Building height prediction using neural networks based on Sentinel multi spectral images

Giedrius Stravinskas[1,*], Vytas Vadapolas[1,*], Arminas Pamakštis[1], Andrius Kriščiūnas[1] and Ingrida Lagzdinytė-Budnikė[1]

[1]Kaunas University of Technology, Faculty of Informatics, Department of Applied Informatics

Abstract

Satellite imagery is a form of data that can be used for many applications, especially those focusing on change over time. In this article, we analyze methods of detecting buildings and predicting their height as well as what key attributes are required for good predictions. Building detection and prediction are done by using neural network algorithms such as convolutional neural networks to estimate their height. Predictions are made based on additional data of the building and its area. In this paper three different building estimation models are implemented. The research showed that using a mixed dataset that takes both Sentinel image patch data and numerical feature input of additional building data performs well even with lower quality images.

Keywords

Building height prediction, Sentinel images, artificial neural networks

1. Introduction

Having a way of quickly predicting building height provides us with essential knowledge for sustainable urban development and plays a vital role in the fields of urban, pollution transmission, building energy consumption, population estimation [1]. Essentially building height information is crucial for the comprehensive understanding of urban development [2].

Determining urban development and its magnitude usually requires the aggregation of many criteria. This is a difficult process due to the time it takes to access this information and the possible changes that might happen while the data is being collected. Even with automated monitoring systems, this creates linked data which is hard to handle [3]. In addition, some things that are not documented or finished will not be collected and this will make the resulting predictions less accurate, as with the degradation of the data accuracy, predictive accuracy goes down too [4]. Therefore, approaching this problem there is a need to use data that is easier to get and reflect the current state precisely. Satellite images that are up to date are a great source of current data that is freely available.

There are many sources of satellite images like Lidar, Sentinel-1, Sentinel-2, InSar, ICESat and others. All of these specialize in different areas, have different spectrums that can be used for different tasks. For example, Lidar and InSar capture ground elevation/deformation and are great for tasks that focus on nature/surface changes [5][6]. ICESat images measure ice sheet balance. The focus is ice, and all-year measurements of clouds and aerosol distributions over land in Polar Regions [7].

Sentinel-1/Sentinel-2 images came from the Copernicus programme which is a European initiative for the implementation of information services dealing with the environment and security, based on observation data received from Earth Observation (EO) satellites and ground-based information. Copernicus API provides access to Satellite images obtained during the Sentinel missions allowing comparison of the same locations within different time frames. The Sentinel-2 satellite is equipped with an opto-electronic multispectral sensor for surveying with a resolution of 10 to 60 m in the visible, near infrared (VNIR), and short-wave infrared (SWIR) spectral zones, including 13 spectral channels. This ensures the capture of differences in vegetation state, including temporal changes, and minimizes impact on the quality of atmospheric photography. The orbit is an average height of 785 km and the presence of two satellites in the mission allows repeated surveys every 5 days at the equator and every 2-3 days at middle latitudes.

An analysis of literature where Satellite images are used has shown a variety of different use cases, such as detecting specific crops and change in the soil structure [8]. Other examples include continuous observation of ships moving on the sea surface [9] and retrieving significant wave height [10].

The focus of this work is on prediction of building height. There are already existing approaches that help to solve similar problems. For example, shadows in combination with gradient formulas are employed in high-resolution images for building height extraction [11], objects that are salient are identified and then their edges are found [12]. Multi-scene building height estimation

IVUS 2022: 27th International Conference on Information Technology

*Corresponding author.

method is using shadow length calculation combined with fish net and Pauta criterion [13]. Other building detection works include using the U-Net model to assign semantic labels to each pixel as building/non-building [14] or using 3D building models in conjunction with satellite imagery to predict building height [15], as well as using deep convolutional neural networks (DCNN) for semantic segmentation and applying filters [16].

Convolution neural networks (CNNs) networks could be one of the most promising options for building detection and height prediction. CNN models have been proven to be good at extracting mid and high-level abstract feature representations from small raw images [17] for classification purposes, by interleaving convolutional and pooling layers, i.e., spatially shrinking the feature maps layer by layer. Recently proposed network architectures also allow for dense per-pixel predictions[18].

