Planet-NeRF

Neural Radiance Fields for 3D Reconstruction on Satellite Imagery in Season Changing Environments

Liv Kåreborn and Erica Ingerstad
Master of Science Thesis in Electrical Engineering

Planet-NeRF: Neural Radiance Fields for 3D Reconstruction on Satellite Imagery in Season Changing Environments

Liv Kåreborn and Erica Ingerstad
LiTH-ISY-EX–24/5631–SE

Supervisor:  
Amanda Berg  
ISY, Linköpings universitet

Tim Yngesjö  
Maxar Technologies

Justus Karlsson  
Maxar Technologies

Examiner:  
Leif Haglund  
ISY, Linköpings universitet

Computer Vision Laboratory  
Department of Electrical Engineering  
Linköping University  
SE-581 83 Linköping, Sweden

Copyright © 2024 Liv Kåreborn and Erica Ingerstad
Abstract

This thesis investigates the seasonal predictive capabilities of Neural Radiance Fields (NeRF) applied to satellite images. Focusing on the utilization of satellite data, the study explores how Sat-NeRF, a novel approach in computer vision, performs in predicting seasonal variations across different months. Through comprehensive analysis and visualization, the study examines the model’s ability to capture and predict seasonal changes, highlighting specific challenges and strengths. Results showcase the impact of the sun on predictions, revealing nuanced details in seasonal transitions, such as snow cover, color accuracy, and texture representation in different landscapes. The research introduces modifications to the Sat-NeRF network. The implemented versions of the network include geometrically rendered shadows, a signed distance function, and a month embedding vector, where the last version mentioned resulted in Planet-NeRF. Comparative evaluations reveal that Planet-NeRF outperforms prior models, particularly in refining seasonal predictions. This advancement contributes to the field by presenting a more effective approach for seasonal representation in satellite imagery analysis, offering promising avenues for future research in this domain.
Acknowledgments

We extend our sincere gratitude to our Maxar supervisors, Sanna Ringqvist, Tim Yngesjö, and Justus Karlsson, for their invaluable guidance and inspiration throughout the thesis. Special appreciation goes to Justus Karlsson for providing an exceptional conda environment to work in and patiently addressing numerous questions. Our heartfelt thanks to Amanda Berg, our Linköping University supervisor, for her profound inspiration and unwavering support during the research and writing process. Amanda, you are our hero, and we couldn’t have done this without you. We also acknowledge Leif Haglund, the Maxar veteran and also our examiner from Linköping University. Lastly, we express our appreciation to Maxar Technologies and the team of 3D Reconstruction for providing us with this enriching and enjoyable thesis opportunity.

Linköping, January 2024
Erica Ingerstad and Liv Kåreborn
# Contents

Notation ix

1 Introduction ........................................ 1
   1.1 Background ........................................ 1
   1.2 Motivation .......................................... 2
   1.3 Aim .................................................. 2
   1.4 Research questions .................................. 2
   1.5 Delimitations ........................................ 3
   1.6 Thesis outline ....................................... 3
   1.7 Author contributions ................................ 3

2 Related work ......................................... 5
   2.1 MLP - Multilayer Perceptron ....................... 5
   2.2 NeRF - Neural Radiance Fields .................... 6
      2.2.1 The principles of neural radiance fields ....... 6
      2.2.2 NeRF applications to satellite imagery ......... 8
      2.2.3 Geometric shadow rendering with EO-NeRF ....... 10
   2.3 NeuS - Learning Neural Implicit Surfaces ........ 11
      2.3.1 NeuS - A neural implicit SDF representation extension to NeRF .... 11
      2.3.2 Presence of varying illumination ............... 12
   2.4 Public methods ..................................... 13
   2.5 Scientific correlation between solar paths and seasons .... 13

3 Method ................................................. 15
   3.1 Sat-NeRF ............................................ 15
      3.1.1 Feature vector mapping ......................... 15
      3.1.2 Result variances in Sat-NeRF source code ...... 16
   3.2 Solar position experiments ....................... 16
   3.3 Sat-NeRF modifications ............................ 17
      3.3.1 Planet-NeRF: Introducing the month embedding vector .... 17
      3.3.2 Solar rays for season predictions ............... 18
      3.3.3 SDF insted of sigma ............................ 19
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>Training</td>
<td>21</td>
</tr>
<tr>
<td>3.5</td>
<td>Evaluation</td>
<td>22</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Metrics</td>
<td>22</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Data</td>
<td>24</td>
</tr>
<tr>
<td>3.6</td>
<td>Experimental setup</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>Results</td>
<td>29</td>
</tr>
<tr>
<td>4.1</td>
<td>Sat-NeRF for predicting scenes undergoing seasonal changes</td>
<td>29</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Quantitative results</td>
<td>30</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Altitude predictions</td>
<td>30</td>
</tr>
<tr>
<td>4.1.3</td>
<td>RGB predictions</td>
<td>32</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Solar impact of Sat-Nerf predictions</td>
<td>36</td>
</tr>
<tr>
<td>4.2</td>
<td>Solar impact on seasonal appearance</td>
<td>38</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Novel sun directions representing each month</td>
<td>38</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Novel equinox sun direction</td>
<td>39</td>
</tr>
<tr>
<td>4.3</td>
<td>Sat-NeRF modifications</td>
<td>41</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Quantitative results</td>
<td>41</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Geometric shadows</td>
<td>42</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Impact of month embedding vector</td>
<td>43</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Impact of solar layer</td>
<td>46</td>
</tr>
<tr>
<td>4.3.5</td>
<td>Signed distance function for altitude prediction</td>
<td>48</td>
</tr>
<tr>
<td>4.3.6</td>
<td>Loss</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>Discussion</td>
<td>51</td>
</tr>
<tr>
<td>5.1</td>
<td>Results, methodology, and future work</td>
<td>51</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Sat-NeRF</td>
<td>51</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Sun impacts</td>
<td>53</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Month embedding</td>
<td>53</td>
</tr>
<tr>
<td>5.1.4</td>
<td>Altitude</td>
<td>54</td>
</tr>
<tr>
<td>5.1.5</td>
<td>GEO+SL</td>
<td>55</td>
</tr>
<tr>
<td>5.1.6</td>
<td>Planet-NeRF+SDF</td>
<td>56</td>
</tr>
<tr>
<td>5.2</td>
<td>Ethical considerations</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>Conclusions</td>
<td>57</td>
</tr>
<tr>
<td>Bibliography</td>
<td></td>
<td>59</td>
</tr>
<tr>
<td>Notation</td>
<td>Meaning</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>$\sigma, \rho$</td>
<td>Volume density</td>
<td></td>
</tr>
<tr>
<td>$x$</td>
<td>3D-location in spatial coordinates</td>
<td></td>
</tr>
<tr>
<td>$d_{\text{cam}}$</td>
<td>2D-viewing direction</td>
<td></td>
</tr>
<tr>
<td>$d_{\text{sun}}$</td>
<td>3D-direction of solar rays</td>
<td></td>
</tr>
<tr>
<td>$t_j$</td>
<td>N-dimensional embedding feature vector of image $j$</td>
<td></td>
</tr>
<tr>
<td>$m_j$</td>
<td>4-dimensional embedding feature vector of month $j$</td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td>Predicted RGB color</td>
<td></td>
</tr>
<tr>
<td>$c_a$</td>
<td>Albedo RGB color</td>
<td></td>
</tr>
<tr>
<td>$c_m$</td>
<td>Albedo month/season RGB color</td>
<td></td>
</tr>
<tr>
<td>$c_{\text{gt}}$</td>
<td>Ground truth RGB color</td>
<td></td>
</tr>
<tr>
<td>$T$</td>
<td>Transmittance scalar</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Opacity scalar</td>
<td></td>
</tr>
<tr>
<td>$s$</td>
<td>Shading scalar</td>
<td></td>
</tr>
<tr>
<td>$s_s$</td>
<td>Season scalar</td>
<td></td>
</tr>
<tr>
<td>$a_{\text{sky}}$</td>
<td>Ambient sky color</td>
<td></td>
</tr>
<tr>
<td>$a_s$</td>
<td>Ambient season color</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>Uncertainty coefficient</td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>Transient scalar</td>
<td></td>
</tr>
<tr>
<td>$A_j, b_j$</td>
<td>Color transformation variables of image $j$</td>
<td></td>
</tr>
</tbody>
</table>
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDF</td>
<td>Signed Distance Function</td>
</tr>
<tr>
<td>MLP</td>
<td>Multi-Layer Perceptron</td>
</tr>
<tr>
<td>MEV</td>
<td>Month Embedding Vector</td>
</tr>
<tr>
<td>SEV</td>
<td>Season Embedding Vector</td>
</tr>
<tr>
<td>SL</td>
<td>Solar Layer</td>
</tr>
<tr>
<td>NeRF</td>
<td>Neural Radiance Fields</td>
</tr>
<tr>
<td>NeRF-W</td>
<td>NeRF in the Wild</td>
</tr>
<tr>
<td>S-NeRF</td>
<td>Shadow NeRF</td>
</tr>
<tr>
<td>SAT-NeRF</td>
<td>Satellite NeRF</td>
</tr>
<tr>
<td>EO-NeRF</td>
<td>Earth Observation NeRF</td>
</tr>
<tr>
<td>NeUS</td>
<td>Learning Neural Implicit Surfaces</td>
</tr>
<tr>
<td>GEO-NeUS</td>
<td>Geometry-Consistent NeuS</td>
</tr>
<tr>
<td>SAT-MESH</td>
<td>Satellite NeuS</td>
</tr>
<tr>
<td>Neuralrecon-W</td>
<td>Neural 3D Reconstruction in the Wild</td>
</tr>
</tbody>
</table>
This master thesis aims to investigate how neural radiance fields, the latest technology for neural 3D reconstruction, can handle seasonal changes in satellite images and how it can be modified for better accuracy. The thesis was carried out at Maxar Technologies, a leading aerospace technology company focused on earth observation and satellite related system services.

1.1 Background

Earth observation satellites are essential for monitoring the environment, and the amount of available data increases rapidly. Beneficial applications are found in e.g. meteorology [25], agriculture [3], forestry [21], biodiversity conservation [9], and regional planning [16]. As technology advances and the number of orbiting satellites increase, their significance grows [31]. However, the world we observe through earth observation satellites is far from static, and using these images for 3D reconstruction faces challenges. Due to environmental variations, including seasonal changes, varying incoming light, and the presence of transient objects, multi-date imagery is frequently disregarded in novel view synthesis. At the same time, these extensive collections of images contain valuable information about how specific regions change over time.

Recent advances in neural 3D reconstruction are neural radiance fields (NeRF) [17], a technique that has demonstrated impressive capabilities of creating accurate representations of 3D objects or scenes by representing them as a 5D-vector valued function or field. This function implicitly handles tasks such as synthesizing new views and reconstructing 3D information, matching geometry and colors with the camera projections from various angles [17]. Since the initial research in this area [17], numerous researchers have explored different aspects [4, 12, 14, 15, 20]. Some have investigated how this approach copes with tran-
sient objects, such as cars [15, 17], while others have explored its application to multi-date satellite images [4, 12, 14, 20].

1.2 Motivation

Although previous works adapt the use of NeRF to satellite imagery by handling the occurrence of transient objects, an intriguing yet underexplored terrain lies in understanding its capabilities within dynamic environments subject to seasonal changes. Seasonal variations include, among other things, varying amounts of snow, and the density and color appearance of vegetation. Handling images from various seasons can pose challenges in creating an accurate 3D representation of the scene since the seasons contribute to changes in geometry, color, and illumination. The core motivation of this project is to explore how NeRF can be used to extract the seasonal appearances in multi-date imagery, enabling the creation of accurate 3D models that contain seasonal states. The objective of this thesis is to evaluate how existing neural rendering methods for 3D reconstruction on satellite imagery handle the complexities of shifting seasons. In particular, this thesis will center its evaluation on Sat-NeRF, the groundbreaking work of Mari et al. [12]. Furthermore, this research will analyze the use of solar rays as inputs to a NeRF network, contributing to the comprehensive understanding of its capability to capture seasonal variations. Additionally, this thesis will investigate the possibility of enhancing the performance of these methods to better address the unique challenges posed by scenes that change with the seasons.

