remote sensing

MDPI

Article

Mapping Building Heights at Large Scales Using Sentinel-1 Radar Imagery and Nighttime Light Data

Mohammad Kakooei [1,*] and Yasser Baleghi [2]

[1] Department of Computer Science, Chalmers University of Technology, 41258 Gothenburg, Sweden
[2] Department of Electronic Engineering, Babol Noshirvani University of Technology, Babol 4714871167, Iran; y.baleghi@nit.ac.ir
* Correspondence: kakooei@chalmers.se

Abstract: Human settlement areas significantly impact the environment, leading to changes in both natural and built environments. Comprehensive information on human settlements, particularly in urban areas, is crucial for effective sustainable development planning. However, urban land use investigations are often limited to two-dimensional building footprint maps, neglecting the three-dimensional aspect of building structures. This paper addresses this issue to contribute to Sustainable Development Goal 11, which focuses on making human settlements inclusive, safe, and sustainable. In this study, Sentinel-1 data are used as the primary source to estimate building heights. One challenge addressed is the issue of multiple backscattering in Sentinel-1's signal, particularly in densely populated areas with high-rise buildings. To mitigate this, firstly, Sentinel-1 data from different directions, orbit paths, and polarizations are utilized. Combining ascending and descending orbits significantly improves estimation accuracy, and incorporating a higher number of paths provides additional information. However, Sentinel-1 data alone are not sufficiently rich at a global scale across different orbits and polarizations. Secondly, to enhance the accuracy further, Sentinel-1 data are corrected using nighttime light data as additional information, which shows promising results in addressing multiple backscattering issues. Finally, a deep learning model is trained to generate building height maps using these features, achieving a mean absolute error of around 2 m and a mean square error of approximately 13. The generalizability of this method is demonstrated in several cities with diverse built-up structures, including London, Berlin, and others. Finally, a building height map of Iran is generated and evaluated against surveyed buildings, showcasing its large-scale mapping capability.

Keywords: deep learning; radar multiple-backscattering correction; Sentinel-1 polarization; nighttime light; SDG 11

check for updates

Citation: Kakooei, M.; Baleghi, Y. Mapping Building Heights at Large Scales Using Sentinel-1 Radar Imagery and Nighttime Light Data. *Remote Sens.* **2024**, *16*, 3371. https://doi.org/10.3390/rs16183371

Academic Editor: Dino Ienco

Received: 25 June 2024
Revised: 7 September 2024
Accepted: 9 September 2024
Published: 11 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The growth of urbanization has brought about significant effects on the planet. Despite the small surface area that urban areas occupy on the earth, they contribute to a large portion of global energy consumption and carbon emission. According to [1], 80% of global energy consumption and 75% of carbon emissions can be attributed to urban areas. Thus, it is essential to obtain more detailed data from urban areas to better understand their impact on the environment. The 3D structure of urban areas, which encompasses the vertical dimension and building height in addition to the horizontal footprint, plays a crucial role in this regard. Studies have shown that the vertical structure of urban areas is a key factor in modeling the urban environment [2] and has a significant impact on various aspects such as weather, biophysics, economy, and social properties [3]. For instance, the density of urban patterns can increase greenhouse gas emissions and the heat island effect [4], while the 3D structure of urban areas can be used as an index to interpret energy consumption [5], greenhouse gas emissions [6], heat islands [7,8], and population distribution [9]. The

combination of both the horizontal and vertical information of buildings in a specific area is known as the occupation density [10]. Therefore, to better understand the impact of urban areas on the environment and achieve Sustainable Development Goals such as SDG 11, it is crucial to consider the 3D structure of urban areas in addition to their horizontal footprint.

Buildings are the key structures in human settlement areas. Over the years, vertical urban analysis and the investigation of vertical urban expansion have been carried out at various resolutions, ranging from local [11] to global scales [12]. These analyses have utilized a variety of data sources, including DSMs derived from Cartosat-1, Ikonos, WorldView-2, LIDAR, and others [11], as well as MODIS-based change detection methods [12]. However, because urban analysis is a multifaceted process, relying solely on two-dimensional vertical building footprints as an independent variable is not sufficiently informative. As a result, there has been a growing trend in using urban height maps for studying local areas [13] and in cities across the world [14,15]. For example, Kedron et al. [13] utilized LIDAR data to study pre- and post-Hurricane Katrina conditions, while Zhang et al. [15] employed Landsat series data to analyze both horizontal and vertical urban growth on a larger scale.

The level of detail (LOD) required for 3D human settlement (HS) information varies depending on the application. Remote sensing (RS) technology is commonly used to extract building information, but its effectiveness is limited by the remote sensor specifications, such as spatial resolution, spectral bands, and coverage area [16]. The CityGML standard outlines different LODs for building modeling, which are illustrated in Figure 1. LOD0, the lowest LOD, depicts the building footprint. LOD1 assumes the building is a cube with height information [17]. Subsequent LODs, starting from LOD2, provide more detailed information on the building structure and type. For example, LiDAR is deployed to generate higher LODs, but it can be expensive and time-consuming to obtain LiDAR data on a large scale [18].



Figure 1. Different levels of detail (LODs) of building models according to Groger et al. [19].

Human-made structures and buildings have various horizontal and vertical properties, and their patterns differ from one country to another [20]. For example, a recent study showed the urban development pattern is distributed in the USA, while it is high-rise and dense in Europe [21]. Therefore, it is hard, time-consuming, and expensive to extract reliable features to generate a fine-resolution height map on a large scale. The purpose of this study is finding robust features from RS data for building height map generation at a large scale, which has the potential to be applied on a country scale. So, the RS data are limited to non-commercial ones that are accessible in the public domain. Therefore, the target is generating the building height map at LOD1. It is worthwhile to mention that the LOD2 level and higher levels require VHR data, Biljecki et al. (2016) [22], which is out of this paper's scope. In Section 2, we review some studies in height map generation, mentioning their pros and cons. Section 3 includes the description of the study areas and the utilized dataset. Section 4 describes the methodology used to find the robust features and large-scale map generation. Section 5 includes the discussion and conclusion. Furthermore, some tables and extra analysis are included in Appendices A–D for interested readers.

The goal of this work is to utilize publicly available data to generate a level of detail 1 (LOD1) building height map that can be produced at a large scale. To achieve this, it is essential to use RS data that are globally available and offer sufficient spatial resolution to distinguish neighborhood areas. Sentinel-1 data have shown high potential for building height estimation, where each 10 m resolution pixel represents the average building height in a neighborhood, demonstrating a positive correlation between the backscattered signal and building height [23]. We analyze the potential of Sentinel-1 SAR data for building height estimation by exploring different polarizations (VV, VH, HH, HV), orbit paths, and directional paths (ascending and descending). However, Sentinel-1 data alone lack the necessary richness at a global scale in terms of polarization, orbit, and direction path to generate a comprehensive building height map. The contribution of this paper lies in fusing multi-source RS images to estimate building heights and generate a large-scale building height map. To achieve this, we combine Sentinel-1 data with the building footprint density and nighttime light data, and deploy an encoder–decoder deep model based on U-Net architecture [24].

2. Literature Review

Recent studies have examined the importance of the horizontal and vertical structure of settlement areas. Specially in urban areas, building height is an important piece of information that is hard to generate. Therefore, most previous works on producing building height maps are limited to local and regional scales [25]. Four main categories are found in RS-based building height estimation, which are,

1. Using adjacent shadow information in optical imagery [26]. But the accuracy of optical imagery decreases in dense areas with many overlapping shadows due to adjacent shadow information [27].
2. High-resolution stereo images [28] that require special imagery instruments.
3. LiDAR data that are known as the most accurate and reliable source. But their usage is limited to a local region as it is costly [29].
4. Radar data, especially Informatory SAR (InSAR), have a high potential in height mapping [30]. But, they have some issues such as noisy data, foreshortening, and shadowing problems [31].

In addition, data fusion has become a popular method to utilize the strengths of different RS platforms for estimating building heights. For example, Geiß et al. [32] proposed a hierarchical method to fuse TanDEM-X and Sentinel-2 data in order to extract 3D structure of buildings. There are also some works that estimated building volume by fusing LiDAR and radar data. Despite providing an accurate map, this fusion is not applicable in large-scale studies [25,33].

Furthermore, fusing SAR and optical data is informative for 3D object reconstruction in urban areas [34]. The main challenge in fusing optical and SAR data is finding the disparity map using a matching algorithm. For example, Hirschmuller [35] deployed the classic SGM algorithm to find the disparity map. In another work, Bagheri et al. [25] proposed a procedure including rational polynomial coefficients (RPCs) extraction for each image, block-based multi-sensor direction adjustment based on RPCs, and searching for matching points by assuming images are at the same plan to fuse SAR and optical data.

Deploying radar data for height mapping has been reported in a number of publications, including Rajpriya et al. [36], who proposed a fusion of Cartosat-1 and Ikonos DEMs to generate a 3D model of buildings at LOD1. In another work, Marconcini et al. [37] proposed a method to deploy TanDEM-X to estimate building height. Furthermore, there are some studies that deployed available DEMs (e.g., SRTM DEM) to reconstruct 3D building map, such as Misra et al. [38], which utilized different data sources including SRTM, ASTER, AW3S, and TanDEM-X, and compared their performance in building height estimation.

Height estimation on large scales has been studied in some recent studies. Li et al. [39] proposed an index by combining VV and VH polarization of Sentinel-1 backscattering signals to develop a building height model. As a result, they generated a 500 m resolution

building height map of seven USA cities including New York, Chicago, and Los Angeles. In another work, Li et al. [40] proposed random forest-based models and generated the first 3D building map at a continental scale (China, USA, and EU) with 1 km resolution. This map includes height and volume information of buildings in addition to a 2D vertical building footprint. Moreover, Frantz et al. [41] fused Sentinel-1 and Sentinel-2 data by support vector machine models to generate a building height map of Germany in 10 m resolution. This study shows the potential of Sentinel data in large-scale building height map generation in a fine spatial resolution. Although these works show the potential of RS data for large-scale building height estimation, their proposed methods are not powerful enough to generate accurate maps.

In summary, this approach highlights the importance of integrating multiple data sources for accurate global building height mapping while maintaining a limited number of input features and a lightweight deep model for simplicity and applicability. Although Sentinel-1 data are effective for building height estimation, their limitations in polarization, orbit, and direction hinder comprehensive global coverage. To overcome these challenges, we combined Sentinel-1 with additional data sources such as building footprint density and nighttime light data, enhancing the accuracy and scalability of building height map generation.

