TU Delft
Delft University of Technology
Faculty of Architecture
and the Built Environment

Leveraging Shadows for Accurate AI Segmentation and Height Estimation

Lars C. Huizer

Delft University of Technology
Faculty of Architecture
and the Built Environment

Delft University of Technology

Leveraging Shadows for Accurate AI Segmentation and Height Estimation

Master's Thesis - Research Proposal

To fulfill the requirements for the degree of
Master of Science in Geomatics
at Delft University of Technology under the supervision of
Dr. A. Rafiee (Geomatics, Delft University of Technology)
and
Ir. E. Verbree (Geomatics, Delft University of Technology)

Lars C. Huizer

February 17, 2025

Contents

1 Introduction

The usage of point clouds, a set of data points whose combined purpose is to represent 3D objects and is often collected by using LiDAR scanners or photogrammetry, has become increasingly ubiquitous within various fields of expertise over the years. Raw point clouds may be collected and converted into formats useful for the creation of digital twins, which are models that seek to digitally mirror a real-world phenomenon to help drive data-centric decision-making processes (AlBalkhy et al., 2024). Such digital twins are used in a wide variety of fields; in the field of Architecture, Engineering and Construction (AEC) and Facility Management (FM), for example, professionals are becoming increasingly interested in employing point clouds for purposes such as geometric quality assessment, a process in which the physical profiles and condition of constructed objects may be investigated to see whether or not they comply with set standards (Kim et al., 2019). In the field of forestry and environmental monitoring, LiDAR scans and subsequent point clouds may be processed to estimate the amount of biomass, forest structure and tree height, which in turn may help in the support of sustainable forest management (Edson & Wing, 2011).

Another example is the use of LiDAR and LiDAR-derived point clouds to measure the height of objects such as buildings. Here, it is relevant to mention the Actueel Hoogtebestand Nederland (freely translated to ”Current Dutch Elevation Record”), often abbreviated to AHN. It is a national elevation dataset based on LiDAR data collected by either helicopter or airplane, with the fourth edition of the dataset sporting a height accuracy of up to 5 cm and an estimated average point density of 10-14 points per square meter (AHN, 2020b; Ministerie van Onderwijs, 2018). While the AHN and similar point cloud datasets work well for determining the height of broad and well-defined objects such as buildings, thin or elongated structures (for example, antennas, masts and overhead lines) pose a greater challenge. Since these objects often have a smaller reflective surface area, they are typically represented less reliably.

The relevance of mapping such thin objects is twofold; firstly, thin objects often are important in an urban context, for instance in monitoring overhead lines in public transportation systems or for mapping out anything that may obstruct flight paths near airfields. Secondly, accurate mapping matters in areas where data collection happens less frequently or where the appropriate tools may be unavailable depending on the accuracy needed, such as in rural or remote regions. Because accurate datasets allow for a better-informed decision-making process and a deeper understanding of the spatial environment, it is evident that there is significant value in improving the quality of our methods for capturing, processing, and interpreting geospatial data. In many regions, particularly in developing countries and remote areas, the access to high-quality spatial data remains a challenge. Whereas countries such as the Netherlands are able to invest heavily in representative geospatial data as it is a relatively small country with a high GDP, other countries may run into financial or technological roadblocks. For example, the procurement of large-scale LiDAR datasets can prove expensive, especially when collected at higher resolutions, accuracies and precisions. In such cases where there is a lack of high-quality data, the need for alternative methods that can be used for data enhancement becomes apparent.

While various methods exist for smoothing out point cloud surfaces and removing outliers, few solutions exist for the issue of missing data. Depending on the method of data collection, points representing a surface may be lost during recording due to environmental factors such as reflectivity or transparency of a given material and procedural issues such as occlusions or viewing angles (Huang et al., 2020). Solutions for the issue of point completion exist in the form of deep-learning-based

procedures, where neural networks are employed to infer missing points (Yu et al., 2021). However, these methods are not yet at a level of maturity where they can be reliably applied to geospatial data if the need for accuracy is high.

Given this, an alternative approach to increasing data quality can be found in exploiting other kinds of geospatial data that may directly or indirectly encode the necessary information required. For example, in the case of the accurate height determination of elongated vertical objects where LiDAR point clouds are sparse or partially occluded, a workable solution for accurate height estimation is to use shadow-based techniques. Shadows encode spatial information in the form of geometry and orientation that may be exploited to learn more about the object that casts them. For example, by combining the length of a shadow and solar angle of the sun at a given time of day, it becomes possible to calculate the height of the original object.

To do so on a large scale requires the automatic segmentation of shadows in aerial photography. Methods for the segmentation of sunlit regions from unlit regions in imagery can be broadly divided into two categories: the property-based methods, where strictly the inherent properties of a given shadow are used, and the model-based methods where additional information about the surrounding region is used to aid in segmentation, e.g. through known geometry of the shadow-casting object and solar angles (Liasis & Stavrou, 2016).

Although the filter-based methods were historically considered to be the most performant as per Liasis and Stavrou (2016), recent developments in the field of Artificial Intelligence and the advent of deep-learning based solutions have led to significant improvements in shadow detection and segmentation accuracy, resulting in comparable accuracies to the filter-based methods (Luo et al., 2020). However, the accurate extraction of shadows from aerial imagery remains difficult for a number of reasons. For example, time of day has a large influence on e.g. contrast, colour and saturation which may lead to different accuracies of segmentation. Additionally, shadows cast by buildings may overlap with one another, leading to a loss of semantic information. Another issue is when a shadow self-overlaps with the object that is casting it, e.g. a church tower overlapping with the underlying primary structure.

Given these challenges, the improvement of shadow segmentation could potentially be achieved by combining multiple data sources to compensate for the individual shortcomings of the data. Conventional neural networks generally rely on large-scale training using input-output data pairs and relies purely on statistical inference of the data in itself. However, in cases where such data is lacking lies another solution in the form of Physics-Informed Neural Networks (PINNs) Raissi et al. (2017). PINs work by not only looking at purely the data, but also the rules and laws that describe the data. This allows similar models to infer well even with lower amounts of training data. As an example, PINNs have been employed to solve fluid dynamics problems such as in the work by Jin et al. (2021), where the addition of the Navier-Stokes equations enabled the model to deliver promising results in flow simulation. Building on this notion, an opportunity potentially lies in employing the concept of PINNs for the regularization of shadow-segmentation using aerial photography. In this case, the use of pre-calculated shadow projections from point clouds can serve as the method of regularization, thus combining two different datasets to reach more accurate and robust segmentation than normal segmentation would have achieved.