1.3 Aim

The thesis investigates how neural radiance fields can model 3D scenes in environments with seasonal changes. Through this research, we seek to enhance the understanding of NeRF applicability to the domain of earth observation, with the ultimate goal of providing a more accurate and adaptable approach for 3D reconstruction in dynamic real-world environments, and offering a foundation for future research endeavors and practical applications that harness the full potential of this groundbreaking technology. Addressing the challenge of seasons will hopefully aid in creating an earth observation tool capable of truthfully representing our planet anywhere at any time of the year.

1.4 Research questions

The thesis investigates the techniques of neural radiance fields on satellite images in seasonal changing areas by answering the following research questions:

1. How well does Sat-NeRF network perform on scenes with seasonal changes according to the chosen metric systems?
2. What is the significance of solar directions as an input to the Sat-NeRF network for the understanding and prediction of seasonal variations?

3. Can the performance of Sat-NeRF be improved on scenes that undergo seasonal changes by altering the network architecture?

1.5 Delimitations

Given that NeRF is a rapidly evolving field with continuous releases of new papers, the thesis focused exclusively on investigating the performance of Satellite NeRF (Sat-NeRF) in environments with seasonal changes.

While concluding this thesis, the authors became aware of the release of Season-NeRF [5]. It is important to note that Season-NeRF was not a source of inspiration for the methodology employed; rather, it was utilized solely in the results section for comparative purposes.

Furthermore, it should be mentioned that the reason for not including a comparison with Earth Observation NeRF (EO-NeRF) in this study is due to the unavailability of EO-NeRF’s source code at the time of this research. Without access to their source code, the authors were unable to run EO-NeRF’s model on the data sets used for this thesis, thus precluding a direct comparison of results.

1.6 Thesis outline

The thesis contains six chapters. In chapter 2 the reader is provided with the important theory, key concepts and related work in the area. The method of investigation to answer the research questions is found in chapter 3. The results are presented in chapter 4 and a deeper discussion of the results can be found in chapter 5. Finally, the conclusions that emerged from the results and discussion are presented in chapter 6.

1.7 Author contributions

The distribution of tasks between the authors was equitable, and a continuous open dialogue was maintained throughout the project. Both authors shared responsibilities in evaluating the performance of Sat-NeRF concerning research question 1, and in modifying the network architecture to enhance seasonal predictions in texture accuracy, as outlined in research question 3. However, Erica Ingerstad assumed primary responsibility for evaluating the significance of solar rays for the learning and prediction of seasonal variations, pertaining to research question 2, while Liv Kåreborn primarily investigated the potential improvement in depth accuracy by adjusting the network to predict a signed distance function instead of a volume density function, concerning research question 3.
The current chapter delves into an exploration and analysis of existing literature and research directly relevant to the subject matter under investigation. The review includes a broad examination of previous studies, methodologies, and findings significant to the research domain. By carefully examining these earlier works, the goal is to place our contributions within the bigger picture of what is already known. The critical evaluation of related research is crucial for identifying gaps, challenges, and opportunities that guide and inform the unique contributions presented in this scientific discourse.

### 2.1 MLP - Multilayer Perceptron

The NeRF and NeuS models, that are described further in Section 2.2 and Section 2.3.1, consist of a deep fully-connected neural network without convolutional layers which is referred to as a multilayer perceptron (MLP), $F$. MLP is a feed-forward neural network that uses backpropagation and is characterized by having a high degree of connectivity, as every neuron in one layer is connected to every neuron in the next layer, which allows learning complex patterns and relationships in the data. The typical operation in a fully-connected layer involves computing a weighted sum of the inputs (including the output from the previous layer) and then applying an activation function to the result.

A MLP network works well on non-linear data and captures the complex relationships between the inputs and the outputs. The training process of NeRF, Sat-NeRF, EO-NeRF, and NeuS consists of minimizing the cost functions related to each method [4, 10, 12, 14, 15, 17, 20, 28].
2.2 NeRF - Neural Radiance Fields

NeRF, first presented by Mildenhall et al. [17] in 2020, shows effective results for generating novel views of complex 3D scenes from 2D input images. Since the method was presented it has been further developed by multiple researchers to handle e.g. sparse inputs [19, 30], slow training time [7, 11, 22], large scale views [26], satellite imagery [12, 14], or inconsistent lighting [4, 15]. This section begins by introducing the Multilayer Perceptron (MLP), which serves as a fundamental component in the prior works mentioned in this section [4, 10, 12, 14, 15, 17, 20, 28]. Subsequently, the foundational principles of NeRF are presented, followed by extensions to NeRF relevant to satellite imagery [4, 12, 14] and considerations for seasonal changes [20].

2.2.1 The principles of neural radiance fields

Neural Radiance Fields, NeRF, was first presented in 2020 by Mildenhall et al. [17]. The main idea of NeRF is to use a deep fully-connected neural network to predict a continuous 3D-representation of a scene at unseen angles using 2D-images. The method of NeRF represents a continuous scene as a 5D-vector valued function $\mathcal{F}$, where the input consists of a 3D-location $\mathbf{x} = (x, y, z)$ and a 2D-viewing direction $\mathbf{d}_{\text{cam}} = (\theta, \phi)$. The output is an emitted color $\mathbf{c} = (r, g, b)$ and a volume density $\sigma$. The continuous 5D-scene representation is approximated with a multi-layer perceptron (MLP) network $\mathcal{F}_\theta : \mathbf{x}, \mathbf{d}_{\text{cam}} \rightarrow (\mathbf{c}, \sigma)$ which maps each input 5D-coordinate to its corresponding volume density and directional emitted color.

Principles from classical volume rendering techniques are used for rendering the color of the rays passing through the scene. Individual rays are traced across the scene and the volume density can be described as a differential probability of the ray hitting an intersection. This approach is illustrated in Figure 2.1. The expected color $\mathbf{c}(\mathbf{r})$ of a camera ray $\mathbf{r}(t) = \mathbf{o} + t \mathbf{d}, t \in [t_n, t_f]$, where $t_n$ and $t_f$ represents the near and far bounds of the scene, can be described by the following equation:

$$
\mathbf{c}(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \sigma(\mathbf{r}(t)) \mathbf{c}(\mathbf{r}(t), d) \, dt,
$$

(2.1)

where $T(t) = \exp \left( - \int_{t_n}^{t_f} \sigma(\mathbf{r}(s)) ds \right)$.

The function $T(t)$ describes the probability that a ray goes from $t_n$ to $t_f$ without hitting an object. In practice, the predicted output of the color approximates the integral by summing sampled points. The following equation describes the approximation for $N$ discrete samples:
2.2 NeRF - Neural Radiance Fields

\[
\mathbf{c}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i \mathbf{c}_i, \text{ where } T_i = \exp \left( - \sum_{j=1}^{i-1} \sigma_j \delta_j \right),
\]

\[
\alpha_i = (1 - \exp (-\sigma_i \delta_i)) \text{ and } \delta_i = t_{i+1} - t_i.
\] (2.2)

NeRF trains two networks, firstly a coarse network which outputs the color prediction \( \mathbf{c}_c(\mathbf{r}) \). The output from the coarse network is used to train the fine network with the color estimation output \( \mathbf{c}_f(\mathbf{r}) \) that becomes the final rendering prediction of the network. The optimization process of the training is performed by taking the total squared error between the true pixel color and the rendered color from both the coarse network and the fine network. The loss function is described by the following equation, where \( \mathbf{c}_{gt}(\mathbf{r}) \) represents the ground truth color of the rays in a set \( \mathcal{R} \).

\[
\mathcal{L} = \sum_{\mathbf{r} \in \mathcal{R}} \left[ \| \mathbf{c}_c(\mathbf{r}) - \mathbf{c}_{gt}(\mathbf{r}) \|^2_2 + \| \mathbf{c}_f(\mathbf{r}) - \mathbf{c}_{gt}(\mathbf{r}) \|^2_2 \right].
\] (2.3)

**Figure 2.1:** The training pipeline of NeRF. (a) The network takes in a 5D input vector with position and direction \((x, y, z, \theta, \phi)\) and passes it through the MLP function \( F_\Theta \) which outputs the predicted color and density \((RGB\sigma)\) for each point (b). Classic volume rendering techniques are used to combine the points along the ray and composite them into an image (c). The rendering is optimized by minimizing the residual of the predicted color and the true color from the synthesized image and the ground truth image (d). Illustration from [17].
2.2.2 NeRF applications to satellite imagery

One major limitation of the original NeRF [17] is that it only uses input images taken around the same time and with similar lighting to get an accurate representation of the scene. This implies that it faces problems when presented with images contain e.g. dynamic objects, seasonal variations, or shadows. An advanced NeRF method, known as Satellite NeRF (Sat-NeRF), was developed by Marí et al. [12] in 2022. Sat-NeRF was inspired by S-NeRF [4], which was the first published NeRF approach that addressed the problem of using satellite imagery for 3D-reconstruction. Sat-NeRF could handle more complex scenes, i.e. scenes with transient objects and inconsistent lights on a global scale. Instead of having the color depend on a 2D-viewing direction, \( d_{\text{cam}} \), Sat-NeRF uses a 3D-direction of the solar rays, denoted as \( d_{\text{sun}} \), as input and assumes the scene is represented by a Lambertian surface. In order to handle the transient objects that usually occur in satellite datasets which contain images captured with large time variation, the authors of Sat-NeRF creates an \( N \)-dimensional embedding feature vector \( t_j \) for each image \( j \). The vector \( t_j \) contains information of the transient elements that cannot be explained by the position of the sun. The method of Sat-NeRF represents a scene using the volumetric function \( F : (x, d_{\text{sun}}, t_j) \rightarrow (\sigma, c_a, s, \alpha_{\text{sky}}, \beta) \), where the outputs \( \sigma \) and \( c_a \) are the same as in NeRF and represent the volume density and the albedo RGB color. Sat-NeRF also has the additional outputs \( s, \alpha_{\text{sky}} \), and \( \beta \), where \( s \) represents the shadows learned from the geometry x together with the solar ray directions \( d_{\text{sun}} \), \( \alpha_{\text{sky}} \) is an ambient sky color learned from the solar ray directions \( d_{\text{sun}} \) independent from the geometry x. Finally \( \beta \) is the uncertainty coefficient related to the probability that the color of a pixel is explained by a transient object [12].

The Sat-NeRF network architecture, illustrated in Figure 2.2, comprises a primary block of fully-connected layers dedicated to predicting static scene properties (volume density \( \sigma \) and albedo color \( c_a \)) using \( h \) channels per layer. A secondary block with fewer layers and half the channels estimates shading scalar \( s \) based on solar rays’ direction \( d_{\text{sun}} \) and \( h \) geometry-related features from the main block. Additionally, two single-layer blocks predict uncertainty coefficient \( \beta \) and ambient sky color \( \alpha_{\text{sky}} \) from transient embedding vector \( t_j \) and \( d_{\text{sun}} \), respectively.