3. Dataset and Study Area

The aim is deploying free RS data with appropriate spatial resolution to estimate building heights. There are hundreds of RS satellites in service today, but only a few satisfy these requirements, such as the Sentinel-1 satellite in the European Copernicus program [42], which includes Sentinel series satellites, which has an attractive potential for estimating building heights (e.g., Sentinel-1 SAR data). The Sentinel-1 mission consists of two satellites, Sentinel-1 A and Sentinel-1 B, orbiting on the same path with a six-day revisiting time. Data from Sentinel-1 are acquired at band C at 10 m spatial resolution regardless of weather conditions or day/night times. With Sentinel-1, the mostly used image acquisition mode is IW (Interferometric Wide Swath), covering a width of 250 km. There are also other modes such as EW (Extra Wide Swath), and SM (Strip Map). Traditionally, backscattered data have been provided in two polarizations, VV and VH, but, lately, HH and HV polarizations have become available in many places [43]. However, Sentinel-1 data are noisy due to the backscattering feature of signals in urban areas, and the combination of single bounce and either double bounce or triple bounce [44]. Furthermore, the relationship between the backscattering signal and building height can be affected by the building direction and its material [45].

3.1. Sentinel-1 as Independent Variable

The main independent variables are Sentinel-1 data, for which a one-year pixel-based median value of 2020 is calculated. The data are accessed through the Google Earth Engine (GEE) platform [46], where Sentinel-1 data are preprocessed through the following steps,

1. Removing thermal noise;
2. Radiometric calibration;
3. Geometric correction using SRTM 30 data [47].

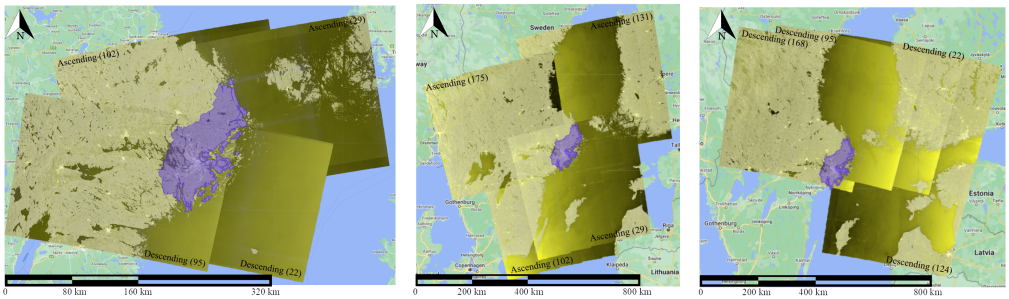
Data are also separated according to different orbits and paths, and their orbit direction including ascending or descending. Despite Koppel et al.'s [23] recommendation to combine different directions, our investigation shows this separation can contribute to a higher accuracy of height estimation.

Table 1 shows the coverage of Sentinel-1 in IW and EW modes over Stockholm, Sweden. According to this table in the IW mode, path numbers 29 and 102 include 45 and 47 images, respectively. Further, there are two paths with 27 and 30 images in the descending direction. Additionally, each ascending or descending direction in the EW mode includes four orbit paths, each of which has 12 to 30 images.

Table 1. Number of available images in different paths in Stockholm, Sweden, 2020.

Mode and Bands	Orbit Direction	Path Number (Number of Images)
IW—(VV, VH)	Ascending Descending	29 (45), 102 (47) 22 (46), 95 (44)
EW—(HH, HV)	Ascending Descending	29 (14), 102 (14), 131 (12), 175 (15) 22 (30), 95 (14), 124 (23), 168 (15)

Figure 2 shows how Sentinel-1 data cover Stockholm, Sweden in 2020. As aforementioned in Table 1, in IW mode, there are two paths in the ascending direction and two paths in the descending direction. Furthermore, in the EW mode there are four paths in the ascending direction and four paths in the descending direction.



(**a**) (**b**) (**c**)

Figure 2. Different Sentinel-1 coverage modes over Stockholm, with the city's boundary overlaid on the raster images. The path number is indicated on each median image. (**a**) Sentinel-1 path orbits in IW mode, displayed in false color (VV, VV, VH). (**b**) Sentinel-1 path orbits in EW mode in the ascending direction, displayed in false color (HH, HH, HV). (**c**) Sentinel-1 path orbits in EW mode in the descending direction, displayed in false color (HH, HH, HV).

3.2. Nighttime Light as Independent Variable

Monthly VIIRS nighttime light data in the infrared band are used in this study. These data are globally available and, if there is no cloud coverage problem, these data are available each month. As a pre-processing step, VIIRS nighttime light is corrected by Mills et al.'s method [48] to reduce errors and the light saturation problem. Figure 3 shows the VIIRS nighttime light data mean value in 2020 over Stockholm, Sweden.

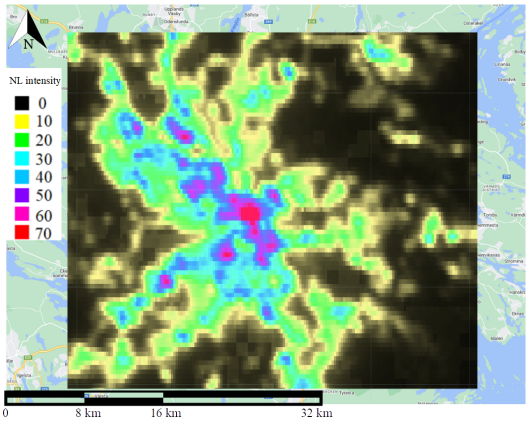


Figure 3. Nighttime light data for Stockholm, Sweden (2020) display varying levels of light intensity. The color palette indicates the intensity values of nighttime illumination.

3.3. Copernicus Land Monitoring Service as Dependent Variable

To train and validate the proposed method, we use data sources from European capitals' building height map, provided by Copernicus through the land monitoring service

program [49]. These data are in 10 m spatial resolution and are produced using a stereo images IRS-P5 and datasets including DSM, DEM, and normalized DSM. This dataset just includes building height, and other things (such as trees) are not included. This dataset is available for 38 cities including Tirana (Albania), Vienna (Austria), Brussels (Belgium), Sarajevo (Bosnia and Herzegovina), Sofia (Bulgaria), Zagreb (Croatia), Nicosia (Cyprus), Prague (Czech Republic), Copenhagen (Denmark), Tallinn (Estonia), Helsinki (Finland), Paris (France), Berlin (Germany), Athens (Greece), Budapest (Hungary), Reykjavík (Iceland), Dublin (Ireland), Rome (Italy), Pristina (Kosovo), Riga (Latvia), Luxembourg City (Luxembourg), Valletta (Malta), Podgorica (Montenegro), Amsterdam (Netherlands), Skopje (North Macedonia), Oslo (Norway), Warsaw (Poland), Lisbon (Portugal), Bucharest (Romania), Belgrade (Serbia), Bratislava (Slovakia), Ljubljana (Slovenia), Madrid (Spain), Stockholm (Sweden), Bern (Switzerland), Ankara (Turkey), and London (United Kingdom).

This dataset is a proper source for building height modeling because these data are combined with an urban area map and just include the building height; other vertical objects are removed. These data are at LOD1, in which each building is supposed to be a cube without any rooftop structure. Figure 4 shows this dataset for Rome, Paris, and Stockholm, and their corresponding ground truth high-resolution map is illustrated.

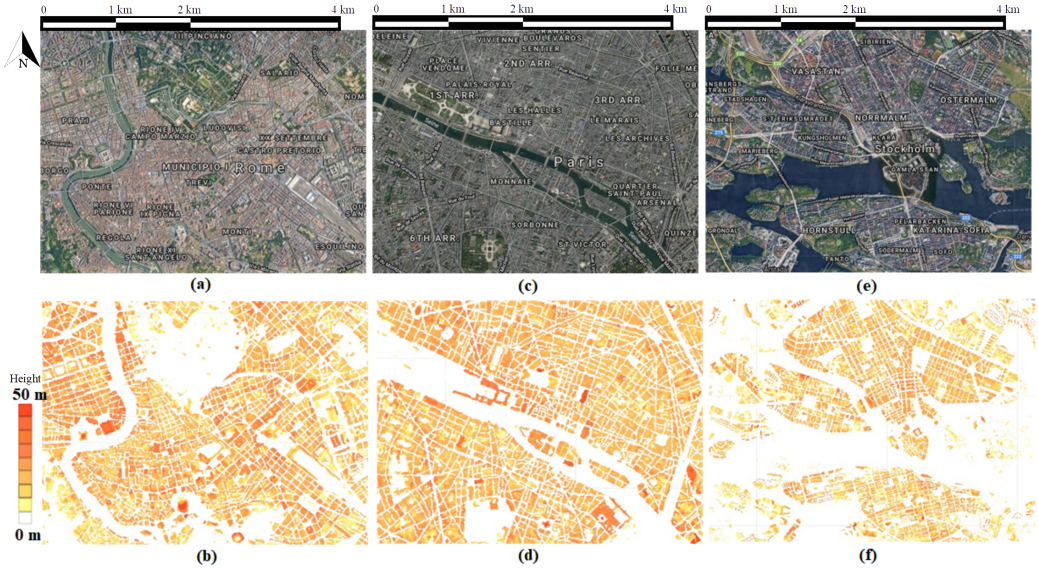


Figure 4. Building height map from Copernicus land monitoring service program [49]. The high-resolution optical images are at the top and the height maps are in the second row. (**a**,**b**) Rome, Italy. (**c**,**d**) Paris, France. (**e**,**f**) Stockholm, Sweden.

3.4. Building Height Surveying in North of Iran as Test Data

The goal of this surveying is providing a new test dataset to evaluate the deep models in a different region. This survey includes about 2000 buildings in two cities in the north of Iran (1000 samples per city). The free Android software CartoDruid version 0.60.4 is used to register the building height information gathered through surveying. This software enables the user to define the region of interest by point or polygon. To reduce the surveying geolocation error, surveys are defined at the center of building rooftops.

3.5. Global Human Settlement Layer as Additional Data

Similar to the work that was carried out by [39,40], the analysis is limited to the human settlement areas provided by the JRC Global Human Settlement Layer (GHSL), which was produced in 2020. This map is used to mask non-human settlement areas and to generate the density of built-up areas in a neighborhood as an independent variable.

4. Methodology

The methodology in Figure 5 is divided into two phases: first, shallow models are used to identify how different features are informative for building height estimation, followed by a deep learning model for generating building height. Since deep models cannot be interpreted as easily as shallow models, shallow models are used during the first phase to test the building height estimation accuracy of each configuration.