Following the ever-existing need for alternative methods for data enhancement, this thesis will explore the use of PINNs for an enhanced segmentation of shadows from aerial photography by including shadow projections generated from point clouds as regularization. The goal is that the increased robustness and accuracy of shadow segmentation will lead to more accurate height estimation of objects; in particular in those that contain thin and elongated objects which may be captured sparsely in LiDAR datasets.

2 Related Work

This section will review existing methods for the estimation of object heights, with a particular focus usage of shadow segmentation techniques in both traditional and deep learning approaches. Firstly, LiDAR-based reconstruction of buildings will be discussed as this is relevant due their limitations in representing thin and elongated objects. Then, the segmentation of shadows from aerial photography will be discussed, as well as the derivation of an object's height through these shadows. Finally, a summary of the research gap is provided and based thereon the research questions will be formalized as well as the scope of this thesis.

2.1 LiDAR-based Reconstruction Algorithms

In a paper written by Brenner (2005) where he discussed the reconstruction of buildings from imagery and laser scanning, he highlighted how the then-current manual workflows led to high costs when having to process large amounts of data. Here, he highlighted the need for more "automated, efficient extraction systems" assisting in the extraction of objects from laser scans and photography. Since then, LiDAR has grown to be one of the most popular methods for capturing high-density 3D data (Nex & Rinaudo, 2011). Whereas photogrammetry may have issues if the terrain or lighting are not favourable, LiDAR is able to perform relatively well even in rugged terrain. However, that does not mean LiDAR is without its challenges: for example, LiDAR has issues with accurate boundary extraction depending on the resolution (Cheng et al., 2008). A figure illustrating this problem can be seen in appendix A.3.

A variety of papers discuss the reconstruction of boundaries for buildings. For example, (Dorninger & Pfeifer, 2008) describe an automated approach for the extraction, reconstruction and regularization of buildings that exploits the fact that buildings are often made up of planar surfaces.(Kwak & Habib, 2014) recreated the geometry of buildings by recursively applying a minimum bounding rectangle algorithm to collected LiDAR data. (R. Wang et al., 2016) use LiDAR data to first generate a Digital Elevation Model (DEM), after which a series of raster-based operations are applied to generate a boundary representation of building objects.

However, while these methods are able to deal with issues in boundary representation in cases where enough contextual information is available (i.e. plane fitting relies on recognizable geometric patterns and a certain completeness of a dataset), they may struggle in cases with occlusions or insufficient LiDAR hits when the objects have thin edges. This limitation is particularly present for objects such as antennae or ornaments with small reflective surfaces. A practical example of this issue can be found in Figure 1 Since these objects reside on top of buildings more often than not, accurate knowledge of such geometries are relevant when accurate knowledge on the height of a building is required. As such, there is a need for approaches that are able to compensate for these issues when LiDAR data is lacking.

2.1.1 Point Cloud Completion

A potential option for missing LiDAR data is through point cloud completion by neural networks. Fei et al. (2022) provide a large-scale review of many different deep-learning completion methods and their architectures. For example, point cloud completion networks such as PointNet++ (Qi et al., 2017) and PF-Net (Huang et al., 2020) use hierarchical learning, a process where (in the context of

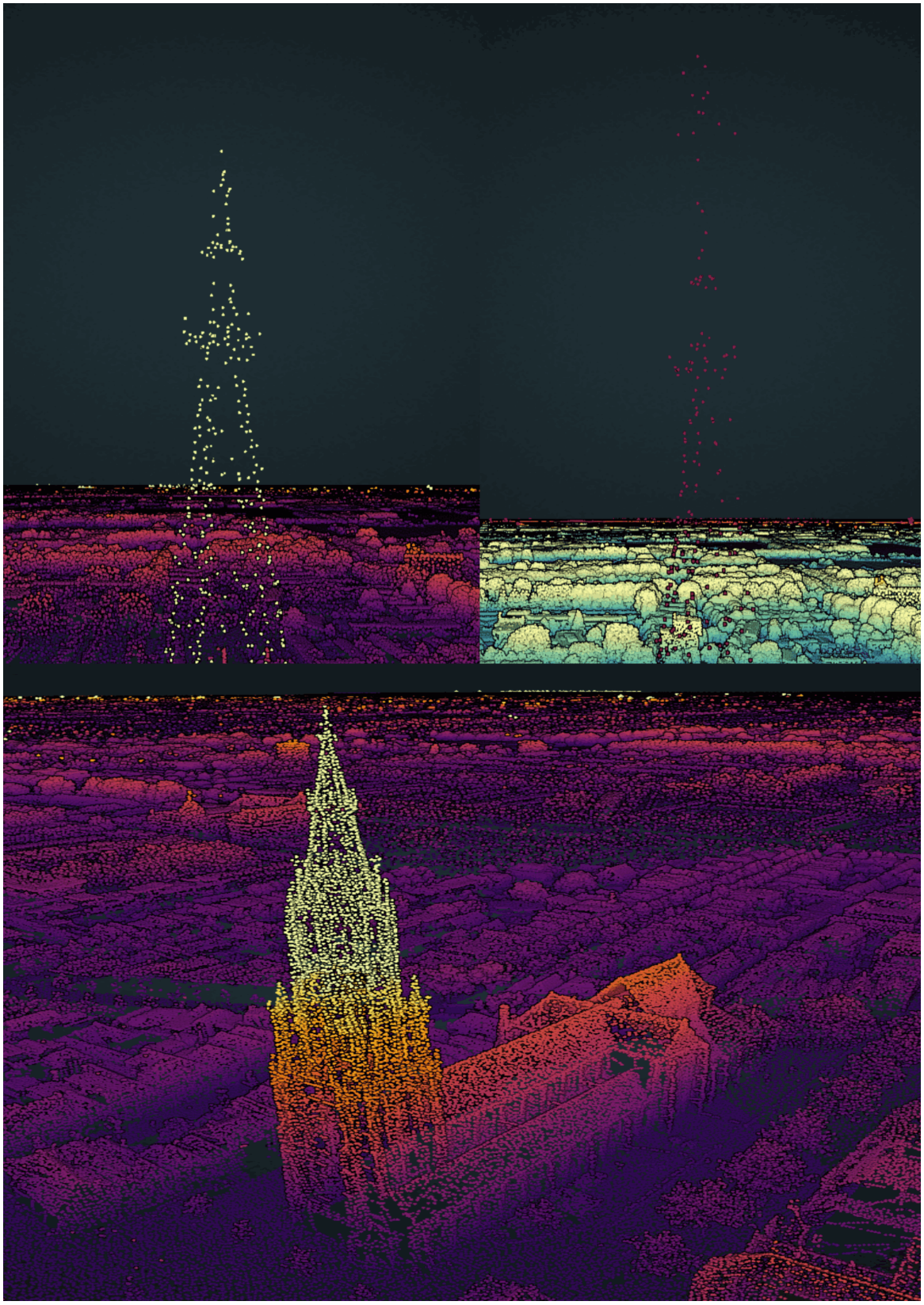

Figure 1: Point Cloud representations of the ”Nieuwe Kerk” in the city of Delft, the Netherlands. In the top left, the AHN2 point cloud dataset is loaded, with the AHN5 visible in the top right. As can be seen, there is a major difference in the representation of the church’s cross. Manual inspection reveals that there is a difference of 3.45 meters in height between the the tops of the two datasets, which would be a significant difference if an accurate height estimation of the church is needed.

