD) with elevation values representing the height in meters relative to the World Geodetic System 1984 (WSG 84) ellipsoid. Point clouds were derived from the LiDAR DSMs for use in evaluating the sparse point clouds produced by each method. To derive a point cloud from a LiDAR DSM, each pixel from the DSM raster is mapped to one ground point in the point cloud. Thus, it is expected that the derived ground-truth point clouds will contain holes in areas with vertical features, such as building edges, where multiple surfaces may overlap. This is a known limitation of LiDAR-derived data products. Overall, this ground-truth data enables a quantitative measurement of the 3D surface reconstruction performance of the explored methods.

## 2.2 Study Areas

Data was collected for two scenes. The first scene is the University of Alaska Anchorage located in Achorage, Alaska. This area is characterized by a mix of urban structures, open spaces, and consistent snow coverage, presenting a rather easier reconstruction problem due to the sharp edges from buildings and varying surface elevations. However, the snow coverage visually homogenizes the scene, reducing the contrast between different surfaces and making it challenging to distinguish between objects and ground features. This uniformity in surface appearance will test the ability of each reconstruction method at accurately capturing and modeling fine details in urban polar landscapes, such as building outlines and elevation changes.

The second scene is Mount Doran, located in the Chugach Range of Alaska. This region presents a complex terrain with minimal surface features and extensive snow cover, making it an ideal testbed for assessing the robustness of satellite-based NeRF methods in modeling mountainous polar environments. The consistent snow coverage at Mount Doran minimizes color variation and surface texture, emphasizing the challenge of reconstructing the finer geometric details of the landscape. The homogeneous appearance of the snow-covered terrain will require the reconstruction methods to rely heavily on subtle elevation cues and shading variations to accurately model the terrain. This scene is particularly valuable for testing the ability of each method at handling the complexities of natural, rugged environments where traditional features like sharp edges and distinct textures are less prominent.

To further illustrate the feature-sparse nature of these terrains, the OpenCV software library [18] was used to detect and extract key features from a sample image in each dataset. These results are compared against a sample image from the Jacksonville dataset originally used to test S-NeRF and Sat-NeRF [15], [21]. OpenCV detected 5,090 features in the Jacksonville scene, 4,786 features in the Anchorage scene, and only 1,467 features in the Mount Doran scene. A visual comparison of these results is provided in Figure 2.1. The Scale Invariant Feature Transform (SIFT) method [20] was used for feature detection in this example, as it is the most widely used method for describing local features in images.



(a) Jacksonville, FL (b) Anchorage, AK (c) Mount Doran, AK

Figure 2.1: Comparison of features detected in the Jacksonville, Anchorage, and Mount Doran scenes. The original images are shown in the top row, and the corresponding features detected are shown in the bottom row. OpenCV detected 5,090 features in Jacksonville, 4,786 in Anchorage, and 1,467 in Mount Doran.

## 2.3 Data Preparation

Datasets were prepared for each scene. First, the RGB bands were extracted from the multi-spectral 8-band satellite images. Then, the images were cropped to the extent of ground-truth LiDAR DSMs. Lastly, the RPC models for each image were updated to account for the cropping. It is important to note that only the images with consistent snow coverage were chosen during the data collection process. This was done to minimize variations in surface depth from snowfall, ensuring that each dataset captures a similar terrain. It is a known limitation that multi-date datasets of polar regions will have some variations in surface depth due to snowfall. However, the goal for this research was to collect a dataset that minimizes any potential source of error from such inconsistencies, while also maintaining a sufficient distribution of viewing angles and dataset size. Therefore, although surface depth variations may exist from image to image, an effort was made to compile the most consistent dataset as possible for each scene.

In addition to the terrain type, urban or mountainous, this research also aims to examine how AOI size influences the accuracy of the reconstructed scene, as S-NeRF and Sat-NeRF were originally tested with urban environments limited to $300\ m^2$. Thus, three datasets were compiled for each scene, covering small, medium, and large AOI sizes. Details on the ablation study examining the influence of AOI size on the best-performing satellite-based NeRF method can be found in Section 3.1.3. For the Anchorage scene, the AOI sizes covered $300\ m^2$, $900\ m^2$, and $1,500\ m^2$. For the Mount Doran scene, the AOI sizes covered $600\ m^2$, $1,200\ m^2$, and $2,000\ m^2$. By exploring various AOI sizes, the optimal dataset parameters for accurate surface reconstruction can be determined, facilitating more effective applications in polar contexts.

### 2.3.1 Pre-processing Satellite Imagery

As mentioned previously, data pre-processing involved extracting the RGB bands from the 8-band multi-spectral satellite images and cropping them to the extent of a LiDAR DSM. Figures 2.2-2.4 show the clipping extent for the small, medium, and large cropped Anchorage datasets, and Figures 2.5-2.7 show the clipping extent for the small, medium, and large cropped Mount Doran datasets. After image pre-processing, the RPC models were updated to represent the extent of the cropped images. The GDAL and RPCM libraries were used to perform these adjustments. GDAL (Geospatial Data Abstraction Library) was utilized for handling the geo-spatial data and performing the image cropping, while the RPCM (Rational Polynomial Coefficient Manipulation) library was employed to embed the updated RPC models into the cropped images. These tools ensured that geo-referencing was preserved for each image after cropping, which is necessary when generating a point cloud in Metashape with RPC models.

In addition to cropping the images and extracting the visible RGB bands, pre-processing was also done to down-sample and rotate the images to be coherent with NeRF training. The original satellite images are north-aligned, but some may be rotated by 90° or more, making them misaligned. In these cases, the images are rotated based on their azimuth value to correct their orientation. After this initial correction, the images are further rotated by 180° plus their azimuth value. This ensures all images are forward-facing and align with the forward direction of the satellite. Finally, the images in each dataset are down-sampled with anti-aliasing to ensure the X and Y cell size of each pixel matches the GSD of the ground-truth LiDAR DSM used during evaluation.

The Anchorage datasets include 10 training images and 2 testing images, with off-nadir angles between $0.9° \leq x \leq 23.5°$ and a GSD of 1.0 m/pxl. The distribution of camera and solar angles are presented in Figure 2.8. Three datasets were collected for small, medium, and large AOIs. After pre-processing, the image resolutions were $300\ pxl^2$, $600\ pxl^2$, and $1,200\ pxl^2$, covering approximately $300\ m^2$, $600\ m^2$, and $1,200\ m^2$ AOIs, respectively. A sample images from each dataset can be seen in Figure 2.9.

The Mount Doran datasets include 8 training images and 2 testing images, with off-nadir angles between $18.4° \leq x \leq 28.8°$ and a GSD of 2.0 m/pxl. The distribution of camera and solar angles are presented in Figure 2.10. Three datasets were collected for small, medium, and large AOIs. After pre-processing, the image resolutions were $300\ pxl^2$, $600\ pxl^2$, and $1,000\ pxl^2$, covering approximately 600

$m^2$, $1,200\ m^2$, and $2,000\ m^2$ AOIs, respectively. A sample image from each dataset can be seen in Figure 2.11.

Figure 2.2: The clipping extent for the small Anchorage dataset. For this dataset, the pushbroom images were cropped to the extent of the small AOI ($300 \times 300$ m).

Figure 2.3: The clipping extent for the medium Anchorage dataset. For this dataset, the pushbroom images were cropped to the extent of the medium AOI (600 × 600 m).

Figure 2.4: The clipping extent for the large Anchorage dataset. For this dataset, the pushbroom images were cropped to the extent of the large AOI ($1,200 \times 1,200$ m).

Figure 2.5: The clipping extent for the small Mount Doran dataset. For this dataset, the pushbroom images were cropped to the extent of the small AOI ($600 \times 600$ m).

Figure 2.6: The clipping extent for the medium Mount Doran dataset. For this dataset, the pushbroom images were cropped to the extent of the medium AOI ($1,200 \times 1,200$ m).

Figure 2.7: The clipping extent for the large Mount Doran dataset. For this dataset, the pushbroom images were cropped to the extent of the large AOI ($2,000 \times 2,000$ m).

### 2.3.2 Adjusting Image RPC Models

A Rational Polynomial Coefficient (RPC) model is a mathematical representation that relates 2D pixels coordinates in an image to 3D ground coordinates in world space. Each RPC is defined as a set of third-degree polynomial coefficients. These coefficients are used to map between positions on the image plane and corresponding points in the world, specified by latitude, longitude, and altitude. The RPC model is widely used for optical satellite imagery, as it can describe complex acquisition systems independently from satellite-specific physical modeling [21]. As opposed to an approximate pinhole camera model, the RPC model can capture complex lens distortions and nonlinear relationships between image and world spaces, allowing for accurate transformations between the two. This improved camera model offers greater flexibility and precision in satellite image geo-referencing and correction.

An RPC model is representative of an entire pushbroom image scan and must be updated if the image is cropped. Failing to update the RPCs after cropping can lead to inaccuracies in geo-referencing and distortions in the spatial positioning of the image. However, there is minimal publicly available information on how exactly the RPC model must be updated after an image crop. Many tools were explored for cropping satellite images and updating their RPC information, including the NASA Ames Stereo Pipeline (ASP) [8], ArcGIS [3], ENVI [29], and GitHub repositories RPCM and RPC Cropper [16], [27]. To confirm if the tools used successfully updated the RPC models, a sparse point cloud was

generated in Metashape using the cropped images and updated RPCs. Unfortunately, in all cases, the tools aforementioned produced either excessively noisy clouds or no cloud at all, indicating that the RPC model was incorrectly adjusted.

After consulting with one of the authors of the 2019 IEEE Data Fusion Competition (DFC) dataset [19], which contains cropped satellite images with correctly adjusted RPC models, it was found that, to update the RPC models based on an image crop, the line and sample offset values in the RPC must be subtracted from the pixel coordinates of the top-left corner of the crop. This method successfully corrected the RPC models, as indicated by significantly less noise and consistent spatial positioning in the Metashape-generated point clouds. These adjustments ensured that the spatial relationships captured by the RPCs remained accurate for the cropped satellite images.

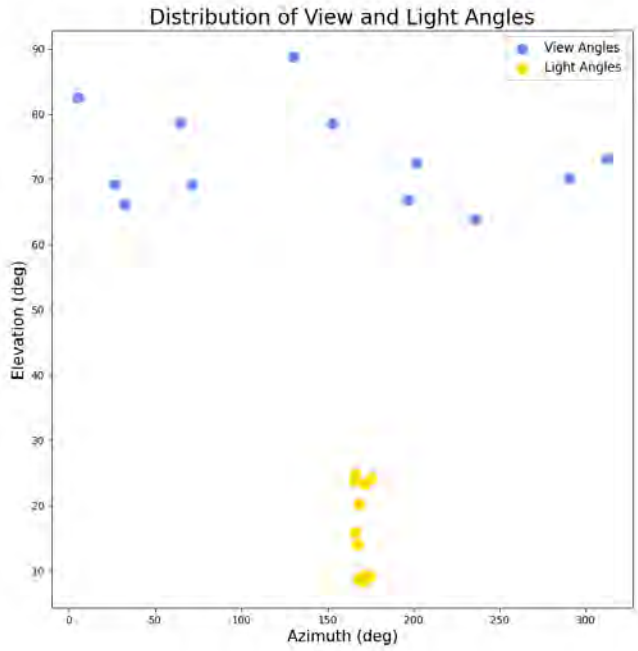



Figure 2.8: The distribution of view and light angles for the Anchorage dataset.

Small Dataset Medium Dataset Large Dataset

Figure 2.9: Sample images from the small, medium, and large Anchorage datasets.

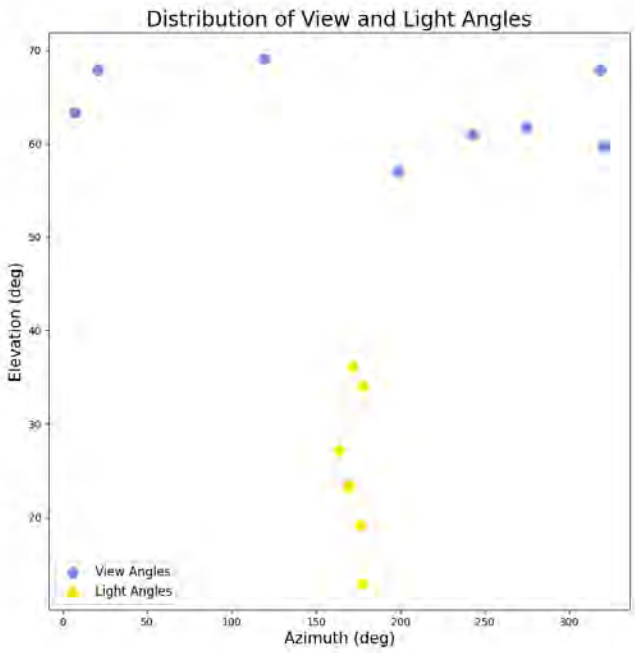
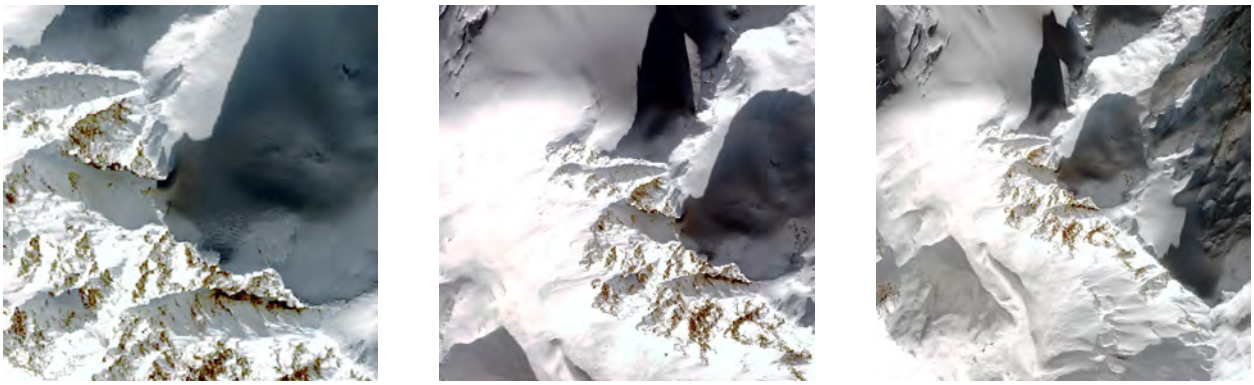



Figure 2.10: The distribution of view and light angles for the Mount Doran dataset.



Small Dataset Medium Dataset Large Dataset

Figure 2.11: Sample images from the small, medium, and large Mount Doran datasets.

# Chapter 3

# Reconstructing Polar Regions Using NeRF

In this section, traditional and NeRF-based photogrammetric methods will be tested with satellite imagery of a mountainous polar region and an urban snow-covered region. This research will reveal how an off-the-shelf, industry-standard MVS method performs at 3D surface reconstruction using satellite imagery of feature-sparse polar terrain, and further, whether NeRF methods can overcome the limitations of MVS methods in keypoint detection caused by snow and ice-covered surfaces. Agisoft Metashape will be used to asses the performance of an industry-standard MVS technique at 3D reconstruction, while NeRF variants S-NeRF and Sat-NeRF will be used to asses the capabilities of neural volumetric rendering techniques at both tasks of 3D reconstruction and novel view synthesis.

To further refine the most effective satellite-based NeRF method, a series of ablation studies will be conducted, focusing on three core variables: scene size (AOI size), activation function, and positional encoding frequency. Scene size directly pertains to the data, as the area of interest influences the scale and variability of the imagery used, potentially affecting how well the model generalizes across different environments. In contrast, the activation function and positional encoding frequency relate to architectural elements within the NeRF model itself. These modifications impact how the model represents and learns intricate surface details, influencing the accuracy of the reconstructions.

By isolating and studying these factors, the ablation studies will reveal how data properties and model architecture influence the accuracy and detail of the reconstructed surfaces. This comprehensive analysis will allow for the identification of optimal configurations, ensuring that the NeRF method is tailored to meet the unique challenges of polar terrain and snow-covered environments, where traditional approaches may fail. The overall goal of this chapter is to provide a comparative assessment of surface reconstruction methods for polar climates, and further, establish a framework for improving neural volumetric rendering approaches in such challenging regions.

## 3.1 Experimental Design

Metashape, S-NeRF, and Sat-NeRF will each be tested tested using WorldView-2 ($\leq$ 2 m/pxl) satellite imagery from two scenes:

1. University of Alaska Anchorage, located in Anchorage Alaska
2. Mount Doran, located in Chugach Alaska

The images in each dataset will contain only the visible RGB bands and be clipped to the extent of a desired AOI. Details on data pre-processing are provided in Section 2.3. In these initial experiments, the AOI size used across all three methods will be 300 $m^2$ for the Anchorage scene and 600 $m^2$ for the Mount Doran scene. Testing both the Mount Doran and Anchorage datasets will reveal if the additional features present in urban scenes improve the final 3D scene reconstruction, and further, whether NeRF-based methods possess the same limitations as traditional MVS methods in reconstructing feature-sparse, snow-covered surfaces. Additionally, testing each method with the same datasets will allow for a direct comparison of their performance, highlighting differences in reconstruction fidelity and robustness to variations in terrain texture.

### 3.1.1 Metashape

First, Metashape will be tested with each dataset. Experiments will be run with and without the RPC camera models to determine their impact on the accuracy of the 3D scene reconstructions. These results will reveal whether the inclusion of RPC models in Metashape enhances the alignment and feature matching processes with satellite imagery of polar terrain. Furthermore, results will be compared across datasets to gain insight into how effectively RPC models facilitate 3D reconstruction in different polar environments, thereby informing best practices for applying traditional MVS photogrammetric techniques in diverse geographical contexts. The performance of Metashape will be quantitatively evaluated using one metric:

1. Mean distance between the predicted and ground-truth point clouds, in meters

The data product output by Metashape is a sparse point cloud of the target scene, thus only the task of 3D surface reconstruction can be evaluated. The information used for evaluation will be a point cloud derived from a ground-truth LiDAR DSM. The LiDAR DSMs have a resolution of 1 m/pxl for the Anchorage scene and 2 m/pxl for the Mount Doran scene. To derive a point cloud from a LiDAR DSM, each pixel from the DSM raster is mapped to one ground point in the point cloud. Thus, it is expected that the derived ground-truth point clouds will contain holes in areas with vertical features, such as building edges, where multiple surfaces may overlap. This is a known limitation of LiDAR-derived data products. Future work should include evaluating Metashape on the task of novel view synthesis, in addition to the task of surface reconstruction, to allow for a complete comparison with NeRF-based methods.

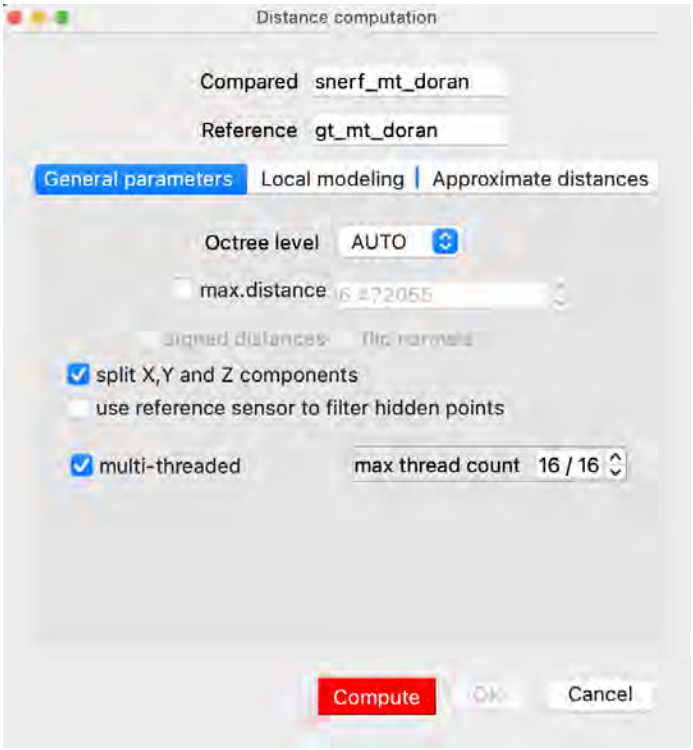



Figure 3.1: Screenshot from the "General Parameters" tab of the Cloud-to-Cloud Distance Computation tool in CloudCompare.

To calculate the mean distance between the predicted and ground-truth point clouds, the Cloud-to-Cloud Distance Computation tool from the CloudCompare software [1], [14] will be used, with the following parameters set in Figures 3.1 and 3.2. The X, Y, and Z components of the distance computation will be split up to generate 3 separate scalar fields. The X, Y, and Z scalar fields correspond to the absolute distance between each compared point and its nearest reference point along each dimension. For evaluation, only the absolute distance along the Z-dimension will be used, as the point clouds output by S-NeRF and Sat-NeRF represent the depth of the scene at its maximal altitude. Additionally, S-NeRF and Sat-NeRF primarily aim to improve vertical accuracy, particularly in reconstructing terrain elevation. This focus is due to the importance of altitude in satellite imagery, where variations in elevation significantly affect surface geometry and shadow patterns. By isolating the Z-dimension, the evaluation will directly assess the ability of each model at capturing elevation change and surface depth.