SIREN layers, initialized as proposed in [24] and recommended in [4], are employed. A softplus function predicts \( \sigma \). The uncertainty \( \beta \) also uses a softplus [15] for smoother optimization compared to the typical ReLU [1]. Outputs related to normalized RGB values are processed with sigmoid functions to constrain them within \([0, 1]\). The value of \( h \) is set to 512, adjustable based on resolution and observed area size.
Sat-NeRF only trains a coarse network, and the output of the network is then rendered using the same techniques as in NeRF [17]. However, Sat-NeRF uses another loss function during training which incorporates a solar correction term $L_{SC}$, a depth supervision term $L_{DS}$ and by weighting the RGB color loss with the uncertainty prediction $\beta$. The evolved loss function is described in the following equation:

$$L = L_{RGB}(\mathcal{R}) + \lambda_{SC} L_{SC}(\mathcal{R}_{SC}) + \lambda_{DS} L_{DS}(\mathcal{R}_{DS}),$$

where:

$$L_{RGB}(\mathcal{R}) = \sum_{r \in \mathcal{R}} \left( \frac{\|c(r) - c_{GT}(r)\|_2^2}{2\beta'(r)^2} + \frac{\log \beta'(r) + \eta}{2} \right),$$

$$L_{SC}(\mathcal{R}_{SC}) = \sum_{r \in \mathcal{R}_{SC}} \left( \sum_{i=1}^{N_{SC}} (T_i - s_i)^2 + 1 - \sum_{i=1}^{N_{SC}} T_i \alpha_i s_i \right),$$

and

$$L_{DS}(\mathcal{R}_{DS}) = \sum_{r \in \mathcal{R}_{DS}} d_{\text{sun}}(r) (d(r) - \|X(r) - o(r)\|_2)^2.$$

The uncertainty prediction $\beta$ is given by: $\beta'(r) = \beta(r) + \beta_{\text{min}}$, where $\beta_{\text{min}} = 0.05$ and $\eta = 3$. The logarithm is included to prevent $\beta$ from approaching infinity. The solar correction term in Equation 2.6 is added to the loss to minimize the shading scalar $s$ to produce unrealistic results for solar ray directions that are not seen in the training data [12].
2.2.3 Geometric shadow rendering with EO-NeRF

Another recent work that handles inconsistent lighting in satellite imagery is Multi-Date Earth Observation NeRF (EO-NeRF) [14]. EO-NeRF represents a scene using the volumetric function \( F: (x, d_{\text{sun}}, t_j) \to (\sigma, c_a, a_{\text{sky}}, \beta, \tau, A_j, b_j) \), where the inputs \( x, d_{\text{sun}}, \) and \( t_j \) and the outputs \( \sigma, c_a, a_{\text{sky}}, \) and \( \beta \) are the same as in Sat-NeRF, while the outputs \( \tau, A_j, \) and \( b_j \) are added. \( \tau \) is a transient scalar learned from the geometry \( x \) and the embedding feature vector \( t_j \). \( A_j \) and \( b_j \) are RGB vectors used for encoding a color transformation between the albedo and the current image \( j \). The Sat-NeRF method predicts the shadows \( s(r) \) as a scalar in the interval \([0,1]\), depending on the solar direction [12]. Points in the shade or points involving transient objects correspond to low values of \( s(r) \) resulting in a large contribution from the predicted ambient sky color \( a_{\text{sky}} \). Points with a high value of \( s(r) \) let the color prediction be explained by the albedo RGB color, \( c_a \), to a large extent. As a result, this method does not generalize well to new solar direction inputs. While Sat-NeRF handles this problem by adding a solar correction term to the loss, the EO-NeRF method renders the shadows based on the geometry and the solar ray directions instead of predicting them. To represent the transient objects they use a transient scalar \( \tau \). The network architecture of EO-NeRF is presented in Figure 2.3 and the loss function used in the network is described in Equation 2.5.

![Figure 2.3: EO-NeRF network architecture, with \( x, d_{\text{sun}}, \) and \( t_j \) as inputs and \( \sigma, c_a, a_{\text{sky}}, \beta, \tau, A_j, \) and \( b_j \) as outputs. In the figure \( a_{\text{sky}} \) is represented as \( a \). Illustration from [14].](image-url)
2.3 NeuS - Learning Neural Implicit Surfaces

The following sections, 2.3.1 and 2.3.2, describe the method of learning neural implicit surfaces and how its research have improved some of the difficulties in NeRF.

2.3.1 NeuS - A neural implicit SDF representation extension to NeRF

The volume rendering in the various methods of NeRF faces problems when it comes to surface reconstruction. The learned implicit representation of a volume density field suffers from bias, inherits geometric errors, and has difficulties extracting high-quality surfaces since it is rather intended for novel view synthesis. A method that proposes a solution to this is NeuS [28].

NeuS represents an object using two functions: one that estimates the signed distance from a point to the object by learning a neural implicit signed distance function (SDF) representation and another that encodes the color based on the position of the point and the viewing direction. These functions are created using a MLP network. The surface of an object is defined as the zero-level set of the SDF. To train the SDF network, a probability density function called S-density is introduced. This function is derived from the Sigmoid function and is used for volume rendering with 2D input images as supervision [28]. By using a SDF for surface representation NeuS manages to achieve an accurate surface representation, and by learning a neural implicit SDF representation from a novel volume rendering scheme they manage to handle the presence of abrupt depth changes. In comparison to NeRF, this method deals with the problems of conspicuous noise that sometimes occur on planar regions. Besides NeRF, NeuS also outperforms other

Figure 2.4: Sat-Mesh overview. Points of the scene are sampled along the emitted rays from the satellite view. The positions and directions of these points are input to the two MLPs which predict the SDF and color. Illustration from [20].
state of the art neural scene representation methods in terms of reconstruction quality [28].

The method used by NeuS has also shown great results on satellite imagery [20]. Following the approach of learning neural implicit surfaces [28], Sat-Mesh [20] is a 2-model method that predicts the SDF and the pixel color. An overview of the Sat-Mesh method is shown in Figure 2.4. The color prediction function in Sat-Mesh will be further described in Section 2.3.2.

2.3.2 Presence of varying illumination

Further extensions to the neural surface reconstruction work is the method of Neuralrecon-W [10], that combines the techniques of NeuS [28] and the latent appearance modeling used in Sat-NeRF [12]. The network represents the scene using two neural implicit functions, \( f(x) = \text{MLP}_{\text{SDF}}(x) \) is used to predict the signed distance to the true surface, and \( c_j = \text{MLP}_{\text{COLOR}}(x, d_{\text{cam}}, t_j) \) is used to predict the color of a point \( x \) as it appears in a given image \( j \). \( \{t\}_{j=1}^{N} \) are appearance embeddings for transient objects corresponding to each input image. A further extension to the color prediction function is Sat-Mesh, presented by Qu and Deng [20]. The Sat-Mesh method incorporates a latent embedding vector \( \{L\}_{j=1}^{N} \) in the training process to encode every image appearance. The method allows the model to learn and render the different seasons and enables to texture the mesh with the corresponding seasonal appearances of each image. It makes it possible to retrospectively apply the features from a specific image without altering the underlying geometry [20].

![Sat-Mesh network architecture. Illustration from [20].](image-url)
2.4 Public methods

In this chapter, source code is accessible for the original NeRF paper [18], SatNeRF [13], and NeuS [29]. The source code from Sat-NeRF is used in this thesis.

2.5 Scientific correlation between solar paths and seasons

In order to understand the phenomena of seasons, one must understand the movement of our planet in the solar system. The Sun’s position seen from Earth is related to the Earth’s axial tilt and its orbit around the Sun. As the Earth completes its elliptical orbit around the Sun in about 365.25 days, the Sun’s apparent motion in the sky, known as solar paths, undergoes perceptible changes [23].

The Earth’s axis is inclined at approximately 23.5 degrees relative to its orbital plane around the Sun, causing the cyclical transition of seasons throughout the year. The axial tilt and orbital motion result in varying amounts of sunlight reaching different regions of the Earth at different times. During the summer solstice that occurs around June 21st in the Northern Hemisphere, the North Pole tilts towards the Sun, leading to prolonged daylight hours, elevated temperatures, and the Sun’s highest path in the sky. Conversely, during the winter solstice around December 21st in the Northern Hemisphere, the North Pole tilts away from the Sun, resulting in shorter daylight hours, cooler temperatures, and a lower solar path. In the Southern Hemisphere, the seasons are reversed, with the summer solstice occurring around December 21st and the winter solstice around June 21st [23].

![Diagram of solar paths and seasons](image)

**Figure 2.6**: Solar paths for June solstice, December solstice, and the equinox in the north hemisphere. Illustration from [23].
Equinoxes, such as the vernal equinox around March 21st and the autumnal equinox around September 23rd, mark times when the Earth's axis is neither tilted towards nor away from the Sun. During these equinoxes, day and night exhibit approximate equality in length globally. Consequently, during these equinoxes, the Sun's path observed from Earth remains consistent, making it identical for both spring and fall [23]. Figure 2.6 shows the solar paths seen from the north hemisphere of Earth during four different months.
This chapter presents the method that has been used to obtain answers to the research questions in this thesis. It begins with a description of a slightly modified version of the Sat-NeRF network. Subsequently, there is a section detailing the implementations for the Planet-NeRF network. The following section provides a detailed description of the metrics employed for evaluating the models, and finally a comprehensive exposition of the conducted experiments is presented.

3.1 Sat-NeRF

The source code from Sat-NeRF [12] is available online and was used in this thesis. This section presents the initial tests and results obtained from the cloned source code of Sat-NeRF.

3.1.1 Feature vector mapping

Sat-NeRF uses an N-dimensional embedding feature vector, \( t_j \) for each image \( j \), which contains information of transient elements. This vector is only used in the loss function, allowing a lower punishment in areas containing transient objects.

In the validation step, the authors of Sat-NeRF hardcoded the mapping of these vectors to align with chosen similar images in the training dataset. As an initial phase of this research, the validation images were mapped to vectors extracted from images captured on nearby dates, with consideration given to the corresponding months.
### 3.1.2 Result variances in Sat-NeRF source code

In this study, the authors of this thesis utilized a publicly available source code of Sat-NeRF, referencing the work of Marí et al. [12, 13]. The tests were replicated on the same data set used in Sat-NeRF. The code was executed on the authors’ computing environment to replicate and extend the findings reported by the original authors. The evaluation metrics are described in Section 3.5.1. Notable variations in results emerged during the implementation compared to the outcomes documented in the referenced work [12]. The observed differences in results between the authors’ tests of the cloned source code and the findings reported by the original authors of the referenced work will not be further addressed or modified in the scope of this study. Subsequent discussions will detail the observed differences and outline the decision not to address or modify these disparities. The focus will be on comparing future results solely to the outcomes of Sat-NeRF obtained through the authors’ computing environment.

### 3.2 Solar position experiments

In order to answer research question 2 about the significance of the solar direction inputs, multiple tests were carried out on the trained models from the original Sat-NeRF network. In these tests, the trained models were fed with unseen sun directions as inputs. The positioning of the sun when observed from a specific location and time is characterized by two coordinates: azimuth and elevation. All images in the training datasets were captured between 17:00 and 18:00 UTC (Coordinated Universal Time). During standard time (UTC-6), this translates to 11:00-12:00 local time, and during daylight saving time (UTC-5), it corresponds to 12:00-13:00 local time. Consequently, 17:00 UTC is designated as the time for determining unseen solar positions. Emphasizing the novelty of these positions, the year 2023 is utilized, given that all training images precede this date. The Ephem library in Python was employed to generate elevation and azimuth for specific dates and locations at 17:00 UTC.

---

**Table 3.1:** Results in PSNR, SSIM, and Alt. MAE obtained by cloning the Sat-NeRF repository [13] and running Sat-NeRF on their preprocessed data for JAX_068, as well as Sat-NeRF’s results from their paper.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR</th>
<th>SSIM</th>
<th>Alt. MAE [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sat-NeRF</td>
<td>24.80</td>
<td>0.903</td>
<td>1.275</td>
</tr>
<tr>
<td>Paper</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloned</td>
<td>24.64</td>
<td>0.91</td>
<td>1.49</td>
</tr>
</tbody>
</table>
3.3 Sat-NeRF modifications

This section introduces modifications aimed at enhancing season prediction. The modifications have been made in the Sat-NeRF network architecture to be able to evaluate how altering the network can improve seasonal handling. The modifications presented in this section served as the foundational experiments for the assessment and conclusion of research question 3. The core element of the upcoming Planet-NeRF network is the Sat-NeRF network architecture, enhanced by the incorporation of a latent month-embedding vector. This architecture will undergo evaluation both independently and in combination with other implementations, including the integration of geometrically rendered shadows (Sat-NeRF+GEO+SL) and signed distance functions (Planet-NeRF+SDF).