point clouds) an original set of points is continuously abstracted by downsampling, increasing the local such that the further along in the hierarchy, the larger the local region becomes. Such a method helps the model understand both global and local contexts. Another neural network architecture used for the generation of points is the Generative Adversarial Network (GAN), which is comprised of a generator that predicts missing points, and a discriminator that assesses the probability of generated points being valid based on real data (Tan et al., 2023; X. Wang et al., 2020)

Although the above mentioned solutions may provide workable outcomes in cases outside of aesthetical purposes, Fei et al. (2022) mention that more work needs to be done to get point cloud completion to a point where it is practical when high accuracies are required. In cases where the models do not have sufficient semantic surrounding information that they may exploit such as e.g. symmetries or other design features, they will have difficulty in generating missing points that would be similar to the ground truth. Considering the relative immaturity of point cloud completion algorithms, the need for alternative methods for height measurement is further stressed.

2.2 Shadow-Based Height Estimation

The idea of using shadows to estimate the height of various objects has already been well-described in the existing body of literature (Kadhim & Mourshed, 2018; Lee & Kim, 2013; Liasis & Stavrou, 2016; Shao et al., 2011; Shettigara & Sumerling, 1998).

2.2.1 Shadow Detection and Deep-Learning Algorithms

In order to accurately calculate the height of an object based on its cast shadow, its shadow needs to be segmented from aerial photography with suitable precision. According to Liasis and Stavrou (2016), the methods by which to detect shadows can be mostly described by two categories, namely the property-based methods and the model-based methods. With property-based methods, the spectral and spatial features of shadowed regions themselves are utilized, whereas model-based methods employ the use of additional site-specific information, such as the solar altitude and the object's geometry.

Filter methods are property-based methods that analyse shadows based on e.g. intensity values or textures. The most simple example are histogram-based techniques that work based on the assumption that sunlit and shadowed regions have a clear separation between them in terms histogram levels (Adeline et al., 2013). A threshold by which to segment the different areas may then either be set automatically through statistical means (Otsu, 1979) or may be selected manually (Yamazaki et al., 2009). Another example of such a filter includes Gabor filters (Granlund, 1978), which are linear filters consisting of a Gaussian function and a sine-wave sensitive to edges and ridges that are oriented in a particular direction.

Model-based approaches consider a priori knowledge about the site of study. If one has pre-gathered geometrical information about the object casting a shadow as well as knowledge on the then-current standing of the sun, segmentation would become easier since its shadows can be precisely calculated instead of needing to interpret images. An example of this is Volumetric Shadow Analaysis as performed by Lee and Kim (2010) which assumes that the ground sample distance, elevation and solar azimuth are known.

In 2012, Adeline et al. (2013) conducted a comparative study where they evaluated and ranked the above mentioned shadow-detection techniques on their F-scores. The conclusion was that histogram thresholding using the methods of Nagao et al. (1979) performed the best, followed by a physics based method by Richter and Müller (2005), a support vector machine-based (SVM) method and respectively in last a K-means clustering method and the SMACC method.

	Property-based methods		**Physics-based methods**		**Machine-learning methods**
	Histogram thresholding method (Nagao et al., 1979)	RGB combination model	Richter and Müller (2005) method	SMACC	SVM
Average F-score for testing dataset	92.5	87.5	90.0	83.9	87.7

Table 1: The results of the comparative study by Adeline et al. (2013)

Since then however, the traditional machine learning methods such as SVM and K-means have largely been outpaced by deep learning algorithms. Deep learning is a subset of machine learning where neural networks with multiple layers are used to interpret data. One example of a deep learning method are Convolutional Neural Networks (CNNs), which are commonly employed for the purposes of computer vision. A CNN consists of multiple layers through which input data is fed, with each layer performing specific operations to extract and process features.

CNNs have been employed in various papers for shadow segmentation. For example, a widely used CNN architecture named U-Net (Ronneberger et al., 2015) was adapted by Jiao et al. (2020) for the segmentation of clouds and shadows. This is then followed by Dense Conditional Random Field (Dense CRF) refinement, where the inference of a given pixel is not only dictated by its local context but the global context (Krähenbühl & Koltun, 2011). In another example, Luo et al. (2020) created a CNN named "DSSDNET" in an encoder-decoder residual structure, which was trained using an auxiliary supervision structure giving each level the ability to train directly on the ground truth, thus avoiding vanishing gradient issues. It differs from conventional CNNs and deep-supervision networks in the sense that the outputs of the intermediate auxiliary layers are combined and refined into a final prediction, instead of only fusing from the last levels of the network. In their testing, the model reached an average F-score of 91.78%, outcompeting other shadow detection methods such as U-Net (in its original form per Ronneberger et al. (2015), which scored an F-score of 87.84%.

While CNNs like DSSDNet have demonstrated a good accuracy in shadow segmentation, their approaches rely solely on the features learned from the input data during training. To date however, there are no CNNs that are trained to use precalculated shadow masks as input for training. The concept of using pre-calculated shadows with CNNs has to the best of my knowledge only been used in the paper by Ufuktepe et al. (Ufuktepe et al., 2021), where they used precalculated shadow masks to generate a ground-truth dataset by which the model could be self-supervised and trained.

2.2.2 Physics-Informed Neural Networks for Shadow Segmentation

In the current body of work, most of the work on shadow segmentation through neural networks is purely driven by a data-centric approach; that is, any inference or training is only based on the relations encoded within the data. However, if one has knowledge on the laws that this data has to abide by (often the laws of physics), it becomes possible to fine-tune the learning process to reach more robust and accurate results. This is the basis of Physics-Informed Neural Networks (PINNs), as originally proposed by Raissi et al. (2017).

