Semi-Supervised Learning from Street-View Images and OpenStreetMap for Automatic Building Height Estimation

Hao Li ✉
Technical University of Munich, Germany

Zhendong Yuan ✉
Utrecht University, The Netherlands

Gabriel Dax ✉
Technical University of Munich, Germany

Gefei Kong ✉
Norwegian University of Science and Technology, Trondheim, Norway

Hongchao Fan ✉
Norwegian University of Science and Technology, Trondheim, Norway

Alexander Zipf ✉
GIScience Chair, Heidelberg University, Germany

Martin Werner ✉
Technical University of Munich, Germany

─── **Abstract** ───

Accurate building height estimation is key to the automatic derivation of 3D city models from emerging big geospatial data, including Volunteered Geographical Information (VGI). However, an automatic solution for large-scale building height estimation based on low-cost VGI data is currently missing. The fast development of VGI data platforms, especially OpenStreetMap (OSM) and crowdsourced street-view images (SVI), offers a stimulating opportunity to fill this research gap. In this work, we propose a semi-supervised learning (SSL) method of automatically estimating building height from Mapillary SVI and OSM data to generate low-cost and open-source 3D city modeling in LoD1. The proposed method consists of three parts: first, we propose an SSL schema with the option of setting a different ratio of "pseudo label" during the supervised regression; second, we extract multi-level morphometric features from OSM data (i.e., buildings and streets) for the purposed of inferring building height; last, we design a building floor estimation workflow with a pre-trained facade object detection network to generate "pseudo label" from SVI and assign it to the corresponding OSM building footprint. In a case study, we validate the proposed SSL method in the city of Heidelberg, Germany and evaluate the model performance against the reference data of building heights. Based on three different regression models, namely Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural Network (CNN), the SSL method leads to a clear performance boosting in estimating building heights with a Mean Absolute Error (MAE) around 2.1 meters, which is competitive to state-of-the-art approaches. The preliminary result is promising and motivates our future work in scaling up the proposed method based on low-cost VGI data, with possibilities in even regions and areas with diverse data quality and availability.

1   Introduction

For decades, the world has been comprehensively mapped in 2D, however a vertical dimension remains underexplored despite its huge potential, which is even more critical in Global South areas due to inherent mapping inequality and diverse data availability. Mapping human settlements as a 3D representation of reality requires an accurate description of vertical dimension besides the 2D footprints and shapes [19, 11, 14, 7, 23]. Such 3D representation of human settlements is of significant importance in many aspects, for instance, quiet and shadow routing [35], environmental exposure modeling [2, 40, 34], architecture and city planning [32, 36] and population capacity estimation [37, 21]. However, it remains challenging to derive low-cost and open-source 3D representation of buildings at scale. In this paper, with "low-cost", we mainly refer to the cost of data acquisition in 3D building modeling.

Given existing methods of photogrammetry and remote sensing, 3D city reconstruction is still a high-cost and time-consuming task, which mostly requires extensive expert knowledge and a large amount of geospatial data (e.g., cadastral data, airborne photogrammetry data). This fact will certainly increase the difficulty of ordinary stakeholders and city governments with limited funding in establishing 3D city modeling systems for their well-being demands. Fortunately, the increasing availability of Volunteer Geographic Information (VGI) together with crowdsourcing technology [16] has provided a low-cost and scalable solution of mapping our world even in a 3D representation. OpenStreetMap (OSM), as the most successful VGI project, was considered as a valuable global data source for creating large-scale 3D city models [14, 12]. For instance, in [10], a joint processing method of OSM and mutli-sensor remote sensing data (e.g., TanDEM-X and Sentinel-2) was developed to generate large-scale 3D urban reconstruction; Milojevic-Dupont et al [27]. demonstrated the capability of accurate building height prediction purely based on morphometric features (or urban forms) extracted from OSM data (e.g., building and street geometry).