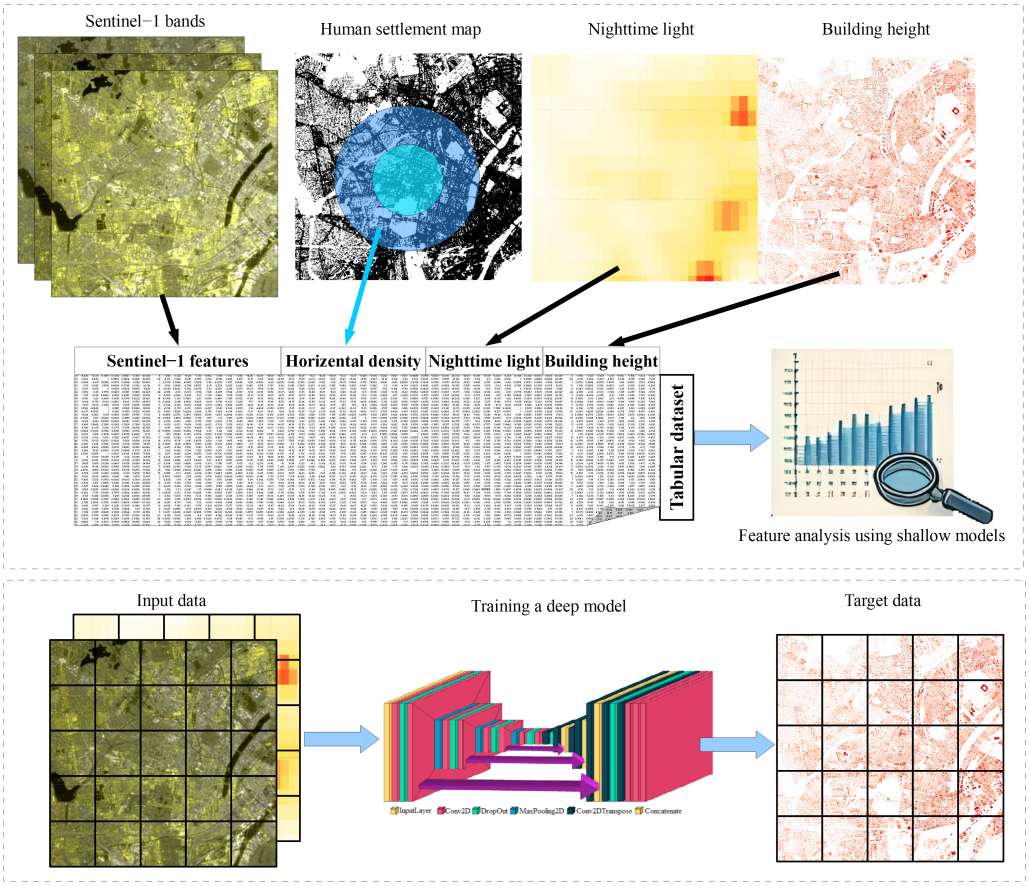


Figure 5. Flowchart of the methodology used in this study. The upper part shows the first phase that studies how different features are informative for the building height estimation. The lower part shows the second phase, which a deep model trained to generate a building height map.

4.1. Feature Analysis Using Shallow Models

During the first phase, the following questions are answered:

1. How much information is there in Sentinel-1's different paths, orbits, and directions for building height estimation?
2. How can the fusion of Sentinel-1 data with other RS data provide higher estimation accuracy?

From another point of view, shallow regression algorithms are used in the first phase to reduce the model uncertainty as much as possible and focus on finding how different configuration of features are informative for building height estimation by comparing different scenarios. In this regard, Ridge Regression (RR), Support Vector Regression with a Linear kernel (SVRL), Multi-Layer Perceptron Neural Networks (NNs), Gradient Boosting (GB), Random Forest (RF) with 100 tree estimators, and Voting (VOT) are deployed as regressors of this phase. Furthermore, the input features to these models are,

- VV_{asc}, VH_{asc}, HH_{asc}, and HV_{asc} are Sentinel-1 VV, VH, HH, and HV bands, respectively, in the ascending direction.

- VV_{desc}, VH_{desc}, HH_{desc}, and HV_{desc} are Sentinel-1 VV, VH, HH, and HV bands, respectively, in the descending direction.
- UM_{50}, UM_{100}, UM_{150}, and UM_{200} are the densities of the vertical building map in circles with radii of 50, 100, 150, and 200 m, respectively.
- NL reveals the nighttime light dataset.

These are defined in various combinations that could be informative in this problem. Then, 9 combinations of these features are listed in Table 2, along with their acronyms. The first four features are dedicated to VV and VH bands of Sentinel-1 in ascending and descending directions. Although Sentinel-1 data have HH and HV polarizations, they are not included in the X1 to X4 feature sets because they are available for limited locations. Furthermore, feature set X5 is similar to X4 but is separated for each path and polarizations. The densities of the vertical building map in circles with radii of 50, 100, 150 and 200 m are included as an input feature in feature sets X6 and X7. These features can be informative to correct double bouncing and triple bouncing, which happen in urban areas where multiple high-rise buildings exist in a neighborhood. In feature sets X8 and X9, the VIIRS nighttime light data are used as additional data that may also be informative to correct the Sentinel-1 data in urban areas.

Table 2. Various combinations of input feature sets are tested with shallow models to assess the informativeness of each feature for building height estimation.

Feature Set	Input Features
X1	VV_{asc}
X2	$VV_{asc} + VH_{asc}$
X3	$VV_{asc} + VV_{desc}$
X4	X3 + $VH_{asc} + VH_{desc}$
X5	X4 + $HH_{asc} + HH_{desc} + HV_{asc} + HV_{desc}$
X6	X5 + UM_{50}
X7	X6 + $UM_{100} + UM_{150} + UM_{200}$
X8	X5 + NL
X9	X7 + NL

Eventually, different combinations of input features are deployed along with different regression methods to find out how much information these features have, such as

- Feature X1 includes just the VV band in the ascending direction, which is used as the base feature for comparison.
- Feature set X2 includes VV and VH polarization, which makes it possible to investigate the importance of various polarizations.
- Feature set X3 includes VV polarization in both ascending and descending orders, which makes it possible to investigate the importance of deploying Sentinel-1 data in different directions.
- Feature set X4 includes VV and VH directions in both ascending and descending orders. This feature set is compared with the X3 feature set to investigate the importance of deploying different polarizations.
- Feature set X5 includes all available Sentinel-1 data within the study area. For example, if in that study area Sentinel-1 data with different paths or polarizations (e.g., HH and HV) are available, they all are included to investigate their impact. This set is the richest set of Sentinel-1 dataset for further investigations.
- Feature set X6 includes all features in X5 and the vertical density building map in a circle with radius 50 m. Therefore, it can be investigated if Sentinel-1 data can be corrected using the vertical building map.
- Feature set X7 includes all features in X6 and the density of the vertical building map in circles with radii of 100, 150 and 200 m. Therefore, more features are included from the vertical building map to investigate its effect in correcting Sentinel-1 backscattering signals.

- Feature set X8 includes all features within X5 and nighttime light data. Both nighttime light and Sentinel-1 data have a high intensity in dense built-up areas. For nighttime light data, this is due to light saturation, and for Sentinel-1 data it is due to double and triple bouncing. So, it is expected to achieve a higher estimation accuracy by combining these two sources.
- Feature set X9 includes all aforementioned features. It means this feature includes all features within X7 and nighttime light data.

4.2. Building Height Map Using Deep Learning

In the second phase, the selected features are used to train a deep model to generate the building height map. Although it will be shown all aforementioned features are informative in the building height estimation problem, the five selected features are VV and VH polarizations of Sentinel-1 data in ascending and descending directions, and the nighttime light data as the additional data to refine the Sentinel-1 backscattering signals. Selecting this feature set is to keep the generalizability of the model to be applicable in different areas, specially in low- and middle-income countries, whose architecture is shown in Figure 6 and is detailed in Appendix D Table A4. It is worth mentioning that it is a lightweight architecture with 409,857 parameters.

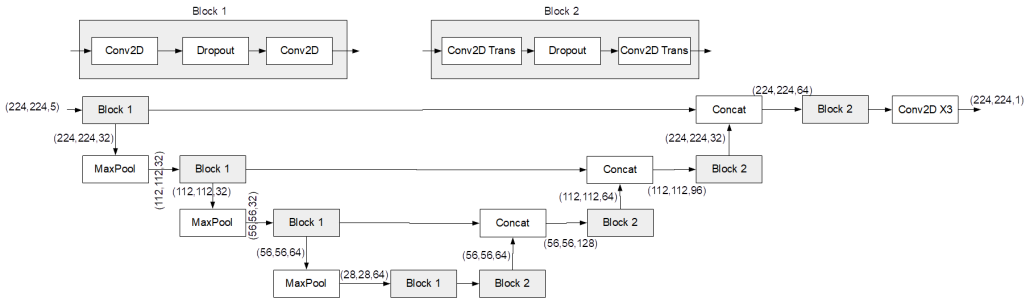


Figure 6. U-Net-based deep model for building height estimation consisting of two main blocks. Block 1 includes two convolutional layers with a dropout layer to prevent overfitting. Block 2 comprises two transposed convolutional layers with an additional dropout layer. MaxPooling is used to reduce the tensor dimensions, while Block 2 restores the dimensions. Residual connections between the encoder and decoder parts are implemented to maintain spatial resolution throughout the network.

To create the building height map, we employ a deep learning model based on the U-Net architecture [24]. U-Net is a convolutional neural network designed specifically for image segmentation tasks. It takes an input image and produces an output of the same size, typically transforming a raw image into a segmented version where each segment corresponds to a specific category or value. The architecture is called "U-Net" because of its U-shaped structure, which includes two main paths: a contracting path and an expanding path. As shown in Figure 6, this design enables the network to capture both detailed local features and broader contextual information within the image, making it particularly effective for segmenting complex images [50]. The contracting path uses convolutional layers combined with pooling to identify and condense key features, progressively reducing the image size. The expanding path then upsamples the condensed image while merging it with high-resolution features from earlier layers, effectively restoring spatial details and improving segmentation accuracy.

The selection of a lightweight deep model builds on our findings from feature analysis using shallow models, where we observed a strong correlation between the selected feature set and building height, even with different shallow models. This indicates that extracting features does not necessitate a complex deep model. In fact, using more complex models could exacerbate overfitting and reduce reliability. Additionally, lightweight deep models are particularly advantageous in developing countries and regions in the Global South,

where data scarcity and infrastructure limitations are common. These models are better suited to operate efficiently under such constraints.