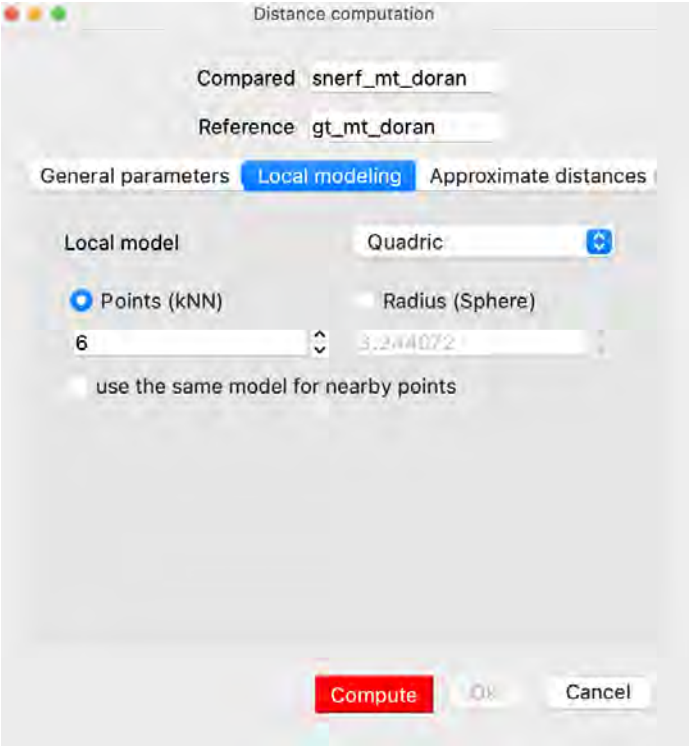



Figure 3.2: Screenshot from the "Local Modeling" tab of the Cloud-to-Cloud Distance Computation tool in CloudCompare.

Additionally, the Quadric local modeling strategy will be used. When no local model is used, the cloud-to-cloud distance is simply the nearest neighbor distance 3.1. However, in most cases, the nearest neighbor is not the actual nearest point on the surface represented by the reference cloud. The Quadric local modeling strategy mitigates inaccuracies in cloud-to-cloud distances by fitting a quadric surface to the local neighborhood of points in the reference cloud. This surface better approximates the underlying structure, allowing for a more accurate calculation of distances between clouds. The Quadric local model is especially useful when the reference cloud has a low density or contains sparse areas, as it can more accurately represent the local surface geometry than nearest neighbor approaches. By approximating the surface curvature in these regions, it enables a more reliable distance calculation even in ares where points are less densely packed. Furthermore, this modeling strategy enhances the accuracy of global statistics, such as the mean distance and standard deviation, by providing a refined estimate of the true surface-to-surface distance across the entire cloud. The official CloudCompare Wiki recommends using the Quadric local model by default, as it is the most versatile [2].

### 3.1.2 S-NeRF and Sat-NeRF

Next, S-NeRF and Sat-NeRF will be tested with each dataset. These initial experiments will have the default configurations set, as described in the papers [15], [21]. The performance of S-NeRF and Sat-NeRF will be quantitatively evaluated using three separate metrics:

1. Mean Peak Signal-to-Noise Ratio (PSNR) of novel views
2. Mean Structural Similarity Index (SSIM) of novel views
3. Mean distance between the predicted and ground-truth point clouds, in meters

The two sources of ground-truth information used for evaluation will be test images in unseen viewing and lighting conditions and point clouds derived from ground-truth LiDAR DSMs. First, S-NeRF and Sat-NeRF will be evaluated on the task of novel view synthesis using the peak signal-to-noise ratio (PSNR). PSNR is a pixel-based metric used to evaluate the quality of a reconstructed image signal. It measures the difference between the original signal and the reconstructed version, in terms of both the signal fidelity and the amount of noise introduced during training. PSNR is calculated as the ratio between the maximum possible power of a signal and the power of the noise that affects the fidelity of the signal. The higher the PSNR value, the better the quality of the signal.

S-NeRF and Sat-NeRF will also be evaluated using the structural similarity index (SSIM) scores. SSIM is a metric used to quantify the similarity between two images. Compared to pixel-to-pixel evaluation metrics like PSNR, SSIM accounts for the way human vision perceives structural information, making it a more reliable measure when evaluating complex scenes with varying illumination. It takes into account the structural information, luminance, and contrast of the images to provide a more comprehensive assessment of their similarity. The SSIM index ranges from -1 to 1, where a value of 1 indicates a perfect match between the two images. SSIM is also used to evaluate the performance of novel view synthesis.

Lastly, S-NeRF and Sat-NeRF will be evaluated on the task of 3D surface reconstruction using the mean distance between the predicted and ground-truth point clouds. To ensure each point in the generated cloud is being matched to the correct point in the ground-truth cloud for distance computation, the S-NeRF and Sat-NeRF DSMs are XY geo-referenced to the ground-truth LiDAR DSMs before evaluation in CloudCompare. This step successfully aligns the NeRF clouds to the ground-truth and minimizes any spatial discrepancies that could affect the accuracy of the evaluation. After XY geo-referencing, the S-NeRF and Sat-NeRF point clouds are evaluated following the same approach outlined in 3.1.1 with the Metashape method.

### 3.1.3 Ablation Studies

The following experiments will investigate the impact of various parameters on novel view synthesis and surface reconstruction performance using the best-performing NeRF method. The best-performing NeRF method will be determined as the method which achieves the lowest mean distance error and highest mean PSNR and SSIM scores on test data, for both the Anchorage and Mount Doran datasets. The first

ablation study will determine the influence of increasing the AOI size on the reconstruction accuracy. For the Anchorage scene, AOIs covering 300 $m^2$, 600 $m^2$, and $1,200$ $m^2$ will be tested, while for the Mount Doran scene, AOIs covering 600 $m^2$, $1,200$ $m^2$, and $2,000$ $m^2$ will be tested. This systematic exploration of different AOI sizes aims to determine how varying the spatial extent of the data influences the accuracy and detail of the 3D reconstructions produced by NeRF, ultimately contributing to a deeper understanding of its generalization to larger scenes.

In addition to conducting an ablation study on AOI size, the impact of various activation functions and Fourier feature frequencies in positional encoding will be explored. Specifically, the study will compare the effects of ReLU and SIREN activation functions for each dataset [4], [28]. These two activation functions will be evaluated to determine how they influence surface reconstruction accuracy and detail in the learned scene volume. Furthermore, various positional encoding frequencies will be tested to assess their ability at capturing fine details and enhancing the rendered image fidelity. Each time a parameter is changed, a new experiment will be run. Parameters will be changed one at a time to isolate their influence on the results. Through these ablation studies, the optimal model configuration for each polar scene will be determined.

It is important to note that the first ablation study presented in this paper, which uses different datasets of varying AOI sizes, will be tested using the default training and network configurations. Additionally, the ablation studies involving direct modifications to the network architecture will be tested using the dataset containing smallest AOI size for each scene (300 $m^2$ and 600 $m^2$ for Anchorage and Mount Doran, respectively). The smallest AOI sizes were chosen for testing because they most closely align with the AOI sizes of datasets originally used to test S-NeRF and Sat-NeRF [15], [21]. In summary, the following ablation studies will be conducted for each scene on the best-performing NeRF model:

1. Scene size (small, medium, large)
2. Positional encoding frequency (0, 8, 16)
3. Activation function (ReLU, SIREN)

## 3.2 Metashape

Metahshape is an industry-standard software for generating 3D reconstructions from 2D remote sensing images. It employs Structure-from-Motion (SfM) under the hood, which is a traditional MVS photogrammetry technique. Thus, it will serve as the baseline method in this research. Metashape was first tested with the Anchorage dataset, covering an AOI of approximately 300 $m^2$. This scene presents an easier reconstruction problem, as it is semi-urban and contains the sharp edges from buildings that SfM relies on. The results for the Anchorage dataset, with and without the RPC models, are presented in Figures 3.3 - 3.4. Metashape achieved a mean distance error of 25.374 m without the RPC models and 2.652 m with the RPC models. These results indicate, both qualitatively and quantitatively, that the inclusion of the RPC data for each image significantly improves the final point cloud. Thus, to generate an accurate

digital surface model of a snow-covered urban landscape in Metashape, the RPC model for each satellite image must be used as additional input.

(a) Ground-Truth (b) Sparse Point Cloud (c) Dense Point Cloud

Figure 3.3: Results for Metashape using the small Anchorage dataset ($300 \times 300$ m) **without** the RPC models. Metashape achieved a mean distance of 25.374 m for the generated sparse point cloud.

(a) Ground-Truth (b) Sparse Point Cloud (c) Dense Point Cloud

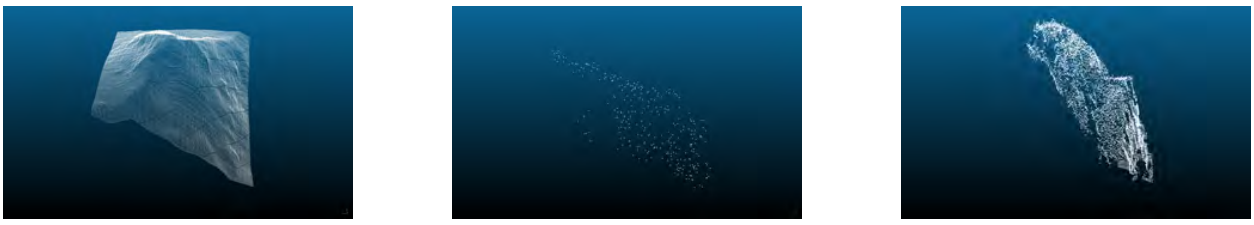
Figure 3.4: Results for Metashape using the small Anchorage dataset ($300 \times 300$ m) **with** the RPC models. Metashape achieved a mean distance of 2.652 m for the generated sparse point cloud.

Metashape was then tested with the Mount Doran dataset, covering an AOI of approximately 600 $m^2$. This scene presents a more challenging reconstruction problem, as it has uniform snow coverage and steep elevation changes, leading to a lack of discernible features for accurate surface reconstruction. The results for the Mount Doran dataset, with and without the RPC models, are presented in Figures 3.5 - 3.6. Metashape achieved a mean distance error of 84.061 m without the RPC models and 15.624 m with the RPC models. Similar to the Anchorage scene, the generated point cloud of the Mount Doran scene contains predominantly noise when only cropped satellite images are used as input. However, the point cloud showed significant improvement with the inclusion of the RPC models. These findings indicate that, regardless of the terrain type, generating an accurate point cloud in Metashape using satellite imagery of polar regions requires the corresponding RPC data.

The higher distance errors for the Mount Doran dataset are likely due to the lack of feature in the scene. Because Metashape uses SfM under-the-hood for its processing, it relies heavily on identifying distinct features to match across multiple images. The homogeneous surface characteristics of the Mount Doran scene, with its snow-covered terrain and minimal texture variation, provide fewer keypoints for accurate feature matching. As a result, the SfM algorithm struggled to generate a dense 3D reconstruction, even with the inclusion of the RPC models. Additionally, the Mount Doran dataset exhibits a stronger non-correlation between images due to changing lighting conditions and variations in snow coverage, compared to the Anchorage dataset. This further prevents SfM from finding reliable keypoint matches, leading to higher distance errors in the final point cloud.

(a) Ground-Truth (b) Sparse Point Cloud (c) Dense Point Cloud

Figure 3.5: Results for Metashape using the small Mount Doran dataset (600 × 600 m) **without** the RPC models. Metashape achieved a mean distance of 84.061 m for the generated sparse point cloud.



(a) Ground-Truth (b) Sparse Point Cloud (c) Dense Point Cloud

Figure 3.6: Results for Metashape using the small Mount Doran dataset (600 × 600 m) **with** the RPC models. Metashape achieved a mean distance of 15.624 m for the generated sparse point cloud.

Table 3.1: Quantitative results for Metashape. The best metrics for each scene are in **bold**. Metashape achieved the highest reconstruction accuracies for both scenes with the use of the RPC models. The generated point cloud of Anchorage had an average distance error of 2.652 m, while the generated point cloud of Mount Doran had an average distance error of 15.624 m.

| Scene | Input | Avg. Dist. (m)↓ | Std. Dev. (m) |
|---|---|---|---|
| Anchorage | Images | 25.374 | 20.6623 |
| | Images + RPC Models | **2.652** | **4.392** |
| Mt Doran | Images | 84.061 | 53.2325 |
| | Images + RPC Models | **15.624** | **16.951** |

## 3.3 S-NeRF

S-NeRF is the first application of NeRFs for the task of satellite photogrammetry. It introduces a shadow-aware irradiance model for accurately reconstructing scenes with varying shadows and an altitude sampling technique better attuned to Earth observation images captured from satellites. The ability of S-NeRF to accurately reconstruct shadowed areas with minimal color variation and surface texture makes it well-suited for reconstructing polar regions with similar characteristics. In this research, S-NeRF was tested with the Anchorage and Mount Doran datasets to asses its abilities at reconstructing polar scenes from multi-vew satellite imagery. The results for the Anchorage dataset are presented in Figures 3.7 - 3.9. For the task of novel view synthesis, the S-NeRF model achieved mean PSNR/SSIM scores of 22.13/0.578 on train data and 10.07/0.139 on test data in unseen viewing and lighting conditions. This indicates that S-NeRF has adequate performance at rendering novel views of snow-covered urban scenes. Noise is observed in the rendered novel views, but the model is capturing the most prominent scene features. However, S-NeRF failed at the task of 3D surface reconstruction. As shown in Figure 3.9, despite achieving a mean distance error of 2.883 m, which is on par with the performance of Metashape for this scene, the generated point cloud contains significant noise. Thus, while S-NeRF has shown moderate performance at the task of novel view synthesis, these results indicate that the model fails at surface reconstruction for snow-covered urban landscapes.

The results for the Mount Doran scene are presented in Figures 3.10 - 3.12. For the task of novel view synthesis, the S-NeRF model achieved mean PSNR/SSIM scores of 22.92/0.699 on train data and 11.541/0.244 on test data in unseen viewing and lighting conditions. S-NeRF shows significant improvement at the task of novel view synthesis for the Mount Doran scene, compared to the Anchorage scene. The reason for this is likely due to the minimal variations in surface color and texture in the Mount Doran scene. In the reference images in Figures 3.10 and 3.11, it is observed that Mont Doran is dominated by a monotonous snow-covered surface. It is suspected that this visual characteristic allows the S-NeRF model to more easily generalize across views without the added complexity of having to render specific colors and textures at specific locations in the scene. Overall, these results indicate that S-NeRF succeeds at rendering visually accurate novel views of mountainous, snow-dominated terrain. However, as with the Anchorage scene, S-NeRF failed at the task of 3D surface reconstruction. The model achieved a mean distance error of 61.616 m for the generated point cloud of the Mount Doran scene, as shown in Figure 3.12. Although S-NeRF was able to capture the overall elevation gradient, it struggled to accurately represent finer topographic details, such as sharp ridges and valleys. An interesting behavior of S-NeRF is that it reconstructed both the Anchorage and Mount Doran scenes as flat planes. Although the generated point clouds were of the correct scale and orientation, all surface details were lost. This suggests that, while S-NeRF demonstrates moderate performance at rendering visually realistic novel views, further refinement is needed to improve its high-fidelity 3D surface reconstruction capabilities, particularly in polar environments.

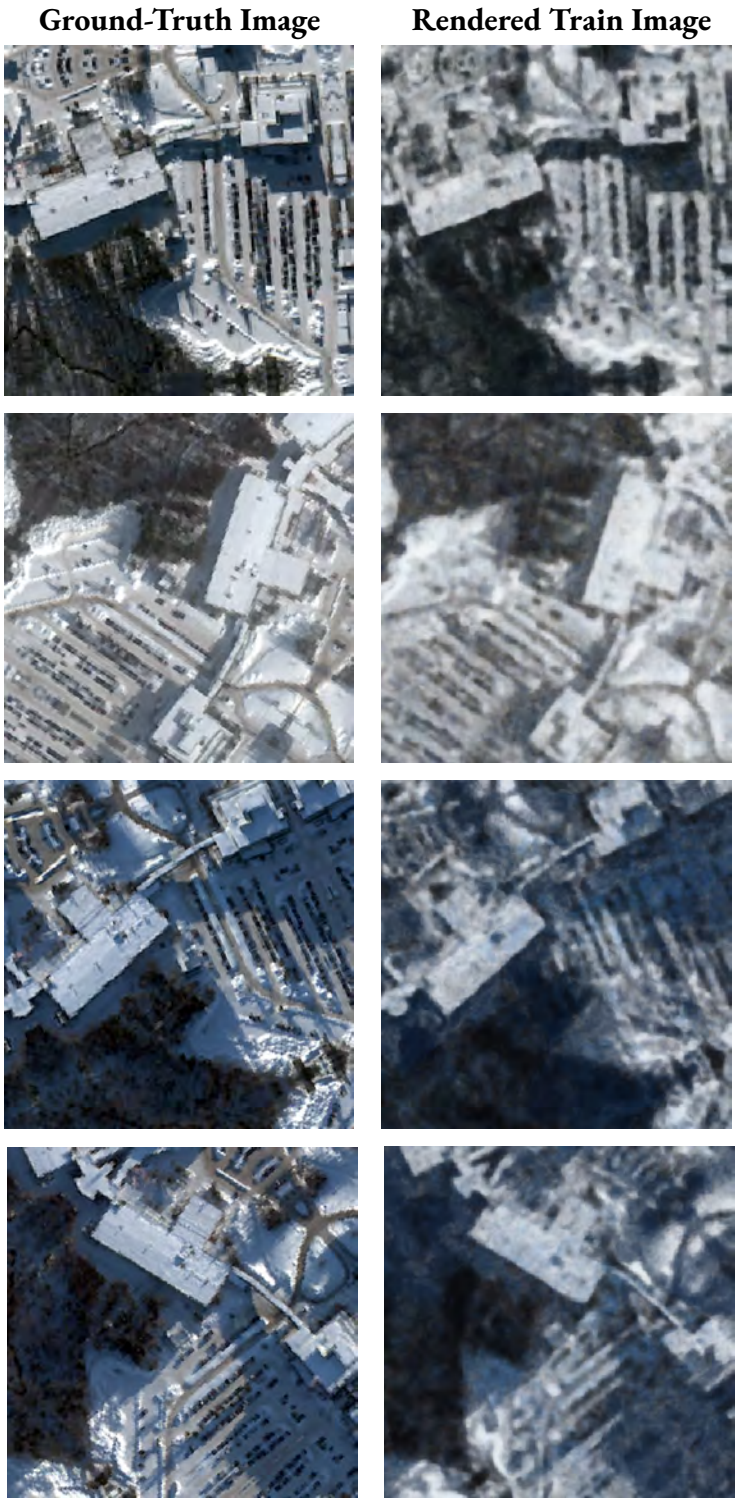



Figure 3.7: A sample of training images rendered from the small Anchorage dataset (300 $\times$ 300 m) using the S-NeRF model. S-NeRF achieved mean PSNR/SSIM scores of 22.13/0.578 on training data.

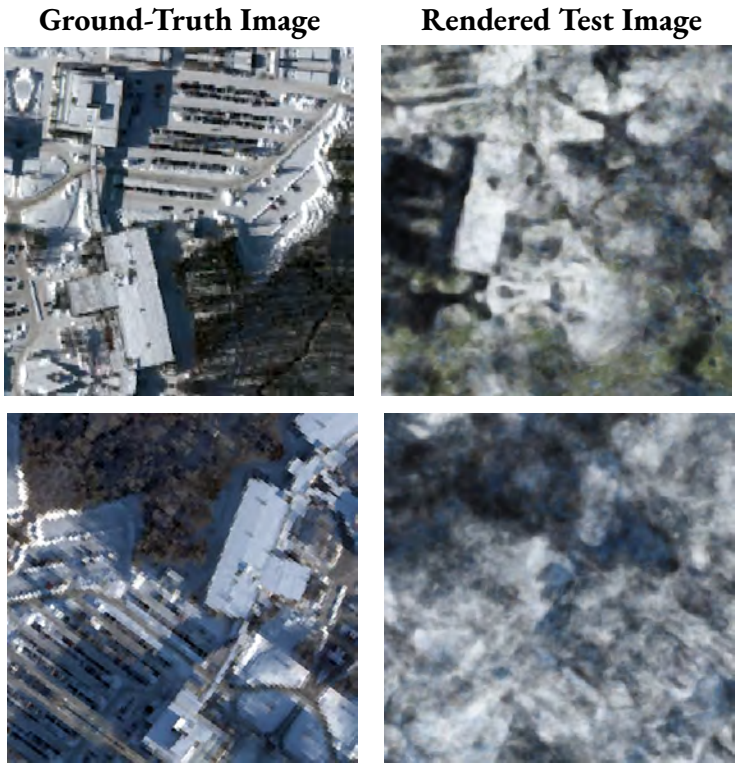



Figure 3.8: Testing images rendered from the small Anchorage dataset (300 × 300 m) using the S-NeRF model. S-NeRF achieved mean PSNR/SSIM scores of 10.07/0.139 on testing data.