Earlier mentioned in the comparative analysis by Adeline et al. (2013), one may have noticed that two physics-based methods were mentioned for the segmentation of shadows, with the method employed by Richter and Müller (2005) coming in second when compared to the other methods. Their method operates by the same philosophy as that a PINN would; both use physics-based constraints. However, they differ greatly since Richter and Müller (2005) in practice is a deterministic method that does not employ neural networks, whereas a PINN relies specifically on a neural network to generalize and infer without needing manual calibration under different circumstances.

While current CNN models for shadow segmentation are relatively performant, there is currently no existing work that explores the usage of physical constrains to regularize the CNN during training and inference. As such, the potential for physics-based constraints to increase the quality of generalization is being left on the table. At the same time, the existing physics-based methods like Richter and Müller (2005) do not make use of the versatility of a CNN which can be applied in more diverse conditions, which means that there is a gap in research where the performance of a CNN enriched with physically informed constraints can be explored.

2.3 Summary of Related Work & Research Gaps

Based on the above literature review, the following research questions were formulated to guide the research:

- The scanning of objects through LiDAR can suffer from issues with boundary representation, which leads to difficulty in height estimation for objects without enough contextual information for inference.

- While many algorithms exist for the purposes of refining a point cloud (e.g. through outlier removal or denoising), few approaches focus on completing missing data outside of deep-based methods or are not usable in precise contexts.

- As Huang et al. (Huang et al., 2020) put it: ”the challenges of [...] missing points [...] have been less addressed and remain unresolved.”.

- Machine learning solutions for point cloud completion have demonstrated some promise, but are ultimately not at a point where they can be widely adopted when accurate results are needed (Fei et al., 2022).

- Shadows offer an alternative way for the derivation of height information, and these can be segmented by various means.

- Although different property-based and model-based methods exist for shadow segmentation, the usage of precalculated shadows as complementary input to aerial photography in convolutional neural networks has to the best of my knowledge not been explored.

- While deep-learning based CNN models have become increasingly popular, they remain purely data-driven without taking into account physical constraints; further assessment into a CNN-model confined by physics-based constraints as used in Physics-Informed Neural Networks (PINNs) for the purposes of shadow segmentation would therefore break novel ground.

3 Research Questions

Based upon the related work and the identified gaps, this thesis will use the following research questions:

Q Main. **How and to what extent can a Physics-Informed Neural Network (PINN) improve the accuracy and reliability of shadow segmentation from aerial photography, and how does this impact height estimation performance for thin, elongated objects that are poorly represented in point clouds when compared to conventional techniques?**

Q1. What is the current state of height-derivation techniques, and what are their respective strengths and limitations?

Q2. How can thin and elongated objects be defined for the purposes of this research?

Q3. What is the current state of shadow detection from aerial photography in the context of thin and elongated objects?

Q4. How does incorporating physics-based constraints into a CNN-based shadow segmentation model impact segmentation accuracy, and how does this improvement translate to better height estimation outcomes compared to conventional LiDAR-based methods?

Q5. Under what environmental and temporal conditions does the proposed method maintain robust performance, and are there scenarios where it outperforms or underperforms conventional approaches?

4 Scope

This thesis will focus on the creation of a Physics-Informed Neural Network designed to segment shadows from aerial photography. The study will be limited to:

- Adapting an existing CNN model (U-Net) for shadow segmentation, thereby using an existing neural architecture.
- The use of precomputed shadow masks based upon LiDAR data as physics-based constraints for fine-tuning.
- The evaluation of performance of the new segmentation model using default metrics such as F-score.
- The usage of height estimation as a validation metric as opposed to a main research goal.
- Aerial photography datasets with known metadata (solar angle, shadow position).

As such, this study will not focus on:

- The improvement of height estimation algorithms
- Object classification except for shadow segmentation
- Reconstruction of buildings or thin objects
- Exploring physics-based constraints other than those that can be generated from point clouds for shadow segmentation.

In this study, the height of a building refers to its absolute highest point. This is often demarcated by elongated objects such as antennae or church crosses, which are frequently underrepresented in LiDAR data

5 Methodology

This section will outline the proposed methodology for shadow segmentation using Physics-Informed Neural Networks (PINNs). Firstly, some relevant theory on shadow-based height estimation will be mentioned as well as the basic theory behind Convolutional Neural Networks (CNNs) and PINNs. Afterwards, the methodology necessary for answering the research questions will be outlined. The approach will use an existing CNN model but will integrate physics-based constraints to effectively turn it into a PINN. The main objective through this is to improve the accuracy of shadow segmentation by enforcing consistency in predicted shadow regions. The estimation of building heights will then be used as a validation metric to assess the effectiveness of the segmentation method, besides traditional F-scores for comparison to other models.

5.1 Theory

5.1.1 Shadow-based height estimation

The estimation of heights from shadows relies on the geometric relationship between shadow length (l_s), solar altitude angle ($\angle Solar_{alt}$) and object height (h_o).

$$h_o = \frac{l_s}{tan(\angle Solar_{alt})} \tag{1}$$

However, this formula makes the assumption that the shadow a given object casts falls on flat ground. As soon as the terrain becomes uneven, or the shadow falls on top of other objects additional measures need to be taken to ensure that the calculation is accurate. For example, adding the difference in height between the base of the building and the tip of shadow is required to come to an accurate height:

$$h_o = \frac{L}{tan(\angle Solar_{alt})} + (h_{tip} - h_{base}) \tag{2}$$

Note in the above equation that L is the planimetric horizontal length of a given shadow. To ensure an accurate height, L should be measured from the tip of the shadow to the point directly vertical from the highest point of the building.

5.1.2 Neural Networks for Shadow Segmentation

A Convolutional Neural Network (CNN) is a deep-learning model that is often used for the purposes of computer vision, or any other field where innate structures in large datasets are exploited for inference. A CNN consists of multiple layers through which data is processed. Generally speaking, the following layers can be identified in a CNN (LeCun et al., 2015):

- **Convolutional Layers**: Layers where input data is scanned using filter banks to detect specific patterns such as textures and shapes.
- **Activation Layer**: The output of each individual node in a convolutional layer here is fed into an activation function, which calculates the output of a neuron. Depending on the type of function used by which to calculate the output, the main goal of this layer is to allow for the approximation of non-linear relations in the data.

- **Pooling Layer**: Downsamples the information coming in to reduce spatial dimension, removing redundant information and reducing computational complexity.
- **Dense Layer**: A fully connected layer that combines all extracted features to provide a final output (Lecun et al., 1998).