3.3.1 Planet-NeRF: Introducing the month embedding vector

To teach the model about key characteristics for different months and, therefore, different seasons, a month-embedding vector $m_j$ was incorporated in the Sat-NeRF network. This network will be referred to as Planet-NeRF in the remainder of the report. The concept behind this approach is to train an embedding layer for each month, where each month's embedding vector has been designed to capture the primary features from images captured during that specific month. The embedding vector is $12 \times K$-dimensional, so that each month contain a $K$-dimensional vector that is trained to represent the appearances of this specific month. $K$ is set to 4, a choice based on evaluations conducted during testing. During the training process, the vector is embedded into the final layer of the MLP but is not integrated until the third epoch. This decision to wait for three epochs was based on extensive testing, which indicated that this approach yielded superior results in terms of PSNR, SSIM, and altitude MAE. These metrics are described in Section 3.5.1. The complete network architecture, together with the embedding vector $m_j$, are presented in Figure 3.1. In the validation step, the network chooses to map the features of this embedding vector corresponding to the month in which the validation image was captured.
3.3.2 Solar rays for season predictions

In the Sat-NeRF framework, the embedding vector $t_j$ is employed in conjunction with shared features to predict transient objects. These shared features are harnessed for the prediction of albedo color. Additionally, the solar direction, in conjunction with these shared features, is utilized for shadow prediction. Notably, the authors of the Sat-NeRF paper exclusively tested their model in areas devoid of seasonal variability. This design choice suggests that the solar direction is configured to capture variations that cannot be explained by a static state but can be explained by the dynamic orientation of the sun. Subsequently, $\beta$ is introduced after two epochs to capture changes that are not accounted for by a static state or the solar direction but can be ascribed to varying image inputs, namely, transient objects.

During the preliminary testing of Sat-NeRF, it was observed that the shading scalar, denoted as $s$, not only detected shadows but also responded to regions with substantial variability between seasons. This observation is logical since both shadow patterns and seasonal variations are influenced by the solar direction. This is due to the sun following distinct paths during different seasons, resulting in varying solar angles.

In light of these empirical findings, the concept of using the solar direction input only for the prediction of seasonal changes emerged. In EO-NeRF [14], shadow computations are based on geometric principles and rely solely on the solar direction as an external input, which is not integrated into the neural network. Inspired by the work of Marí et al. [14], together with the conclusions of the importance of solar rays for seasonal predictions, the authors of this thesis conducted experiments using a modified network architecture. This adaptation focused on the geometric rendering of shadows, thereby enabling the solar layer to exclusively emphasize seasonal appearances. This structural adaptation resulted in the network configuration depicted in Figure 3.2, and this network will be referred to as Sat-NeRF+GEO+SL.

Figure 3.1: The Planet-NeRF network is formed by integrating the Sat-NeRF network architecture with an additional layer representing the latent embedding vector $m_j$. 
3.3 Sat-NeRF modifications

In accordance with the findings presented in Section 2.3.1, it is evident that the NeuS method achieves a notably enhanced level of accuracy in surface reconstruction compared to its predecessor, NeRF. This notable improvement motivated the exploration of a novel approach, aimed at enhancing the NeRF architecture. Specifically, the volume rendering component of the NeRF architecture is replaced with segments from the NeuS network. As a result, the architecture of NeRF underwent significant alterations, including the replacement of NeRF’s fully connected neural network structure, consisting of eight layers as depicted in Figure 2.2, with a network architecture named MLP\textsubscript{SDF} (as illustrated in Figure 3.3). Unlike NeRF, this network focuses on predicting a surface, which is designed to represent the zero-level set of a neural implicit Signed Distance Function (SDF), denoted as $f(x)$, where $x \in \mathbb{R}^3$, as opposed to predicting $\sigma(t)$. The network that resulted from these changes will be referred to as Planet-NeRF+SDF in this thesis.

During the training of the Planet-NeRF+SDF network, a probability density function, denoted as $\phi_s(f(x))$ and referred to as the S-density, is introduced. The S-density is derived from the derivative of the Sigmoid function $\Phi_s(x)$ as indicated in Equations 3.1. The introduction of the S-density significantly influences the learning process of the network.

In the course of training, an opaque density, designated as $\rho$, is calculated. This density is derived from the S-density, as elaborated in Equation 3.2.
Figure 3.3: The Planet-NeRF+SDF network architecture that predict a signed distance function \( f(x) \) instead of \( \sigma \).

The incorporation of the S-density field, volume rendering is employed to train the SDF network exclusively with 2D input images as supervision. This approach, supported by the S-density field, leads to successful minimization of a loss function based on this supervision. This training approach empowers the neural implicit SDF to effectively represent the zero-level surface, that is, to achieve \( f(x(t^*)) = 0 \), where \( x(t^*) \) symbolizes a point on the surface. Additionally, the S-density is expected to exhibit notably high values in the vicinity of the surface.

\[
\phi_s(x) = \frac{se^{-sx}}{(1 + e^{-sx})^2},
\]

\[
\Phi_s(x) = (1 + e^{-sx})^{-1},
\]

\[
\phi_s(x) = \Phi'_s(x).
\]

In terms of the weight function for the SDF network, the conventional volume rendering formulation, which employs the volume density \( \sigma(t) \), is adapted. Instead, the weight function is expressed as \( w(t) = T(t)\rho(t) \), where the S-density is represented as \( \phi(f(r(t))) \) and \( \rho(t) = \phi(f(r(t))) \). The transmittance function, \( T(t) \), is concurrently modified to \( T(t) = \exp\left(-\int_0^t \rho(u)du\right) \). The determination of \( \rho(t) \), an alternative to the volume density \( \sigma \) typically used in volume rendering, is derived from the following equation:

\[
\rho_i = \max\left( \frac{\partial \Phi_s(d(X_i))}{\partial t_i}, 0 \right).
\]

The discrete representation of \( \rho(t) \), denoted as \( \alpha_i \), is obtained through the equation:
\[ \alpha_i = \max \left( \frac{\Phi_s(f(r(t_i))) - \Phi_s(f(r(t_{i+1})))}{\Phi_s(f(r(t_i)))}, 0 \right). \] (3.3)

The calculation of the transmittance is accomplished using the equation:

\[ T_i = \prod_{j=1}^{i-1} (1 - \alpha_j). \] (3.4)

The final step in the volume rendering process involves the extraction of the depth for all surface points. This is achieved by multiplying each ray by its corresponding weight, as illustrated in the function below:

\[ d(r) = \sum_{i=1}^{N} T_i \alpha_i t_i. \] (3.5)

### 3.4 Training

In this chapter, the training process focusing on loss components is described. The RGB loss, representing color misprediction, is defined both with and without an uncertainty coefficient (\( \beta \)). Additionally, an Eikonal loss term is introduced when using Signed Distance Functions (SDF) instead of sigma in the network architecture of Planet-NeRF+SDF.

**RGB loss**

The predicted color is represented as a 3-dimensional vector corresponding to the red, green, and blue color channels. The RGB color values are constrained to the [0, 1] interval, where 0 corresponds to no intensity in the channel, and 1 represents full intensity. This normalization allows the model to output colors that can be easily interpreted and visualized in the standard RGB color space.

The loss associated with color misprediction, denoted as \( \mathcal{L}_{\text{RGB}} \), can be computed both with and without the inclusion of the uncertainty coefficient \( \beta \). When calculating the color loss without considering \( \beta \), no additional penalty is applied to transient objects. In this case, the loss is determined as follows:

\[ \mathcal{L}_{\text{RGB}} = \sum_{r \in \mathcal{R}} \| c(r) - c_{\text{GT}}(r) \|^2_2. \] (3.6)

Here, the loss is calculated as the sum of the squared second norm of the difference between the predicted color \( c(r) \) and the ground truth color \( c_{\text{GT}}(r) \). Alternatively, when incorporating the uncertainty coefficient \( \beta \), defined as:

\[ \beta(r) = \sum_{i=1}^{N} T_i \alpha_i \beta(x_i, t_i), \] (3.7)
where \( \beta(x_i, t_i) \) corresponds to the predicted uncertainty of the \( i \)-th point along the ray \( r \), the RGB loss equation is modified as follows:

\[
\mathcal{L}_{\text{RGB}} = \sum_{r \in \mathcal{R}} \frac{\|c(r) - c_{\text{gt}}(r)\|^2}{2\beta'(r)^2} + \left( \frac{\log \beta'(r) + \eta}{2} \right).
\] (3.8)

Here, \( \beta'(r) = \beta(r) + \beta_{\text{min}} \), where \( \beta_{\text{min}} = 0.05 \) and \( \eta = 3 \) to avoid negative values in the logarithm. The logarithm is included to prevent \( \beta \) from approaching infinity. In this research, the uncertainty coefficient \( \beta \) is introduced after the initial two epochs. Consequently, Equation 3.6 characterizes the ultimate loss during the first two epochs, while Equation 3.8 encapsulates the final loss for the subsequent epochs. This is the loss used in all networks except Planet-NeRF+SDF.

### Eikonal loss

In the case of using sdf instead of sigma as describes in Section 3.3.3, an eikonal loss term is introduced

\[
\mathcal{L}_{\text{REG}} = \frac{1}{nm} \sum_{k,i} (\| \Delta f(\hat{r}_{k,i}) \|_2 - 1)^2,
\] (3.9)

where \( n \) is the sampling size and \( m \) is the batch size. The Eikonal term in the loss function encourages the gradients of the SDF along each ray, \( f(\hat{r}_{k,i}) \), to have a unit 2-norm. This to ensure an accurate representation of the distances to surfaces within the 3D space. This concept is motivated by the Eikonal partial differential equation, which specifies that the magnitude of the gradient of a function should equal 1 to accurately represent distances [6]. In Planet-NeRF+SDF the final loss becomes:

\[
\mathcal{L}_{\text{SDF}} = \mathcal{L}_{\text{RGB}} + \beta_{\text{REG}} \mathcal{L}_{\text{REG}},
\] (3.10)

where \( \beta_{\text{REG}} \) is a weight set to 0.1 and \( \mathcal{L}_{\text{RGB}} \) is given by Equation 3.6 for the first two epochs, and by Equation 3.8 for the later epochs.

## 3.5 Evaluation

In this section the metrics and data used for evaluation is presented.

### 3.5.1 Metrics

The metrics presented in this section have been used to evaluate the models. **Mean absolute error** (MAE), **peak signal to noise ratio** (PSNR), and **structural similarity index measure** (SSIM) are commonly used metrics in previous works of NeRF [4, 5, 12, 14, 20]. To gain a more comprehensive understanding of the deviation, both the **linear error 90%** and **linear error 50%** scores were also calculated for the mean absolute error. Moreover, the **mean error** was computed and utilized as an additional metric to gain an understanding whether the model exhibits a bias towards predicting altitude as either too deep or too shallow. It is important to note
that all of the metrics presented later in this report are mean values, representing an average over all images in the test dataset, thereby providing a holistic view of the model performance. Furthermore, visual examination was necessary when evaluating novel views without ground truth.

**Errors in altitude estimations**

NeRF is trained to predict the altitude of each point in the scene based on the information available in the 2D images. The mean absolute error (MAE) and mean error (ME) is used in NeRF to measure the accuracy of altitude estimation. The ground truth is depth maps provided by Maxar that are converted to altitude maps. MAE quantifies the absolute difference between the predicted altitude by NeRF and the ground truth altitude at each point $x$, in the scene. The MAE is calculated as follows:

$$\text{MAE}(x, x_{gt}) = \frac{\sum_{i=1}^{N} |x_{gt,i} - x_i|}{N}.$$  \hspace{1cm} (3.11)

The ME quantifies the discrepancy at each point $x$, in the scene between the altitude predicted by NeRF and the ground truth altitude. This measurement includes both positive and negative values, reflecting the extent to which the predicted altitude deviates above or below the true altitude. ME is calculated as:

$$\text{ME}(x, x_{gt}) = \frac{\sum_{i=1}^{N} x_{gt,i} - x_i}{N}.$$  \hspace{1cm} (3.12)

As mentioned earlier, the linear error at 90% (LE90) and 50% (LE50) are calculated for the MAE. These metrics are determined by first sorting the MAE values for each pixel $x$, according to their magnitude. Then, identifying the error value below which 90% and 50% of the data points, respectively, fall. Note that the LE50 metric essentially represents the median value since it corresponds to the point at which half of the data points are below this value and half are above. Therefore mirroring the concept of the median in statistical analysis.