Many of these models rely on high-resolution satellite imagery, the detail of which makes it easier for the models to identify and detect extremely small elements [19][20]. The spatial resolution of high-resolution satellite images is about 1m/pixel [20]. The downside, however, is that these high-resolution images are not taken very often. This creates the preconditions for working with old data, which could potentially give a false picture of the current situation. Therefore, lower quality satellite imagery is a more appropriate choice in this case.

In this work Sentinel satellite imagery with medium-resolution 10 m/pixel Sentinel images [21] has been used which allows to see the development of housing and infrastructure. The authors of this study performed experiments combining Sentinel imagery with additional geographic and time data in order to achieve feasible building height estimation accuracy.

Finally work is composed as follows: the details described in Chapter II include the process of collecting images of the buildings, their height and additional data, as well as the peculiarities of forming a complete data set. Chapter III describes three different models that were used to estimate the height of the buildings. Chapter IV provides a comparative analysis of the results of these models. Chapter V provides summaries and insghts from the study.

2. Dataset creation

For the sake of building variety, an area of interest was bound to Europe. This gives the ability to have different types of buildings without covering the whole world. For the administrative information (building location, height, area) "Planet OSM" database was used which contains complete copies of the "OpenStreetMap" database. "OpenStreetMap" (OSM) is a collaborative project to create a free editable geographic database of the world. For getting the building imagery data "Copernicus Open Access Hub" applied programming interface (API) was used [22]. This allows getting satellite image patches with additional information (latitude, longitude, time of the year and the day that the pictures were taken). To further simplify the process "Copernicus Open Access Hub" applied programming interface (API) was used with "Python" programming language integration. This has facilitated the download and processing of Sentinel-2 images. An example of taken image can be seen below (see **Figure 1**).

Figure 1: Sentinel-2 example satellite image of Barcelona

Areas in Europe were selected by bounding mentioned geographical zone in a two-point rectangle (58°59'42.0"N 10°14'20.4"W x 36°57'00.0"N 36°25'51.6"E). An example can be seen below (see **Figure 2**). According to bounds "Planet OSM" data query was adapted to only select buildings in the bounded area. Also, buildings that were less than 20m in height were filtered out.

Figure 2: Bound geographical zone with "geojson.io"

The "geojson.io" and building polygon (coming from "Planet OSM" database) coordinate data was used to download large Sentinel images of the region. Polygon

data from “OpenStreetMap” was iterated and patches of 32x32px were cut out by centering patch to the center of building polygon. This helped to build image dataset consisting of 32x32px image patches with building in the center. Also, metadata was taken from the image (latitude, longitude, time of the year and the day the pictures were taken). Then image patch data was concatenated with administrative building data. Each image of a building is accompanied by its height and area in square meters (for non-linear model). The dataset creation flowchart can be seen below (see **Figure 3**).

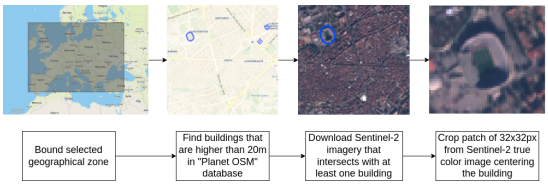

Figure 3: Dataset creation flowchart

Created dataset contained 9988 images. To have more evenly distributed dataset, heights that have lower amount of images were removed. Building height 0.5 quantile was calculated to be 30m and therefore buildings of height above 30m were removed from the dataset (see histogram in **Figure 4**).

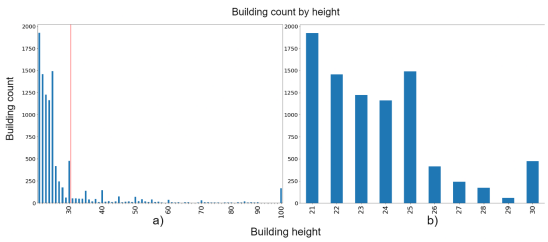

Figure 4: Data filtered under 0.5 quantile – 30m. a) part histogram contains all dataset buildings, red line marks 0.5 quantile; b) part histogram contains only building counts in range from 21-30m

After removing mentioned buildings dataset containing 8622 images and was split to train, validation, and test datasets. The distribution of heights after split stayed the same. Train set contained 5980 images of 2887 unique buildings, validation set contained 2237 images of 1352 unique buildings and test set contained 405 images of 399 unique buildings (see **Table 1**). The splitting was made to prevent having the same building in train, validation or test sets.