For example, an image inserted into a CNN is first processed by the convolutional layers where features such as edges and textures are detected. As these features flow further into the model through more pooling and convolutional layers, the data is progressively abstracted. In the end, a dense fully connected layer maps these features back into specific output (Lecun et al., 1998; LeCun et al., 2015).

In order to train a basic CNN, a training dataset is fed into the network and the kernels in the convolutional layers are randomly initialized. After passing through the layers, the performance of the neural

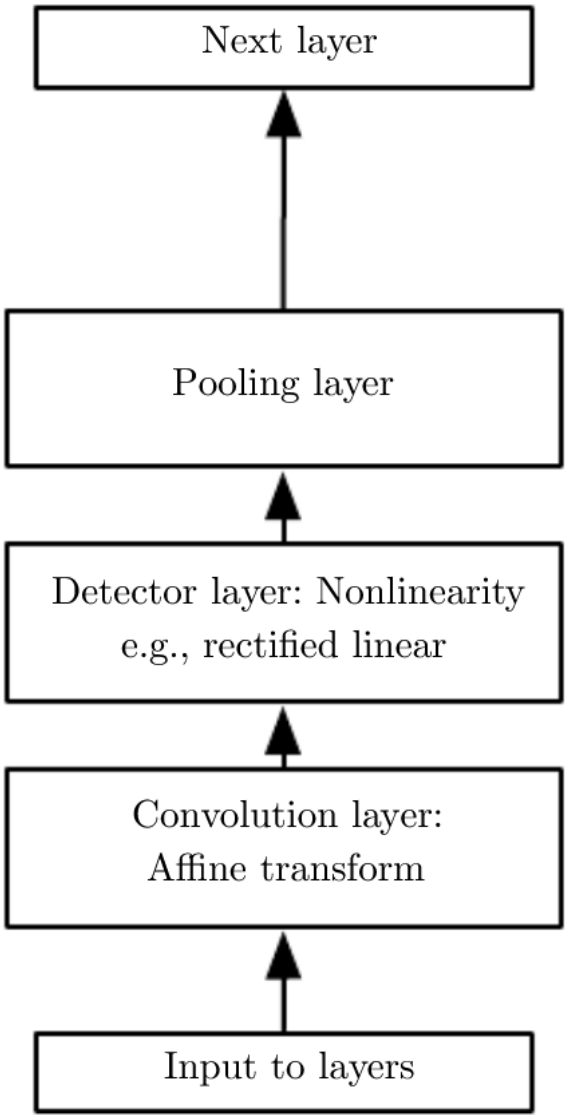

Figure 2: Schema showing the layout of a typical CNN from Goodfellow et al. (2016) . In the figure, the ”Detector layer” is the same as the ”Activation Layer” mentioned above.

network is calculated through a loss function, which measures the difference between the predicted output and the ground truth. As an example of a loss function, in cases where the output is binary (i.e. a pixel is shadowed or non-shadowed), binary cross-entropy can be used which measures difference between the output and the ground truth:

$$l_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(\hat{y}_i) + (1-y_i) log(1-\hat{y}_i)] \tag{3}$$

where i is a pixel in image N, y_i is the binary label of a given pixel i, and $\hat{y}_i$ the probability that pixel i belongs to the ground-truth label. Through a process called backpropagation, this loss function is then used to define how much each individual parameter in the convolutional layers has contributed to the loss. Based on the result of backpropagation, the weights are then finally adjusted through an algorithm like gradient descent to minimize the error (LeCun et al., 2015)

The ways in which a neural network can be trained can be split up into three categories depending on the data available; Firstly, supervised training relies on labeled datasets. This means that for a given input, the output is compared to a validation set. Secondly, unsupervised training relies on the algorithm to learn patterns from unlabeled data, meaning that there is no known mapping from input to output. Finally, there is semi-supervised learning which utilizes both labeled and unlabeled datasets (Nan, 2023). Such a method is advantageous, as fully supervised training requires large amounts of data which is either expensive or time consuming to procure, which is not feasible in all cases.

However, transfer learning offers an alternative method for reducing the costs of training. Originally introduced by Pratt et al. (1991), transfer learning relies on the concept of taking a neural network originally made for a similar purpose and using the same weights instead of training a new model with randomly initialized weights. For example, in order to train a neural network that can recognize cats, it might make sense to start out with the weights of a CNN already trained for dogs, as they are morphologically relatively similar. From there, the model can be further trained and fine-tuned into specifically recognizing cats, and will require relatively less paired input-output data than if the weights were randomly initialized.

5.1.3 Physics-Informed Neural Networks for Shadow Segmentation

A CNN can be turned into a Physics-Informed Neural Network (PINN) by integrating physical constraints into the loss function used during training. This can be achieved by simply adding the loss function to any given prior loss function. For example, if binary cross-entropy was used in a prior model:

$$l_{total} = l_{BCE} + l_{Phys} \tag{4}$$

where l_{Phys} is a loss function that penalizes based upon a given physical constraint. For the segmentation of shadows, this loss can be based upon a precomputed shadow mask that abides by real-world constraints. The exact formulation of l_{Phys} will be defined later during the research phase, and will likely include some form of scaling parameter to balance its strength against the pre-existing loss function.

5.2 Method

Here, the general methodology that will be used for this study is described. A simple schematic overview of the methodology can be found in Appendix 4.

5.2.1 Dataset Selection and Preprocessing

This study will use aerial imagery datasets annotated with shadows as well as LiDAR point clouds. An overview is provided below:

- **Point cloud**: The AHN5 AHN (2020a), a high resolution LiDAR dataset will serve as the primary source of point cloud data for non-terrain objects. It will be used for generating physics-based shadow masks and also serve as a baseline comparison for height-estimation metrics.
- **Digital Surface Model**: A DSM based upon the AHN4 describing the entirety of Netherlands.
- **25CM Aerial Photography**: a low-resolution dataset containing RGB imagery of the Netherlands taken in the summer
- **7.5CM Aerial Photography**: A higher resolution dataset containing RGB imagery of the Netherlands taken in the winter.
- **Annotated Aerial Photography Dataset**: A pre-made publicly available dataset containing annotated aerial photography indicating shadows, as per Luo et al. (2020).

This data will be processed in the following way prior to training and evaluation of the model:

- **Point Cloud**: The point cloud will be clustered to gather distinct buildings. Then, for each building, a shadow mask will be calculated using the method described in section 5.2.2. This mask will then be used as input for the PINN loss function. Additionally, variations on the quality of the data (occlusions, lower resolution) will be simulated to assess the robustness of the method later.
- **Digital Surface Model**: The DSM will be used to extract ground height information necessary for height estimation as described in section 5.1.1.
- **Aerial Photography**: The 8CM high-resolution aerial photography will be downsampled to 25CM to ensure that equal resolution data will be available of both summer and wintertime. Based on these images, a training dataset will be created for the purposes of transfer learning.