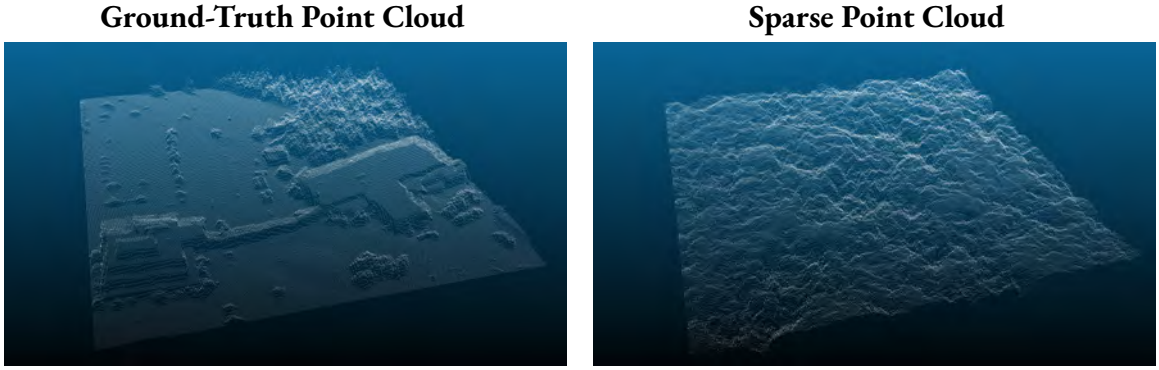



Figure 3.9: Surface reconstruction results for S-NeRF using the small Anchorage dataset (300 × 300 m). S-NeRF achieved a mean distance of 2.883 m for the generated sparse point cloud.

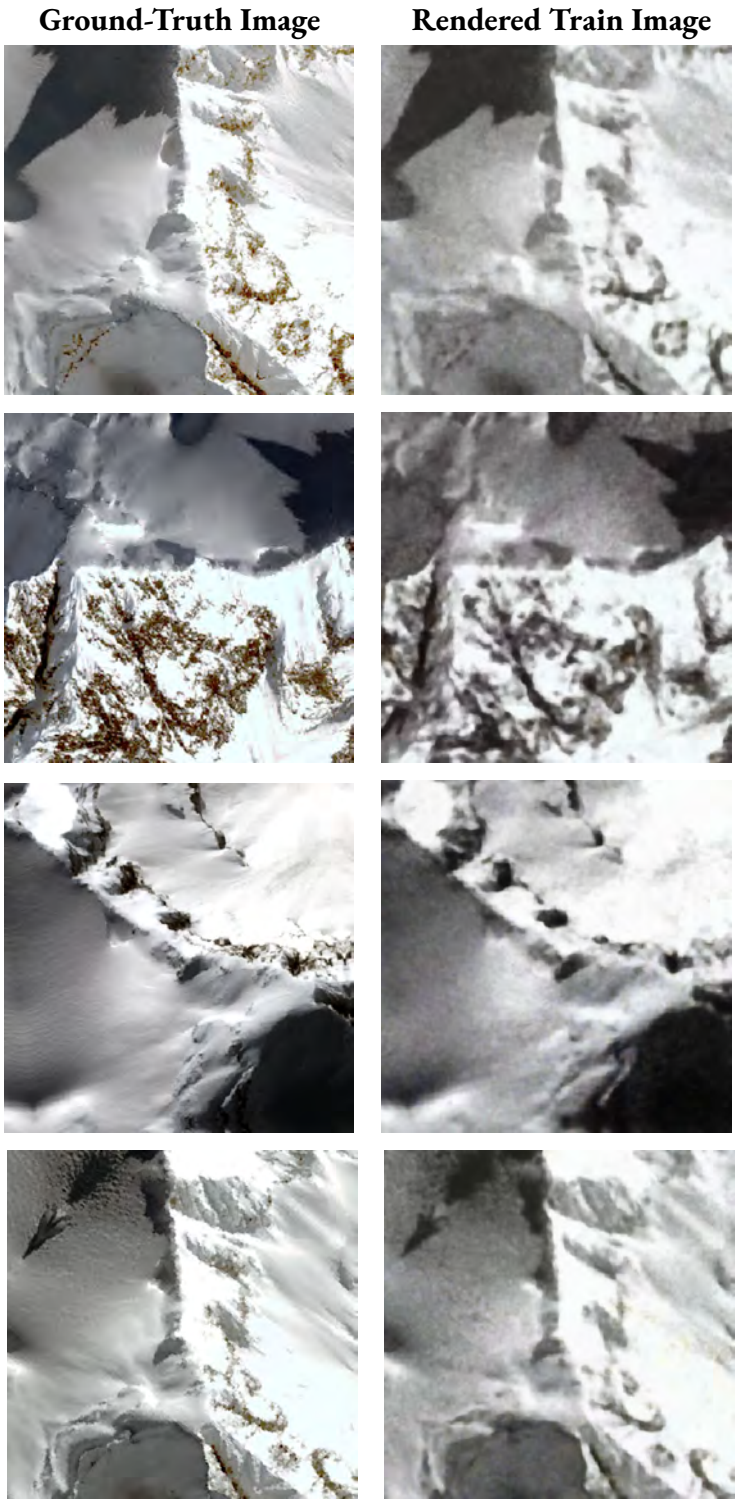



Figure 3.10: A sample of training images rendered from the small Mount Doran dataset ($600 \times 600$ m) using the S-NeRF model. S-NeRF achieved mean PSNR/SSIM scores of 22.92/0.699 on training data.

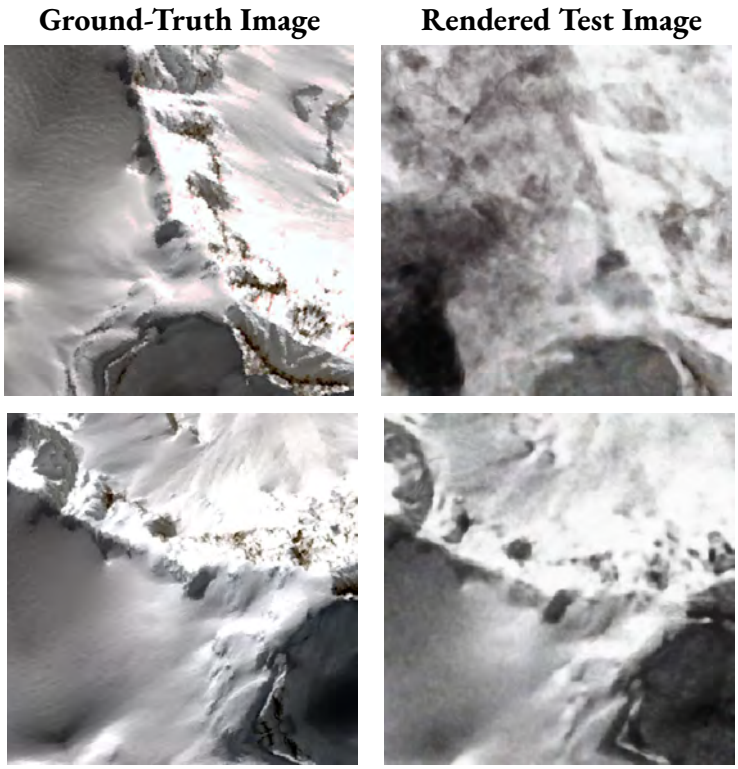



Figure 3.11: Testing images rendered from the small Mount Doran dataset (600 × 600 m) using the S-NeRF model. S-NeRF achieved mean PSNR/SSIM scores of 11.541/0.244 on testing data.

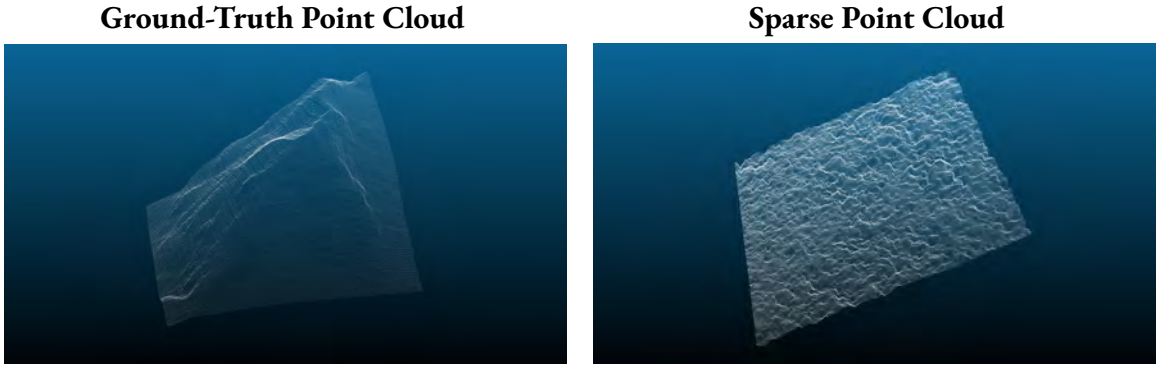



Figure 3.12: Surface reconstruction results for S-NeRF using the small Mount Doran dataset (600 × 600 m). S-NeRF achieved a mean distance of 61.616 m for the generated sparse point cloud.

## 3.4 Sat-NeRF

Sat-NeRF improves upon S-NeRF by leveraging the RPC camera models for each input image. As opposed to an approximate pinhole camera model, an RPC camera model is widely used for optical satellite imagery, as it can provide independence from the satellite-specific physical model [21]. In this research, Sat-NeRF was tested with the Anchorage and Mount Doran datasets to asses its abilities at reconstructing polar scenes from multi-vew satellite imagery. The results for the Anchorage dataset are presented in Figures 3.13 - 3.15. For the task of novel view synthesis, the Sat-NeRF model achieved mean PSNR/SSIM scores of 23.1/0.846 on train data and 12.173/0.249 on test data in unseen viewing and lighting conditions. Sat-NeRF demonstrated more accurate novel view synthesis performance for the Anchorage scene, compared to S-NeRF, as indicated by higher PSNR/SSIM scores on both train and test data. However, visually speaking, the rendered testing views still appear significantly noisy. Despite this, the rendered training views appear much more visually precise, compared to those obtained with S-NeRF. Figure 3.19 compares the output of S-NeRF and Sat-NeRF side-by-side, and it can see that the rendered view of Anchorage from S-NeRF appears fuzzier compared to the rendered view from Sat-NeRF. Artifacts are observed in the Sat-NeRF output, however, the overall scene was rendered with much more detail. From these results, it can be concluded that the use of RPC camera models improves the performance of satellite-based NeRF methods at novel view synthesis of snow-covered urban scenes. However, Sat-NeRF failed at the task of 3D surface reconstruction. As shown in Figure 3.15, the model achieved a mean distance error of 5.724 m for the generated point cloud of the Anchorage scene. These results suggest that, while the use of RPC models improves novel view synthesis performance, they do not significantly improve the surface reconstruction performance for snow-covered urban scenes.

The results for the Mount Doran scene are presented in Figures 3.16 - 3.18. For the task of novel view synthesis, the Sat-NeRF model achieved mean PSNR/SSIM scores of 23.573/0.783 on train data and 12.797/0.461 on test data in unseen viewing and lighting conditions. As with the S-NeRF model, a significant improvement at the task of novel view synthesis is observed for the Mount Doran scene, compared to the Anchorage scene. This is likely attributed to the homogeneous surface characteristics of the Mount Doran scene, which may make it easier for satellite-based NeRF methods to generalize across views, as stated previously in the S-NeRF results. Furthermore, Sat-NeRF demonstrated more accurate novel view synthesis performance, compared to S-NeRF, as indicated by higher PSNR/SSIM scores on both train and test data. This is illustrated in Figure 3.19, which compares the output of S-NeRF and Sat-NeRF side-by-side. Again, it can see that the rendered view of Mount Doran from S-NeRF appears fuzzier compared to the rendered view from Sat-NeRF. Because similar noise patterns are observed with the S-NeRF model for both scene types, it can be concluded that this is a characteristic of S-NeRF with polar terrain. The model has a tendency to produce grainier image renderings at lower fidelities. Sat-NeRF, on the other hand, excelled at view synthesis with both datasets. The rendered images appear sharper, especially in snow-dominated areas of the scene. Overall, these results indicate that Sat-NeRF succeeds at rendering visually accurate novel views of mountainous, snow-dominated terrain. However, despite its success in novel view synthesis, Sat-NeRF struggled with accurate 3D surface reconstruction of the Mount

Doran scene, as shown in Figure 3.18. The generated point cloud contained significant noise, with a mean distance error of 58.031 m. While S-NeRF has a tendency to reconstruct the scenes as flat planes, the point clouds generated by Sat-NeRF exhibit severe distortions, likely due to the use of the RPC camera models. This suggests that, while the model excels in rendering photorealistic 2D images, it encounters difficulties in capturing the geometric details required for precise 3D reconstruction of polar environments.

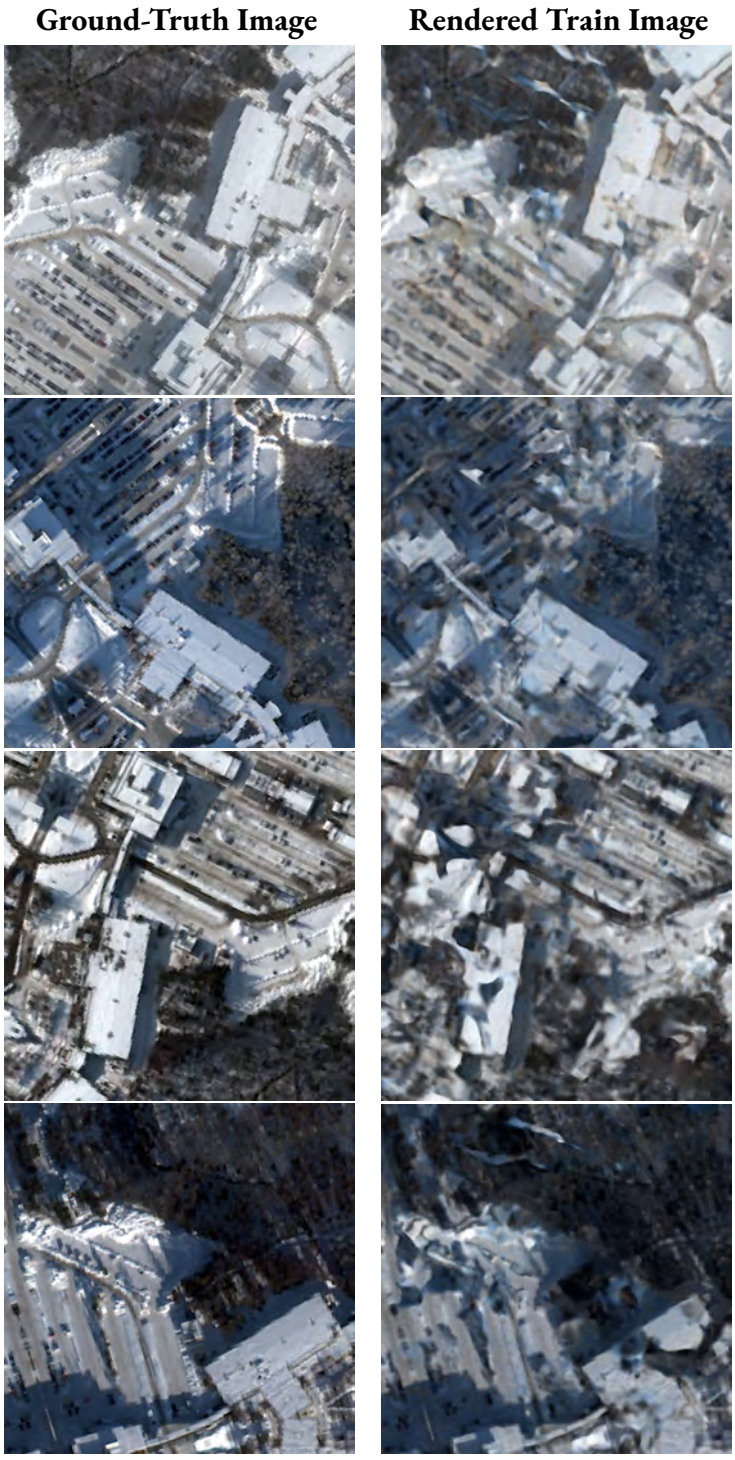



Figure 3.13: A sample of training images rendered from the small Anchorage dataset (300 × 300 m) using the Sat-NeRF model. Sat-NeRF achieved mean PSNR/SSIM scores of 23.099/0.846 on training data.

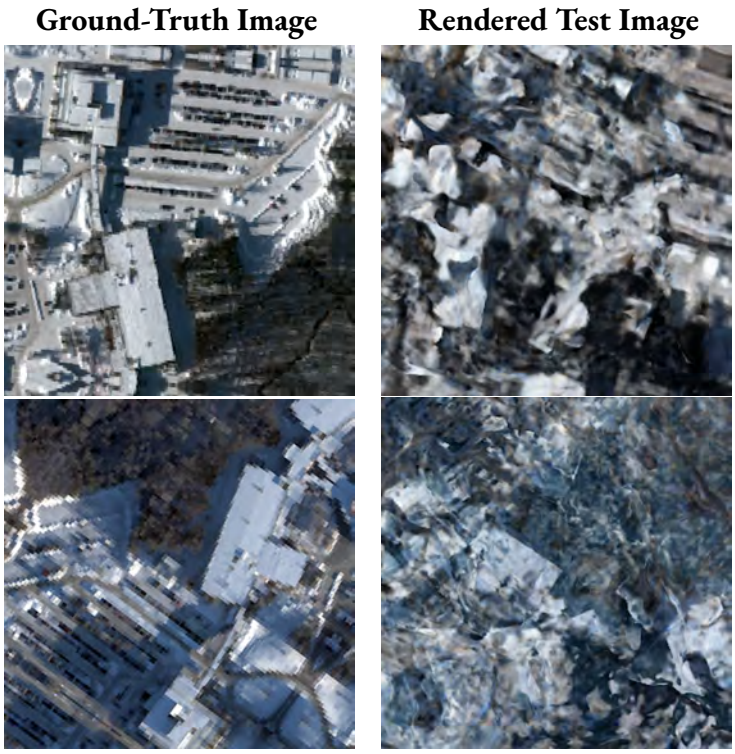



Figure 3.14: Testing images rendered from the small Anchorage dataset (300 × 300 m) using the Sat-NeRF model. Sat-NeRF achieved mean PSNR/SSIM scores of 12.173/0.249 on testing data.

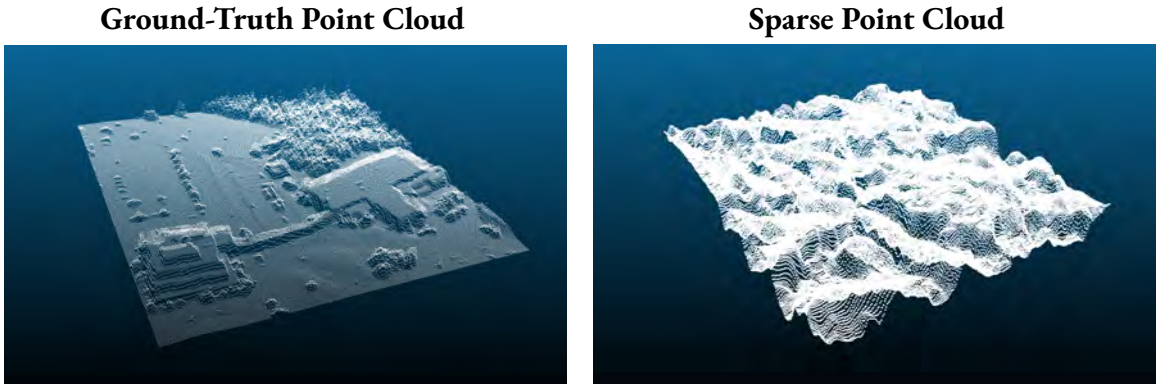



Figure 3.15: Surface reconstruction results for Sat-NeRF using the small Anchorage dataset (300 × 300 m). Sat-NeRF achieved a mean distance of 5.724 m for the generated sparse point cloud.

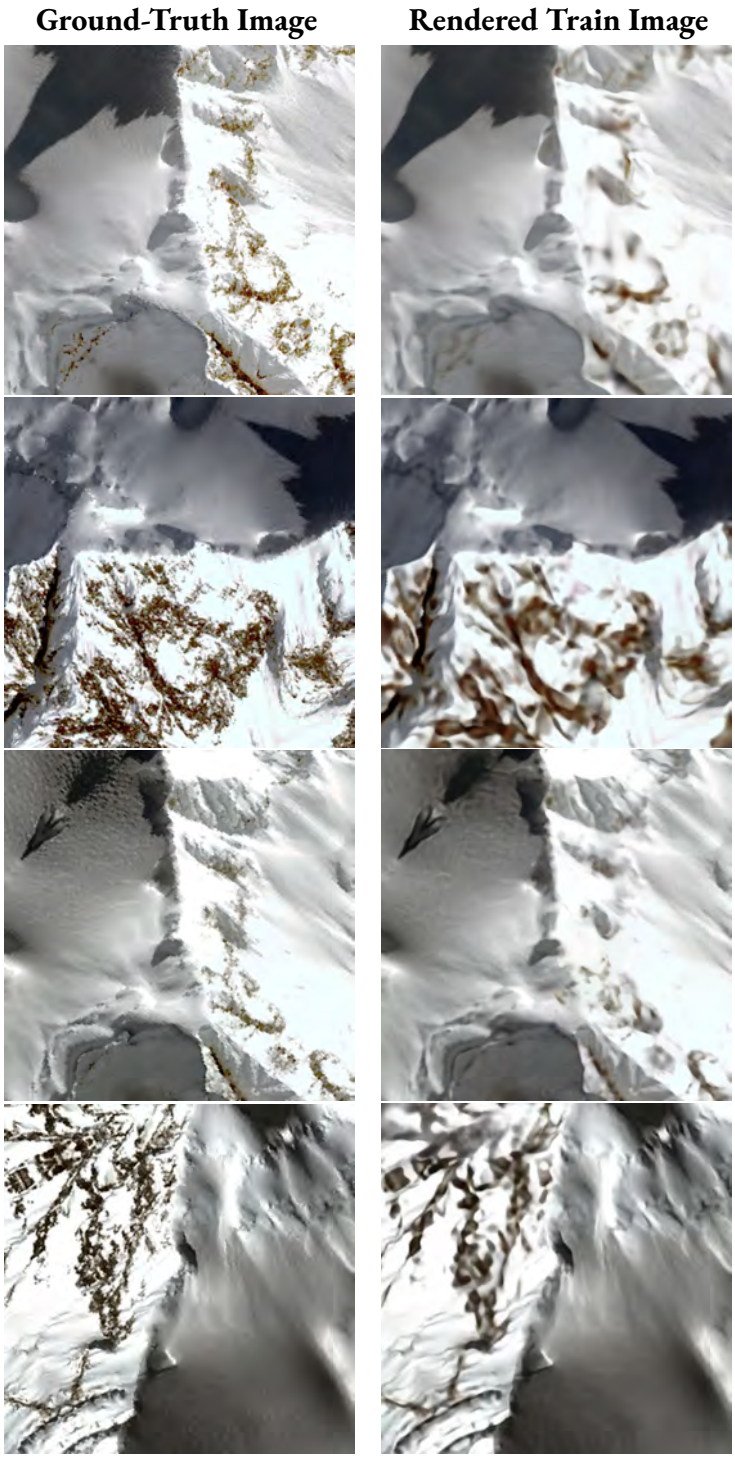



Figure 3.16: A sample of training images rendered from the small Mount Doran dataset (600 × 600 m) using the Sat-NeRF model. Sat-NeRF achieved mean PSNR/SSIM scores of 23.573/0.783 on training data.