**PSNR - Peak Signal to Noise Ratio**

In NeRF, peak signal-to-noise ratio (PSNR) is a key metric for assessing the fidelity of rendered images. It measures the similarity between generated and real images, with higher PSNR values indicating a closer match. The PSNR equation quantifies the ratio of the peak possible signal power to the power of the noise that affects the image. Higher PSNR values indicate a closer match between the rendered and ground truth images. The PSNR is calculated using the following equation:

$$\text{PSNR}(x, x_{gt}) = 20 \cdot \log_{10} \left( \frac{\text{MAX}(x_{gt})}{\sqrt{\text{MSE}}} \right),$$  \hspace{1cm} (3.13)
where \( \text{MSE}(x, x_{\text{gt}}) = \frac{1}{nm} \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \|x_{\text{gt}}(i, j) - x(i, j)\|^2. \) (3.14)

\( \text{MAX}(x_{\text{gt}}) \) represents the maximum possible pixel value when summing over all pixels in the image, \( x_{\text{gt}} \) is the ground truth image, and \( x \) is the predicted image [8].

**SSIM - Structural Similarity Index Measure**

The structural similarity index measure (SSIM) was employed to quantify the resemblance between predicted images and ground truth images. SSIM measures the perceptual dissimilarity between two images: the predicted image from the model output, \( x \), and the ground truth image, \( x_{\text{gt}} \). This assessment is determined by the following equation:

\[
\text{SSIM}(x, x_{\text{gt}}) = \frac{(2\mu_x \mu_{x_{\text{gt}}} + c_1)(2\sigma_{x,x_{\text{gt}}} + c_2)}{\left(\mu_x^2 + \mu_{x_{\text{gt}}}^2 + c_1\right)\left(\sigma_x^2 + \sigma_{x_{\text{gt}}}^2 + c_2\right)}.
\] (3.15)

In this context, \( \mu_x \) and \( \mu_{x_{\text{gt}}} \) represent the pixel sample mean of \( x \) and \( x_{\text{gt}} \), while \( \sigma_x^2 \) and \( \sigma_{x_{\text{gt}}}^2 \) correspond to the variances of \( x \) and \( x_{\text{gt}} \). Additionally, \( \sigma_{x,x_{\text{gt}}} \) denotes the covariance between \( x \) and \( x_{\text{gt}} \). To ensure a stable division, two constants, \( c_1 \) and \( c_2 \), are introduced, with \( L = 2^{\#\text{bits/pixel}} - 1 \) signifying the dynamic range of pixel values. The default values for \( k_1 \) and \( k_2 \) are 0.01 and 0.03, respectively. The SSIM score falls within the range \([-1, 1]\), with a score of 1 indicating perfect structural similarity [2].

Moreover, SSIM serves as a model for assessing image quality by considering how changes are perceived in structural information. It takes into account perceptual effects such as luminance and contrast masking. Unlike the metrics MAE and PSNR, which estimate absolute errors, SSIM factors in the inter-dependencies of neighboring pixels that convey crucial information about the object structure in the image. Luminance masking means that distortions are less noticeable in brighter areas, while contrast masking implies that distortions are less visible in textured regions [2].

### 3.5.2 Data

The evaluation in this study utilizes data sets from the 2019 IEEE GRSS Data Fusion Contest, encompassing five distinct areas located in Omaha, Nebraska, USA, and Jacksonville, Florida, USA (refer to Table 3.2). All images employed are sourced from Maxar WorldView-03, possessing a size of 2048×2048 pixels and covering an approximate area of 65000 m².

The Omaha data sets are specifically employed to address research questions 1 through 3, given their inclusion of seasonal variations. Conversely, the Jacksonville data serves as a benchmark for 1, facilitating a comparison of Sat-NeRF performance across areas with and without seasonal changes.
For the Jacksonville data sets, the training and test image split aligns with the methodology adopted by Sat-NeRF. In the case of Omaha, four images are extracted from each dataset for testing. These images are selected to ensure representation of each distinct season (see Table 3.3). The choice of the fourth image aims to maintain a sufficient number of test images, thereby enabling a fair and robust evaluation.

Additionally, to uphold fairness in evaluation, images from months with only one representation are excluded from the test set. This decision is grounded in the absence of corresponding training data for such months. The ground truth for these datasets comprises both RGB images, akin to those used for training, and lidar data. The lidar data exhibits a resolution of 0.3 m/pixel at nadir.

Furthermore, to optimize training time, all images are down-sampled by a factor of four, resulting in a size of 512×512 pixels for training purposes. This down-sampling process contributes to a more efficient training pipeline while preserving essential features for accurate model learning.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>#Img</th>
<th>Train/Test</th>
<th>Height range [m]</th>
<th>AOI size [km²]</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAX_004</td>
<td>11</td>
<td>9/2</td>
<td>25</td>
<td>0.66</td>
</tr>
<tr>
<td>JAX_068</td>
<td>19</td>
<td>17/2</td>
<td>57</td>
<td>0.66</td>
</tr>
<tr>
<td>OMA_132</td>
<td>41</td>
<td>37/4</td>
<td>31.9</td>
<td>0.62</td>
</tr>
<tr>
<td>OMA_212</td>
<td>38</td>
<td>34/4</td>
<td>34.6</td>
<td>0.61</td>
</tr>
<tr>
<td>OMA_374</td>
<td>30</td>
<td>26/4</td>
<td>59.4</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Table 3.2: Description of the data sets utilized in the experiments. Each data set is identified by an ID, where IDs beginning with ‘JAX’ indicate that the data set originates from Jacksonville. Similarly, IDs starting with ‘OMA’ denote that the data set was collected from an area in Omaha.
3.6 Experimental setup

The experiments were designed to assess the models in areas exhibiting seasonal variations, with details of all evaluated models provided in Table 3.4. Initially, it was crucial to gauge the performance of Sat-NeRF by applying their network to areas featured in their publications. This step ensured that the results reported in their paper were achievable and that the evaluation code used in this thesis closely aligned with the one employed by the authors of Sat-NeRF. This evaluation utilized Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) metrics. The evaluation focused on models trained on the dataset used in their paper for JAX_068, obtained from the link in their GitHub repository.

Subsequently, Sat-NeRF was tested on image sets with both substantial and minimal seasonal variations. This step aimed to address research question 1, investigating the performance of the network in scenes with seasonal changes. Comparison of the results on datasets with minimal seasonal variations was undertaken to discern potential performance variations. The rationale behind not directly comparing Sat-NeRF’s results on areas with seasonal variations to those in the Sat-NeRF paper lies in the possibility of overoptimistic results and the disparate testing conditions, particularly on cropped areas in Jacksonville with full resolutions.

The Sat-NeRF models, trained on three distinct areas in Omaha, underwent testing using a set of unseen sun directions as input. This testing was designed to address research question 2 and provide insights into the influence of the sun in learning seasons. Given that all the generated images are synthetic, a visual examination was conducted for evaluation purposes.

<table>
<thead>
<tr>
<th>#img</th>
<th>Jan</th>
<th>Feb</th>
<th>Mars</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAX_004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Test</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>JAX_068</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Test</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OMA_132</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Test</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OMA_212</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Test</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OMA_374</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Train</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Test</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Table 3.3: Number of images from different months in train and test data sets.*
Following this, all models featuring the modified Sat-NeRF architecture underwent execution on all Omaha datasets. Firstly, a model inspired by EO-NeRF, emphasizing the geometric rendering of shadows instead of prediction, was implemented to address research question 2. Subsequent modifications included incorporating a month embedding vector and a combination of a month embedding vector with the signed distance function (sdf) to address the queries outlined in research question 3.

Consistent parameters across all models included network parameters, activation functions, number of layer samples, and batch size (with the exception of ME+SDF with batch size 512), mirroring those employed by Sat-NeRF. The training time was determined through trial experiments that monitored convergence. Figure 3.4 illustrates the progression of MAE, PSNR, and SSIM. The training time for all models was standardized to 20 epochs after a thorough evaluation.

**Figure 3.4:** The graph visually depicts the convergence patterns of key metrics, including PSNR, SSIM, and MAE, across different epochs. Convergence and stability is achieved after approximately 20 epochs.
Table 3.4: Description of the models that were trained for evaluation in this thesis. Experiment number 1 was utilized in order to respond to research questions 1 and 2, and experiment number 2 refers to research question 3.
This chapter details the outcomes of the experiments. The initial section illustrates the results achieved with Sat-NeRF in predicting scenes subject to seasonal changes. Following that, the subsequent section examines the impact of the sun in predicting seasonal appearances. Lastly, the third section discloses the results from the different Sat-NeRF modifications.

4.1 Sat-NeRF for predicting scenes undergoing seasonal changes

This section will demonstrate the performance of Sat-NeRF in predicting scenes undergoing seasonal changes. The evaluation will involve comparing the results of Sat-NeRF on different areas in Omaha, which experience seasonal changes, and contrasting them with Jacksonville, a location characterized by minimal seasonal variations.
4.1.1 Quantitative results

<table>
<thead>
<tr>
<th>AOI</th>
<th>PSNR ↑</th>
<th>SSIM ↑</th>
<th>Alt. MAE [m] ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAX</td>
<td>004</td>
<td>068</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>23.25</td>
<td>22.02</td>
<td>22.64</td>
</tr>
</tbody>
</table>

*Table 4.1: Values for PSNR, SSIM, and altitude MAE on two Jacksonville areas using Sat-NeRF.*

<table>
<thead>
<tr>
<th>AOI</th>
<th>PSNR ↑</th>
<th>SSIM ↑</th>
<th>Alt. MAE [m] ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMA</td>
<td>212</td>
<td>132</td>
<td>374</td>
</tr>
<tr>
<td></td>
<td>21.87</td>
<td>19.87</td>
<td>18.04</td>
</tr>
</tbody>
</table>

*Table 4.2: Values for PSNR, SSIM, and altitude MAE on three Omaha areas using Sat-NeRF.*

As observed in Table 3.1, the results reported by Sat-NeRF in their paper surpassed those obtained by cloning the Sat-NeRF repository and running it on the data sets they utilized [27], particularly in terms of altitude prediction performance.

When running Sat-NeRF on all areas in the data set, it is evident that the model performs much better on Jacksonville areas than on Omaha areas in terms of PSNR and SSIM scores, as seen in Table 4.1 and 4.2. Notably, Sat-NeRF excels at altitude prediction for Omaha areas. However, it is important to note that the Omaha data sets have roughly twice the number of training images compared to the Jacksonville data sets (see Table 3.2), which potentially could influence the altitude predictions.

4.1.2 Altitude predictions

Figures 4.2 and 4.3 illustrate altitude predictions using Sat-NeRF for Omaha and Jacksonville areas, respectively. It is evident that Sat-NeRF encounters challenges in predicting altitude for OMA_374 and JAX_068 compared to other areas, as also depicted in Table 4.1 and 4.2. As mentioned in Section 3.5.2, lidar data provides the ground truth altitude measurements. However, this data only covers a subset of the entire scene. Consequently, the assessment of altitude accuracy is only evaluated on this subset. The ground truth lidar altitudes for all datasets can be seen in Figure 4.1.
4.1 Sat-NeRF for predicting scenes undergoing seasonal changes

Generally, Sat-NeRF appears to face the greatest difficulty in estimating altitude for areas with trees and water. In JAX_068, where these features are nearly absent, the altitude predictions appear clearer. However, as indicated in Table 4.1 and 4.2, the numerical results are worse for JAX_068 compared to both OMA_212 and OMA_132. It is crucial to reiterate that the Jacksonville models are trained on a considerably smaller image set than the Omaha models. Additionally, the altitude range for JAX_068 and OMA_374 is almost double that of the other regions, as described in Table 3.2. Considering these factors provides context for understanding the observed performance disparities.