5.2.2 Shadow Mask Generation

The generation of shadow masks will be executed as follows:

1. **Point Cloud Processing**:
    - Extraction of non-ground points to separate buildings from terrain.
    - Cluster individual buildings and other structures using DBSCAN.
    - Assign heights to clustered objects based on the difference between the highest and lowest point.
2. **Solar Position Calculation**:

- Using date-time metadata combined with geographic coordinates, obtain solar azimuth and altitude.
- Conversion of these angles into vectors.

3. **Shadow Projection**

- For each point of building object, compute its projection onto DEM by using solar vector.
- If the projection intersects with another building or object first, a test will be performed to see if the point needs to be invalidated (e.g. when the shadow ends on the side of a wall).
- Apply edge detection to projected points to obtain mask boundaries.

5.2.3 CNN Model Pretraining and Transfer Learning

For this study, U-Net (Ronneberger et al., 2015) as it is a popular, tried and tested image classification model. In the pretraining phase, the model will be trained on the dataset by Luo et al. (2020) to function as a baseline. The training loss function here will be:

$$l_{\text{pretrain}} = l_{\text{BCE}} \tag{5}$$

where Binary Cross-Entropy (l_{BCE}) will change depending on the difference per pixel between the ground-truth and the predicted shadow labels. Afterwards, the total loss function will be updated to include the shadow-mask based loss function prior to fine-tuning the CNN, which will turn it into a PINN.

$$l_{\text{total}} = l_{\text{BCE}} + l_{\text{Phys}} \tag{6}$$

5.2.4 Evaluation Metrics and Occlusion Testing

To evaluate the described PINN-model, the following two metrics will be used:

- **F-Score**: The F-score will be calculated based upon a separate dataset, which will allow for the evaluation of the model based on precision and recall.
- **Height Estimation**: Height values obtained from the calculation from segmented shadows will be compared to the AHN5-determined clustered point cloud heights.

These metrics will be evaluated using various versions of the input datasets. For example, simulated occlusions and lowered resolution of the point clouds can be used to see whether the method is robust enough in situations where only lower quality data is available. In a similar fashion, the metrics will also be assessed over different times of year using the winter and summer aerial imagery, to see whether or not the model is robust to spectral changes. Finally, the newly created physics-informed model will be compared to its pretrained counterpart to assess the impact of the fine-tuning procedure.

6 Time Planning

Date	Range	To-Do
22 Jan	22 Jan	Hand in Research Proposal
22 Jan	28 Jan	Prepare for P2 Presentation
29 Jan	29 Jan	P2 Presentation
29 Jan	5 Feb	Research CNN architectures
5 Feb	12 Feb	Data Collection & Processing
12 Feb	5 Mar	Shadow Simulation
5 Mar	2 Apr	CNN Training & Inference
2 Apr	16 Apr	Height Calculation & Experimental Setup
16 Apr	11 May	Finish Writing Draft
12 May	28 May	P4 Go / No Go
2 Jun	15 Jun	Implement Draft Feedback
16 Jun	4 July	Final Presentations

References

Adeline, K. R. M., Chen, M., Briottet, X., Pang, S. K., & Paparoditis, N. (2013). Shadow detection in very high spatial resolution aerial images: A comparative study. *ISPRS Journal of Photogrammetry and Remote Sensing*, *80*, 21–38. https://doi.org/10.1016/j.isprsjprs.2013.02.003

AHN. (2020a, February 4). *Dataroom* [AHN] [Publisher: AHN]. Retrieved January 22, 2025, from https://www.ahn.nl/dataroom

AHN. (2020b, February 18). *Kwaliteitsbeschrijving* [AHN] [Publisher: AHN]. Retrieved December 13, 2024, from https://www.ahn.nl/kwaliteitsbeschrijving

AlBalkhy, W., Karmaoui, D., Ducoulombier, L., Lafhaj, Z., & Linner, T. (2024). Digital twins in the built environment: Definition, applications, and challenges. *Automation in Construction*, *162*, 105368. https://doi.org/10.1016/j.autcon.2024.105368

Brenner, C. (2005). Building reconstruction from images and laser scanning. *International Journal of Applied Earth Observation and Geoinformation*, *6*(3), 187–198. https://doi.org/10.1016/j.jag.2004.10.006

Cheng, L., Gong, J., Chen, X., & Han, P. (2008). BUILDING BOUNDARY EXTRACTION FROM HIGH RESOLUTION IMAGERY AND LIDAR DATA.

Dorninger, P., & Pfeifer, N. (2008). A comprehensive automated 3d approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds [Number: 11 Publisher: Molecular Diversity Preservation International]. *Sensors*, *8*(11), 7323–7343. https://doi.org/10.3390/s8117323

Edson, C., & Wing, M. G. (2011). Airborne light detection and ranging (LiDAR) for individual tree stem location, height, and biomass measurements [Number: 11 Publisher: Molecular Diversity Preservation International]. *Remote Sensing*, *3*(11), 2494–2528. https://doi.org/10.3390/rs3112494

Fei, B., Yang, W., Chen, W.-M., Li, Z., Li, Y., Ma, T., Hu, X., & Ma, L. (2022). Comprehensive review of deep learning-based 3d point cloud completion processing and analysis [Conference Name: IEEE Transactions on Intelligent Transportation Systems]. *IEEE Transactions on Intelligent Transportation Systems*, *23*(12), 22862–22883. https://doi.org/10.1109/TITS.2022.3195555

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Granlund, G. H. (1978). In search of a general picture processing operator. *Computer Graphics and Image Processing*, *8*(2), 155–173. https://doi.org/10.1016/0146-664X(78)90047-3

Huang, Z., Yu, Y., Xu, J., Ni, F., & Le, X. (2020). PF-net: Point fractal network for 3d point cloud completion. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7659–7667. https://doi.org/10.1109/CVPR42600.2020.00768

Jiao, L., Huo, L., Hu, C., & Tang, P. (2020). Refined UNet: UNet-based refinement network for cloud and shadow precise segmentation [Number: 12 Publisher: Multidisciplinary Digital Publishing Institute]. *Remote Sensing*, *12*(12), 2001. https://doi.org/10.3390/rs12122001

Jin, X., Cai, S., Li, H., & Karniadakis, G. E. (2021). NSFnets (navier-stokes flow nets): Physics-informed neural networks for the incompressible navier-stokes equations. *Journal of Computational Physics*, *426*, 109951. https://doi.org/10.1016/j.jcp.2020.109951

Kadhim, N., & Mourshed, M. (2018). A shadow-overlapping algorithm for estimating building heights from VHR satellite images [Conference Name: IEEE Geoscience and Remote Sensing Letters]. *IEEE Geoscience and Remote Sensing Letters*, *15*(1), 8–12. https://doi.org/10.1109/LGRS.2017.2762424

Kim, M.-K., Wang, Q., & Li, H. (2019). Non-contact sensing based geometric quality assessment of buildings and civil structures: A review. *Automation in Construction*, *100*, 163–179. https://doi.org/10.1016/j.autcon.2019.01.002

Krähenbühl, P., & Koltun, V. (2011). Efficient inference in fully connected CRFs with gaussian edge potentials. *Advances in Neural Information Processing Systems*, *24*.