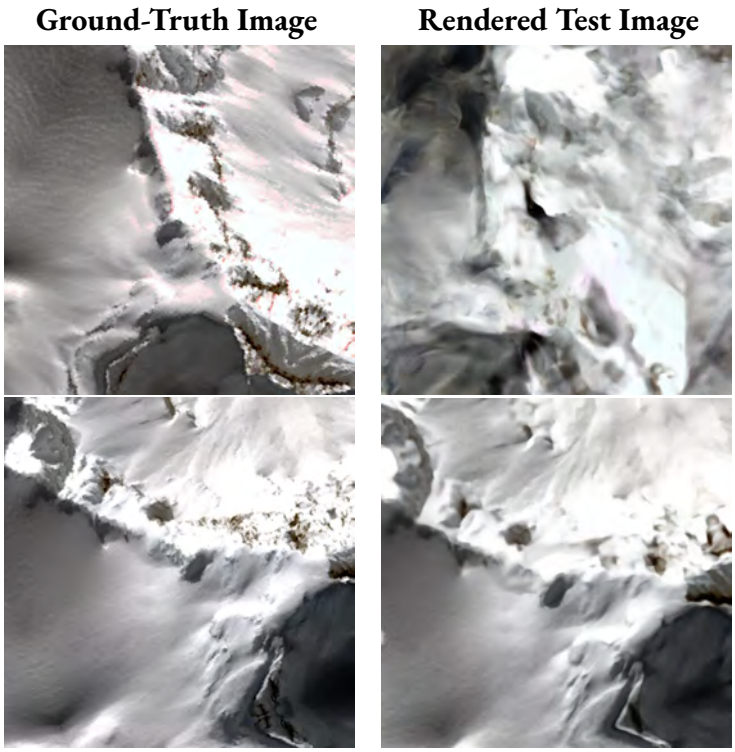



Figure 3.17: Testing images rendered from the small Mount Doran dataset (600 × 600 m) using the Sat-NeRF model. Sat-NeRF achieved mean PSNR/SSIM scores of 12.797/0.461 on testing data.

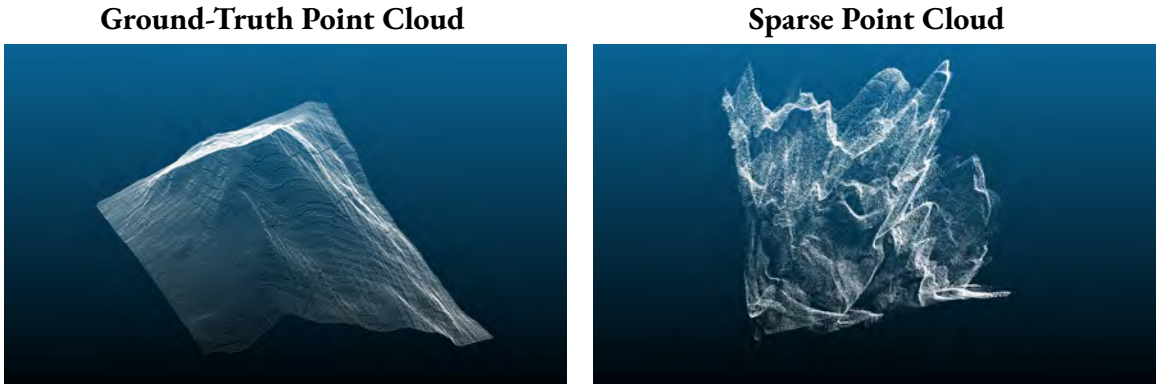



Figure 3.18: Surface reconstruction results for Sat-NeRF using the small Mount Doran dataset (600 × 600 m). Sat-NeRF achieved a mean distance of 58.031 m for the generated sparse point cloud.

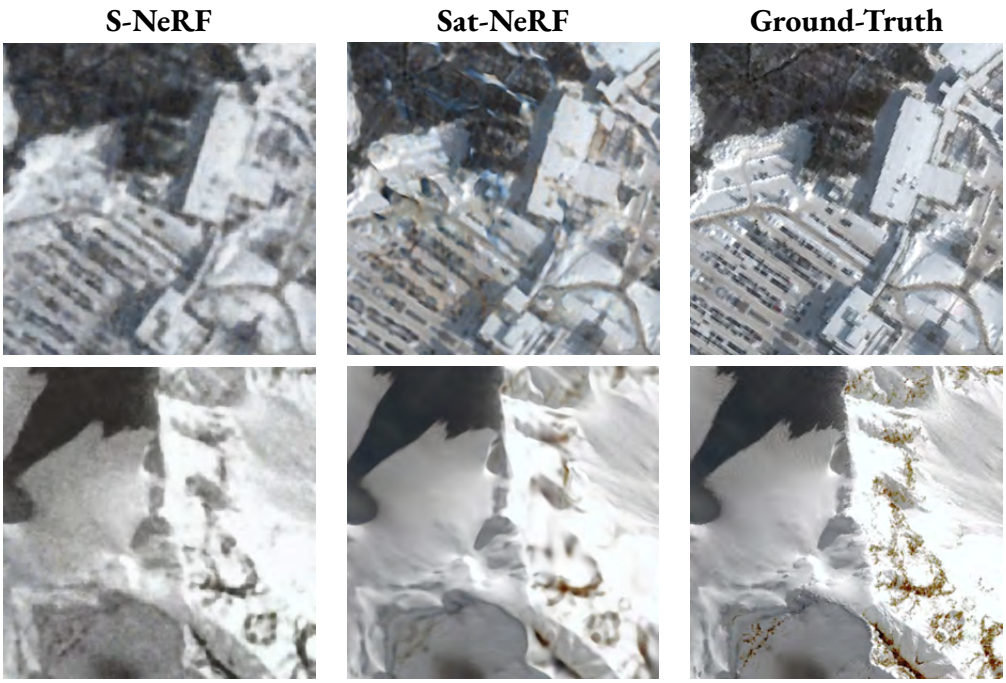



Figure 3.19: Qualitative view synthesis results for each method, with WorldView-2 satellite imagery of Anchorage (top row) and Mount Doran (bottowm row). Sat-NeRF produced the most visually accurate novel views, achieving a mean PNSR/SSIM of 12.173/0.249 for the Anchorage dataset and 12.797/0.461 for the Mount Doran dataset.

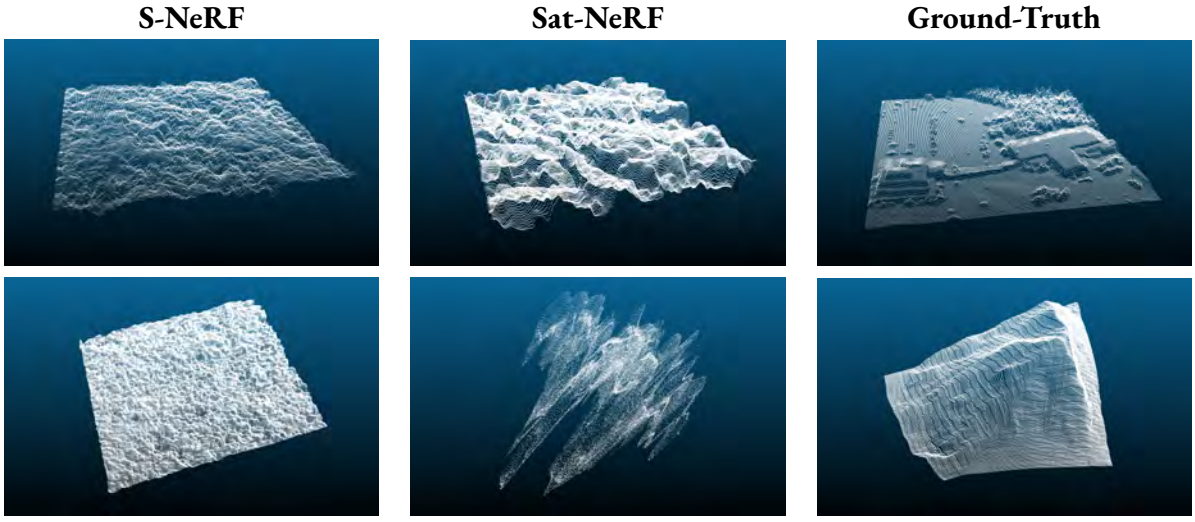



Figure 3.20: Qualitative surface reconstruction results for each method, with WorldView-2 satellite imagery of Anchorage (top row) and Mount Doran (bottowm row). S-NeRF produced the most geometrically accurate surface model for the Anchorage scene, achieving a mean distance error of 2.883 m. Sat-NeRF produced the most geometrically accurate surface model the for the Mount Doran scene, achieving a mean distance error of 58.031 m.

Table 3.2: Quantitative results for each method. The best metrics for each scene are in **bold**. Metashape performed best at surface reconstruction, achieving a mean distance error of 2.652 m for Anchorage and 15.624 m for Mount Doran. Sat-NeRF performed best at novel view synthesis, achieving mean PSNR/SSIM scores of 12.173/0.249 for Anchorage and 12.797/0.461 for Mount Doran.

| Scene | Method | PSNR↑ | SSIM↑ | Avg. Dist. (m)↓ | Std. Dev. (m) |
|---|---|---|---|---|---|
| Anchorage | Metashape | X | X | **2.652** | 4.392 |
| | S-NeRF | 10.07 | 0.139 | 2.883 | **2.317** |
| | Sat-NeRF | **12.173** | **0.249** | 5.724 | 4.081 |
| Mt Doran | Metashape | X | X | **15.624** | **16.951** |
| | S-NeRF | 11.541 | 0.244 | 61.616 | 41.083 |
| | Sat-NeRF | **12.797** | **0.461** | 58.031 | 44.813 |

## 3.5 Ablation Studies

Sat-NeRF outperformed S-NeRF at both tasks of novel view synthesis and 3D surface reconstruction, as indicated by lower average distance errors for the Mount Doran scene and higher PSNR/SSIM scores for both the Anchorage and Mount Doran scenes. Thus, a series of ablation studies were conducted with the Sat-NeRF model to determine the optimal configurations for each dataset. The variables of interest included scene size (i.e., AOI size), positional encoding frequency, and network activation function.

### 3.5.1 Scene Size

The first study investigated how scene size (i.e., AOI size) influences the accuracy of the neural scene representation. S-NeRF and Sat-NeRF were originally designed for urban environments limited to 300 $m^2$. Therefore, this ablation study aimed to determine if satellite-based NeRF methods could generalize to Earth observation datasets with larger AOIs. Experiments were run with AOI sizes covering 300 $m^2$, 600 $m^2$, and $1,200$ $m^2$ for the Anchorage scene and 600 $m^2$, $1,200$ $m^2$, and $2,000$ $m^2$ for the Mount Doran scene. Qualitative view synthesis results from this ablation study are presented in Figures 3.21 - 3.24. The corresponding quantitative results can be seen in Table 3.3. It is observed that Sat-NeRF succeeds at rendering novel views of polar scenes with high visual accuracy, for all the AOI sizes tested. For the task of novel view synthesis, Sat-NeRF achieved the best results with the small dataset for both scenes, reaching mean PSNR/SSIM scores of 12.173/0.249 and 12.797/0.462 for Anchorage and Mount Doran, respectively. It is observed that as AOI size decreased, PSNR/SSIM scores increased for both scenes. These results suggest that Sat-NeRF performs best at view synthesis with small AOI sizes covering approximately 300 $m^2$ to 600 $m^2$.

However, the reconstructed surface models contained excessive noise for all AOI sizes tested across both scenes, indicating that Sat-NeRF is failing to reconstruct the scene geometry. Qualitative results of the surface reconstruction performance are provided in Figures 3.25 - 3.26. The corresponding quantitative results can be seen in Table 3.3. Sat-NeRF achieved an average distance error of 3.687 m for the medium Anchorage dataset and 58.031 m for the small Mount Doran dataset. Although these results suggest that Sat-NeRF more accurately reconstructs small to mid-range AOIs, the generated point clouds contained significant noise for all AOI sizes tested. Thus, based on the qualitative surface reconstruction results, it can be concluded that Sat-NeRF fails at 3D surface reconstruction of polar landscapes, no matter the scene size or terrain type.

An ablation study on AOI size was also conducted with Metashape to determine if it could successfully reconstruct larger scenes from satellite remote-sensed imagery. The same small, medium, and large datasets of Anchorage and Mount Doran were used. For these experiments, the RPC camera models for each image were used as input to Metashape, as the results in Section 3.2 illustrate that including the RPC model for each image significantly improves the surface reconstruction accuracy. Qualitative results for this ablation study are provided in Figures 3.27 - 3.28. The corresponding quantitative results can be see in Table 3.4. It is observed that Metashape can successfully reconstruct polar scenes of varying AOI sizes with high geometric accuracy. Metashape achieved a mean distance error of 2.652 m for the small Anchorage dataset and 13.979 m for the large Mount Doran dataset. It is observed that, as AOI size increased, the surface reconstruction performance decreased for the Anchorage dataset. However, for the Mount Doran dataset, as AOI size increased, the surface reconstruction performance also increased. These results indicate that, for snow-covered urban scenes, Metashape shows an increase in performance with smaller AOI sizes. However, for mountainous polar scenes, Metashape shows an increase in performance with larger AOI sizes.

Table 3.3: Quantitative results for Sat-NeRF with datasets covering various scene sizes. The best metrics for each scene are in **bold**. For the task of novel view synthesis, the highest accuracy was achieved with the small Anchorage scene and small Mount Doran scene. For the task of 3D surface reconstruction, the highest accuracy was achieved with the medium Anchorage scene and small Mount Doran scene.

| Scene | Scene Size | PSNR↑ | SSIM↑ | Avg. Dist. (m)↓ | Std. Dev. (m) |
|---|---|---|---|---|---|
| Anchorage | Small | **12.173** | **0.249** | 5.724 | 4.081 |
| | Medium | 11.024 | 0.123 | **3.687** | **2.776** |
| | Large | 10.963 | 0.119 | 8.368 | 7.412 |
| Mt Doran | Small | **12.797** | **0.462** | **58.031** | **44.813** |
| | Medium | 11.718 | 0.443 | 109.716 | 88.918 |
| | Large | 11.246 | 0.402 | 155.99 | 120.089 |

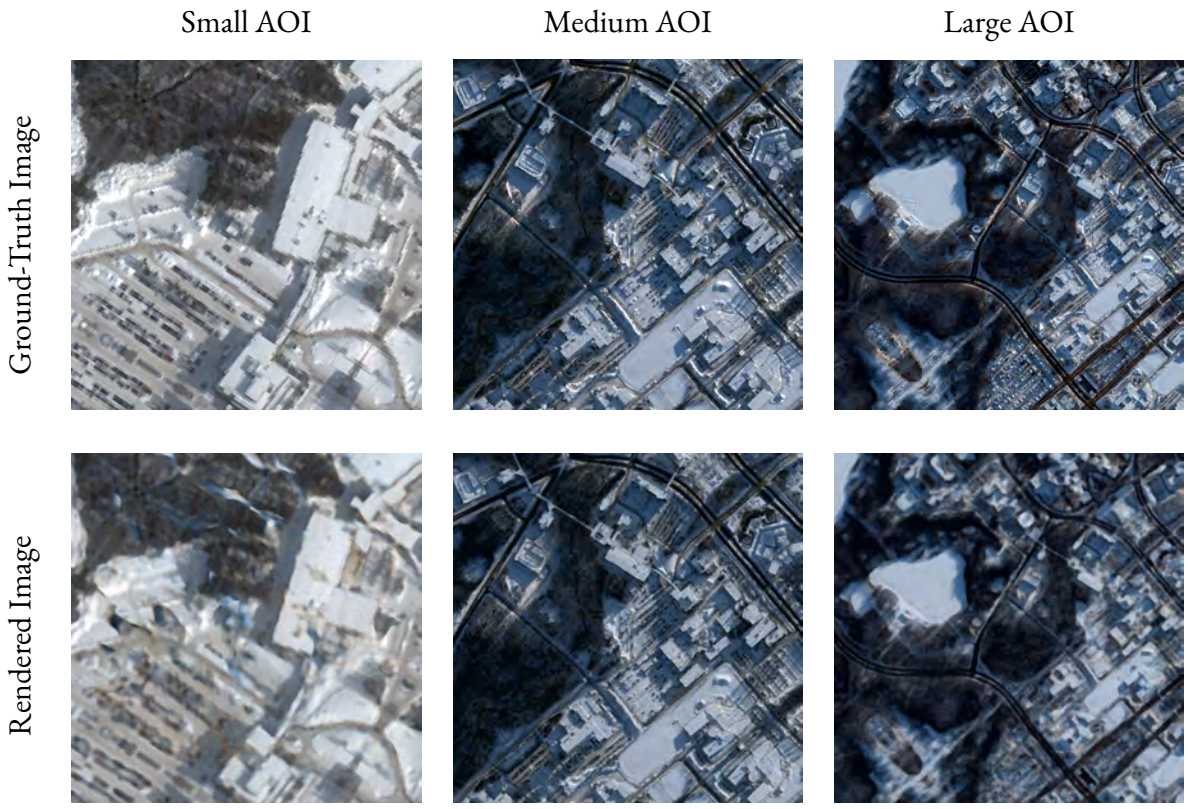



Figure 3.21: View synthesis results on training data with Anchorage scenes of varying AOI sizes. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 23.099/0.846 for the small Anchorage dataset.

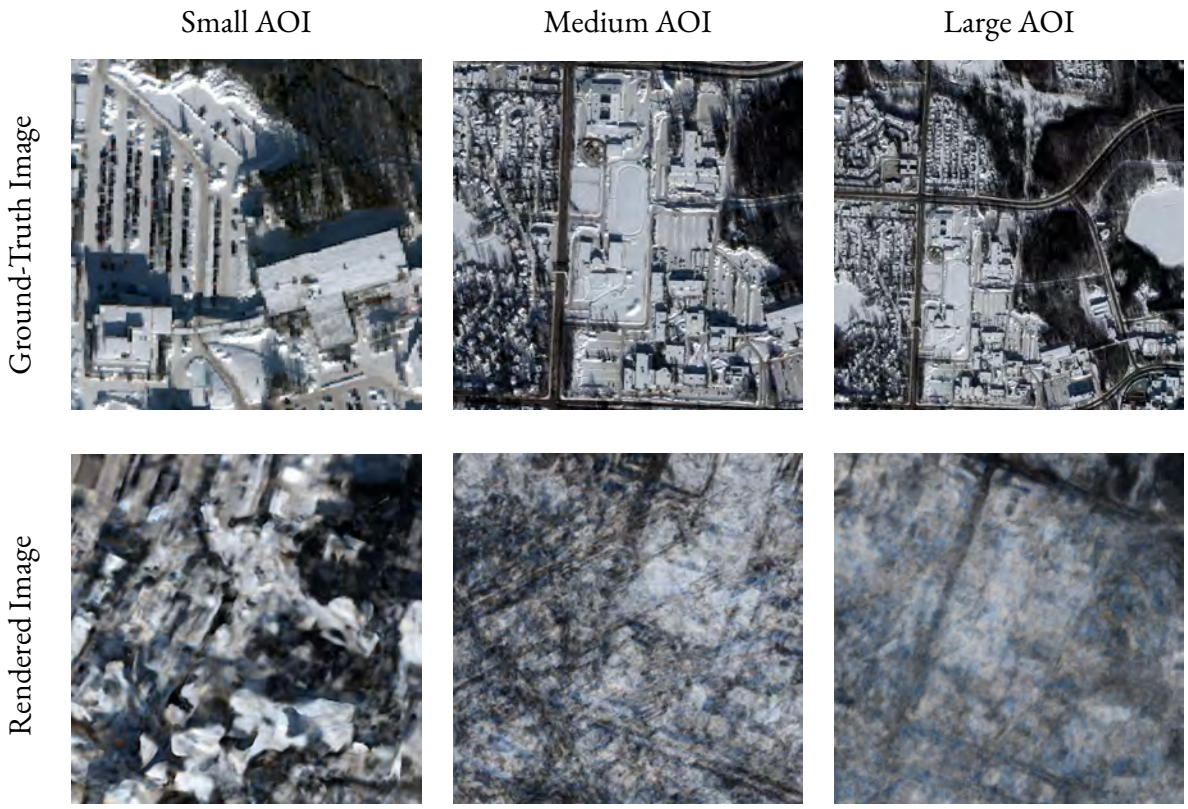



Figure 3.22: View synthesis results on testing data with Anchorage scenes of varying AOI sizes. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 12.173/0.249 for the small Anchorage dataset.