Most of the buildings were taken from southern Europe area. Higher than 30 m buildings were removed from the dataset and are marked in red (see **Figure 5**).

Table 1
Final dataset distribution

Dataset	Image count	Unique building count
Train	5980	2887
Validation	2237	1352
Test	405	399

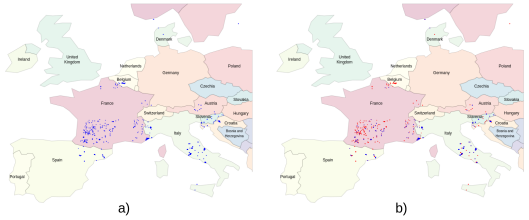

Figure 5: Data distribution on the map and removed buildings. a) part of the image shows initial dataset where buildings are marked as blue dots; b) part of the image shows an initial dataset, where blue dots are buildings lower than 30m and red dots are buildings that were removed as they are above 30m height

3. Modeling and validation

In general, building height from sentinel images may be calculated analytically if such information as Sentinel location in space, sun position, building and building shadow contour are well known. Unfortunately, while sun and Sentinel position in space may be extracted from image metadata, the shape of image buildings and its shadow information are generally unknown because of relatively rough resolution of Sentinel images. In order to define the model possibilities to approximate the building height of rough resolution images and validate that model itself is able to approximate building height from all theoretically necessary data features, three types of models were implemented. Firstly, the baseline Non-linear model that predicts building height from building’s area was created. This allows to get nonrandom initial height prediction which does not depend on Sentinel data. Secondly, Convolutional Neural Network (CNN) with DenseNet201 backbone model was used to extract complex features from Sentinel-2 satellite imagery in order to find relation between building shape presented in rough sentinel image and height.

Finally, a mixed data model has been created, which expands the second model by including such information as latitude, longitude, day of the year and time of the day. This information enables the model to approximate sun position. By taking into account, that Sentinel position from the ground at the image is always similar, the third model has all information necessary for building height

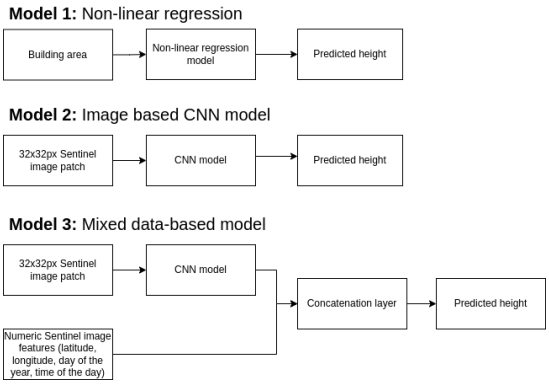

Figure 6: Prediction model schemes

detection. Different models schemes provided in above (see **Figure 6**).

Neural network based models architectures can be seen in below (see **Figure 7**). On the left side (Model 2) - the CNN model that takes an image of size 32x32 pixels and returns a height prediction. On the right side (Model 3) - the Mixed data model is shown, similarly takes a 32x32 pixels image but additionally takes additional numerical data latitude, longitude, day of the year and time of the day (input3 of size 4) and concatenates image and numerical data layers to single layer that returns predicted height. Both models try to minimise the mean squared error loss function (MSE).

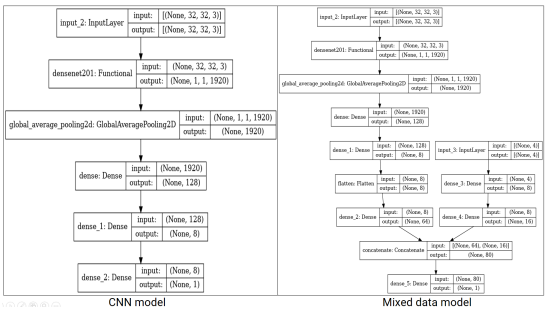

Figure 7: Neural network based models architectures

To measure the accuracy between models three error metrics, namely, Mean Squared Error (MSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were used. These metrics were chosen as the most popular error metrics for regression [23].