To train our model, we utilized the mean square error (MSE) loss function, a popular and reliable choice for regression problems. The selection of an optimizer significantly affects model performance; for this study, we chose the Adam optimizer [51]. Unlike Stochastic Gradient Descent (SGD) [52], Adam uses a momentum-based approach that helps to smooth out the training curve by minimizing oscillations and enhancing stability. Moreover, Adam is easy to implement and generally less sensitive to hyperparameters, facilitating the achievement of optimal results. We set a learning rate of 0.001 and used a batch size of 32. Over the course of 100 epochs, we tracked the validation MSE after each epoch, ultimately selecting the model with the lowest validation MSE as the final one.

Wan et al. [53] investigated the effects of model size on hyperparameter sensitivity by proposing a lightweight deep model. Their findings suggest that smaller models tend to be less affected by changes in hyperparameters. This reduced sensitivity arises from a simpler loss landscape and a lower number of parameters, which minimizes the likelihood of problems such as vanishing or exploding gradients and overfitting. As a result, lightweight models generally require less extensive hyperparameter tuning and sensitivity analysis than their larger counterparts. The stability and reduced complexity of these models also make the training process more straightforward and less susceptible to the instabilities often encountered with more complex networks.

4.3. Accuracy Assessment

In the first phase of our analysis, we generate data samples for each city available in our dataset. Each city's dataset is treated as independent from the others, allowing us to analyze the performance metrics on a per-city basis. This approach ensures that our results are consistent across various cities and provides a deeper understanding of informative features.

For each dataset, a five-fold cross-validation process is employed. This involves training the shadow model on four of the five folds and testing it on the remaining fold, with this process being repeated until each fold has served as the test set once. The average evaluation metrics, including the R^2 score and mean absolute error (MAE), are then calculated. The standard deviation across the five folds is also computed to assess the stability of the model's performance.

In the next section, we visualize these results using bar plots to facilitate comparison and discussion. Detailed information on the average R^2 scores, MAE, and their standard deviations from the five-fold cross-validation are available in the Appendices A–C.

The R^2 score in Equation (1) is a key evaluation metric, which provides a measure of how well the predicted values approximate the actual values, with a score of 1 indicating perfect predictions, defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

where y_i is the actual building height, $\hat{y}_i$ is the predicted building height, $\bar{y}$ is the mean of the actual building height, and n is the number of predictions.

We also use the MAE in Equation (2) as a secondary evaluation metric to gain insight into the accuracy of the predictions in terms of absolute meters. The MAE provides a direct interpretation of the error magnitude in the same units as the output variable, making it particularly useful for understanding how far off the predictions are in practical terms. The MAE is defined as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

where n is the number of predictions, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

For evaluating the building height map, in addition to the MAE, we specifically use the mean square error (MSE) in Equation (3), which is calculated as:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

This formula measures the average squared difference between the building height ground truth and the generated height map, providing a robust assessment of the model's accuracy in generating the building height map. By combining these metrics, we can thoroughly evaluate the performance and reliability of the deep model in mapping building heights.

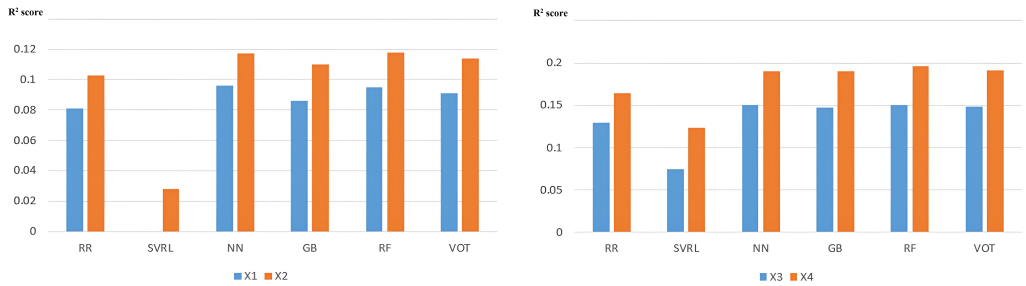
5. Discussion and Results

Based on different feature sets and shallow models, different scenarios were defined in the previous section and evaluated accordingly. For this section, only some evaluation is provided, while the full results on both training and test data can be found in Appendices A–C. First, the importance of combining different polarizations is discussed. Second, the importance of separating ascending and descending directions are examined, which is followed by an examination of deploying data from different orbits. Additionally, the performance of combining Sentinel-1 data with vertical building density and nighttime light data are evaluated.

5.1. Combination of Polarizations

Investigating the importance of combining different polarizations of Sentinel-1 data (VV and VH) in a definite orbit path is based on comparing X1 with X2, and X3 with X4. Feature set X1 just includes VV polarization in the ascending direction and is compared with feature set X2 that includes both VV and VH polarizations in the ascending direction. Similarly, X3 includes VV polarization in both ascending and descending directions and is compared with X4 that includes VV and VH polarizations in ascending and descending directions.

Figure 7 illustrates comparison between X1 and X2, and also between X3 and X4 in test data points of Rome, Italy. The plots verify that the combination of the two polarizations VV and VH in all regression methods results in a higher R^2 coefficient of determination. Therefore, it is necessary to combine all available polarizations of Sentinel-1 for building height estimation.



(**a**) (**b**)

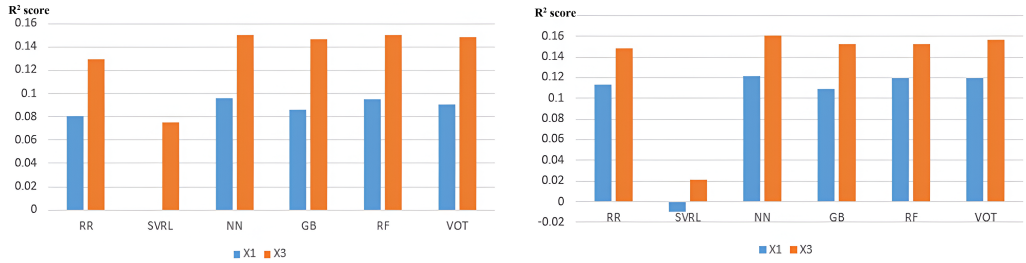
Figure 7. Examining the impact of using combined polarizations at test points in Rome, Italy. The vertical axis represents the R^2 value, while the horizontal axis shows the different models: (**a**) comparing X1 with X2, and (**b**) comparing X3 with X4. The results indicate that integrating both VV and VH polarizations across all regression methods yields a higher R^2 coefficient of determination.

5.2. Combination of Ascending and Descending Paths

The goal of this subsection is investigating the importance of combining ascending and descending directions. As aforementioned, in contrast to Koppel et al. [23] who merged ascending and descending images to achieve the maximum spatial coverage, here each ascending and descending directions of the Sentinel-1 data are deployed as separate

independent variables. In this regard, feature sets X1 and X3 are compared, in which X1 includes VV polarization in the ascending direction and X3 includes VV polarization in both ascending and descending directions.

Figure 8 shows the comparison between X1 and X2 in Rome, Italy, and Paris, France. As a result, the combination of using ascending and descending directions increases the building height estimation accuracy. This also matches with the inherent specification of radar data in which the backscattering signal is dependent on different factors, and thus combining the data samples that are gathered from different backscattering angles increases the building height estimation accuracy.

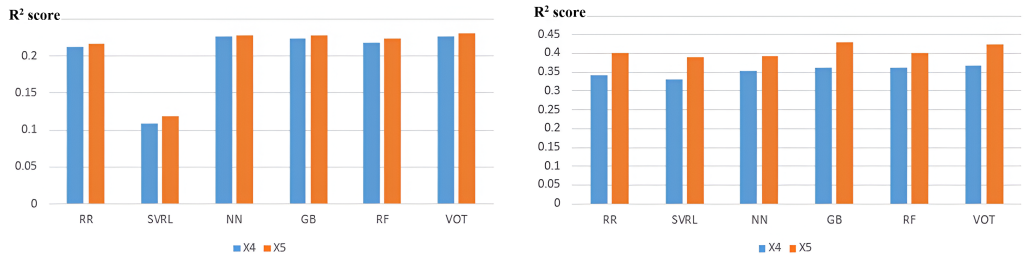


(**a**) (**b**)

Figure 8. Assessing the significance of merging different directional paths by comparing X1 and X3 at test points in (**a**) Rome and (**b**) Paris. The vertical axis represents the R^2 value, while the horizontal axis displays the different models. The findings show that incorporating both ascending and descending directions improves the accuracy of building height estimation.

5.3. Combination of Different Orbit Paths

To show combining different orbit paths is informative for building high estimation, feature sets X4 and X5 are compared in Paris and the Stockholm in Figure 9. Figure 9a shows X4- and X5-based height estimation in Paris, where Sentinel-1 data has one ascending and two descending directions. It shows a little increase in the estimation accuracy, but it is not significant. Figure 9b shows a significant increase in the coefficient of determination in Stockholm, which is rich in having different orbit paths (12 orbit paths according to Table 1). In conclusion, using different direction paths increases the building height estimation accuracy, and the more paths there are available, the higher is the estimation accuracy.



(**a**) (**b**)

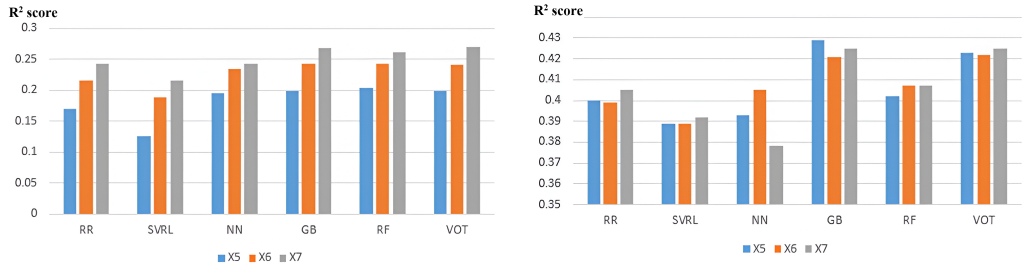
Figure 9. Examining the effect of combining different orbit paths by comparing X4 and X5 at test points in (**a**) Paris and (**b**) Stockholm. The vertical axis denotes the R^2 value, while the horizontal axis represents the different models. The results suggest that incorporating multiple directional paths enhances the accuracy of building height estimation, with greater accuracy achieved as more paths are included.

5.4. Built-Up Density as Addition Data

The goal of this section is investigating the importance of the vertical building density in refining Sentinel-1 backscattering signals. In this regard, feature sets X6 and X7 are compared with feature set X5, which includes Sentinel-1 data in all available paths and directions. Feature set X6 includes all features that are included in X5 and the vertical

building density in a circle with radius 50 m. Feature set X7 is similar to feature set X6 but includes some variations in the radius of the circle of built-up density in which the radius varies by 50, 100, 150 and 200 m.