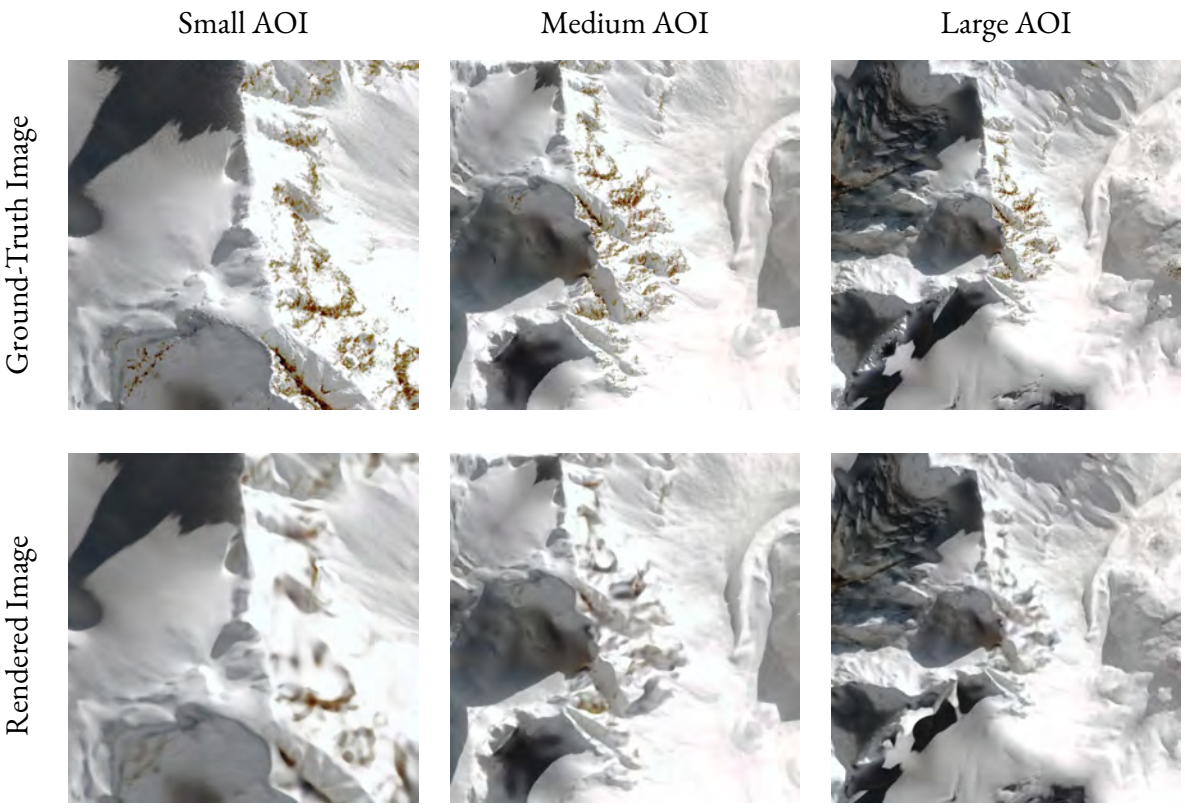



Figure 3.23: View synthesis results on training data with Mount Doran scenes of varying AOI sizes. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 23.573/0.783 for the small Mount Doran dataset.

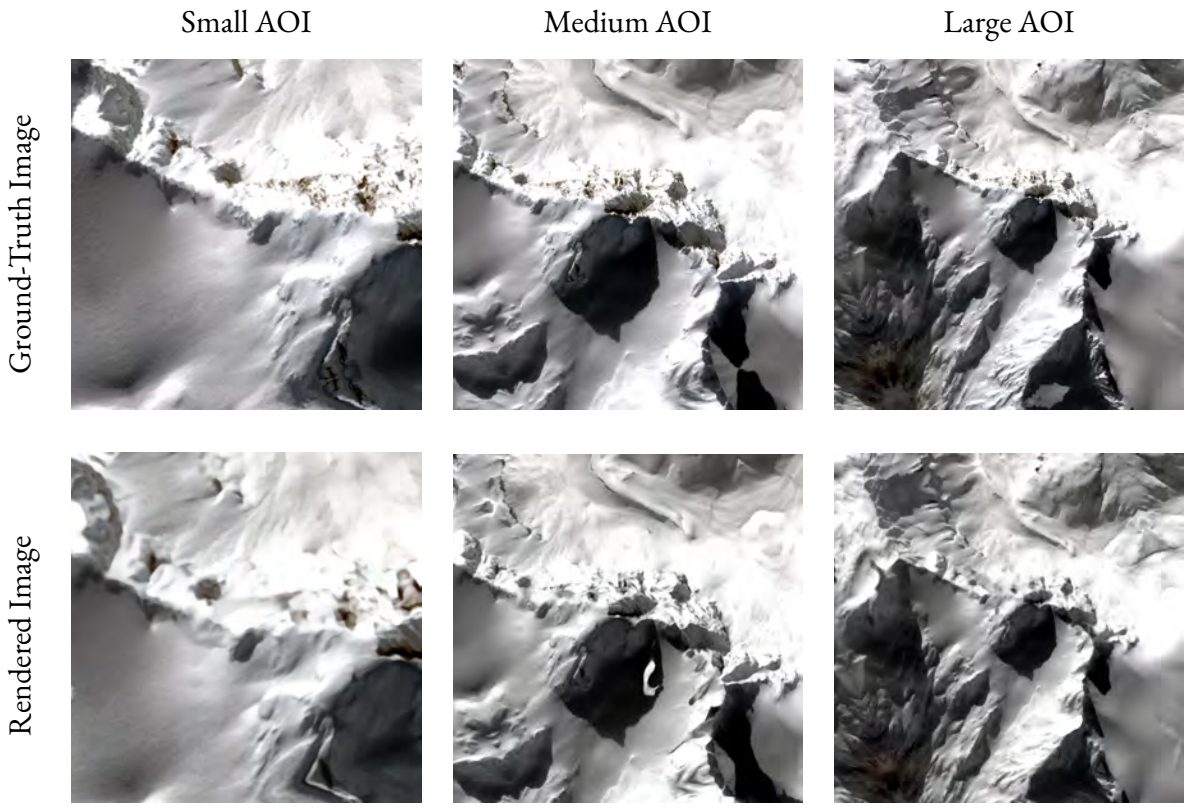



Figure 3.24: View synthesis results on testing data with Mount Doran scenes of varying AOI sizes. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 12.797/0.462 for the small Mount Doran dataset.

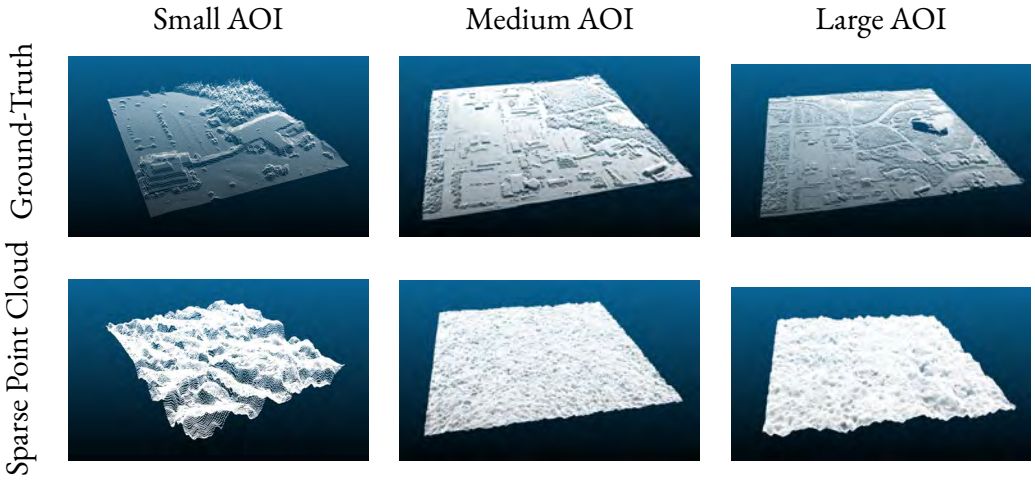



Figure 3.25: Surface reconstruction results for Sat-NeRF with Anchorage scenes of varying AOI sizes. The model achieved a mean distance error of 3.687 m for the medium Anchorage scene.

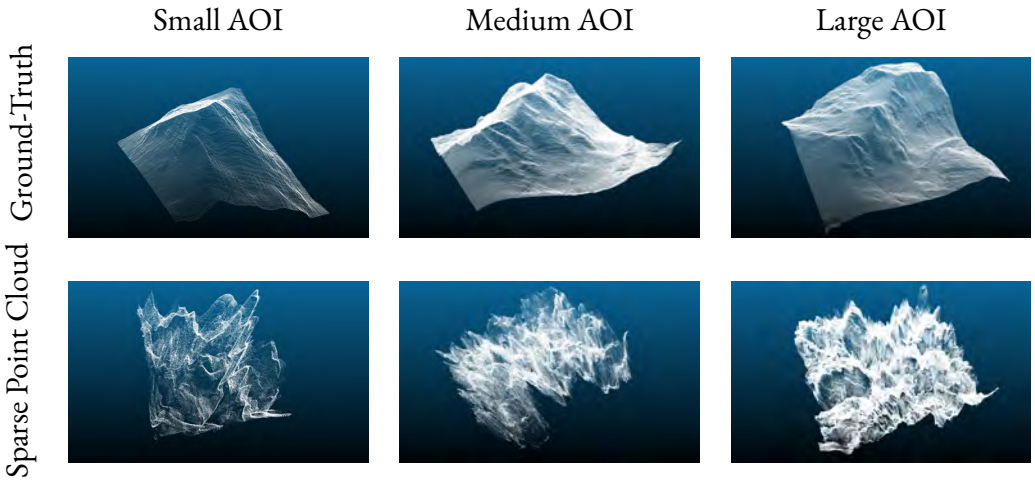



Figure 3.26: Surface reconstruction results for Sat-NeRF with Mount Doran scenes of varying AOI sizes. The model achieved a mean distance error of 58.031 m for the small Mount Doran scene.

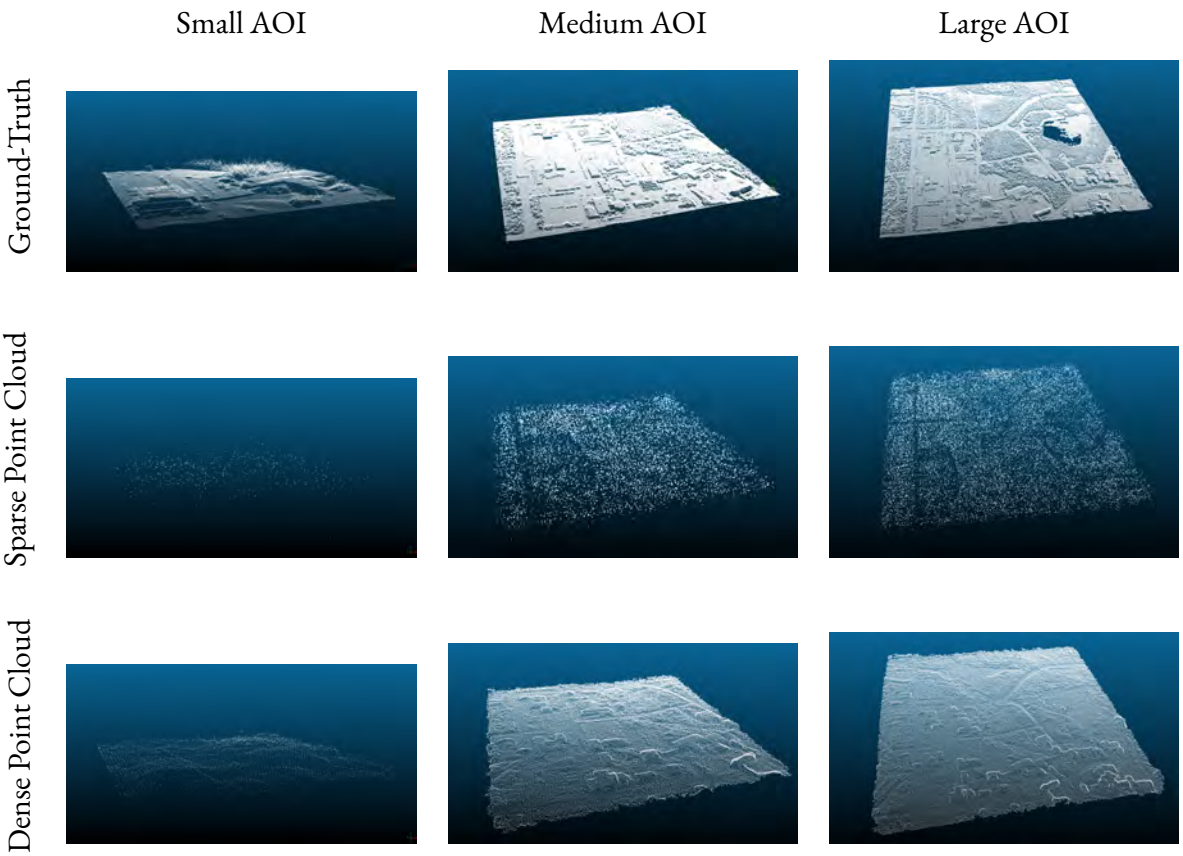



Figure 3.27: Surface reconstruction results for Metashape with Anchorage scenes of varying AOI sizes. The method achieved a mean distance error of 2.652 m for the sparse point cloud of the small Anchorage scene.

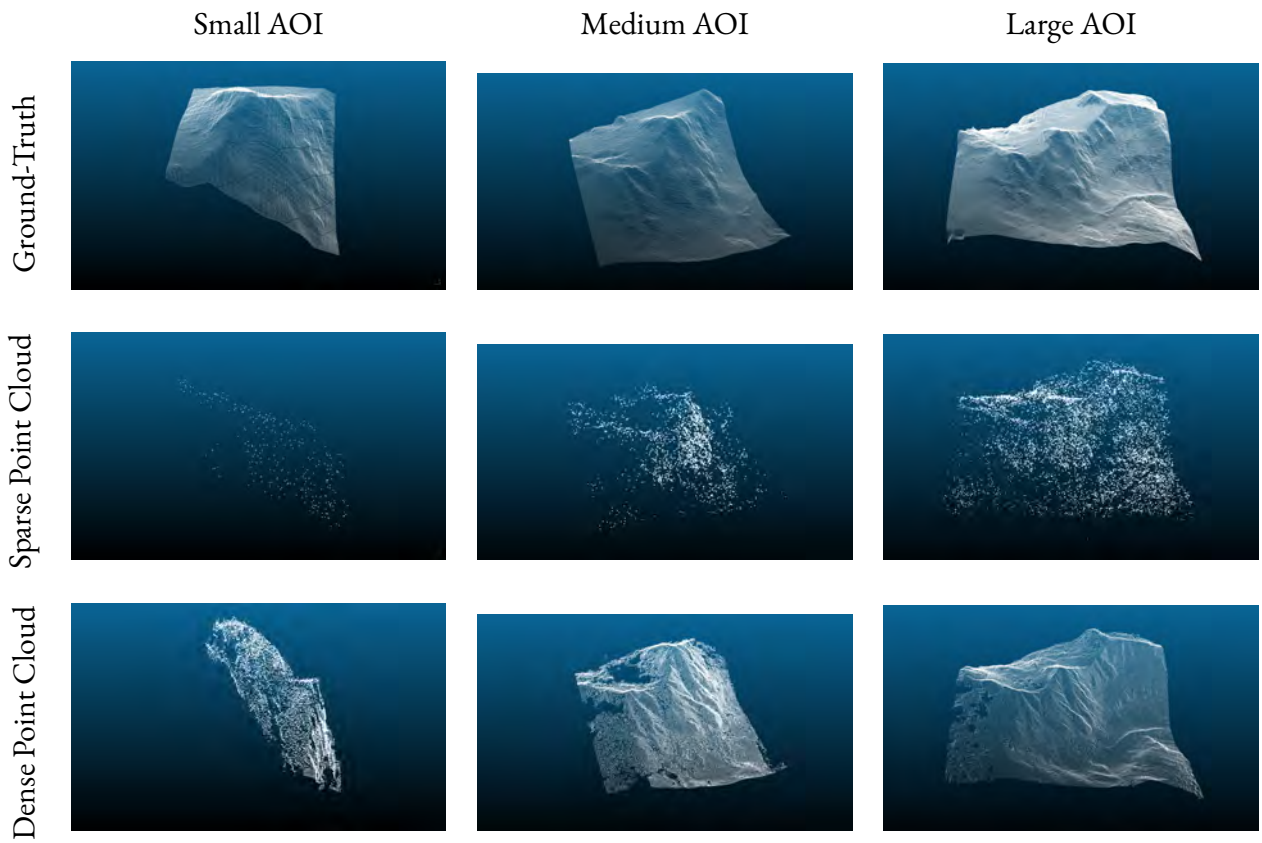



Figure 3.28: Surface reconstruction results for Metashape with Mount Doran scenes of varying AOI sizes. The method achieved a mean distance error of 13.979 m for the sparse point cloud of the large Mount Doran scene.

Table 3.4: Quantitative results for Metashape with datasets covering various scene sizes. The best metrics for each scene are in **bold**. Metashape achieved the highest surface reconstruction accuracy with the small Anchorage scene and large Mount Doran scene.

| Scene | Scene Size | Avg. Dist. (m)↓ | Std. Dev. (m) |
|---|---|---|---|
| Anchorage | Small | **2.652** | 4.392 |
| | Medium | 2.841 | **3.438** |
| | Large | 3.591 | 5.428 |
| Mt Doran | Small | 15.624 | 16.951 |
| | Medium | 18.232 | 17.581 |
| | Large | **13.979** | **15.824** |

### 3.5.2 Positional Encoding Frequency

The second ablation study investigated how the Fourier feature frequency in positional encoding influences the accuracy of the neural scene representation. In traditional NeRF methods, the low-dimensional input coordinates are projected to a higher dimension via positional encoding. This enables the MLP to learn higher frequency functions [30]. A higher positional encoding frequency improves the capacity of the NeRF model at capturing finer details in the scene. However, increasing the Fourier feature frequency can lead to over-fitting, particularly in areas of the scene with less detail or more homogeneous surfaces. In contrast, a lower positional encoding frequency tends to smooth out fine details, which can improve generalization but may result in blurring or loss of intricate scene features. The optimal Fourier feature frequency in positional encoding will vary depending on the complexity of the scene. Thus, the goal of this ablation study is to determine the frequency that leads to the highest novel view synthesis and surface reconstruction accuracy for the Anchorage and Mount Doran datasets.

S-NeRF and Sat-NeRF both employ a basic positional encoding scheme. In Sat-NeRF, the number of Fourier feature frequencies for positional encoding is set to 0 by default. Thus, experiments were run with Fourier feature frequencies 0, 8, and 16 for each dataset. The qualitative novel view synthesis results for the Anchorage scene are presented in Figures 3.29 - 3.30. It can be seen that the use of positional encoding in the Sat-NeRF model improves novel view synthesis results for this scene. The rendered images appear sharper and capture more detail with Fourier feature frequencies of 8 and 16, compared to 0. As with previous Sat-NeRF experiments, significant noise is observed in the rendered test views. However, some areas of the Anchorage scene were resolved with a higher positional encoding frequency, particularly the forested area and parking lot. The quantitative novel view synthesis results for the Anchorage scene are presented in Table 3.5. The PSNR/SSIM scores peaked with a Fourier feature frequency of 8, indicating that a moderate frequency balances the capacity of the Sat-NeRF model at capturing fine details in the

scene without over-fitting to the homogeneous snow-covered areas. Further increasing the Fourier feature frequency beyond 8 lead to a decrease in model performance, as PSNR/SSIM were lowest with a frequency of 16. Overall, these results indicate that the use of positional encoding in the Sat-NeRF model improves novel view synthesis accuracy for snow-covered urban scenes. However, as shown in 3.33, the 3D surface reconstructions of the Anchorage scene contained significant noise for all positional encoding frequencies tested. Thus, it can be concluded that the Fourier feature frequency has no influence on the geometric accuracy of the reconstructed scene.

The qualitative novel view synthesis results for the Mount Doran scene are presented in Figures 3.31 - 3.32. The corresponding quantitative results are presented in Table 3.5. It is observed that the use of positional encoding in the Sat-NeRF model improves novel view synthesis results for the Mount Doran scene. Sat-NeRF achieved the best view synthesis performance with a Fourier feature frequency of 8 in positional encoding, for both train and test data, as indicated the highest PSNR/SSIM scores. These results indicate that a moderate positional encoding frequency effectively captures finer details in the Mount Doran scene. A low-frequency Fourier feature caused blurring in the areas of the scene with more complex details, whereas a higher-frequency Fourier feature caused the model to over-fit to the snow-covered terrain. This pattern aligns with the findings for the Anchorage scene. The use of a moderate Fourier feature frequency in positional encoding may, therefore, be optimal for accurately capturing visual details in polar landscapes while avoiding excessive noise. However, as shown in Figure 3.34, the use of positional encoding in the Sat-NeRF model did not improve the 3D surface reconstruction results. All of the point clouds of the Mount Doran scene contained significant noise for all the Fourier feature frequencies tested. Thus, as with the results for the Anchorage scene, it can be concluded that the Fourier feature frequency in positional encoding does not significantly influence the geometric accuracy of the 3D reconstructions.