After training implemented models, the results showed that when comparing MAE (which can be measured in meters), the worst performing model is the Sentinel Image patch CNN-based regression with an error of 2.31 m, and the best performing model is a mixed data prediction model with an error of 1.89 m. The nonlinear regression model was in the middle among Sentinel-based models with an error of 2.02 m (see **Table 2**). Here, the mean absolute error (MAE) refers to how many meters each forecast differed from the actual height of the building.

Table 2
Building height prediction errors on test set using Logistic regression with weighted classes, Image patch CNN regression and mixed data prediction models

Model	MSE	MAE	MAPE
Logistic regression with weighted classes	7.71	2.02	0.087
Image patch CNN based regression	9.11	2.31	0.098
Mixed data prediction model	6.2	1.89	0.079

According to the results of models predictions on test set of buildings in range 20-30m it is visible that without having any additional information (only that comes with Sentinel imagery - latitude, longitude, time of the year (0-365) and the time of day (0-24) that the pictures were taken) it is possible to predict building height with 1.89m MAE. This is on average 13cm more precise than using additional data of building area which can be not available if building is newly built or not yet registered. Also, the mixed data model is on average 42cm more precise than CNN image model.

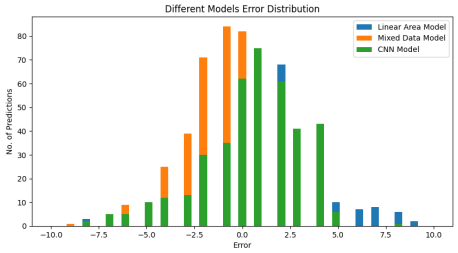

Figure 8: Error distribution of models by no of predictions according to error

As seen in error distributions of different models in **Figure 8**. Most of the predictions of mixed data model are scattered around 0 error. The two other models have more widely scattered error. This means that mixed data model has better accuracy as it makes smaller deviations from ground truth. Also it can be seen that mixed data model seems to predict lower height than the buildings original height.

In order to further investigate the performance of each model in predicting height, three buildings from Nice and France were analyzed in detail (see **Figure 9**). Based on the height prediction results for Building 1 (see **Figure 9**), it can be assumed that for larger buildings, such as sports

arenas, the nonlinear regression model provides a more accurate prediction than other models based on image recognition. In the photo shown, the shadow angles the building, making it larger. It is possible that for this reason, the CNN model predicts a higher height for the building. As seen in the prediction results of building 2 example (see **Figure** 9), the mixed data model is the most accurate. This could be that additional data of latitude and longitude helps the model to consider what other buildings are in the surrounding area and thus making the prediction more accurate. By the results on building 3 (see **Figure** 9) the most accurate prediction is also made by mixed data model.

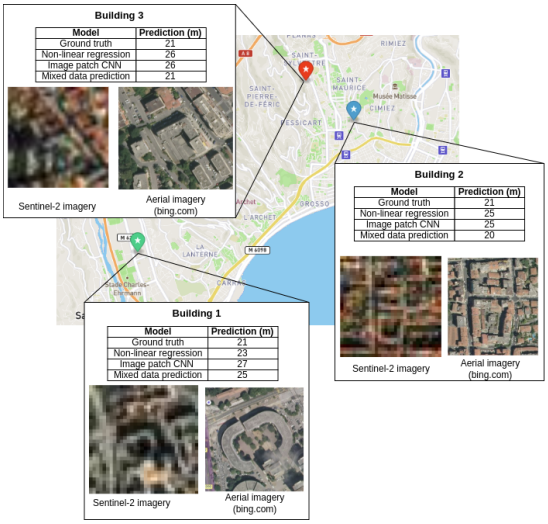

Figure 9: Image prediction test on three buildings in Nice (France). Each building marked on the map and has ground truth height (GT) and model prediction

This could be due to the fact that there is only one clearly separated building from other buildings in the area with visible shadow and there are some surrounding buildings for mixed data reference.