Figure 10 shows the comparison between feature sets X6 and X7 with X5 in test points in Rome and Stockholm. The built-up density is very informative in Rome, but it does not show a meaningful trend in Stockholm. It is concluded due to a rich availability of Sentinel-1 orbit paths in Stockholm that they themselves can correct the Sentinel-1 backscattering signal for building height estimation, and thus it does not require additional data such as building density for the refinement step. On the other hand, Sentinel-1 data are not rich enough in Rome to refine themselves and thus the vertical building density is much more informative for building height estimation.

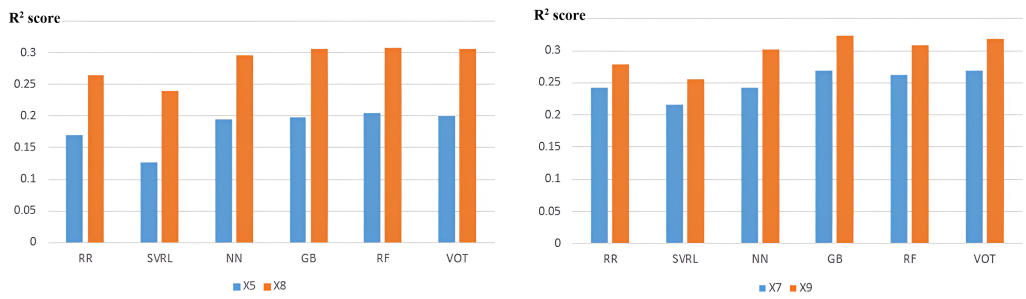


(**a**) (**b**)

Figure 10. Analyzing the impact of built-up density as additional data by comparing X5, X6, and X7 at test points in (**a**) Rome and (**b**) Stockholm. The vertical axis shows R^2, while the horizontal axis represents different models. The built-up density is highly informative in Rome but less so in Stockholm, where the abundant Sentinel-1 orbit paths already provide sufficient data for accurate building height estimation. In contrast, Rome's less comprehensive Sentinel-1 coverage benefits significantly from the inclusion of built-up density data.

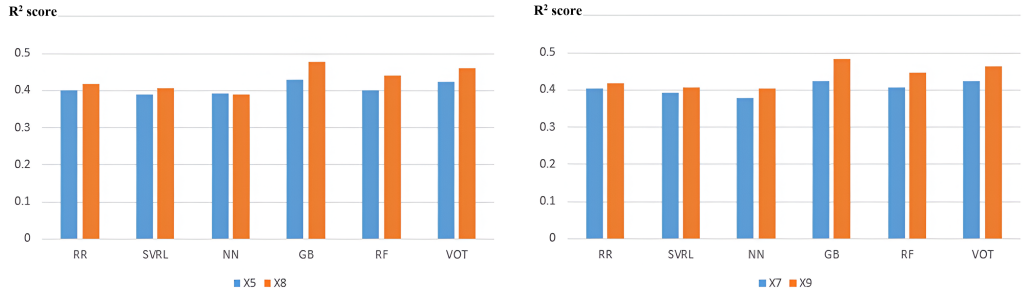
5.5. Nighttime Light as Addition Data

This section investigates the potential of nighttime light data as an additional source to refine Sentinel-1 backscattering signals. To this end, feature set X8 is compared with X5, and feature set X9 is compared with X7. Feature set X5 contains all orbit paths of Sentinel-1 data, while X8 includes both the nighttime light data and the features from X5. Similarly, X7 consists of all available Sentinel-1 paths combined with the vertical building density at various radii, while X9 integrates nighttime light data with the features of X7. As observed with the built-up density in Section 5.4, the inclusion of nighttime light data significantly improves the results in Rome (Figure 11), but shows only a slight enhancement in Stockholm (Figure 12).



(**a**) (**b**)

Figure 11. Analyzing the impact of nighttime light data as an additional feature by comparing ((**a**) X5 with X8 and ((**b**) X7 with X9 at test points in Rome. The vertical axis represents R^2, while the horizontal axis indicates different models. Nighttime light data prove to be highly valuable in Rome, where the limited Sentinel-1 coverage benefits greatly from their inclusion.



(**a**) (**b**)

Figure 12. Examining the role of nighttime light data as an additional feature by comparing (**a**) X5 with X8, and (**b**) X7 with X9 at test points in Stockholm. The vertical axis denotes R^2 values, while the horizontal axis displays various models. In Stockholm, where the extensive coverage from Sentinel-1 already delivers sufficient data, the additional contribution of nighttime light data is minimal.

5.6. Evaluating Building Height Map

Here the building height estimation using the deployed deep model is evaluated. To this end, some visual investigations are provided and estimation accuracies are compared in London (England), Berlin (Germany), and Babol and Amirkola (Iran), according to mean absolute error (MAE) and mean square error (MSE) metrics. Furthermore, the London (England) height map is compared with Wang et al.'s [54] proposed map in which combined Landsat and DSM data generate a building height map of England. Moreover, the Berlin (Germany) map is compared with Frantz et al.'s [41] proposed map in which combined time serious Sentinel-1 and Sentinel-2 data generate a building height map of Germany in 10 m resolution. Finally, the estimation accuracy of the proposed map is evaluated in Babol and Amirkola (Iran) using the surveyed points.

Figure 13 Shows different building map areas in London. This figure shows an overall view in the first column and two zoomed areas in the second and third columns. The first row shows the high-resolution optical image. The second row shows the ground truth map that is taken from the EU Copernicus program. The building height that is generated by Wang et al. [54] is shown in the 3rd row, and the last row shows the generated map. This figure verifies that the generated building map has a high resolution in 10 m in which the building blocks are detectable. It is worthwhile to mention that the visual comparison verifies the higher resolution of our generated map. It is very important to consider that using low-resolution nighttime light data does not ruin the final product spatial resolution while it increases the estimation accuracy by refining the Sentinel-1 backscattering values.

Table 3 shows the estimation accuracy comparison of the proposed method in different locations. Compared with Wang et al.'s [54] method, the proposed method is more accurate and provides a higher spatial resolution map. Compared with Frantz et al.'s [41] method, the proposed method relies on the nighttime light data instead of multispectral Sentinel-2 data. However, the nighttime light data with a coarse spatial resolution are not informative themselves for building height estimation and are used as additional data to correct the Sentinel-1 signal. So, the proposed method does not achieving a much higher estimation accuracy than Frantz et al.'s [41], but achieves a little higher estimation accuracy with fewer data. Furthermore, investigating the height map in Babol and Amirkola (Iran) is of great important and verifies the generalizability of the proposed method, as the reported MAE and MSE show similar values in London and Berlin.

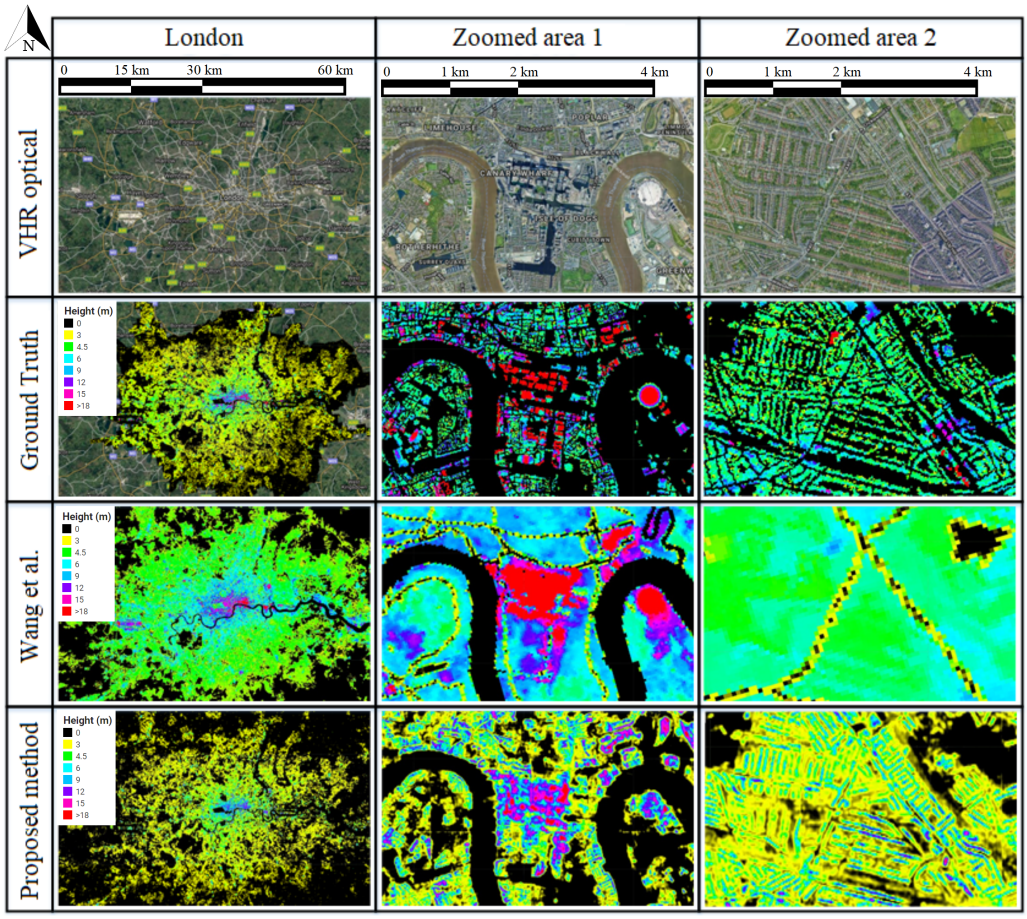


Figure 13. Building height maps in London. The first row shows a high-resolution optical image of the study areas. The second row shows the ground truth. Wang et al.'s [54] generated maps are shown in the third row, following by the generated map in the last row. The first column shows the large-scale image, and the next two columns show two zoomed areas for visual investigation.

Table 3. Comparing the generated height maps in London, Berlin, Babol, and Amirkola.

Location	Method	Dataset	Mean Absolute Error	Mean Square Error
London (England)	Wang et al. [54]	Landsat, DSM	4.22	31.36
	Proposed	Sentinel-1, Nightlight	2.32	13.02
Berlin (Germany)	Frantz et al. [41]	Sentinel-1, Sentinel-2	2.76	29.89
	Proposed	Sentinel-1, Nightlight	2.38	15.74
Babol (Iran)	Proposed	Sentinel-1, Nightlight	2.63	12.24
Amirkola (Iran)	Proposed	Sentinel-1, Nightlight	2.3	10.25

5.7. Scalability

The scalability and applicability of this method for large-scale building height mapping can be discussed from several perspectives:

1. Input data availability: the method primarily relies on the VV and VH bands of Sentinel-1 data in both ascending and descending orders, along with nighttime light

data. These datasets are widely available on a global scale, making it feasible to generate large-scale building height maps across various regions.