The results from this ablation study support existing research which has found that the optimal Fourier feature frequency depends on scene complexity and texture variability. Previous studies have shown that higher frequencies enable NeRF models to capture fine-grained details, while lower frequencies promote generalization in simpler or less textured scenes [37]. Although the Anchorage and Mount Doran scenes contain minimal surface features, a lower positional encoding frequency did not necessarily yield better results. A Fourier feature frequency of 0 was insufficient for recovering finer scene details, while a moderate frequency of 8 achieved a balance in representing both detailed and homogeneous surface features for these scenes. This demonstrates that tuning the positional encoding frequency based on the complexity of the scene can lead to improved performance in novel view synthesis tasks. Additionally, it was found that the Fourier feature frequency in positional encoding had no significant influence on 3D surface reconstruction tasks for these datasets. The point clouds of the Anchorage and Mount Doran scenes contained significant noise for all the positional encoding frequencies tested. Overall, these findings suggest that while novel view synthesis benefits from frequency tuning in positional encoding, 3D surface reconstruction in snow-dominant terrains may require alternative strategies to mitigate noise and improve structural accuracy.

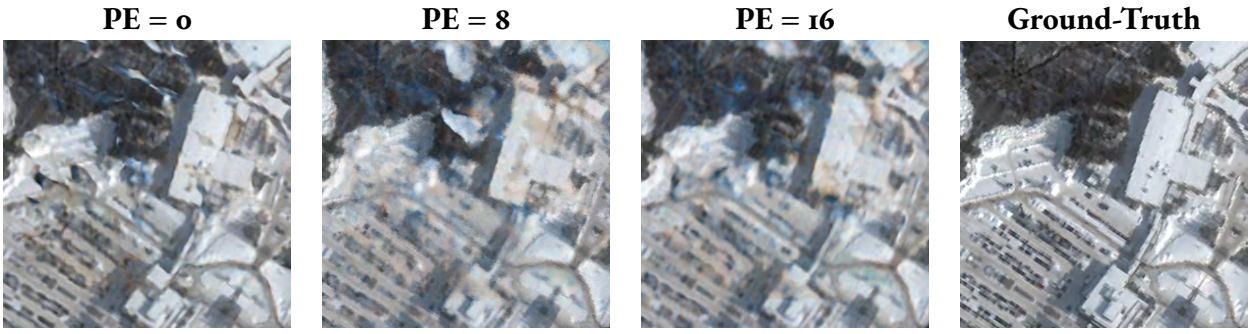



Figure 3.29: View synthesis results on training data with various positional encoding (PE) frequencies, for the small Anchorage scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 23.75/0.888 with a positional encoding frequency of 8.

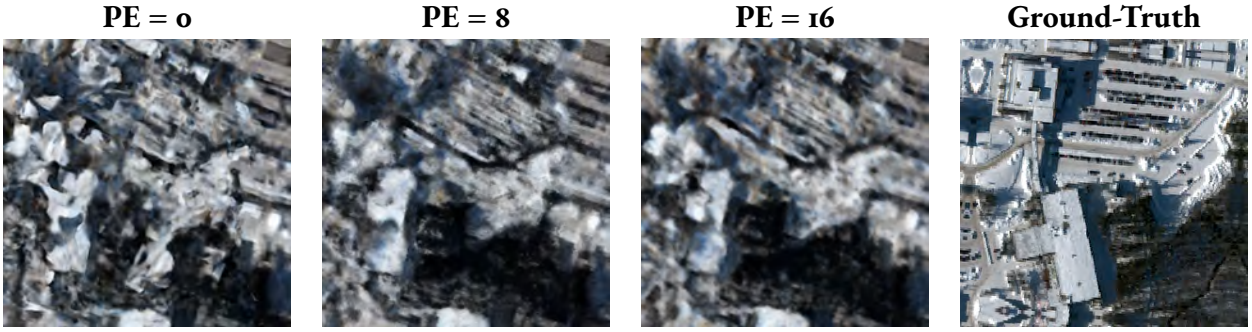



Figure 3.30: View synthesis results on testing data with various positional encoding (PE) frequencies, for the small Anchorage scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 15.573/0.657 with a positional encoding frequency of 8.

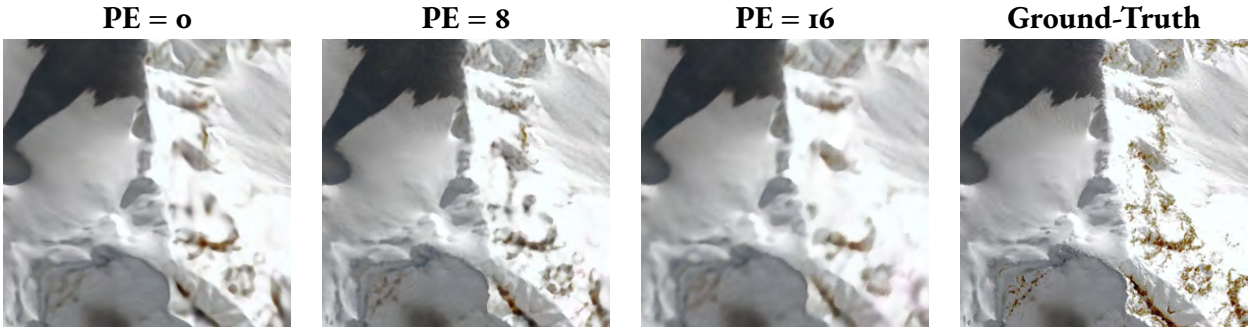



Figure 3.31: View synthesis results on training data with various positional encoding (PE) frequencies, for the small Mount Doran scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 24.014/0.773 with a positional encoding frequency of 8.

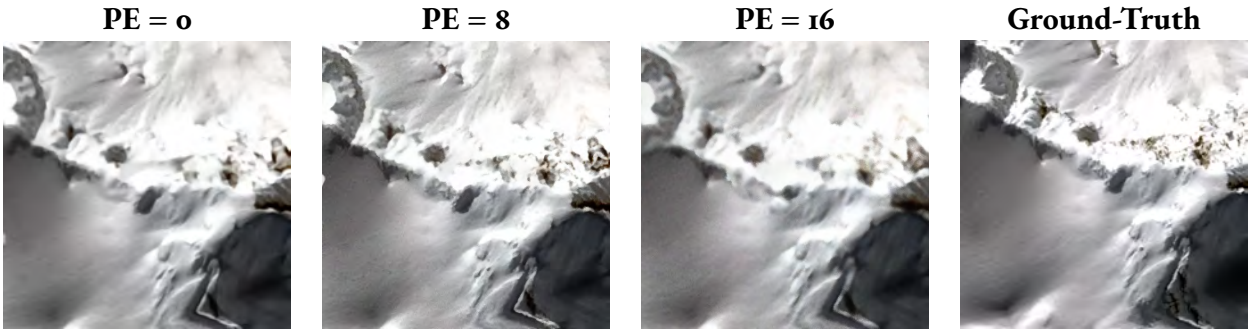



Figure 3.32: View synthesis results on testing data with various positional encoding (PE) frequencies, for the small Mount Doran scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 12.986/0.483 with a positional encoding frequency of 8.

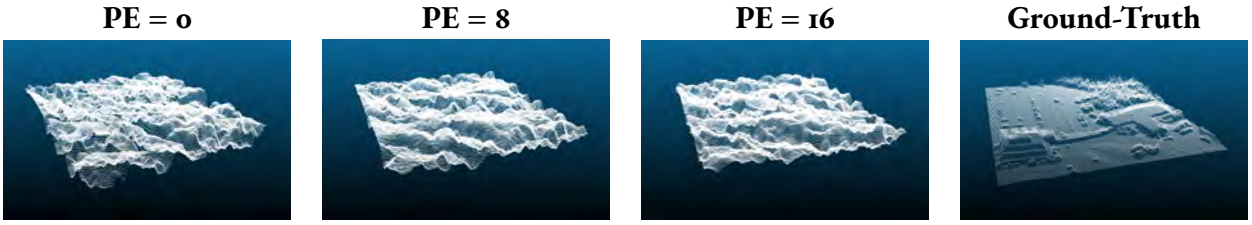



Figure 3.33: Surface reconstruction results for Sat-NeRF with various positional encoding (PE) frequencies, for the small Anchorage scene. The model achieved a mean distance error of 4.553 m with a positional encoding frequency of 16.

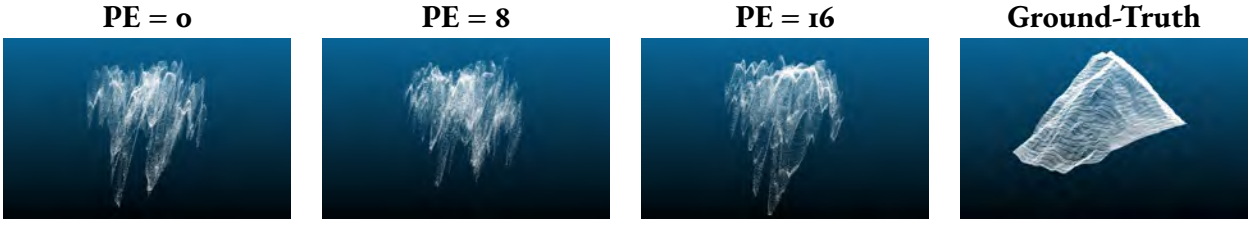



Figure 3.34: Surface reconstruction results for Sat-NeRF with various positional encoding (PE) frequencies, for the small Mount Doran scene. The model achieved a mean distance error of 52.934 m with a positional encoding frequency of 8.

Table 3.5: Quantitative results for Sat-NeRF with various positional encoding (PE) frequencies. The best metrics for each scene are in **bold**. For the task of novel view synthesis, the highest accuracy was achieved using a frequency 8 for both the Anchorage and Mount Doran scenes. For the task of 3D surface reconstruction, the highest accuracy was achieved using a frequency of 16 for the Anchorage scene and 8 for the Mount Doran scene.

| Scene | PE Frequency | PSNR↑ | SSIM↑ | Avg. Dist. (m)↓ | Std. Dev. (m) |
|---|---|---|---|---|---|
| Anchorage | 0 | 12.173 | 0.249 | 5.724 | 4.081 |
| | 8 | **15.573** | **0.657** | 4.576 | 3.538 |
| | 16 | 10.752 | 0.195 | **4.553** | **3.474** |
| Mt Doran | 0 | 12.797 | 0.462 | 58.031 | 44.813 |
| | 8 | **12.986** | **0.483** | **52.934** | **40.346** |
| | 16 | 12.837 | 0.464 | 61.0745 | 61.1343 |

### 3.5.3 Activation Function

The third ablation study investigated how the network activation function in the NeRF MLP influences the accuracy of the neural scene representation. Activation functions play a key role in determining how the MLP models complex relationships between inputs and outputs. Traditional NeRF methods often use ReLU activations, as they introduce non-linearity to the network, enabling it to model complex scene representations, such as surface geometry, occlusions, and lighting variations. However, while ReLU is effective for general scene encoding, its piecewise linear nature can limit the representation of high-frequency details. Current research has found that classical deep ReLU networks are incapable of accurately modeling first-order derivatives of the target signal, resulting in noisy artifacts [13]. An alternative to ReLU activations are sinusoidal activations, also called SIREN activations, which leverage periodic functions to better capture high-frequency details. SIREN activations allow the model to represent complex natural signals and their derivatives by enabling smoother gradient transitions [28]. This helps reduce noise in regions with intricate textures and surface geometry. However, networks that use SIREN activations are very sensitive to model weight initialization [13].

S-NeRF and Sat-NeRF both use SIREN activations by default. Thus, the goal of this ablation study was to determine whether the use of ReLU activations adversely impacts NeRF performance in novel view synthesis and surface reconstruction, as suggested in the literature. Although SIREN activations generally yield superior results by capturing high-frequency scene details, they are prone to training instability, particularly with scenes that contain fewer high-frequency details. Due to the feature-sparse nature of the Mount Doran and Anchorage scenes, this ablation study explores whether ReLU activations might

enhance training stability while still supporting fine-detail scene reconstruction. Experiments were run with ReLU and SIREN activations in the Sat-NeRF model. Qualitative view synthesis results from this ablation study are presented in Figures 3.35 - 3.38. The corresponding quantitative results can be seen in Table 3.5.

For both the Anchorage and Mount Doran scenes, the use of SIREN activations in the Sat-NeRF model dramatically improves novel view synthesis performance, as indicated by higher PSNR/SSIM scores on train and test data. When SIREN activations are used, the rendered images appear much sharper and capture more detail. However, when ReLU activations are used, all surface details are lost across both scenes. It is observed that ReLU activations have a tendency to smooth out the fine details in the rendered images. In the rendered test views of the Anchorage scene, noise is still observed, as shown in Figure 3.36, but it appears much more exaggerated and wide-spread with the use of ReLU activations, compared to SIREN activations. Overall, these results indicate that the use of SIREN activations in the Sat-NeRF model contributes to more accurate novel view synthesis performance for polar scenes.

Conversely, ReLU activations were found to achieve better 3D surface reconstructions results, as indicated by lower average distance errors between the predicted and ground-truth point clouds. However, despite the improvement in surface reconstruction performance, the Sat-NeRF model with ReLU activations generated point clouds with significant noise for both scenes. Notably, the use of ReLU activations caused the point cloud of the Anchorage scene to appear entirely flat with no noticeable topographic features, as shown in Figure 3.39. This is suspected to be attributed to the smoothing effect that ReLU activations introduce. On the other hand, the point cloud of Mount Doran generated with ReLU activations contained less fluctuations in surface geometry compared to the point cloud generated with SIREN activations. However, neither of the reconstructed scenes resemble the ground-truth, as shown in Figure 3.40. Thus, while ReLU activations appeared to improve the average distance error, this metric is ultimately misleading, as neither ReLU nor SIREN activations contributed to visually accurate scene reconstructions of Anchorage or Mount Doran. Overall, it is recommended, based on the results from this ablation study, that SIREN activations be used in satellite-based NeRF MLPs, as they dramatically improve novel view synthesis performance.

Figure 3.35: View synthesis results on training data with each activation function, for the small Anchorage scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 23.099/0.846 with the SIREN activation function.

Figure 3.36: View synthesis results on testing data with each activation function, for the small Anchorage scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 12.173/0.249 with the SIREN activation function.

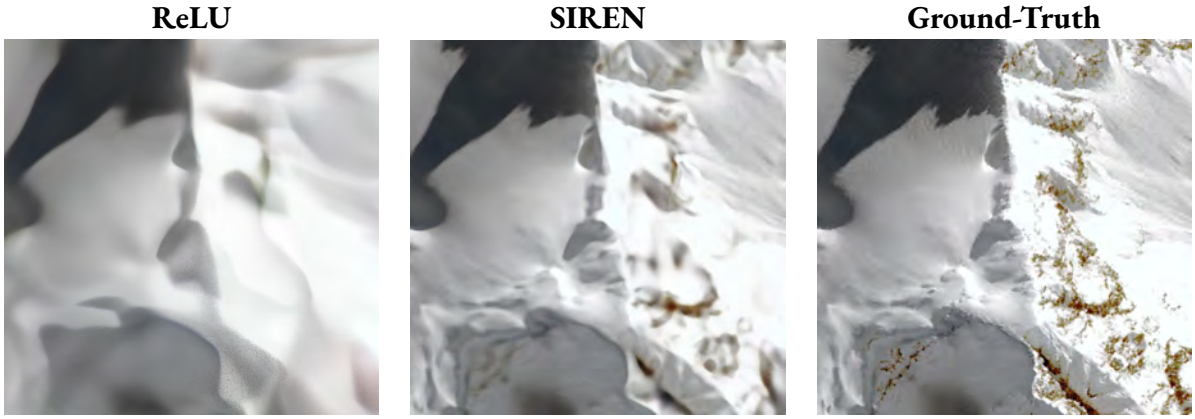



Figure 3.37: View synthesis results on training data with each activation function, for the small Mount Doran scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 23.573/0.783 with SIREN activations.

Figure 3.38: View synthesis results on testing data with each activation function, for the small Mount Doran scene. The Sat-NeRF model achieved high mean PSNR/SSIM scores of 12.797/0.461 with SIREN activation.

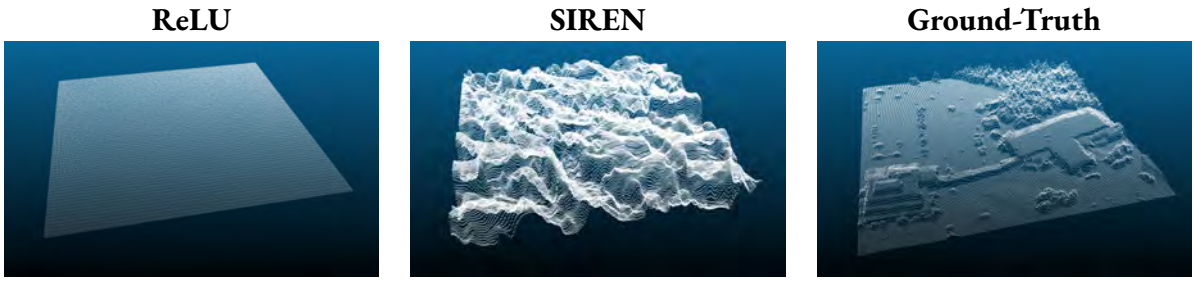



Figure 3.39: Surface reconstruction results for Sat-NeRF with each activation function, for the small Anchorage scene. The model achieved a mean distance error of 2.136 m with the ReLU activation function.

Figure 3.40: Surface reconstruction results for Sat-NeRF with each activation function, for the small Mount Doran scene. The model achieved a mean distance error of 51.854 m with the ReLU activation function.

Table 3.6: Quantitative results for Sat-NeRF with ReLU and SIREN activation functions. The best metrics for each scene are in **bold**. For the task of novel view synthesis, the highest accuracy was achieved using SIREN activations for both the Anchorage and Mount Doran scenes. For the task of 3D surface reconstruction, the highest accuracy was achieved using ReLU activations for both scenes.

| Scene | Activation Function | PSNR↑ | SSIM↑ | Avg. Dist. (m)↓ | Std. Dev. (m) |
|---|---|---|---|---|---|
| Anchorage | ReLU | 10.59 | 0.102 | **2.136** | **2.087** |
| | SIREN | **12.173** | **0.249** | 5.724 | 4.081 |
| Mt Doran | ReLU | 12.228 | 0.402 | **51.854** | **38.143** |
| | SIREN | **12.797** | **0.461** | 58.031 | 44.813 |

## 3.6 Discussion & Conclusion

A quantitative comparison of Metashape, S-NeRF, and Sat-NeRF is presented in Table 3.2. Metahshape outperformed S-NeRF and Sat-NeRF at the task of 3D surface reconstruction for both the Anchorage and Mount Doran scenes, achieving an average distance error of 2.652 m for the small Anchorage dataset and 13.979 m for the large Mount Doran dataset, with the included RPC camera models. The results in Figures 3.3 and 3.5 indicate that the use of RPC models is critical in 3D scene reconstruction. Without them, Metashape fails. The distribution of error in the point clouds generated by Metashape for the small Anchorage and Mount Doran scenes are presented in Figures 3.41 - 3.44. These results revealed that, when RPC models are used as input, Metashape has a tendency to under-estimate the surface altitude for both scenes. Specifically, areas in the Anchorage scene with transient objects and consistent snow coverage were reconstructed approximately 5-10 meters below the true surface. Conversely, areas with variations in shape and color, such as buildings and vegetation, were more accurately reconstructed, with distance errors closer to 0. There were no obvious patterns in the error distribution for the Mount Doran point cloud, although it is worth noting that Metashape successfully captured the steep elevation gradient throughout the scene. Overall, it can be concluded from these results that Metashape is successful at reconstructing polar scenes from satellite imagery with high geometric accuracy, when the RPC models for each image are included as input. Higher accuracies were observed for the urban scene, likely due to it containing a greater number of surface features from buildings and vegetation. Lower accuracies were observed for the mountainous scene, likely due to its homogeneous scene characteristics and complex surface topography.