To furthermore check the correctness of the results more diverse dataset should be made. This dataset should contain images of larger range of heights not only 20-30m with similar distribution of images between heights. Also it would be wise to test how models perform on areas that have dense/sparse building distribution. Feature importance tool checker, such as [24], can be used to determine whether part of a CNN model identifies a building shadow as an important feature/property.

For additional information we compared the Mixed data model to the IM2ELEVATION model used for Building Height Estimation from Single-View Aerial Imagery [25]. Both models use different datasets, the IM2ELEVATION model uses a multisensory fusion of aerial optical and aerial light detection and ranging (Lidar) data to prepare the training data. Both models use a convolutional neural network (CNN) architecture. The IM2ELEVATION model takes a single optical image as input and produces an estimated DSM image as output. The IM2ELEVATION model achieved a mean absolute error of 1.46 while the mixed data prediction model achieved a mean absolute error of 1.89. The Mixed data model, however, makes use of satellite photos of lesser quality than the areal ones. The Mixed data model outperforms the IM2ELEVATION model where buildings are sparsely distributed in the scene. As it performs better when building are not as close together and have more distinct features.

4. Conclusion

The analysis of the literature showed that Convolution neural networks are one of the most promising options at extracting mid and high-level abstract feature representations from small raw images and can be used effectively for building height estimation.

A building dataset consisting of 8622 different buildings was created using images and metadata collected from Sentinel-2 from the Copernicus program to train and test the model. The "OpenStreetMap" database was used to add height and additional information for each building.

During modeling and validation phase, three different models were created. The Non-Linear logistic regression model was trained using the building area for initial data and was used as baseline for other models. The CNN DenseNet201 backbone model was trained only on a sentinel image patches dataset containing additional data about the buildings. It performed worse than the baseline model. The Mixed data model consisting of CNN DenseNet201 backbone feature extractor was trained on a dataset that takes both Sentinel image patch data and numerical feature input of additional building data. It performed better than both the baseline model and CNN DenseNet201 backbone model that used satellite imagery alone.

Comparing the performance of the trained models, it turned out that the mixed data prediction model provides the best results. A mean absolute error of 1.89 and a mean absolute percentage error of 0.079 were achieved. The reason for better performance could be that additional data of latitude and longitude helps the model to consider what other buildings are in the surrounding area and thus make the prediction more accurate.

A limitation of the models is that they are trained mainly only on data from the southern Europe area and feature buildings only up to 30 meters. When estimating buildings from different regions, this could result in inconsistent results. The area was chosen due to the

abundance of data, while the height limit was decided to ensure a more evenly distributed dataset, as smaller building data greatly outnumbers large building data.

Testing the model with low buildings showed that to apply the same model to smaller buildings a more high-quality dataset is needed. To improve predictions, images of buildings that are not as close together are required, as they have more distinct features.

References

[1] Lívia Tomás, Leila Fonseca, Cláudia Almeida, Fernando Leonardi, Madalena Pereira (2016) Urban population estimation based on residential buildings volume using IKONOS-2 images and lidar data, International Journal of Remote Sensing, 37:sup1, 1-28, DOI: 10.1080/01431161.2015.1121301.

[2] Mahtta, Richa, Anjali Mahendra, and Karen C. Seto. "Building up or spreading out? Typologies of urban growth across 478 cities of 1 million+." Environmental Research Letters 14.12 (2019): 124077.

[3] Lim, Chiehyeon, Kwang-Jae Kim, and Paul P. Maglio. "Smart cities with big data: Reference models, challenges, and considerations." Cities 82 (2018): 86-99.

[4] Bansal, Arun, Robert J. Kauffman, and Rob R. Weitz. "Comparing the modeling performance of regression and neural networks as data quality varies: A business value approach." Journal of Management Information Systems 10.1 (1993): 11-32.

[5] Cheng Wang and Nancy F. Glenn Integrating LiDAR Intensity and Elevation Data for Terrain Characterization in a Forested Area. IEEE Geoscience and Remote Sensing Letters · August 2009.