2. Efficient use of Sentinel-1 data: although we initially analyzed multiple paths, orbits, and directions of Sentinel-1 data to determine their usefulness in generating building height maps, the final model utilizes only the ascending and descending directions. By focusing on these two directions, we significantly reduce the computational load and the variety of data needed for height map generation. Additionally, using both directions helps mitigate issues related to shadowing and foreshortening [55], which are common in mountainous areas. However, due to the limited representation of mountainous regions in the dataset, caution is advised when applying this approach in such areas.
3. Prioritizing nighttime light data: the decision to prioritize nighttime light data over human settlement density maps is strategic. Settlement maps are often derived from satellite data and may have varying levels of accuracy, particularly in developing countries and regions in the Global South, where these maps tend to be less reliable. Using nighttime light data as a refinement tool for Sentinel-1 enhances the accuracy of building height estimation in these areas.
4. Model efficiency: the deep model selected for this task is intentionally lightweight to ensure it can be run on moderate hardware. This consideration enhances the model's applicability in resource-limited environments. The rationale behind the choice of model architecture is further detailed in the methodology section.

6. Conclusions

The focus of this study is to investigate the potential of Sentinel-1 data in combination with other data sources for building height estimation. Different combinations of independent variables are explored and analyzed to find how much information they can provide for accurate building height estimation. The findings suggest that combining different polarizations of Sentinel-1 data significantly improves the estimation accuracy. Therefore, VV and VH polarizations are used in generating the large-scale building height map. In addition, it is highly recommended to deploy HH and HV polarizations where available.

The analysis also reveals that combining ascending and descending orbit paths of Sentinel-1 can significantly lead to higher building height estimation accuracy. However, a high number of orbit paths are required to achieve a significant increase in estimation accuracy. For instance, using two or three orbit paths does not yield a significant improvement, while deploying 12 orbit paths results in a much more accurate building height estimation. Moreover, it is found that the vertical building density has a high potential in refining Sentinel-1 backscattering signals and increasing building height estimation accuracy. However, when a large number of orbit paths are used (e.g., 12 orbit paths), deploying vertical building density information does not provide much additional information.

Furthermore, this study demonstrates that nighttime light data have a high potential in refining Sentinel-1 backscattering signals and increasing building height estimation accuracy. Therefore, a deep learning method is developed to combine Sentinel-1 and nighttime light data for generating building height maps. Sentinel-1 backscattering signals are relatively high in highly dense built-up areas due to the double and triple bouncing phenomena. Additionally, nighttime light data are saturated in highly dense built-up areas, which can lead to incorrect high values in those locations. Therefore, using nighttime light data as a proxy data to refine Sentinel-1 data is an innovative approach. The mean absolute error (MAE) and mean square error (MSE) of the generated maps are approximately 2 and 13, respectively. The generated maps are compared with previous studies conducted in London and Berlin. Finally, the large-scale applicability of the method is demonstrated by producing building height maps for human settlement areas in Iran.

The usability of the proposed method may be limited in locations with diverse specifications due to certain constraints of Sentinel-1 data. Particularly, in mountainous regions, the presence of layover, foreshortening, and shadow issues in Sentinel-1 data poses signif-

icant challenges when applying the proposed method. Moreover, the paper introduced a novel approach to separate ascending and descending orbit data, thereby reducing the applicability of Sentinel-1 in mountainous areas. This limitation does not substantially impact the proposed method in predominantly urban and rural regions, as the utilization of a human settlement layer effectively filters and refines the predictions. Nevertheless, for the overarching objective of generating global-scale building height maps, careful consideration of mountainous areas is essential.

Author Contributions: Conceptualization, M.K. and Y.B.; Methodology, M.K.; Investigation, M.K.; Writing—original draft, M.K.; Writing—review & editing, Y.B.; Supervision, Y.B.; Funding acquisition, Y.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Iran National Science Foundation (INSF) through the Grant program No. 99022825.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

Appendix A. Shallow Models in Rome

Implementation in Rome, Italy, is shown in Table A1. This dataset includes 10,860 training data and 2716 test data points.

Table A1. Estimation accuracy of test data in different scenarios in Rome, Italy.

Regression	Metrics	X1	X2	X3	X4	X5	X6	X7	X8	X9
RR	R^2	0.084 ± 0.008	0.1 ± 0.013	0.132 ± 0.006	0.164 ± 0.011	0.169 ± 0.012	0.215 ± 0.01	0.243 ± 0.012	0.278 ± 0.013	0.28 ± 0.013
	MAE	5.99 ± 0.04	5.93 ± 0.06	5.76 ± 0.03	5.61 ± 0.045	5.58 ± 0.044	5.347 ± 0.05	5.207 ± 0.05	5.09 ± 0.045	5.075 ± 0.044
SVRL	R^2	0.003 ± 0.003	0.025 ± 0.01	0.076 ± 0.004	0.12 ± 0.008	0.123 ± 0.008	0.186 ± 0.008	0.215 ± 0.009	0.25 ± 0.009	0.256 ± 0.01
	MAE	5.71 ± 0.05	5.66 ± 0.076	5.53 ± 0.065	5.42 ± 0.07	5.4 ± 0.07	5.236 ± 0.069	5.107 ± 0.06	5.008 ± 0.05	4.989 ± 0.05
NN	R^2	0.094 ± 0.006	0.108 ± 0.014	0.15 ± 0.004	0.17 ± 0.016	0.185 ± 0.014	0.219 ± 0.015	0.251 ± 0.01	0.299 ± 0.014	0.308 ± 0.012
	MAE	5.908 ± 0.04	5.83 ± 0.04	5.65 ± 0.06	5.59 ± 0.14	5.456 ± 0.101	5.33 ± 0.16	5.125 ± 0.12	4.893 ± 0.12	4.862 ± 0.055
GB	R^2	0.086 ± 0.008	0.101 ± 0.008	0.145 ± 0.003	0.18 ± 0.01	0.193 ± 0.012	0.234 ± 0.01	0.264 ± 0.013	0.318 ± 0.012	0.322 ± 0.011
	MAE	5.94 ± 0.05	5.87 ± 0.06	5.65 ± 0.038	5.486 ± 0.05	5.42 ± 0.05	5.21 ± 0.057	5.066 ± 0.06	4.82 ± 0.042	4.799 ± 0.044
RF	R^2	0.092 ± 0.006	0.107 ± 0.01	0.15 ± 0.005	0.18 ± 0.01	0.194 ± 0.11	0.229 ± 0.013	0.256 ± 0.01	0.307 ± 0.01	0.312 ± 0.01
	MAE	5.92 ± 0.04	5.85 ± 0.06	5.64 ± 0.04	5.489 ± 0.065	5.43 ± 0.06	5.255 ± 0.06	5.12 ± 0.065	4.886 ± 0.04	4.86 ± 0.046
VT	R^2	0.091 ± 0.007	0.107 ± 0.01	0.148 ± 0.004	0.183 ± 0.01	0.195 ± 0.011	0.235 ± 0.011	0.265 ± 0.012	0.314 ± 0.011	0.319 ± 0.012
	MAE	5.94 ± 0.04	5.87 ± 0.06	5.668 ± 0.036	5.51 ± 0.05	5.45 ± 0.05	5.246 ± 0.054	5.098 ± 0.058	4.88 ± 0.04	4.86 ± 0.044

Appendix B. Shallow Models in Paris

Implementation in Paris, France. This dataset includes 26,098 training data and 6525 test data points.

Table A2. Estimation accuracy of test data in different scenarios in Paris, France.

Regression Method	Metrics	X1	X2	X3	X4	X5	X6	X7	X8	X9
RR	R^2	0.11 ± 0.007	0.159 ± 0.01	0.149 ± 0.007	0.21 ± 0.01	0.216 ± 0.01	0.215 ± 0.01	0.22 ± 0.01	0.245 ± 0.01	0.245 ± 0.01
	MAE	4.85 ± 0.05	4.64 ± 0.026	4.695 ± 0.04	4.444 ± 0.03	4.423 ± 0.03	4.42 ± 0.03	4.42 ± 0.03	4.34 ± 0.027	4.34 ± 0.028
SVRL	R^2	-0.02 ± 0.007	0.049 ± 0.008	0.026 ± 0.006	0.11 ± 0.006	0.122 ± 0.007	0.122 ± 0.007	0.125 ± 0.007	0.152 ± 0.006	0.152 ± 0.006
	MAE	4.35 ± 0.05	4.26 ± 0.04	4.28 ± 0.05	4.15 ± 0.04	4.131 ± 0.04	4.13 ± 0.041	4.126 ± 0.04	4.087 ± 0.04	4.087 ± 0.041
NN	R^2	0.12 ± 0.006	0.17 ± 0.01	0.162 ± 0.006	0.226 ± 0.009	0.23 ± 0.01	0.23 ± 0.01	0.24 ± 0.01	0.255 ± 0.012	0.26 ± 0.01
	MAE	4.8 ± 0.07	4.44 ± 0.035	4.62 ± 0.057	4.24 ± 0.06	4.258 ± 0.146	4.29 ± 0.187	4.103 ± 0.05	4.188 ± 0.126	4.17 ± 0.085
GB	R^2	0.11 ± 0.009	0.17 ± 0.009	0.156 ± 0.01	0.221 ± 0.006	0.235 ± 0.008	0.234 ± 0.008	0.24 ± 0.016	0.277 ± 0.01	0.276 ± 0.011
	MAE	4.79 ± 0.05	4.5 ± 0.03	4.59 ± 0.05	4.25 ± 0.03	4.21 ± 0.035	4.2 ± 0.037	4.16 ± 0.03	4.077 ± 0.03	4.058 ± 0.03
RF	R^2	0.12 ± 0.007	0.17 ± 0.008	0.162 ± 0.01	0.22 ± 0.007	0.23 ± 0.01	0.23 ± 0.01	0.23 ± 0.008	0.26 ± 0.01	0.258 ± 0.01
	MAE	4.78 ± 0.05	4.5 ± 0.026	4.58 ± 0.05	4.26 ± 0.036	4.23 ± 0.04	4.22 ± 0.04	4.21 ± 0.039	4.133 ± 0.04	4.13 ± 0.037
VT	R^2	0.118 ± 0.007	0.17 ± 0.01	0.163 ± 0.008	0.226 ± 0.007	0.226 ± 0.007	0.24 ± 0.007	0.243 ± 0.01	0.274 ± 0.008	0.274 ± 0.01
	MAE	4.8 ± 0.05	4.52 ± 0.027	4.6 ± 0.05	4.28 ± 0.03	4.284 ± 0.03	4.24 ± 0.037	4.22 ± 0.034	4.131 ± 0.03	4.123 ± 0.03

Appendix C. Shallow Models in Stockholm

Implementation in Stockholm, Sweden. These data includes 4764 training data and 1192 test points. Sentinel 1 data provide various imagery in paths and polarizations over Stockholm. Therefore, this shows the power of Sentinel 1 data in building height estimation. In VV and VH polarization, Sentinel 1 data include two orbit paths in ascending directions and two orbit paths in descending directions. Furthermore, in HH and HV polarizations, Sentinel 1 data have four ascending paths and four descending over Stockholm.