Kwak, E., & Habib, A. (2014). Automatic representation and reconstruction of DBM from LiDAR data using recursive minimum bounding rectangle. *ISPRS Journal of Photogrammetry and Remote Sensing*, *93*, 171–191. https://doi.org/10.1016/j.isprsjprs.2013.10.003

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition [Conference Name: Proceedings of the IEEE]. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, T., & Kim, T. (2013). Automatic building height extraction by volumetric shadow analysis of monoscopic imagery [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01431161.2013.796434 *International Journal of Remote Sensing*, *34*(16), 5834–5850. https://doi.org/10.1080/01431161.2013.796434

Lee, T., & Kim, T. (2010). GENERATION OF 3d BUILDING MODELS FROM COMMERCIAL IMAGE DATABASE THROUGH SAHDOW ANALYSIS. *San Diego*.

Liasis, G., & Stavrou, S. (2016). Satellite images analysis for shadow detection and building height estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *119*, 437–450. https://doi.org/10.1016/j.isprsjprs.2016.07.006

Luo, S., Li, H., & Shen, H. (2020). Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset. *ISPRS Journal of Photogrammetry and Remote Sensing*, *167*, 443–457. https://doi.org/10.1016/j.isprsjprs.2020.07.016

Ministerie van Onderwijs, C. e. W. (2018, October 29). *Actueel Hoogtebestand Nederland - Bronnen en kaarten - Rijksdienst voor het Cultureel Erfgoed* [Last Modified: 2019-07-01T14:26 Publisher: Ministerie van Onderwijs, Cultuur en Wetenschap]. Retrieved December 13, 2024, from https://www.cultureelerfgoed.nl/onderwerpen/bronnen-en-kaarten/overzicht/actueel-hoogtebestand-nederland

Nagao, M., Matsuyama, T., & Ikeda, Y. (1979). Region extraction and shape analysis in aerial photographs. *Computer Graphics and Image Processing*, *10*(3), 195–223. https://doi.org/10.1016/0146-664X(79)90001-7

Nan, L. (2023, February 15). *Introduction to machine learning**. Retrieved January 22, 2025, from https://3d.bk.tudelft.nl/courses/geo5017/handouts/01_notes_Introduction.pdf

Nex, F., & Rinaudo, F. (2011). LiDAR or photogrammetry? integration is the answer. *Italian Journal of Remote Sensing*, 107–121. https://doi.org/10.5721/ItJRS20114328

Otsu, N. (1979). A threshold selection method from gray-level histograms [Conference Name: IEEE Transactions on Systems, Man, and Cybernetics]. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Pratt, L. Y., Mostow, J., & Kamm, C. A. (1991). Direct transfer of learned information among neural networks. *Proceedings of the ninth National conference on Artificial intelligence - Volume 2*, 584–589.

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). PointNet++: Deep hierarchical feature learning on point sets in a metric space.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017, November 28). Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. https://doi.org/10.48550/arXiv.1711.10561

Richter, R., & Müller, A. (2005). De-shadowing of satellite/airborne imagery [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01431160500114664]. *International Journal of Remote Sensing*, *26*(15), 3137–3148. https://doi.org/10.1080/01431160500114664

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation [Series Title: Lecture Notes in Computer Science]. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2015* (pp. 234–241, Vol. 9351). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Shao, Y., Taff, G. N., & Walsh, S. J. (2011). Shadow detection and building-height estimation using IKONOS data [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01431161.2010.517226]. *International Journal of Remote Sensing*, *32*(22), 6929–6944. https://doi.org/10.1080/01431161.2010.517226

Shettigara, V. K., & Sumerling, G. M. (1998). Height determination of extended objects using shadows in SPOT images.

Tan, L., Lin, X., Niu, D., Wang, D., Yin, M., & Zhao, X. (2023). Projected generative adversarial network for point cloud completion [Conference Name: IEEE Transactions on Circuits and Systems for Video Technology]. *IEEE Transactions on Circuits and Systems for Video Technology*, *33*(2), 771–781. https://doi.org/10.1109/TCSVT.2022.3204771

Ufuktepe, D. K., Collins, J., Ufuktepe, E., Fraser, J., Krock, T., & Palaniappan, K. (2021). Learning-based shadow detection in aerial imagery using automatic training supervision from 3d point clouds [ISSN: 2473-9944]. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 3919–3928. https://doi.org/10.1109/ICCVW54120.2021.00439

Wang, R., Hu, Y., Wu, H., & Wang, J. (2016). Automatic extraction of building boundaries using aerial LiDAR data [Publisher: SPIE]. *Journal of Applied Remote Sensing*, *10*(1), 016022. https://doi.org/10.1117/1.JRS.10.016022

Wang, X., Ang, M. H., & Hee Lee, G. (2020). Point cloud completion by learning shape priors [ISSN: 2153-0866]. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10719–10726. https://doi.org/10.1109/IROS45743.2020.9340862

Yamazaki, F., Liu, W., & Takasaki, M. (2009). Characteristics of shadow and removal of its effects for remote sensing imagery [ISSN: 2153-7003]. *2009 IEEE International Geoscience and Remote Sensing Symposium*, *4*, IV–426–IV–429. https://doi.org/10.1109/IGARSS.2009.5417404

Yu, X., Rao, Y., Wang, Z., Liu, Z., Lu, J., & Zhou, J. (2021). PoinTr: Diverse point cloud completion with geometry-aware transformers. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 12478–12487. https://doi.org/10.1109/ICCV48922.2021.01227