**Figure 4.1:** Ground truth altitudes from lidar data for all datasets.

**Figure 4.2:** Altitude predictions for Sat-NeRF on areas with seasonal variations.
4.1.3 RGB predictions

Figures 4.4, 4.5, and 4.6 present the rendered RGB and albedo produced by Sat-NeRF in three distinct Omaha regions. These visuals capture the seasonal variations winter, spring, and summer, providing a comprehensive representation of the network’s performance across varying seasonal appearances. Notably, the predicted color for OMA_132 in January closely resembles the corresponding albedo color, a similarity that is also observed in smaller portions of the other two January images.

However, this phenomenon is not consistently present in the predictions for March and September, even though the albedo predictions exhibit similarities across all months. Overall, it appears that Sat-NeRF well captures seasonal variations, particularly evident in the March predictions. Conversely, for the September images, the predicted image for OMA_374 bears a striking resemblance to the ground truth March images. Additionally, the predicted image for OMA_132 exhibits hints of the brownish ground color present in the ground truth March images. Notably, the predicted images for March display indications of a green color, absent in the ground truth images but present in the ground truth September images. This confusion for March and September can also be seen in Figure 4.11.

When examining the albedo predictions for various scenes, a notable observation emerges. The albedo predictions for all Omaha regions deviate significantly from the true colors of the scenes, predominantly adopting a white/pink hue. In contrast, the albedo predictions for the two Jacksonville areas exhibit a closer resemblance to the ground truth colors, indicating a more accurate representation of the actual scene colors.

Despite the aforementioned points, it is clear that Sat-NeRF successfully captures a lot of the key characteristics from the different season in most of the cases.
Figure 4.4: Ground truth as well as RGB and albedo predictions by Sat-NeRF for the Omaha areas with seasonal variations.
Figure 4.5: Ground truth as well as RGB and albedo predictions by Sat-NeRF for the Omaha areas with seasonal variations.
Figure 4.6: Ground truth as well as RGB, and albedo predictions by Sat-NeRF for the Omaha areas with seasonal variations.


Figure 4.7: Ground truth as well as RGB, and albedo predictions by SatNeRF for the Omaha areas with seasonal variations.

4.1.4 Solar impact of Sat-Nerf predictions

In Figure 4.8, it is evident that the shading scalar intended to have high values in shadowed areas for applying the ambient sky color also exhibits high values in areas with significant seasonal variations (such as grass and trees). Additionally, the predicted sky color does not seem to resemble typical sky colors; rather, it appears to align more closely with the dominant color in the image.

Contrasting, when examining Figure 4.9, this phenomenon is less pronounced. In the images for JAX_004, there are trends of the shading scalar having larger values for grass and trees than expected since they in many cases seem to be in direct sunlight. However, even though there are no mayor seasonal variation in Jacksonville some difference in the color of grass and trees can still be observed. For JAX_068, where there are almost no trees and grass present, the shading scalar almost predominantly attains high values in shadowed areas.

When inspecting the predicted shading scalars and ambient sky colors for the Jacksonville regions, it is noticeable that the ambient sky color is consistently blue, despite the dominant colors in the images being green or grey. Which indicated that the ambient color seems to accurately predict the sky color, in contrasts with the situation in Omaha.
4.1 Sat-NeRF for predicting scenes undergoing seasonal changes

<table>
<thead>
<tr>
<th>132</th>
<th>212</th>
<th>374</th>
<th>132</th>
<th>212</th>
<th>374</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Ground truth" /></td>
<td><img src="image2" alt="Shading scalar" /></td>
<td><img src="image3" alt="Sky ambient" /></td>
<td><img src="image4" alt="Ground truth" /></td>
<td><img src="image5" alt="Shading scalar" /></td>
<td><img src="image6" alt="Sky ambient" /></td>
</tr>
</tbody>
</table>

**Figure 4.8:** Shading scalar and ambient sky color predictions when running Sat-NeRF on the different Omaha areas during two distinct seasons.

<table>
<thead>
<tr>
<th>004</th>
<th>004</th>
<th>068</th>
<th>068</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image7" alt="Ground truth" /></td>
<td><img src="image8" alt="Shading scalar" /></td>
<td><img src="image9" alt="Sky ambient" /></td>
<td><img src="image10" alt="Ground truth" /></td>
</tr>
</tbody>
</table>

**Figure 4.9:** Shading scalar and ambient sky color predictions when running Sat-NERF on the different Jacksonville areas.
4.2 Solar impact on seasonal appearance

The Sat-NeRF models trained on the three different sites in Omaha, presented in the previous section 4.1, were used to evaluate the impact of the sun direction input in the prediction of seasons. The evaluation consisted of generating novel sun directions for specified times of the year.

4.2.1 Novel sun directions representing each month

<table>
<thead>
<tr>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>Jul</td>
<td>Aug</td>
<td>Sep</td>
<td>Oct</td>
<td>Nov</td>
<td>Dec</td>
</tr>
<tr>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
</tbody>
</table>

Figure 4.10: One rendered image from each month of the year using Sat-NeRF. The different months are generated using solar-directions from the 15th of each month. Note that the months with red labels are missing in the training data and the model has, therefore, not been trained on these solar directions.

In Figure 4.10 the image predictions resulting from the introduction of novel sun directions generated for each month are depicted. These sun directions are specifically generated to emulate the authentic solar position for the 15th day of each month, precisely at 17:00 in the year 2023, as described in section 3.2. The visual representation underscores the substantial impact of the sun on predicting seasons, notwithstanding the fact that the model’s performance may not be impeccable.

The results reveal season-specific details, such as the presence of snow in October to January, and accurate color predictions from September to April. Noteworthy details in texture representation are observed in forested areas, demonstrating sparser growth in November to February, and fuller growth in April to September. However, it is important to note that the model encounters challenges in accurately capturing lighting variations, a limitation particularly evident in the representation of May, June, July, and August.

The influence of the sun on the predicted outcomes is particularly noteworthy, indicating its pivotal role in shaping the model’s ability to capture seasonal variations. It is crucial to note that the model was not trained on data from
4.2 Solar impact on seasonal appearance

June, July, and November. This absence of training data for these months adds a layer of complexity to comprehending solar directions during this period, potentially contributing to challenges in accurately predicting outcomes for June, July, and November. Despite this limitation, the observed influence of the sun on the model’s predictions highlights the importance of considering solar dynamics in understanding the model’s behavior across different months.

The model’s comprehension of the unique sun direction, symbolizing the solar position in November, can be elucidated by the resemblances between this solar position and those of its adjacent months. Specifically, the elevation and azimuth of the solar path in November are anticipated to fall within the range defined by the values observed in October and December. In contrast, June introduces an intricate scenario due to the occurrence of the summer solstice, resulting in a more distinct solar path that poses challenges for accurate prediction.

4.2.2 Novel equinox sun direction

<table>
<thead>
<tr>
<th>Equinox months</th>
<th>2015-03-04</th>
<th>2015-03-23</th>
<th>2014-09-08</th>
<th>2014-09-20</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OMA 374</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-03-04</td>
<td>2015-03-30</td>
<td>2015-09-16</td>
<td>2015-09-28</td>
<td></td>
</tr>
<tr>
<td><strong>OMA 132</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-03-04</td>
<td>2015-03-30</td>
<td>2015-09-16</td>
<td>2015-09-28</td>
<td></td>
</tr>
<tr>
<td><strong>OMA 212</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Figure 4.11: The ground truth images of the equinox months used in the training of the models.](image)

Figure 4.11: The ground truth images of the equinox months used in the training of the models.
The evaluation consisted of generating novel sun directions from the equinox months, i.e. Mars and September, as input to the models. The three Omaha data sets contain different amounts of training images from these months (see Table 3.3). Figure 4.11 shows the first and last image captured in Mars and September that were used in training for each dataset. The images captured in March distinctly exhibit a brown hue, whereas those taken in September distinctly portray a green color.

<table>
<thead>
<tr>
<th></th>
<th>March</th>
<th>September</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td><img src="image1" alt="March Image" /></td>
<td><img src="image2" alt="September Image" /></td>
</tr>
<tr>
<td>15th</td>
<td><img src="image3" alt="March Image" /></td>
<td><img src="image4" alt="September Image" /></td>
</tr>
<tr>
<td>30th</td>
<td><img src="image5" alt="March Image" /></td>
<td><img src="image6" alt="September Image" /></td>
</tr>
</tbody>
</table>

Figure 4.12: Output images generated from the Sat-NeRF models of the three Omaha areas based on input solar directions, ensuring an accurate representation of the geographic area’s solar position. The solar directions are generated to reflect the 1st, 15th and 30th of March (on the left) and of September (on the right), specifically at 17:00 year 2023.

In order to conduct a more in-depth analysis of the models’ ability to discern distinctive features during equinox months, sun directions representing the 1st, the 15th, and the 30th of each month were generated. The outcomes, depicted in Figure 4.12, reveal a noteworthy trend where the models exhibit confusion between March and September. September is portrayed with an overly brown tone for all datasets, which is not an accurate representation when looking at the ground truth images in Figure 4.11. March, on the other hand, is more faithfully represented in its true colors, but shows an excessively green hue in the last generated image from OMA_374, which is not representative this time of the year. Upon examination of the generated images from the OMA_374 area (Figure 4.12), it becomes apparent that these images are nearly identical. This confusion is evident to the extent that, if the image sets were interchanged, distinguishing whether the images were generated in March or September would become a challenging task/not be possible.
4.3 Sat-NeRF modifications

These findings underscore a significant limitation in the model’s capacity to provide an accurate representation throughout the entire year. While the network demonstrates an aptitude for learning intricate details from the solar path, it appears insufficient to consistently capture the nuanced variations exhibited during equinox months. The observed confusion between March and September highlights the complexity of modeling seasonal transitions accurately, emphasizing the need for further refinement and considerations to enhance the model’s overall performance.

4.3 Sat-NeRF modifications

In this section, the results from the Sat-NeRF modifications; Planet-NeRF, Planet-NeRF+SDF, and Sat-NeRF+GEO+SL will be presented and compared with the results obtained by Sat-NeRF [12] and Season-NeRF [5].

4.3.1 Quantitative results

Table 4 displays the mean PSNR, SSIM, and altitude MAE scores for all models across all data sets. It is evident that Planet-NeRF excels in both PSNR and SSIM, securing the top two scores in all cases except one. Additionally, Planet-NeRF achieves the best performance in altitude MAE, albeit with a narrower margin compared to the other metrics. In most instances, Planet-NeRF outperforms Planet-NeRF+SDF, although not consistently across all cases. Notably, the primary objective of SDF was to surpass Sat-NeRF’s altitude prediction by predicting a signed distance function instead of volume density. Despite this goal, SDF outperforms volume density prediction in only one out of three cases. Conversely, Season-NeRF consistently performs worse across all above mentioned metrics compared to other models.

<table>
<thead>
<tr>
<th>Model</th>
<th>AOI: OMA</th>
<th>PSNR ↑</th>
<th>SSIM ↑</th>
<th>Alt. MAE [m] ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>212</td>
<td>132</td>
<td>374</td>
<td>Mean</td>
</tr>
<tr>
<td>Sat-NeRF</td>
<td>21.87</td>
<td>19.87</td>
<td>18.04</td>
<td>19.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.74</td>
<td>0.71</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.84</td>
</tr>
<tr>
<td>Season-NeRF</td>
<td>17.79</td>
<td>19.82</td>
<td>20.48</td>
<td>19.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.57</td>
<td>0.62</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.94</td>
</tr>
<tr>
<td>Planet-NeRF</td>
<td>22.40</td>
<td>21.15</td>
<td>20.29</td>
<td>21.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.76</td>
<td>0.78</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.82</td>
</tr>
<tr>
<td>Planet-NeRF+SDF</td>
<td>22.39</td>
<td>21.08</td>
<td>20.36</td>
<td>21.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.77</td>
<td>0.73</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.91</td>
</tr>
<tr>
<td>Sat-NeRF+GEO+SL</td>
<td>20.77</td>
<td>18.97</td>
<td>18.62</td>
<td>19.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.72</td>
<td>0.65</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.42</td>
</tr>
</tbody>
</table>

Table 4.3: Displayed are the values for PSNR, SSIM, and altitude MAE across all Omaha areas and various model types. The value reported for Season-NeRF was sourced from their paper, specifically referencing the results from their full model. In the table bold text highlight the best-performing model for that specific area and metric.