Table A3. Estimation accuracy of test data in different scenarios in Stockholm, Sweden.

Regression Method	Metrics	X1	X2	X3	X4	X5	X6	X7	X8	X9
RR	R^2	0.233 ± 0.005	0.279 ± 0.01	0.324 ± 0.015	0.361 ± 0.02	0.4 ± 0.016	0.4 ± 0.016	0.404 ± 0.016	0.417 ± 0.014	0.420 ± 0.016
	MAE	4.907 ± 0.137	4.677 ± 0.121	4.478 ± 0.144	4.275 ± 0.127	4.068 ± 0.114	4.068 ± 0.114	4.05 ± 0.12	4.02 ± 0.112	4.014 ± 0.118
SVRL	R^2	0.209 ± 0.004	0.257 ± 0.01	0.313 ± 0.013	0.348 ± 0.02	0.394 ± 0.015	0.394 ± 0.015	0.396 ± 0.015	0.41 ± 0.014	0.412 ± 0.015
	MAE	4.74 ± 0.144	4.538 ± 0.123	4.397 ± 0.149	4.199 ± 0.117	4.02 ± 0.113	4.021 ± 0.112	4.014 ± 0.112	3.97 ± 0.108	3.96 ± 0.11
NN	R^2	0.266 ± 0.012	0.307 ± 0.014	0.349 ± 0.021	0.374 ± 0.017	0.363 ± 0.028	0.37 ± 0.022	0.378 ± 0.021	0.397 ± 0.03	0.39 ± 0.02
	MAE	4.74 ± 0.16	4.536 ± 0.164	4.389 ± 0.13	4.199 ± 0.108	4.2 ± 0.11	4.135 ± 0.078	4.14 ± 0.119	4.101 ± 0.101	4.16 ± 0.0107
GB	R^2	0.25 ± 0.018	0.301 ± 0.018	0.345 ± 0.03	0.383 ± 0.025	0.418 ± 0.02	0.421 ± 0.017	0.425 ± 0.016	0.453 ± 0.025	0.455 ± 0.028
	MAE	4.768 ± 0.161	4.524 ± 0.161	4.331 ± 0.154	4.143 ± 0.137	3.949 ± 0.11	3.923 ± 0.112	3.932 ± 0.114	3.865 ± 0.104	3.854 ± 0.117
RF	R^2	0.26 ± 0.014	0.313 ± 0.012	0.351 ± 0.026	0.381 ± 0.018	0.412 ± 0.025	0.414 ± 0.026	0.411 ± 0.022	0.437 ± 0.03	0.438 ± 0.03
	MAE	4.744 ± 0.157	4.491 ± 0.142	4.304 ± 0.153	4.16 ± 0.106	4.005 ± 0.105	4.0 ± 0.102	4.011 ± 0.112	3.94 ± 0.11	3.94 ± 0.125
VT	R^2	0.26 ± 0.009	0.31 ± 0.011	0.353 ± 0.022	0.387 ± 0.02	0.423 ± 0.016	0.424 ± 0.016	0.427 ± 0.016	0.45 ± 0.02	0.454 ± 0.02
	MAE	4.779 ± 0.147	4.523 ± 0.14	4.339 ± 0.149	4.151 ± 0.124	3.964 ± 0.104	3.95 ± 0.107	3.947 ± 0.112	3.88 ± 0.11	3.87 ± 0.12

Appendix D. Deep Model

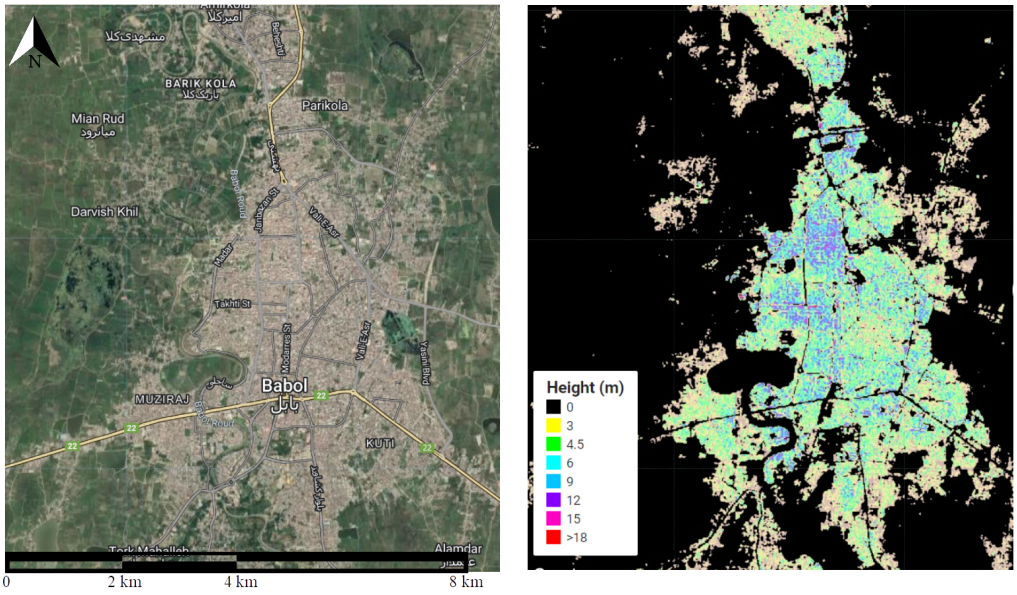
A detailed description of the deep model is provided in Table A4.

Table A4. Detailed description of deep model for building height estimation.

Layer	Input	Output	Parameters
Input layer	-	(224, 224, 5)	0
Conv2D + dropout	(224, 224, 5)	(224, 224, 32)	1472
Conv2D + MaxPooling	(224, 224, 32)	(112, 112, 32)	9248
Conv2D + dropout	(112, 112, 32)	(112, 112, 32)	9248
Conv2D + MaxPooling	(112, 112, 32)	(56, 56, 64)	9248
Conv2D + dropout	(56, 56, 64)	(56, 56, 64)	18,496
Conv2D + MaxPooling	(56, 56, 64)	(28, 28 64)	36,928
Conv2D + dropout	(28, 28 64)	(28, 28 64)	36,928
Conv2D	(28, 28 64)	(28, 28 64)	36,928
Conv2DTranspose + Dropout	(28, 28 64)	(28, 28 64)	36,928
Conv2DTranspose	(28, 28 64)	(56, 56, 64)	36,928
Concatenate	(56, 56, 64)	(56, 56, 128)	0
Conv2DTranspose + Dropout	(56, 56, 128)	(56, 56, 64)	73,792
Conv2DTranspose	(56, 56, 64)	(112, 112, 64)	36,928
Concatenate	(112, 112, 64)	(112, 112, 96)	0
Conv2DTranspose + Dropout	(112, 112, 96)	(112, 112, 32)	27,680
Conv2DTranspose	(112, 112, 32)	(224, 224, 32)	9248
Concatenate	(224, 224, 32)	(224, 224, 64)	0
Conv2DTranspose + Dropout	(224, 224, 64)	(224, 224, 32)	18,464
Conv2DTranspose	(224, 224, 32)	(224, 224, 32)	9248
Conv2D	(224, 224, 32)	(224, 224, 32)	1056
Conv2D	(224, 224, 32)	(224, 224, 32)	1056
Conv2D	(224, 224, 32)	(224, 224, 1)	33
			Total params = 409,857

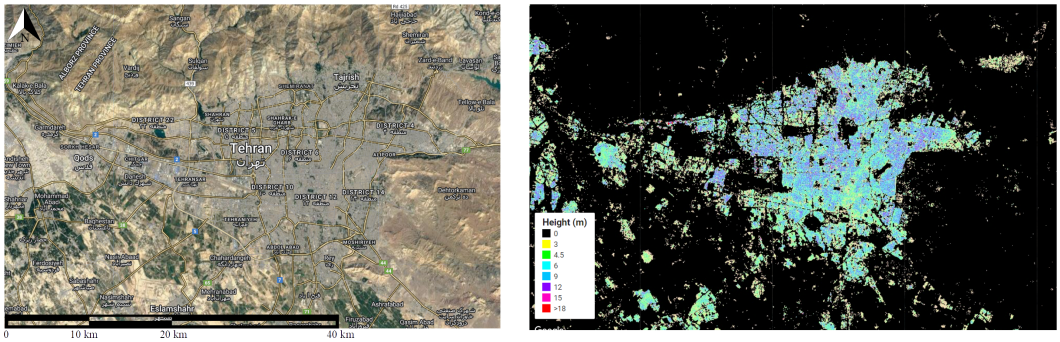
Appendix E. Height Map

Here, more visualizations of the building height map over Iran are provided.



(**a**) (**b**)

Figure A1. Building height map in Babol (Iran). (**a**) High-resolution optical image, (**b**) height map.



(**a**) (**b**)

Figure A2. Building height map in Tehran (Iran). (**a**) High-resolution optical image, (**b**) height map.