Appendices

A Figures

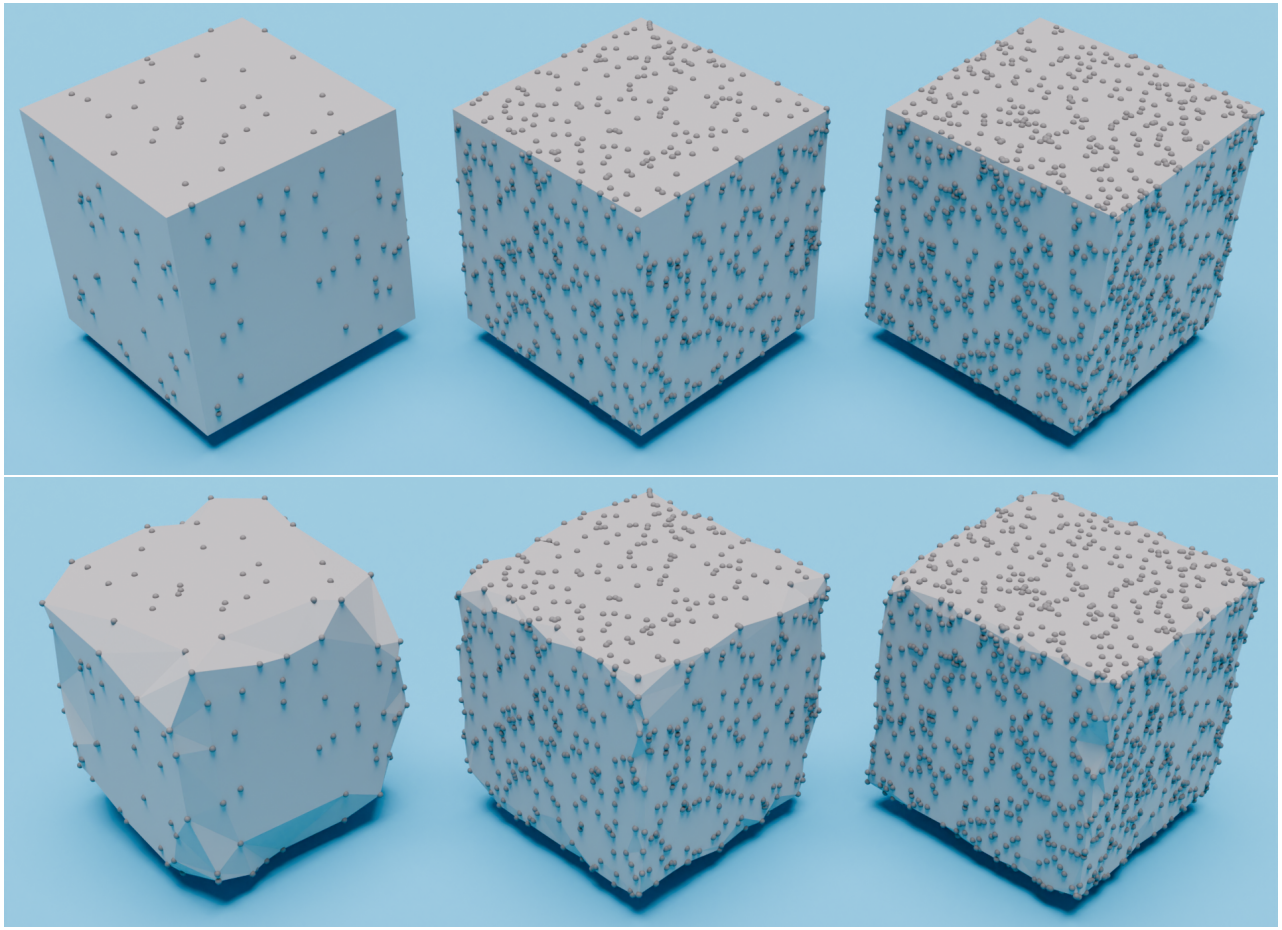

Figure 3: Illustration of point cloud density and its impact on surface reconstruction quality. The top row shows an initial cube with simulated point clouds of increasing density from left to right, while the bottom row demonstrates the corresponding reconstructed surfaces using Blender's "Nearest Vertex" shrink-wrapping method. Lower-density point clouds result in coarser surfaces and boundary representations, whereas higher densities result in an representation more accurate to the original.

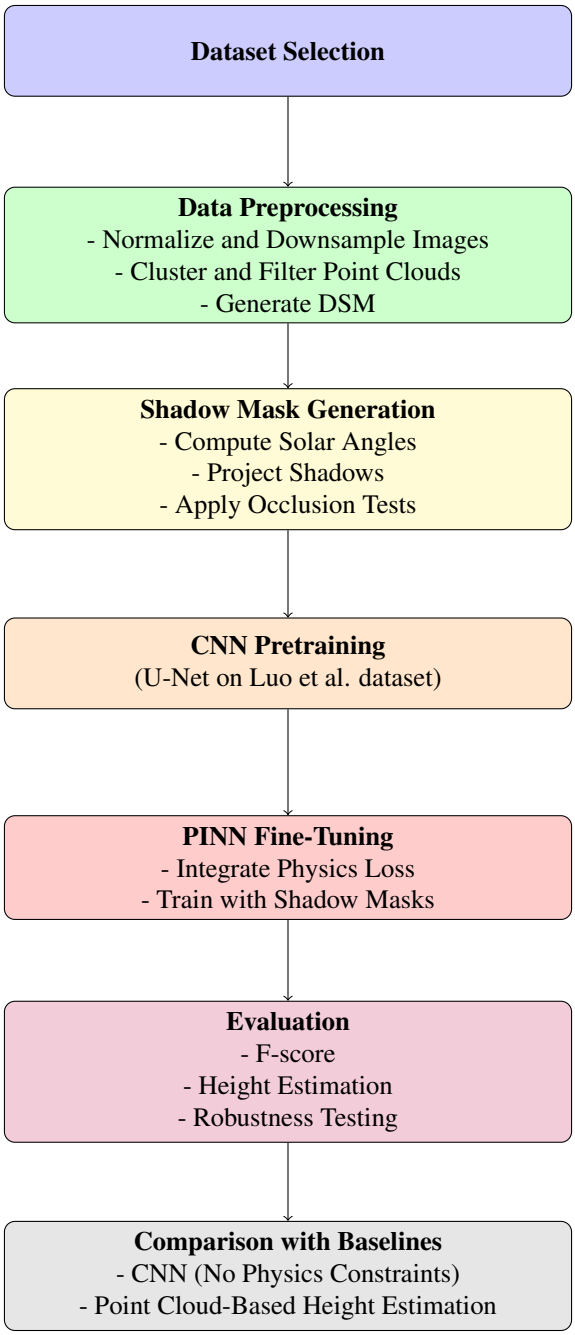

Figure 4: Flowchart indicating the various steps in the methodology.