[6] Kang, Y.; Lu, Z.; Zhao, C.; Xu, Y.; Kim, J.-W.; Gallegos, A.J InSAR monitoring of creeping landslides in mountainous regions: A case study in Eldorado National Forest, California. Remote Sensing of Environment 258 (2021): 112400.D. Harel, First-Order Dynamic Logic, volume 68 of Lecture Notes in Computer Science, Springer-Verlag, New York, NY, 1979. doi:10.1007/3-540-09237-4.

[7] Zwally, H.J.; Schutz, B.; Abdalati, W.; Abshire, J.; Bentley, C.; Brenner, A.; Bufton, J.; Dezio, J.; Hancock, D.; Harding, D.; et al. ICESat's Laser Measurements of Polar Ice, Atmosphere, Ocean, and Land. J. Geodyn. 2002, 34, 405–445.

[8] Lang, Nico, Konrad Schindler, and Jan Dirk Wegner. "Country-wide high-resolution vegetation height mapping with Sentinel-2." Remote Sensing of Environment 233 (2019): 111347.

[9] Yu, W.; You, H.; Lv, P.; Hu, Y.; Han, B. A Moving Ship Detection and Tracking Method Based on Optical Remote Sensing Images from the Geostationary Satellite. Sensors 2021, 21, 7547. https://doi.org/10.3390/s21227547.

[10] Xue, Sihan, et al. "Significant wave height retrieval from Sentinel-1 SAR imagery by convolutional neural network." Journal of Oceanography 76.6 (2020): 465-477.

[11] Raju, P. L. N., Himani Chaudhary, and A. K. Jha. "SHADOW ANALYSIS TECHNIQUE FOR EXTRACTION OF BUILDING HEIGHT USING HIGH RESOLUTION SATELLITE SINGLE IMAGE AND ACCURACY ASSESSMENT." International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2014).

[12] X. Cai, H. Sui, R.Lv, and Z. Song, "Automatic circular oil tank detection in high-resolution optical image based on visual saliency and Hough transform" In: Proc. of IEEE Workshop on Electronics, Computer and Applications, pp.408-411, 2014.

[13] Xie, Yakun, et al. "Multi-Scene Building Height Estimation Method Based on Shadow in High Resolution Imagery." Remote Sensing 13.15 (2021): 2862.

[14] Irwansyah, Edy, Heryadi, Yaya, Agung, Alexander. (2021). Semantic Image Segmentation for Building Detection in Urban Area with Aerial Photograph Image using U-Net Models. 10.1109/AGERS51788.2020.9452773.

[15] David Frantz, Franz Schug, Akpona Okujeni, Claudio Navacchi, Wolfgang Wagner, Sebastian van der Linden, Patrick Hostert. "National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series." Remote Sensing of Environment 252 (2021): 112128.

[16] Niemeyer, Joachim, Franz Rottensteiner, and Uwe Soergel. "Contextual classification of lidar data and building object detection in urban areas." ISPRS journal of photogrammetry and remote sensing 87 (2014): 152-165.

[17] Zhou Feiyan, Jin Linpeng and Dong Jun, "Review of convolutional neural networks", Journal of computer science, vol. 40, no. 6, pp. 1229-1251, 2017.

[18] Bearman, Amy, et al. "What's the point: Semantic segmentation with point supervision." European conference on computer vision. Springer, Cham, 2016.

[19] Tobler W. "Measuring Spatial Resolution" 1987. https://www.researchgate.net/publication/291877360 Measuring spatial resolution.

[20] IKONOS-2. https://earth.esa.int/web/eoportal/satellite-missions/i/ikonos-2.

[21] Resolution and swath. https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath.

[22] Copernicus Open Access Hub. https://scihub.copernicus.eu/.

[23] Botchkarev A., "Performance Metrics (Error Mea-

sures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology" 2018. https://arxiv.org/abs/1809.03006.

[24] Ribeiro M. T., Singh S., Guestrin C. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. 2016. https://arxiv.org/pdf/1602.04938v1.pdf.

[25] Liu C-J, Krylov VA, Kane P, Kavanagh G, Dahyot R. IM2ELEVATION: Building Height Estimation from Single-View Aerial Imagery. Remote Sensing. 2020; 12(17):2719. https://doi.org/10.3390/rs12172719