An examination of the linear error 90% (LE90) and 50% (LE50) values in Table 4.4 suggests that none of the models exhibit significant outliers compared to others, as their performance across these metrics closely mirrors the results observed for mean absolute error (MAE). Furthermore, when analyzing the mean
error (ME) in altitude from the same table, it appears that all models demonstrate a balanced distribution of prediction errors in terms of altitude. This is indicated by the ME values being closely centered around zero, suggesting an almost equal incidence of predictions that are either too deep or too shallow.

The downsampling factor and resolution of the images used in training of Planet-NeRF, Planet-NeRF+SDF, and Sat-NeRF+GEO+SL (described in Section 3.5.2) are the same as Season-NeRF uses.

<table>
<thead>
<tr>
<th>Model</th>
<th>LE90 ↓</th>
<th>LE50 ↓</th>
<th>Alt. ME [10^-6] → 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOI: OMA</td>
<td>212 132 374 Mean</td>
<td>212 132 374 Mean</td>
<td>212 132 374</td>
</tr>
<tr>
<td>Sat-NeRF</td>
<td>1.59 2.78 9.28 4.55</td>
<td>0.50 0.65 3.25 1.47</td>
<td>6.26 -4.08 6.01</td>
</tr>
<tr>
<td>Season-NeRF</td>
<td>* * * *</td>
<td>0.45 0.73 3.23 1.47</td>
<td>* * *</td>
</tr>
<tr>
<td>Planet-NeRF</td>
<td>1.51 2.78 9.25 4.51</td>
<td>0.47 0.62 3.33 1.47</td>
<td>-2.93 -3.47 -6.49</td>
</tr>
<tr>
<td>Planet-NeRF+SDF</td>
<td>1.74 2.65 10.08 4.82</td>
<td>0.55 0.64 3.52 1.57</td>
<td>-4.00 1.52 -9.80</td>
</tr>
<tr>
<td>Sat-NeRF+GEO+SL</td>
<td>3.06 3.94 10.38 5.79</td>
<td>0.98 1.62 4.34 2.31</td>
<td>-33.0 -5.38 -7.92</td>
</tr>
</tbody>
</table>

*Table 4.4:* Displayed are the values for LE90, LE50, and Alt. ME, all in meters, across all Omaha areas and various model types. The value reported for Season-NeRF was sourced from their paper, specifically referencing the results from their full model. (*) denotes missing values. In the table bold text highlight the best-performing model for that specific area and metric.

### 4.3.2 Geometric shadows

Figure 4.13 serves as a good illustration of desired outcomes obtained by rendering geometric shadows as opposed to predicting them. In this context, it is evident that the geometrically rendered shading scalar primarily has high values in areas with shadows. Concurrently, the season scalar obtains high values in areas of the image that are subject to significant seasonal variations, such as grass and trees. In addition to that the season scalar also captures shadows, which could be desirable given that the shade color results from a blend of the sky and ground colors.

In the absence of geometric shadows, as in Sat-NeRF for example, both shadows and overall texture seem to be predicted by the shading scalar, as seen in the same figure, and therefore uniformly multiplied by the same ambient sky color. However, as observed in Figure 4.13, this anticipated relationship does not hold. The ambient sky color expected to reflect the sky color for the Sat-NeRF+GEO+SL model, appears to follow the color scale of the season, akin to Sat-NeRF’s behavior.

Quantitative results indicate that geometric shadows have a detrimental impact on the altitude Mean Absolute Error (MAE), implicating a challenge in altitude prediction. This is likely attributed to these authors difficulties in approximating an accurate bounding box of the scene. Despite the poor numeric results archived by this method, the qualitative results, shed light on how and what the network is learning using this method.
4.3 Sat-NeRF modifications

4.3.3 Impact of month embedding vector

When examining the final predicted images generated by Planet-NeRF in Figures 4.14, 4.15, and 4.16, a striking resemblance to the ground truth images is evident across all cases, with the exception that the predicted images displaying slightly intensified colors. A direct visual comparison with the results obtained by Sat-NeRF (see Figures 4.4, 4.5, and 4.6) indicates a superior performance in prediction achieved by Planet-NeRF. This visual assessment is further supported by the quantitative results for SSIM and PSNR found in Table 4.

Analyzing the seasonal albedo predictions within the same figures reveals the capability of the month embedding vectors to capture the essential characteristics of distinct seasons. However, an issue arises in the seasonal albedo color due to the nature of the month embedding vector attempting to capture the appearance of a scene during a specific month. Notably, shadows are included in these images. This occurrence is likely attributed to the fact that all images taken of
each scene in this data set are captured within a time frame of less than an hour, resulting in nearly identical shadows in images from the same month and introducing them as a seasonal appearance.

**Figure 4.14:** Ground truth as well as RGB and seasonal albedo predictions by Planet-NeRF for the Omaha areas with seasonal variations.
### Figure 4.15: Ground truth as well as RGB and seasonal albedo predictions by Planet-NeRF for the Omaha areas with seasonal variations.
4.3.4 Impact of solar layer

Even with the integration of the month embedding vector (MEV) in Planet NeRF, the influential role of the sun in seasonal predictions remains evident. Figure 4.17 illustrates that the shading scalar retains high values for areas that exhibit seasonal variations, even for Planet-NeRF that utilizes MEV. A comparative analysis of the shading scalar for Planet-NeRF and Sat-NeRF suggests that it has slightly lower values for Planet-NeRF. This observation is supported by the lighter gray tones in the shading scalar image, indicating a somewhat diminished role of the sun in seasonal prediction when employing the MEV.

Examining the predicted ambient sky color for Planet-NeRF, it aligns with the dominant image color in most cases. Notably, there is an exception where the ambient sky color appears blue, despite the main image color being brownish, suggesting that the ambient sky color has successfully captured the accurate sky color.

Figure 4.18 further demonstrates that the rendered images change dynamically, even with a static MEV. This variability arises by altering the solar directions to represent the beginning, middle, and end of February. The ability of solar direction adjustments to impact predictions suggests that it remains a potent fac-
tor, providing more flexibility than relying solely on texturing each month based on the MEV.

![Table](image)

<table>
<thead>
<tr>
<th></th>
<th>Planet-NeRF</th>
<th></th>
<th>Sat-NeRF</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground truth</td>
<td>Shading scalar</td>
<td>Sky ambient</td>
<td>Shading scalar</td>
<td>Sky ambient</td>
</tr>
<tr>
<td>January</td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>March</td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
</tr>
<tr>
<td>September</td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
</tr>
</tbody>
</table>

**Figure 4.17:** Predicted RGB for Planet-NeRF and Sat-NeRF, compared to the ground truth RGB for OMA_212.

![Images](image)

**Figure 4.18:** Three rendered images of OMA_374 in February with three different input solar directions, from the beginning, middle and end of February.
4.3.5 Signed distance function for altitude prediction

As seen in Table 4.3 the numerical results for Planet-NeRF+SDF are slightly worse than for Planet-NeRF, with exception for some cases. When examining the images presented in Figure 4.19, it’s apparent that the predictions from both models are visually similar to some extent. Nonetheless, Planet-NeRF+SDF appears to struggle more with accurately capturing the complex tree structures in OMA_374, in comparison to Planet-NeRF. For a detailed examination, the specific predictions made by Planet-NeRF and Planet-NeRF+SDF within the segments where there is lidar data, and therefore where the altitude MAE is being computed are depicted in Figures 4.20 through 4.22.

As seen in Figure 4.23 the final RGB images are very similar, specifically the colors which seem almost identical. Which also can be confirmed by the very similar PSNR and SSIM scores for the two models.

![Figure 4.19: An altitude comparison between Planet-NeRF and Planet-NeRF+SDF for three different Omaha areas.](image-url)
4.3 Sat-NeRF modifications

![Lidar Planet-NeRF Planet-NeRF +SDF](image1)

**Figure 4.20:** Lidar ground truth altitude as well as the predicted altitudes for Planet-NeRF and Planet-NeRF+SDF on the subset where lidar data exists.

![Lidar Planet-NeRF Planet-NeRF +SDF](image2)

**Figure 4.21:** Lidar ground truth altitude as well as the predicted altitudes for Planet-NeRF and Planet-NeRF+SDF on the subset where lidar data exists.

![Lidar Planet-NeRF Planet-NeRF +SDF](image3)

**Figure 4.22:** Lidar ground truth altitude as well as the predicted altitudes for Planet-NeRF and Planet-NeRF+SDF on the subset where lidar data exists.
4.3.6 Loss

As stated in Section 3.4, there is a notable change in the final loss calculation when using Planet-NeRF+SDF compared to other methods. Specifically, the loss function transitions from $L_{\text{RGB}}$ to a more complex form $L_{\text{SDF}}$ (see Equations 3.8-3.10). When observing the values for final losses in Table 4.5, it becomes evident that this modification results in a substantially higher loss value for Planet-NeRF+SDF compared to the loss values observed with Sat-NeRF and Planet-NeRF.

<table>
<thead>
<tr>
<th></th>
<th>Sat-NeRF</th>
<th>Planet-NeRF</th>
<th>Planet-NeRF+SDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_{\text{RGB}}$</td>
<td>1.17</td>
<td>0.95</td>
<td>-</td>
</tr>
<tr>
<td>$L_{\text{SDF}}$</td>
<td>-</td>
<td>-</td>
<td>1446.01</td>
</tr>
</tbody>
</table>

*Table 4.5*: The final loss of three different models trained on OMA_212 at epoch 20.
This chapter thoroughly discusses the results presented in Section 4, exploring the methodologies employed and identifying potential areas for future work. Subsequently, ethical considerations are addressed.

5.1 Results, methodology, and future work

This section integrates results, methodology, and future prospects for the diverse aspects undertaken in the project. Certain outcomes are influenced by limitations within the applied methods, prompting a thorough exploration by the authors to understand the root causes and contemplate avenues for future refinement. Conducting a comprehensive review, this section aims to shed light on the intricacies of the findings, offer insights into methodological nuances, and deepen understanding of the implications, all while encouraging potential avenues for improvement moving forward.

5.1.1 Sat-NeRF

In Section 4.1, it is evident that Sat-NeRF struggles more with predicting scenes undergoing significant seasonal changes than with those experiencing minimal seasonal variations. This observation is supported by higher PSNR and SSIM scores for Jacksonville scenes, which experience minimal seasonal changes, compared to Omaha scenes with notable seasonal variations. Since PSNR and SSIM evaluate the texturing of predicted images against the ground truth, the seasonal variations appear to introduce considerable challenges.

Despite these challenges, a visual examination of the predicted images reveals that Sat-NeRF can indeed register seasonal variations. In many instances, the model successfully captures key aspects of different seasons. This capability is
particularly evident when analyzing the shading scalar and the predicted sky ambient color. It is observed that the shading scalar attains higher values in areas with significant seasonal variability, suggesting the sun’s influential role in these variations. Additionally, the sky ambient color seems to adopt a hue more reflective of seasonal changes than the intended sky color.

When examining the predicted albedo colors for Omaha regions generated by both Sat-NeRF and Planet-NeRF (see Figures 4.4, 4.5, 4.6, 4.14, 4.15, and 4.16), a notable observation is their failure to accurately capture the true albedo colors of the depicted scenes. This discrepancy may stem from the fact that the albedo prediction solely relies on the spatial coordinates, $x$, and the transient embedding vector, $t$, as input, neglecting the influence of solar directions. This fact also strengthens the theory that the sun plays an important role in seasonal predictions.

The omission of solar directions as an input parameter implies that the network lacks the contextual understanding of why images exhibit variations across different seasons. Consequently, the network may strive to find a universal solution that, when combined with the shading scalar and ambient sky color (which now seem to be more of a seasonal scalar and season color), yields satisfactory predictions for all seasons. This tendency could lead the network to adapt an almost entirely white albedo color. It is reasonable to argue that white, being a neutral and versatile color, might be perceived as more generalizable to diverse images that with substantial differences in appearance across seasons.