References

1. Al-Zu'bi, M.; Radovic, V. *SDG11-Sustainable Cities and Communities: Towards Inclusive, Safe, and Resilient Settlements*; Emerald Group Publishing: Leeds, UK, 2018.
2. Wentz, E.A.; York, A.M.; Alberti, M.; Conrow, L.; Fischer, H.; Inostroza, L.; Jantz, C.; Pickett, S.T.; Seto, K.C.; Taubenböck, H. Six fundamental aspects for conceptualizing multidimensional urban form: A spatial mapping perspective. *Landsc. Urban Plan.* **2018**, *179*, 55–62. [CrossRef]
3. Hudeček, T.; Hnilička, P.; Dlouhý, M.; Leňo Cutáková, L.; Leňo, M. Urban structures, population density and municipal expenditures: An empirical study in the Czech Republic. *Urban Stud.* **2019**, *56*, 3450–3465. [CrossRef]
4. Berger, C.; Rosentreter, J.; Voltersen, M.; Baumgart, C.; Schmullius, C.; Hese, S. Spatio-temporal analysis of the relationship between 2D/3D urban site characteristics and land surface temperature. *Remote Sens. Environ.* **2017**, *193*, 225–243. [CrossRef]
5. Resch, E.; Bohne, R.A.; Kvamsdal, T.; Lohne, J. Impact of urban density and building height on energy use in cities. *Energy Procedia* **2016**, *96*, 800–814. [CrossRef]
6. Borck, R. Will skyscrapers save the planet? Building height limits and urban greenhouse gas emissions. *Reg. Sci. Urban Econ.* **2016**, *58*, 13–25. [CrossRef]
7. Perini, K.; Magliocco, A. Effects of vegetation, urban density, building height, and atmospheric conditions on local temperatures and thermal comfort. *Urban For. Urban Green.* **2014**, *13*, 495–506. [CrossRef]
8. Kakooei, M.; Baleghi, Y. Spatial-Temporal analysis of urban environmental variables using building height features. *Urban Clim.* **2023**, *52*, 101736. [CrossRef]
9. Alahmadi, M.; Atkinson, P.; Martin, D. Estimating the spatial distribution of the population of Riyadh, Saudi Arabia using remotely sensed built land cover and height data. *Comput. Environ. Urban Syst.* **2013**, *41*, 167–176. [CrossRef]
10. Xia, C.; Yeh, A.G.O.; Zhang, A. Analyzing spatial relationships between urban land use intensity and urban vitality at street block level: A case study of five Chinese megacities. *Landsc. Urban Plan.* **2020**, *193*, 103669. [CrossRef]

11. Taubenböck, H.; Esch, T.; Felbier, A.; Wiesner, M.; Roth, A.; Dech, S. Monitoring urbanization in mega cities from space. *Remote Sens. Environ.* **2012**, *117*, 162–176. [CrossRef]
12. Mertes, C.M.; Schneider, A.; Sulla-Menashe, D.; Tatem, A.; Tan, B. Detecting change in urban areas at continental scales with MODIS data. *Remote Sens. Environ.* **2015**, *158*, 331–347. [CrossRef]
13. Kedron, P.; Zhao, Y.; Frazier, A.E. Three dimensional (3D) spatial metrics for objects. *Landsc. Ecol.* **2019**, *34*, 2123–2132. [CrossRef]
14. Straka, M.; Sodoudi, S. Evaluating climate change adaptation strategies and scenarios of enhanced vertical and horizontal compactness at urban scale (a case study for Berlin). *Landsc. Urban Plan.* **2019**, *183*, 68–78. [CrossRef]
15. Zhang, W.; Li, W.; Zhang, C.; Hanink, D.M.; Liu, Y.; Zhai, R. Analyzing horizontal and vertical urban expansions in three East Asian megacities with the SS-coMCRF model. *Landsc. Urban Plan.* **2018**, *177*, 114–127. [CrossRef]
16. Biljecki, F.; Ledoux, H.; Stoter, J. Generating 3D city models without elevation data. *Comput. Environ. Urban Syst.* **2017**, *64*, 1–18. [CrossRef]
17. Ledoux, H.; Meijers, M. Topologically consistent 3D city models obtained by extrusion. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 557–574. [CrossRef]
18. Stoter, J.; Roensdorf, C.; Home, R.; Capstick, D.; Streilein, A.; Kellenberger, T.; Bayers, E.; Kane, P.; Dorsch, J.; Woźniak, P.; et al. 3D modelling with national coverage: Bridging the gap between research and practice. In *3D Geoinformation Science*; Springer: Cham, Switzerland, 2015; pp. 207–225.
19. Gröger, G.; Kolbe, T.H.; Nagel, C.; Häfele, K.H. *OGC City Geography Markup Language (CityGML) Encoding Standard*; Open Geospatial Consortium: Arlington, VA, USA, 2019.
20. Grace Wong, K. Vertical cities as a solution for land scarcity: The tallest public housing development in Singapore. *Urban Des. Int.* **2004**, *9*, 17–30. [CrossRef]
21. Dong, T.; Jiao, L.; Xu, G.; Yang, L.; Liu, J. Towards sustainability? Analyzing changing urban form patterns in the United States, Europe, and China. *Sci. Total Environ.* **2019**, *671*, 632–643. [CrossRef]
22. Biljecki, F.; Ledoux, H.; Stoter, J.; Vosselman, G. The variants of an LOD of a 3D building model and their influence on spatial analyses. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 42–54. [CrossRef]
23. Koppel, K.; Zalite, K.; Voormansik, K.; Jagdhuber, T. Sensitivity of Sentinel-1 backscatter to characteristics of buildings. *Int. J. Remote Sens.* **2017**, *38*, 6298–6318. [CrossRef]
24. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015; pp. 234–241.
25. Bagheri, H.; Schmitt, M.; d'Angelo, P.; Zhu, X.X. A framework for SAR-optical stereogrammetry over urban areas. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 389–408. [CrossRef] [PubMed]
26. Wang, X.; Yu, X.; Ling, F. Building heights estimation using ZY3 data—A case study of Shanghai, China. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 1749–1752.
27. Kakooei, M.; Baleghi, Y. Shadow detection in very high resolution RGB images using a special thresholding on a new spectral–spatial index. *J. Appl. Remote Sens.* **2020**, *14*, 016503. [CrossRef]
28. Buyukdemircioglu, M.; Kocaman, S.; Isikdag, U. Semi-automatic 3D city model generation from large-format aerial images. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 339. [CrossRef]
29. Kim, C.; Habib, A.; Chang, Y.C. Automatic generation of digital building models for complex structures from LiDAR data. *Int. Arch. Photogram. Remote Sens.* **2008**, *37*, 456–462.
30. Sirmacek, B.; Taubenbock, H.; Reinartz, P.; Ehlers, M. Performance evaluation for 3-D city model generation of six different DSMs from air-and spaceborne sensors. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 59–70. [CrossRef]
31. Stilla, U.; Soergel, U.; Thoennessen, U. Potential and limits of InSAR data for building reconstruction in built-up areas. *ISPRS J. Photogramm. Remote Sens.* **2003**, *58*, 113–123. [CrossRef]
32. Geiß, C.; Leichtle, T.; Wurm, M.; Pelizari, P.A.; Standfuß, I.; Zhu, X.X.; So, E.; Siedentop, S.; Esch, T.; Taubenböck, H. Large-area characterization of urban morphology—mapping of built-up height and density using TanDEM-X and Sentinel-2 data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2912–2927. [CrossRef]
33. Mathews, A.J.; Frazier, A.E.; Nghiem, S.V.; Neumann, G.; Zhao, Y. Satellite scatterometer estimation of urban built-up volume: Validation with airborne lidar data. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *77*, 100–107. [CrossRef]
34. Wegner, J.D.; Ziehn, J.R.; Soergel, U. Combining high-resolution optical and InSAR features for height estimation of buildings with flat roofs. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 5840–5854. [CrossRef]
35. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 328–341. [CrossRef]
36. Rajpriya, N.; Vyas, A.; Sharma, S. Generation of 3D Model for Urban area using Ikonos and Cartosat-1 Satellite Imageries with RS and GIS Techniques. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 899. [CrossRef]
37. Marconcini, M.; Marmanis, D.; Esch, T.; Felbier, A. A novel method for building height estmation using TanDEM-X data. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 4804–4807.
38. Misra, P.; Avtar, R.; Takeuchi, W. Comparison of digital building height models extracted from AW3D, TanDEM-X, ASTER, and SRTM digital surface models over Yangon City. *Remote Sens.* **2018**, *10*, 2008. [CrossRef]

39. Li, X.; Zhou, Y.; Gong, P.; Seto, K.C.; Clinton, N. Developing a method to estimate building height from Sentinel-1 data. *Remote Sens. Environ.* **2020**, *240*, 111705. [CrossRef]
40. Li, M.; Koks, E.; Taubenböck, H.; van Vliet, J. Continental-scale mapping and analysis of 3D building structure. *Remote Sens. Environ.* **2020**, *245*, 111859. [CrossRef]
41. Frantz, D.; Schug, F.; Okujeni, A.; Navacchi, C.; Wagner, W.; van der Linden, S.; Hostert, P. National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series. *Remote Sens. Environ.* **2021**, *252*, 112128. [CrossRef] [PubMed]
42. Aschbacher, J.; Milagro-Pérez, M.P. The European Earth monitoring (GMES) programme: Status and perspectives. *Remote Sens. Environ.* **2012**, *120*, 3–8. [CrossRef]
43. Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M.; et al. GMES Sentinel-1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24. [CrossRef]
44. Dong, Y.; Forster, B.; Ticehurst, C. Radar backscatter analysis for urban environments. *Int. J. Remote Sens.* **1997**, *18*, 1351–1364. [CrossRef]
45. Li, H.; Li, Q.; Wu, G.; Chen, J.; Liang, S. The impacts of building orientation on polarimetric orientation angle estimation and model-based decomposition for multilook polarimetric SAR data in Urban areas. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5520–5532. [CrossRef]
46. Amani, M.; Ghorbanian, A.; Ahmadi, S.A.; Kakooei, M.; Moghimi, A.; Mirmazloumi, S.M.; Moghaddam, S.H.A.; Mahdavi, S.; Ghahremanloo, M.; Parsian, S.; et al. Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5326–5350. [CrossRef]
47. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
48. Mills, S.; Weiss, S.; Liang, C. VIIRS day/night band (DNB) stray light characterization and correction. In Proceedings of the Earth Observing Systems XVIII, San Diego, CA, USA, 26–29 August 2013; Volume 8866, pp. 549–566.
49. Building Height 2012. Available online: https://land.copernicus.eu/local/urban-atlas/building-height-2012 (accessed on 9 September 2022).
50. Lv, Z.; Huang, H.; Gao, L.; Benediktsson, J.A.; Zhao, M.; Shi, C. Simple multiscale UNet for change detection with heterogeneous remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
51. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization (2014). *arXiv* **2017**, arXiv:1412.6980.
52. Robbins, H.; Monro, S. A stochastic approximation method. *Ann. Math. Stat.* **1951**, *22*, 400–407. [CrossRef]
53. Wan, G.; Yao, L. LMFRNet: A Lightweight Convolutional Neural Network Model for Image Analysis. *Electronics* **2023**, *13*, 129. [CrossRef]
54. Wang, P.; Huang, C.; Tilton, J.C. Mapping Three-dimensional Urban Structure by Fusing Landsat and Global Elevation Data. *arXiv* **2018**, arXiv:1807.04368.
55. Kakooei, M.; Nascetti, A.; Ban, Y. Sentinel-1 global coverage foreshortening mask extraction: An open source implementation based on Google Earth Engine. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6836–6839.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.