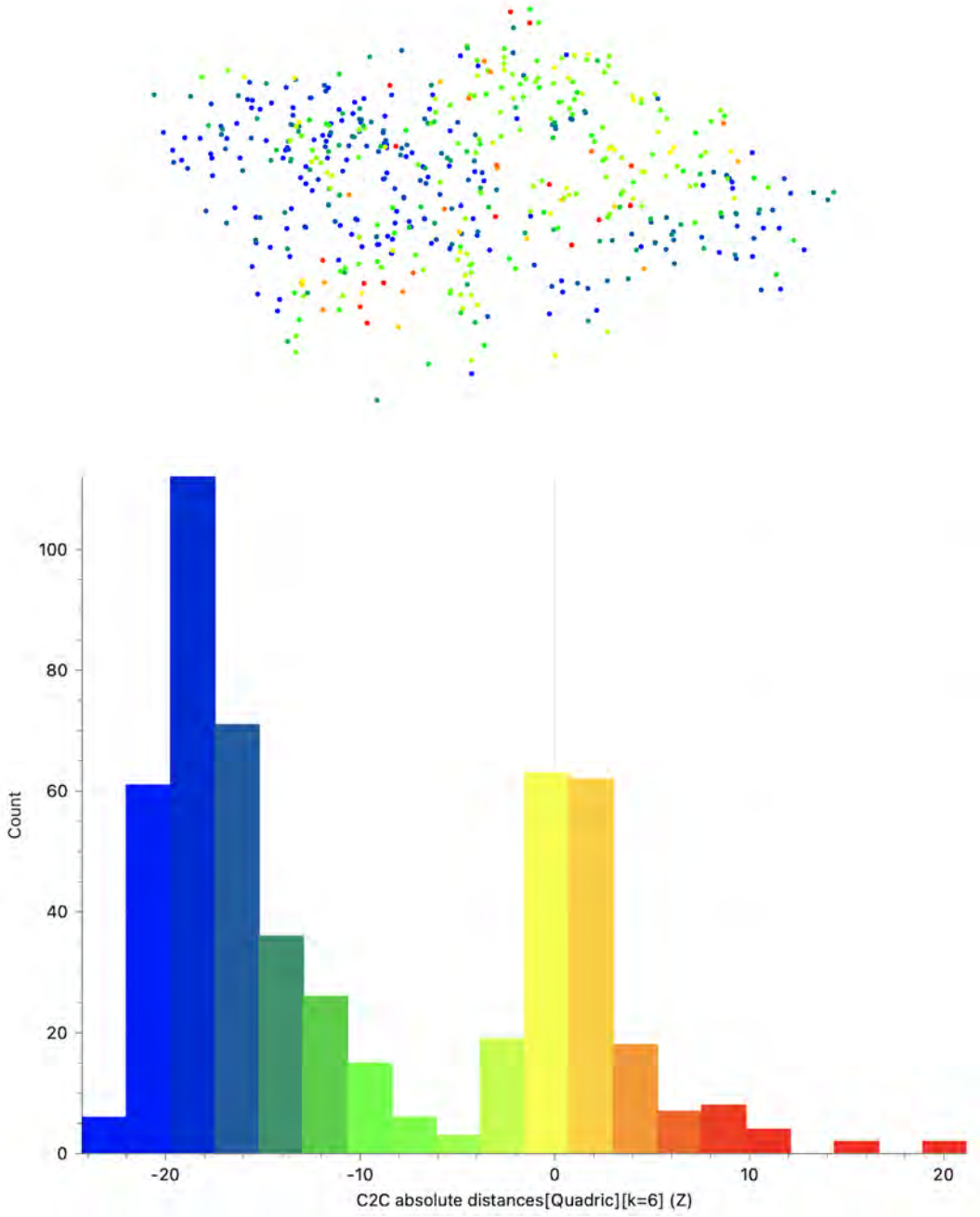



Figure 3.41: Metashape error distributions for the sparse point cloud of the small Anchorage scene (300 × 300 m). The images and RPC models were used as input. The sparse point cloud (top) and histogram (bottom) are colorized by the absolute distance error along the Z-dimension, with saturation bounds set to (-10.0, 10.0) m. Metashape achieved a mean distance error of 2.652 m for the sparse point cloud of the small Anchorage scene.

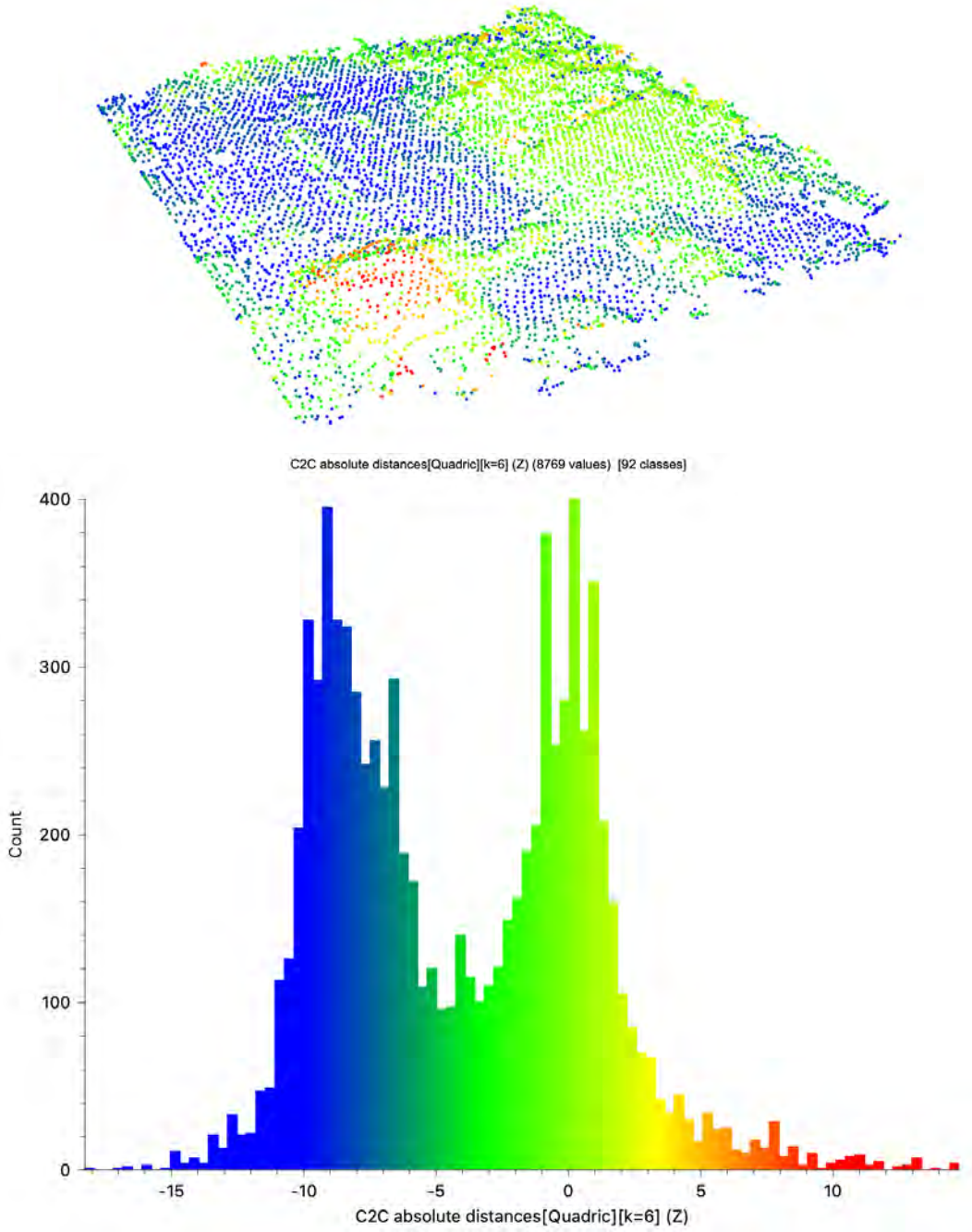



Figure 3.42: Metashape error distributions for the dense point cloud of the small Anchorage scene ($300 \times 300$ m). The images and RPC models were used as input. The dense point cloud (top) and histogram (bottom) are colorized by the absolute distance error along the Z-dimension, with saturation bounds set to ($-10.0$, $10.0$ m). Metashape achieved a mean distance error of 2.024 m for the dense point cloud of the small Anchorage scene.

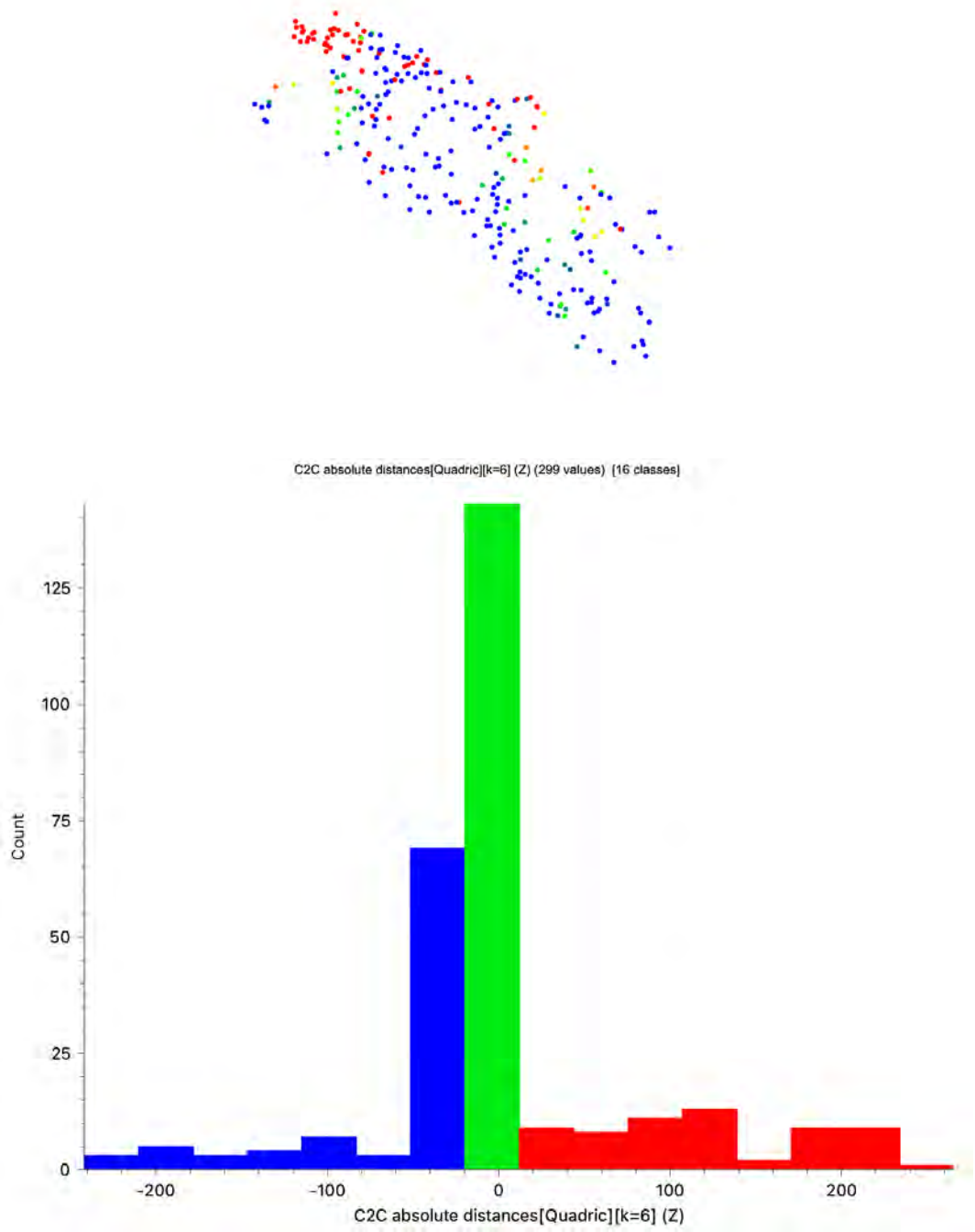



Figure 3.43: Metashape error distributions for the sparse point cloud of the small Mount Doran scene (600 × 600 m). The images and RPC models were used as input. The sparse point cloud (top) and histogram (bottom) are colorized by the absolute distance error along the Z-dimension, with saturation bounds set to (−10.0, 10.0 m). Metashape achieved a mean distance error of 15.624 m for the sparse point cloud of the small Mount Doran scene.

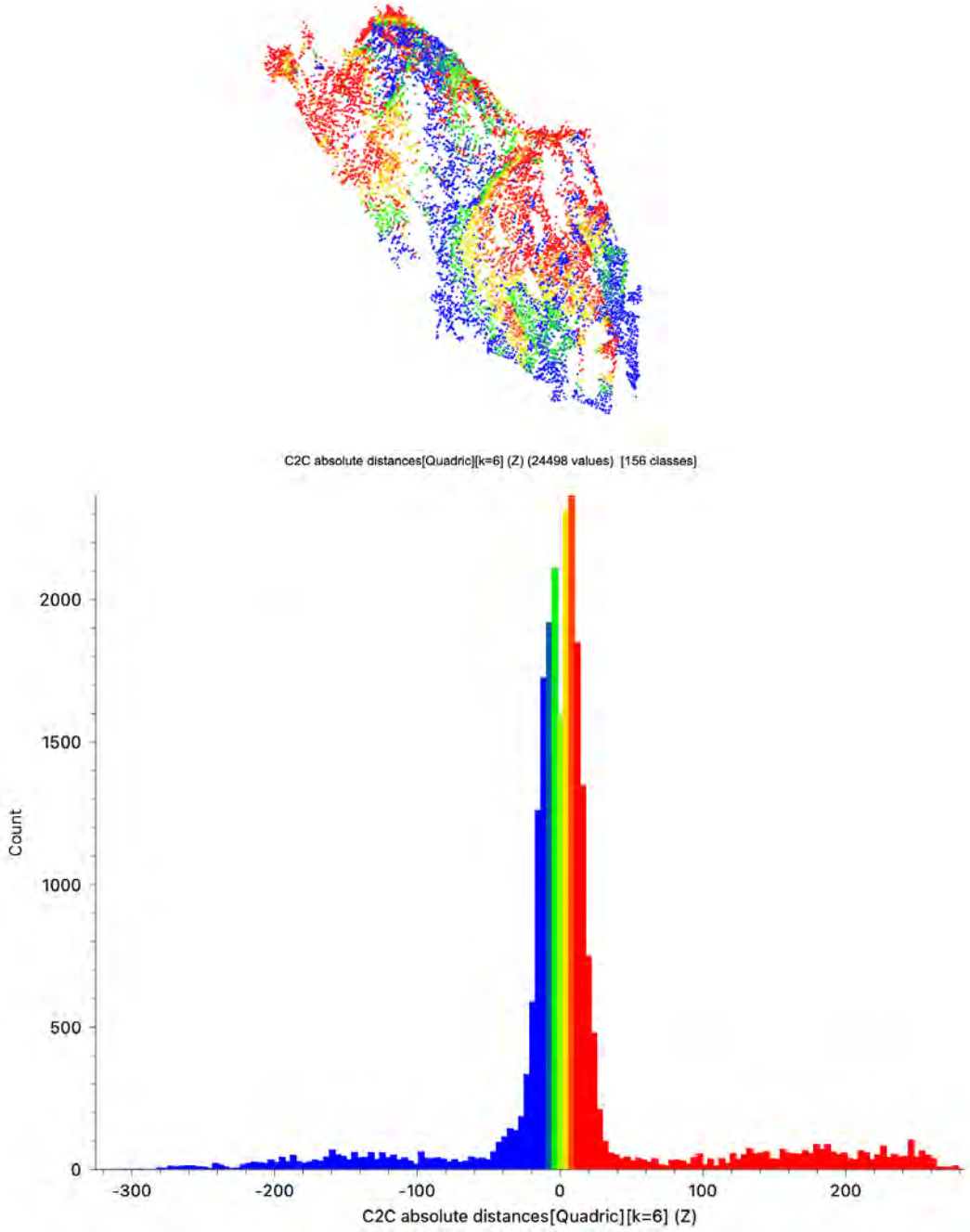



Figure 3.44: Metashape error distributions for the dense point cloud of the small Mount Doran scene ($600 \times 600$ m). The images and RPC models were used as input. The dense point cloud (top) and histogram (bottom) are colorized by the absolute distance error along the Z-dimension, with saturation bounds set to ($-10.0$, $10.0$ m). Metashape achieved a mean distance error of 13.034 m for the sparse point cloud of the small Mount Doran scene.

S-NeRF and Sat-NeRF failed at the task of 3D surface reconstruction with all polar datasets tested, as indicated by the significant noise in the generated point clouds. No matter the terrain type, AOI size, positional encoding frequency, or network activation function, the reconstructed 3D scenes had no visual resemblance to the ground-truth. These results highlight the inability of S-NeRF and Sat-NeRF at accurately reconstructing polar landscapes. For the S-NeRF model, the Anchorage and Mount Doran scenes were reconstructed as flat planes. Although the generated point clouds were of the correct scale and orientation, all surface details were lost, as shown in Figures 3.9 and 3.12. The reconstructed 3D scenes contained significant noise and failed to capture the intricate topographic details. Likewise, the point clouds generated from the Sat-NeRF model exhibited severe distortions, as shown in Figures 3.15 and 3.18. For both Anchorage and Mount Doran, the true surface features could not be discerned from the noise in the reconstructed scenes. Thus, the results from this research suggest that S-NeRF and Sat-NeRF are not well-suited for reconstructing high-resolution surface details in polar environments. The failure to capture any true surface topography emphasizes the limitations of these models in handling diverse scene complexities.

However, despite the limitations of S-NeRF and Sat-NeRF at 3D surface reconstruction, these methods performed well at novel view synthesis, especially with mountainous polar scenes. Both models struggled with rendering novel view of the Anchorage scene but showed dramatic improvement for the Mount Doran scene, as seen in Figure 3.19. Thus, it can be concluded that satellite-based NeRF methods perform best with homogeneous scenes that contain minimal textures and surface features. S-NeRF achieved PSNR/SSIM scores of 10.07/0.139 and 11.541/0.244 on test data for the small Anchorage and Mount Doran scenes, respectively. Sat-NeRF demonstrated even higher performance, achieving PSNR/SSIM scores of 12.173/0.249 and 12.797/0.461 for the small Anchorage and Mount Doran scenes, respectively. These results highlight the effectiveness of both models in generating visually accurate representations, even in challenging polar environments. Sat-NeRF achieved higher PSNR/SSIM scores compared to S-NeRF, suggesting that the use of RPC camera models improves the performance of NeRF methods at novel view synthesis with satellite data. As shown in Figures 3.13 and 3.16, views rendered by Sat-NeRF exhibited sharper details and more consistent color accuracy. This improvement underscores the importance of leveraging the native satellite camera models in scene reconstruction methods. In all cases, the use of RPC camera models improved performance across all methods tested in this research, both NeRF-based and SfM-based. In the context of satellite-based NeRF models, while challenges remain in 3D surface reconstruction, the advancements in novel view synthesis, through the integration of RPC models, offer promising avenues for further research and application in polar climate studies.

The ablation studies on the Sat-NeRF model revealed that, for the both the Anchorage and Mount Doran scenes, a small AOI size (300-600 $m^2$), a positional encoding frequency of 8, and the use of SIREN activation functions contributed to the most accurate novel view synthesis results. A moderate Fourier feature frequency in positional encoding improved novel view synthesis performance, as it enabled the Sat-NeRF model to accurately capture finer details in scene without over-fitting to the snow-covered areas. While higher frequencies have been found to aid in representing complex surfaces, the results from this research demonstrate that they can also introduce noise when using snow-covered polar datasets with a

mixture of intricate and homogeneous surface details. Additionally, a smaller AOI size benefited novel view synthesis performance for both the Anchorage and Mount Doran scenes, likely because it enables the Sat-NeRF model to focus on capturing finer local details. Although the highest PSNR/SSIM scores were achieved with the small Anchorage and Mount Doran scenes, Sat-NeRF successfully rendered novel views for all AOI sizes tested. Lastly, the use of SIREN activations further enhanced the ability of the Sat-NeRF model at capturing high-frequency details essential for photorealistic novel view synthesis, particularly in areas with complex textures, such as buildings and vegetation. SIREN activations allowed for sharper and more detailed renderings, compared to ReLU activations. Overall, it can be concluded from this research that a small AOI size, moderate positional encoding frequency, and SIREN activations provide an optimal configuration for novel view synthesis in both urban and mountainous snow-covered polar scenes. This combination enables Sat-NeRF to capture intricate textures and maintain a realistic representation of complex details, without excessive over-fitting to homogeneous surfaces. The findings from these ablation studies underscore the importance of adjusting model parameters based on both the scene characteristics and the specific task objectives, ensuring that NeRF is fine-tuned to balance accuracy and stability in the reconstruction of high-frequency details.

Overall, for the task of novel view synthesis, the Mount Doran scene achieved higher PSNR/SSIM scores than the Anchorage scene, for all ablation studies presented in this research. The reason why is likely due to there being less variations in surface color and texture throughout the scene. Mount Doran is dominated by a monotonous snow-covered surface, allowing the NeRF model to more easily generalize across views. In contrast, the diverse textures and colors in the Anchorage scene introduces a higher level of variability, and thus complexity, making it more challenging for the model to synthesize novel views with consistent accuracy. For this reason, slightly lower PSNR/SSIM scores are observed for the Anchorage scene, compared to the Mount Doran scene.

For 3D surface reconstruction, however, Anchorage achieved lower mean distance errors, likely due to the more consistent surface elevation across the scene. The Anchorage scene has a relatively narrow altitude range, from approximately 20 to 70 meters, allowing for more accurate depth estimation and reducing potential reconstruction errors. In contrast, the Mount Doran scene presents a challenging altitude range, varying widely from 270 to 1,108 meters. This significant elevation variability introduces complexities in accurately capturing depth, leading to higher mean distance errors in the 3D reconstruction of the scene. These findings suggest that scene-specific features such as elevation range, surface texture, and environmental conditions should be carefully evaluated when selecting NeRF parameters, as these factors significantly impact model performance across different tasks.