In contrast, when examining Sat-NeRF’s predictions for albedo colors in the context of Jacksonville scenes, this deviation from accurate color representation is notably absent. In these cases, the network appears to generate albedo colors that align more closely with the actual ground truth colors of the scenes. This difference in behavior can be attributed to the higher similarity among images from the two distinct Jacksonville data sets, where minimal seasonal variations exist.

In the Sat-NeRF network architecture, the solar layer operates by predicting an ambient sky color intended to be included in parts of the image that contains shadows. In regions with distinct seasons, the network forecasts a seasonal color rather than a sky color. Additionally, it identifies areas more likely to be influenced by seasons, either independently or in conjunction with shadow-affected regions. However, seasonal characteristics can exhibit greater visual diversity than shadows. For instance, an area may feature snow-covered ground while the trees remain green. In such cases, the network faces the challenge of determining whether the ambient color should be white or green. A prospective avenue for exploration could involve testing the prediction of pixel-wise colors for the sky, instead of one sky color, which could function more like a pixel-wise season color. This approach may enhance the network’s ability to handle intricate scenarios where seasonal color attributes vary significantly.
5.1 Results, methodology, and future work

5.1.2 Sun impacts

As the Earth’s axial tilt and orbital position around the Sun are responsible for the changing seasons, it is reasonable to assume that a deep neural network learns about seasons in a similar manner. In March and September, the Sun reaches the equinox, following a solar trajectory that shows minimal deviation between these months. Consequently, the network’s input remains relatively consistent, despite the potential significant distinctions between March and September. Addressing this challenge, the Planet-NeRF network has successfully solved this problem introducing a month embedding solution. While the Sun encapsulates many seasonal features, the incorporation of a month embedding proves to be crucial in accurately characterizing dates near the equinoxes.

The solar position undergoes changes not only throughout the year but also within a day, from sunrise to sunset. However, it is worth noting that all images in the datasets were specifically captured between 17:00-18:00 UTC. This restricted timeframe might have made it more convenient to utilize the sun in season predictions, especially when contrasted with datasets featuring a broader daytime interval. To gain a more comprehensive understanding, further investigation is warranted on datasets with a larger time span. It is crucial to acknowledge that the effectiveness of the sun in season predictions cannot be universally guaranteed and may vary depending on the context.

5.1.3 Month embedding

The incorporation of the month embedding vector has yielded improved results based on the metrics utilized (PSNR, SSIM, and altitude MAE). However, it remains crucial to conduct a visual assessment of the models’ performance in season prediction. Figure 5.1 showcases novel view images of the OMA_374 area. While the prediction for November appears inaccurate, the other images demonstrate a more faithful representation of their corresponding months compared to the images in Figure 4.10, generated using Sat-NeRF. Sat-NeRF exhibits confusion between unseen months like June and July and even misinterprets neighboring months (May and August) in dark tones. In contrast, Planet-NeRF more accurately captures the distinctive attributes of these months, despite the absence of June and July in the training dataset.

The less favorable outcomes in the months of June, July, and November may depend on their absence in the training data. A potential solution to this issue could involve restructuring the annual timeline into broader segments, like Spring, Summer, Autumn, and Winter. Examining a more detailed approach to seasonal representation, such as training a distinct vector for each week, presents intriguing possibilities. However, to implement this effectively, an increased amount of training data is necessary.

The observation that the albedo color in the predictions of both Sat-NeRF and Planet-NeRF consistently appears as white or pinkish across all scenes prompts a reevaluation of the role of albedo color in the models. This is particularly relevant when adapting Planet-NeRF to areas with significant seasonal variability. Given Planet-NeRF’s capability to predict a season-specific albedo color using the MEV,
5.1 Discussion

Figure 5.1: Rendered images of the OMA_374 area using Planet-NeRF. The different months are generated using solar-directions from the 15th of each month year 2023, and month embedding vector of each month. Note that the months with red labels are missing in the training data.

it raises an important question: Could this feature potentially eliminate the need for a separate albedo prediction? This consideration stems from the fact that the season-specific albedo prediction might be more effective in accounting for the dynamic changes in appearance characteristic of different seasons. This potential approach could simplify the model while potentially enhancing its accuracy in regions with diverse seasonal characteristics, suggesting a potential avenue for future research.

5.1.4 Altitude

The altitude predictions generated by all tested models rely solely on spatial coordinates, $x$. This input configuration leaves the network without contextual information regarding the specific season it is predicting altitude for. For instance, the dynamic nature of scenes, such as changes in tree appearance due to seasonal variations, introduces a considerable variability in true altitudes from one season to another. The current approach does not account for this temporal aspect, possibly contributing to the observed challenges in predicting altitude for scenes rich in trees. This difficulty manifests in both numerical metrics and visual quality.

Another obstacle in altitude prediction arises during the evaluation process against ground truth altitude obtained from lidar data. Firstly, since the lidar data only is provided for a subset of the scene for each dataset, it is possible that the results would vary if the altitude could be evaluated on the entire scene. Second, the uncertainty surrounding the timing of data collection adds complexity in identifying potential sources of error. A plausible assumption is that the lidar data is collected at a specific moment, possibly during any season, making the ground truth less representative for other seasons. In regions with substantial altitude variations across seasons, like forests or areas with deep snow, having
multiple lidar models as ground truth would be advantageous and give better accuracy in the measurements. This approach is expected to enhance accuracy in altitude measurements.

The lidar data might contain transient objects, posing challenges in precisely evaluating altitude predictions. Unfortunately, the authors of this thesis were unable to find information on this matter.

Also one can note that the altitude predictions are very similar for Sat-NeRF and Planet-NeRF. This is probably because they have the exact network structure for predicting sigma.

A potential avenue for future work involves introducing a dedicated month embedding vector to inform the prediction of volume density, $\sigma$, and the signed distance function, $f(x)$. This enhancement would enable the network not only to capture seasonal texture variations with a month embedding but also to discern altitude variations at each pixel, addressing the complexities introduced by seasonal changes in scenes.

### 5.1.5 GEO+SL

The visualization in Figure 4.13 confirms the idea of employing geometrically rendered shadows and incorporating the solar layer to identify seasonally affected regions. The shading scalar showed high values for pixels in the shade, while the season scalar showed high values for pixels containing seasonal attributes, such as the vegetation. However, despite the visual success of representing the idea, the quantitative results from Sat-NeRF+GEO+SL indicate inferior performance compared to the other models across various metrics in all the areas. Consequently, it cannot be conclusively established that the approach of Sat-NeRF+GEO+SL surpasses the method of allowing the solar layer to integrate season and shadow predictions, as in Sat-NeRF and Planet-NeRF, rather the opposite.

The idea of incorporating geometrically rendered shadows is inspired by EO-NeRF [14]. Unfortunately, the unavailability of the source code for EO-NeRF by Mari et al. [14] prevented its utilization in this study. The less-than-ideal results in altitude prediction raise the possibility of implementation issues, especially considering that EO-NeRF, as reported by Mari et al. [14], achieved superior results compared to our implementation. To better understand the efficacy of the solar layer for seasonal predictions, it would have been valuable to test this concept in an environment with greater altitude accuracy. This suggests that the results encountered from Sat-NeRF+GEO+SL may stem from implementation intricacies rather than flaws in the underlying concept. Despite these challenges, the authors believe in the potential of combining geometric shadow rendering with the solar layer for seasonal predictions, with the aim to assign a shadow-specific color to shadows, while areas affected by seasonal changes will adopt a season-specific color. This perspective identifies it as a promising direction for future research in this domain.
5.1.6 Planet-NeRF+SDF

When employing Signed Distance Function (SDF) for altitude prediction instead of volume density, a notable shift occurs in the total loss, escalating from approximately 1 to around 1500. This stark increase suggests that the dominance of the eikonal loss, associated with predicting altitude using SDF potentially overwhelms the impact of the RGB loss. This shift might imply that the network tends to downplay the influence of pixels with potential transient objects during scene learning, possibly affecting color learning as well.

However, an intriguing observation arises when examining the quantitative results of Planet-NeRF and Planet-NeRF+SDF. Despite the substantial increase in total loss, the PSNR and SSIM scores exhibit remarkable similarity, and the visual comparison of predicted RGB images reveals a close resemblance. Consequently, the role played by the eikonal loss overshadowing the RGB loss remains unclear.

It is essential to highlight a critical aspect concerning the utilization of SDF for altitude prediction in this thesis. The parameters associated with Planet-NeRF+SDF, namely the initial learning rate, $\beta_{\text{REG}}$, number of samples, and the strategy for updating the learning rate, have not undergone comprehensive investigation. Consequently, there exists a significant possibility that these parameters may be suboptimal, potentially impacting the overall performance of the model.

Given the suboptimal performance of Sat-NeRF+GEO+SL in altitude estimation, which significantly impacted overall results, it would have been beneficial for the authors to either investigate and address the underlying causes of Sat-NeRF+GEO+SL’s inadequate altitude predictions or to concentrate on fine-tuning the parameters of Planet-NeRF+SDF. This focused approach could potentially have led to one superior model instead of two that both needs further work.

5.2 Ethical considerations

Considering the global reach and the potential invasiveness of satellite images, an ethical discussion regarding them feels inevitable. Satellite images, by their very design, capture extensive geographical areas, often including residential zones, private properties, and sensitive regions. This capability inherently raises concerns about privacy and the ethical implications of observing and recording these areas without explicit consent from the individuals affected.

However, it is important to acknowledge the limitations of satellite imagery in terms of resolution and detail. Current satellite technologies, while advanced, do not possess the resolution to identify individuals or specific personal identifiers such as license plates. This inherent limitation significantly mitigates the potential for overly invasive surveillance. The images captured are generally not detailed enough to infringe on personal privacy in a direct manner, as they cannot reveal identifiable features of private individuals or their personal belongings.

Despite this, the ethical use of satellite imagery requires a careful and considered approach.
This thesis has undertaken a comprehensive exploration into the utilization and enhancement of NeRF for 3D reconstruction of satellite imagery, with a specific focus on accommodating seasonal changes.

The first research question (1) directed attention to evaluating the existing Sat-NeRF network’s performance in scenes characterized by seasonal changes. The assessment brought to light significant challenges, particularly in predicting scenes with substantial seasonal variations. The disparities in PSNR and SSIM scores across scenes with varying seasonal dynamics underscored the intricate nature of texturing predictions when confronted with pronounced seasonal changes. Despite these challenges, a visual examination highlighted instances where Sat-NeRF exhibited a commendable capacity to capture seasonal variations, treating season-affected areas akin the way of transient objects.

The investigation into Sat-NeRF’s ability to capture seasonal attributes further delved into the crucial role played by sun direction inputs, as explored in research question 2. Recognizing the sun’s varying orbit from winter to summer, these inputs are crucial for the network to learn and comprehend different seasons. However, the equinox months introduce complexities, challenging the reliability of solar direction inputs for an accurate representation of the entire year.

Addressing research question 3 on the enhancement of the Sat-NeRF network, affirmative progress has been made. This thesis introduced three altered network structures, resulting in the development of Planet-NeRF, Planet-NeRF+SDF and Sat-NeRF+GEO+SL. The refined networks Planet-NeRF and Planet-NeRF+SDF demonstrated improvements as validated by both quantitative metrics (PSNR, SSIM, and depth MAE) and visual assessments. Intriguingly, the incorporation of geometrically rendered shadows and a solar layer for seasonal predictions (Sat-NeRF+GEO+SL) did not yield improvements in handling seasonal variations ac-
conceding the metrics. On the contrary, the results saw a deterioration from the original Sat-NeRF network.

In conclusion, addressing the dynamic nature of seasons poses a significant challenge for NeRF methodologies, encompassing complexities not only in the method and development but also in its evaluation. The existing datasets fall short in providing ground truth data of sufficient quality to thoroughly assess the impact of seasonal variations. However, this thesis, with a particular focus on the Planet-NeRF network, adeptly navigated these challenges and revealed noteworthy improvements compared to existing works. These efforts resulted in achieving state-of-the-art results on publicly available datasets, highlighting the effectiveness of the proposed approach in handling the intricacies associated with seasonal variations.