# Chapter 4

# Conclusion

This research explored the world of NeRF-based satellite photogrammetry with polar region datasets. All existing research on this topic has used the same data from the 2019 IEEE GRSS Data Fusion Contest (DFC2019) and 2016 IARPA Multi-View Stereo 3D Mapping Challenge (IARPA2016), which consists of WorldView-3 satellite images over urban scenes in Jacksonville, Florida, Omaha, Nebraska, and San Fernando, Argentina [11], [19]. In this research, satellite-based NeRF methods S-NeRF and Sat-NeRF were tested with newly collected datasets of WorldView-2 satellite images over a snow-covered urban scene in Anchorage, Alaska and a mountainous polar scene in the Chugach, Alaska. Results were compared against an industry-standard method for MVS photogrammetry, and ablation studies were conducted to determine the optimal dataset and model configurations for achieving the most accurate results. This research is believed to be the first to test satellite-based NeRF methods S-NeRF and Sat-NeRF with in-the-wild satellite image datasets of polar landscapes, thus revealing their performance in a broader context and ultimately paving the way for real-world applications of these methods in polar region scenarios.

Metashape outperformed S-NeRF and Sat-NeRF at the task of 3D surface reconstruction, generating point clouds of Anchorage and Mount Doran scenes with high degrees of accuracy. The RPC models of each input image played a crucial role in generating accurate reconstructions, as Metashape failed without them. Thus, this research concludes that a geometry-based method like Metashape can successfully generate accurate surface models of polar landscapes, with the use of satellite images and their corresponding RPC models. S-NeRF and Sat-NeRF, on the other hand, failed at the task of 3D surface reconstruction with the Anchorage and Mount Doran datasets. In all cases, the generated point clouds contained significant noise and captured no true topographic details. However, despite these results, the methods performed well at the task of novel view synthesis, especially for snow-dominated mountainous polar scenes, achieving PSNR/SSIM scores similar to those obtained with the original Jacksonville and Omaha datasets [15], [21]. Sat-NeRF performed best overall at novel view synthesis, due to it leveraging the RPC model of each image during the learning process. Overall, a considerable gap remains in current research regarding the generalization of satellite-based NeRF methods to sparsely-textured polar datasets. Based on the results presented in this research, it can be concluded that traditional MVS photogrammetric approaches remain more reliable for 3D surface reconstruction in these contexts.

## 4.1 Limitations and Future Work

This research presents the first step in exploring the potential of satellite-based NeRF methods for scene reconstruction of polar regions. However, there remain a number of limitations. First, it is important to note that all of the results presented in this research are highly dependent on the data used. The poor surface reconstruction performance of S-NeRF and Sat-NeRF is partially due to the data itself, and not necessarily the methods being used. There are many attributes of the Anchorage and Mount Doran datasets that could be to blame, the most influential perhaps being the restricted viewing elevation angles. As seen in Figure **??**, the distribution of viewing elevation angles in the Anchorage and Mount Doran datasets are more restricted than the Jacksonville dataset, originally used to test S-NeRF and Sat-NeRF. This is further illustrated in Figure 4.4, which compares the images with the lowest and highest viewing elevation angles from each dataset. It can be seen that the image from the Jacksonville dataset with the lowest elevation view sufficiently captures the sides of the buildings in the scene. However, the image from the Anchorage dataset with the lowest elevation view does not. This is due to the fact that the satellite images of Anchorage and Mount Doran come from high-latitude locations on Earth. Even though the low elevation views from the Anchorage and Jacksonville scenes are within a few degrees, the Anchorage scene is unable to capture the same viewing elevation range.

The restricted viewing elevation range is a known limitation of high-latitude satellite imagery. At high latitudes, satellite orbits converge, leading to significant overlap in adjacent imaging swaths [10]. This limits the viewing angle perspectives in polar satellite image datasets. Based on results presented in this research, it can be concluded that lower elevation views are critical for successfully reconstructing scenes with the S-NeRF and Sat-NeRF methods. The lack of oblique views hinders the ability of these methods at capturing the vertical structure and finer topographic details in homogeneous polar landscapes. Future work should prioritize the collection of satellite image datasets at lower latitudes to ensure a sufficient distribution of viewing angles. In particular, highly oblique views that capture the sides of prominent features in the scene will greatly improve the capability of satellite-based NeRF methods at representing detailed 3D structures in snow-covered and high-relief regions.

Additionally, it should be noted that the low solar elevation angles in the Anchorage and Mount Doran datasets are due to the images being captured around the winter season. In high-latitude regions on Earth, like Alaska, the Sun remains low on the horizon during the wintertime, which significantly affects the lighting conditions in satellite images. These low solar elevation angles result in longer shadows and reduced illumination, leading to increased contrast variations that could be negatively impacting S-NeRF and Sat-NeRF at generating accurate surface reconstructions. Shadows cast across large portions of the scene can obscure key features and create challenges in differentiating between actual surface contours and shadowed regions. Addressing these illumination constraints could enhance the performance of satellite-based NeRF methods in polar applications.

Finally, although there exist other NeRF methods which more accurately model reflectance and illumination than those explored here, such methods were primarily designed for casual, close-range image capture [12], [34], [35]. The effectiveness of these methods in reconstructing a scene from satellite remote

sensing images will inevitably be poor due to the highly constrained nature of the data. Therefore, only NeRF methods specifically tailored to satellite imagery were explored. S-NeRF and Sat-NeRF are the most suitable candidates for the datasets presented in this research, as they incorporate a shadow-aware irradiance model that can accurately reconstruct shadowed areas with minimal color variation and surface texture. The source code for EO-NeRF has not been made publicly available, so this method was not tested. However, it is important to note that S-NeRF, Sat-NeRF, and EO-NeRF all use a Lambertian function to model surface reflectance, which greatly oversimplifies light-reflectance interactions in a scene. This is a major limitation to consider when applying these methods to highly specular polar datasets. Nonetheless, these methods show promise in reconstructing monotone, repetitive scenes due to their ability at successfully modeling shadows, both visually and geometrically. Once an optimal sparsely-textured low-latitude dataset has been determined, future research should focus on integrating reflectance-aware and satellite-aware NeRF methods, to create a NeRF method capable of reconstructing highly specular scenes with varying material properties using satellite imagery alone.

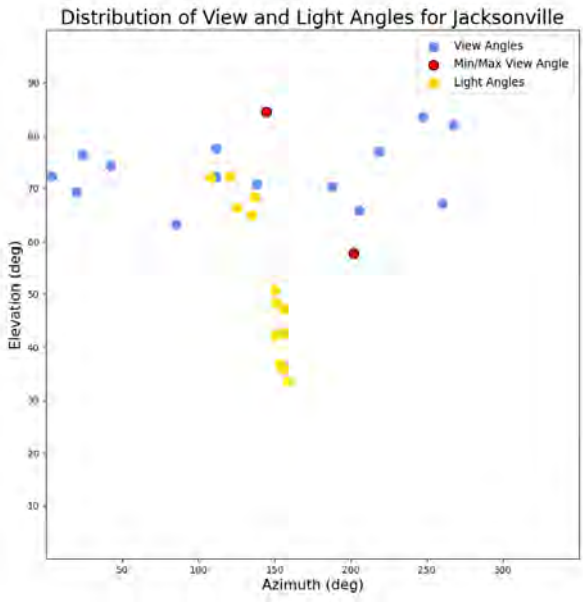



Figure 4.1: Plotted view and light angles for the Jacksonville dataset. The minimum and maximum viewing elevation angles are plotted in red.

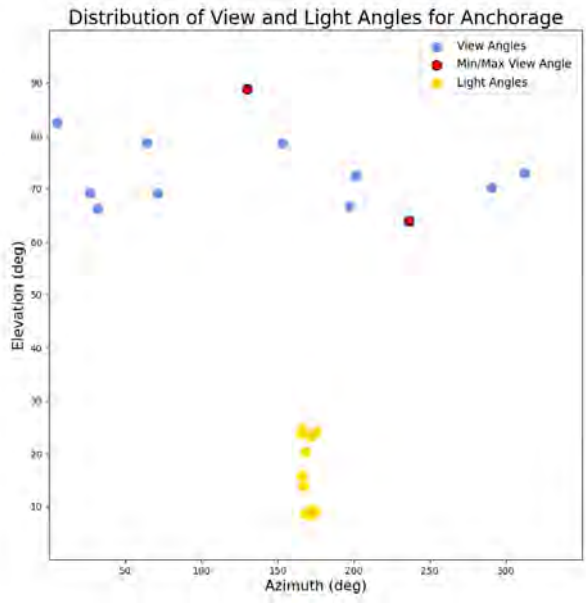



Figure 4.2: Plotted view and light angles for the Anchorage dataset. The minimum and maximum viewing elevation angles are plotted in red.

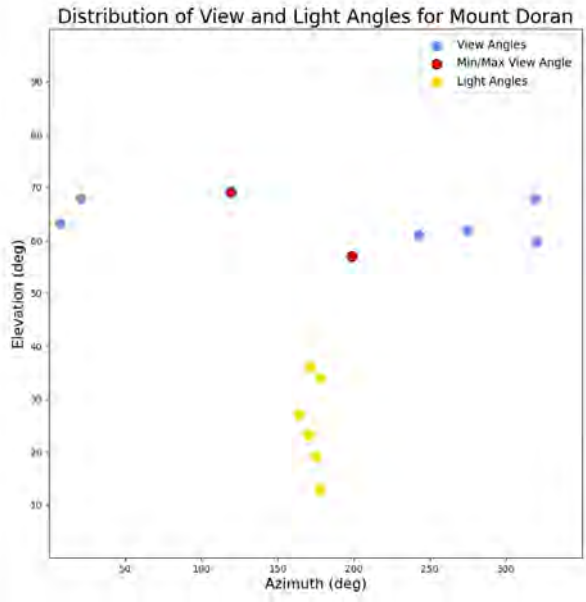



Figure 4.3: Plotted view and light angles for the Mount Doran dataset. The minimum and maximum viewing elevation angles are plotted in red.

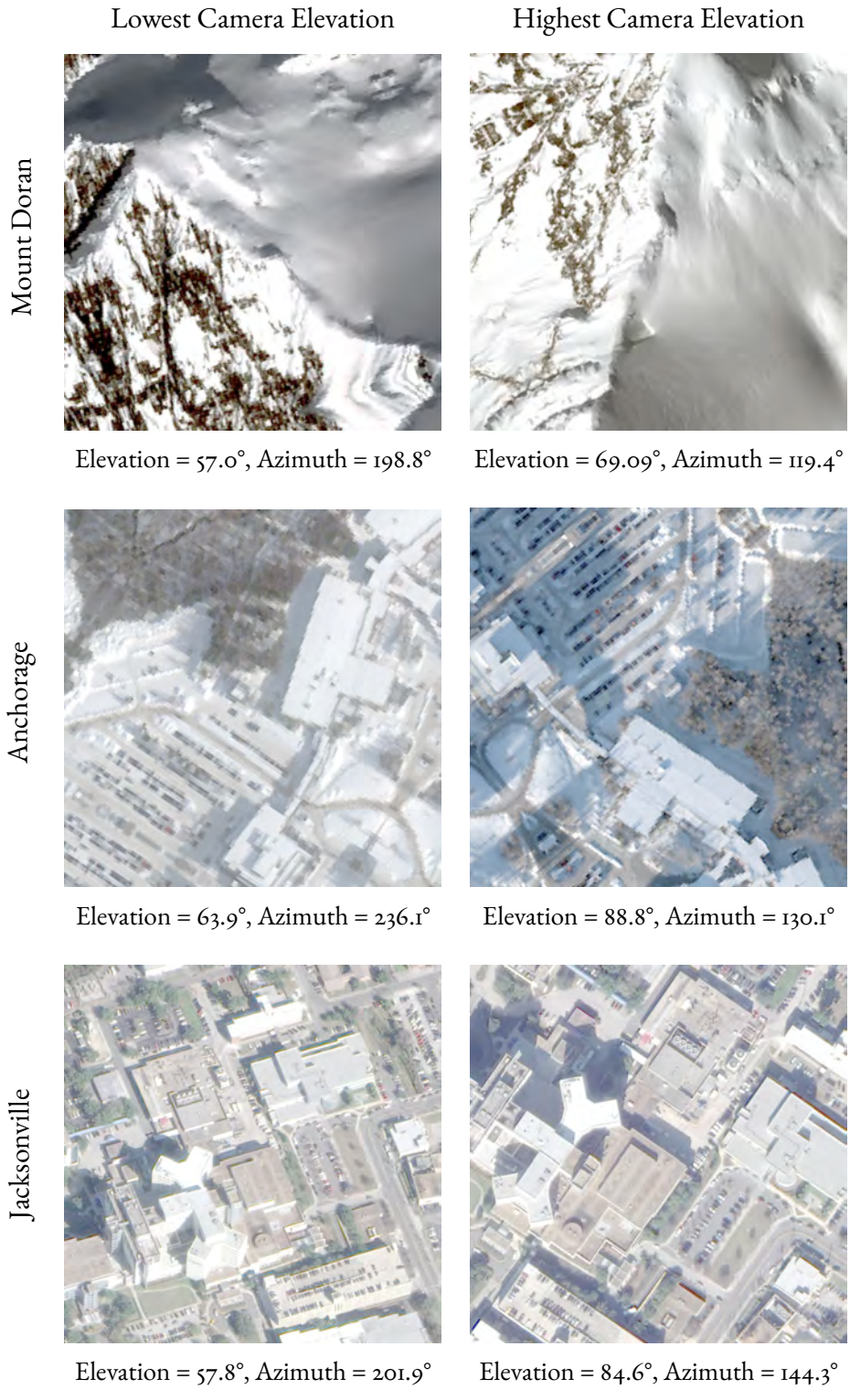



Figure 4.4: The images with the lowest and highest camera elevation angles from the Jacksonville, Anchorage, and Mount Doran datasets.

# Bibliography


[1] Oct. 2015. [Online]. Available: https://www.cloudcompare.org/doc/wiki/index.php/Cloud-to-Cloud_Distance.

[2] Nov. 2021. [Online]. Available: https://www.cloudcompare.org/doc/wiki/index.php/Distances_Computation.

[3] E. S. R. I. (Esri), *Arcgis*, version 10.8.2, 2020. [Online]. Available: https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview.

[4] A. F. Agarap, *Deep learning using rectified linear units (relu)*, 2019. arXiv: 1803.08375 [cs.NE]. [Online]. Available: https://arxiv.org/abs/1803.08375.

[5] E. S. Agency. [Online]. Available: https://earth.esa.int/eogateway/missions/worldview-2.

[6] A. H. Ahmadabadian, S. Robson, J. Boehm, M. Shortis, K. Wenzel, and D. Fritsch, "A comparison of dense matching algorithms for scaled surface reconstruction using stereo camera rigs," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 78, pp. 157–167, 2013, ISSN: 0924-2716. DOI: https://doi.org/10.1016/j.isprsjprs.2013.01.015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271613000452.

[7] R. Akiki, R. Mari, C. De Franchis, J.-M. Morel, and G. Facciolo, "Robust rational polynomial camera modelling for sar and pushbroom imaging," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, Jul. 2021. DOI: 10.1109/igarss47720.2021.9554583. [Online]. Available: http://dx.doi.org/10.1109/IGARSS47720.2021.9554583.

[8] O. Alexandrov, ScottMcMichael, M. Broxton, *et al.*, *NeoGeographyToolkit/StereoPipeline: 2024-09-04-daily-build*, version 2024-09-04-daily-build, Sep. 2024. DOI: 10.5281/zenodo.13690906. [Online]. Available: https://doi.org/10.5281/zenodo.13690906.

[9] S. Bianco, G. Ciocca, and D. Marelli, "Evaluating the performance of structure from motion pipelines," *Journal of Imaging*, vol. 4, no. 8, 2018, ISSN: 2313-433X. DOI: 10.3390/jimaging4080098. [Online]. Available: https://www.mdpi.com/2313-433X/4/8/98.

[10] R. Bindschadler, "Landsat coverage of the earth at high latitudes," *Photogrammetric Engineering Remote Sensing*, vol. 69, pp. 1333–1339, Dec. 2003. DOI: 10.14358/PERS.69.12.1333.

[11] M. Bosch, Z. Kurtz, S. Hagstrom, and M. Brown, "A multiple view stereo benchmark for satellite imagery," Oct. 2016, pp. 1–9. DOI: 10.1109/AIPR.2016.8010543.

[12] M. Boss, R. Braun, V. Jampani, J. T. Barron, C. Liu, and H. P. A. Lensch, *Nerd: Neural reflectance decomposition from image collections*, 2021. arXiv: 2012.03918 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2012.03918.

[13] S.-F. Chng, S. Ramasinghe, J. Sherrah, and S. Lucey, *Garf: Gaussian activated radiance fields for high fidelity reconstruction and pose estimation*, 2022. arXiv: 2204.05735 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2204.05735.

[14] *Cloudcompare (version 2.13.0)*, Retrieved from http://www.cloudcompare.org/, 2024. [Online]. Available: http://www.cloudcompare.org/.

[15] D. Derksen and D. Izzo, "Shadow neural radiance fields for multi-view satellite photogrammetry," *CoRR*, vol. abs/2104.09877, 2021. arXiv: 2104.09877. [Online]. Available: https://arxiv.org/abs/2104.09877.

[16] C. de Franchis, *Rpc_cropper*, https://github.com/carlodef/rpc_cropper, 2019.

[17] GDAL Development Team, *Gdal/ogr geospatial data abstraction software library*, Version 3.6.2, Open Source Geospatial Foundation, 2024. [Online]. Available: https://gdal.org/.

[18] Itseez, *Open source computer vision library*, https://github.com/itseez/opencv, 2015.

[19] B. Le Saux, N. Yokoya, R. Hänsch, and M. Brown, *Data fusion contest 2019 (dfc2019)*, 2019. DOI: 10.21227/c6tm-vw12. [Online]. Available: https://dx.doi.org/10.21227/c6tm-vw12.

[20] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–, Nov. 2004. DOI: 10.1023/B:VISI.0000029664.99615.94.

[21] R. Marí, G. Facciolo, and T. Ehret, "Sat-nerf: Learning multi-view satellite photogrammetry with transient objects and shadow modeling using rpc cameras," 2022. arXiv: 2203.08896 [cs.CV].

[22] R. Marí, G. Facciolo, and T. Ehret, "Multi-date earth observation nerf: The detail is in the shadows," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 2035–2045. DOI: 10.1109/CVPRW59228.2023.00197.

[23] R. Martin-Brualla, N. Radwan, M. S. M. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, *Nerf in the wild: Neural radiance fields for unconstrained photo collections*, 2021. arXiv: 2008.02268 [cs.CV].

[24] B. Mildenhall, P. Hedman, R. Martin-Brualla, P. Srinivasan, and J. T. Barron, *Nerf in the dark: High dynamic range view synthesis from noisy raw images*, 2021. arXiv: 2111.13679 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2111.13679.

[25] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *CoRR*, vol. abs/2003.08934, 2020. arXiv: 2003.08934. [Online]. Available: https://arxiv.org/abs/2003.08934.

[26] V. Mousavi, M. Khosravi, M. Ahmadi, *et al.*, "The performance evaluation of multi-image 3d reconstruction software with different sensors," *Measurement*, vol. 120, pp. 1–10, 2018, ISSN: 0263-2241. DOI: https://doi.org/10.1016/j.measurement.2018.01.058. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S026322411830071X.

[27] RPCM Developers, *Rpcm: Rational polynomial coefficient model library for python*, Version 1.0.0, 2024. [Online]. Available: https://github.com/cmla/rpcm.

[28] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, *Implicit neural representations with periodic activation functions*, 2020. arXiv: 2006.09661 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2006.09661.

[29] E. V. I. Solutions, *Envi*, version 5.6, 2020. [Online]. Available: https://www.nv5geospatialsoftware.com/Products/ENVI.

[30] M. Tancik, P. Srinivasan, B. Mildenhall, *et al.*, "Fourier features let networks learn high frequency functions in low dimensional domains," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33, Curran Associates, Inc., 2020, pp. 7537–7547. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/55053683268957697aa39fba6f231c68-Paper.pdf.

[31] P. Tarolli, "High-resolution topography for understanding earth surface processes: Opportunities and challenges," *Geomorphology*, vol. 216, pp. 295–312, 2014, ISSN: 0169-555X. DOI: https://doi.org/10.1016/j.geomorph.2014.03.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169555X14001202.

[32] T. Toutin, "Elevation modelling from satellite visible and infrared (vir) data," *International Journal of Remote Sensing*, vol. 22, no. 6, pp. 1097–1125, 2001. DOI: 10.1080/01431160117862. eprint: https://doi.org/10.1080/01431160117862. [Online]. Available: https://doi.org/10.1080/01431160117862.

[33] H. Turki, D. Ramanan, and M. Satyanarayanan, *Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs*, 2022. arXiv: 2112.10703 [cs.CV].

[34] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. T. Barron, and P. P. Srinivasan, *Ref-nerf: Structured view-dependent appearance for neural radiance fields*, 2021. arXiv: 2112.03907 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2112.03907.

[35] Z. Wang, T. Shen, J. Gao, *et al.*, "Neural fields meet explicit geometric representations for inverse rendering of urban scenes," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023.

[36] M. Westoby, J. Brasington, N. Glasser, M. Hambrey, and J. Reynolds, "'structure-from-motion' photogrammetry: A low-cost, effective tool for geoscience applications," *Geomorphology*, vol. 179, pp. 300–314, 2012, ISSN: 0169-555X. DOI: https://doi.org/10.1016/j.geomorph.2012.08.021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169555X12004217.

[37] Y. Xiangli, L. Xu, X. Pan, *et al.*, *Bungeenerf: Progressive neural radiance field for extreme multi-scale scene rendering*, 2023. arXiv: 2112.05504 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2112.05504.