Moreover, several recent works in [41] and [28] highlight the huge potential of low-cost street-view images (SVI) in increasing the efficiency of large-scale 3D city modeling. The idea is intuitive as SVI provides a low-cost and close-range observation of urban buildings, therefore contains key information needed for 3D reconstruction, such as facade elements, shapes, and building heights. Given the fast development of geospatial machine learning and artificial intelligence (GeoAI) [17], automatic interpretations of SVI have become more efficient than ever before. Hence, the geospatial ML method, which can integrate building height information derived from SVI with existing 2D building footprints from OSM, presents a promising solution for creating large-scale and open-source 3D city models.

Figure 1 An overview of building height estimation via semi-supervised learning from OpenStreetMap data and street-view images.

In this paper, we propose a semi-supervised learning (SSL) method (as shown in Figure 1) to accurately estimate building height based on open-source SVI and OSM data. As a case study, we implement the proposed method by training three different machine learning (ML) models, namely Random Forests (RF), Support Vector Machine (SVM), and Convolutional Neural Network (CNN), in the city of Heidelberg, Germany. Specifically, we first extract multi-level urban morphometric features from existing OSM data (i.e., buildings, streets, street blocks) as a feature space to the regression of building height, then we collect SVI with metadata via the Mapillary platform (`https://www.mapillary.com`) and design a building floor estimation workflow with a pre-trained facade object detection network to generate "pseudo label" for the SSL of building height estimation models. As a result, we create an open-source LoD1 3D city models for selected areas in Heidelberg using the low-cost SVI data and OSM 2D building footprints.

2   Related Work

2.1   Building Height Estimation

Existing methods of building height estimation generally rely on Light Detection and Ranging (LiDAR) [15, 29], Synthetic Aperture Radar (SAR) [25], and high-resolution remote sensing image data [26]. In these data sources, LiDAR data provides highly accurate information of building height but is difficult to estimate building height in large scale, considering its collection cost. For SAR, the estimation result is often affected by the mixture of different microwave scattering, thus have high uncertainties [33]. To avoid these problems, many researchers also investigate remote sensing image data. For these methods, considering that remote sensing image data does not contain 3D information directly, existing works select stereo/multi-view images as the data source to achieve the estimation of building height [1, 8, 42].

However, although SAR and remote sensing image data have a relatively low collection cost than LiDAR data, the complex data processing of these data source causes their high time and labor costs. Compared with these three data, SVI data and 2D building footprint data are easier and cheaper to be collected and processed, especially with the support of VGI (e.g., Mapillary and OpenStreetMap). There have been some early efforts to estimate building height based on these new data sources. Biljecki et al. [6], Milojevic-Dupont et al. [27], and Bernard et al. [4] proposed several methods based on RF or other ML approaches to analyze the relationship between building heights and their features (such as building area and type), and finally achieve the building height estimation from 2D footprint data. Yan and Huang [39] proposed a deep learning-based method to estimate building height from SVI. Zhao et al. [43] combined 2D building footprints and SVI to estimate building heights, which also used deep learning technology. These methods also achieved good performance but require a large amount of training data, which limits their generalization and practicality. Currently, there is little work on how to accurately estimate building height from 2D building footprint and SVI with only limited training data.

2.2   VGI and 3D Building Models

CityGML is a well-known international standard for 3D building modeling. In CityGML 2.0, 3D building models are divided into five levels of detail (LoD). In LoD0, only the 2D footprint information is involved in the model. In LoD1, the LoD0 model is extruded by their building heights, and the obtained cuboid after extrusion are the LoD1 model. In

LoD2, the 3D roof structure information is added into the LoD2 model. The LoD3 model further contains the facade element information, such as windows and doors. The LoD4 model is more complicated and contains both external and internal building elements. To meet the requirements of the abovementioned CityGML standard, many cities like New York, Singapore, and Berlin have created and freely released 3D city models with different LoDs in the past years. However, most of these 3D city building models are constructed in LoD1 or LoD2 for urban area, while large-scale and fine-grained (LoD3 and LoD4) models with semantic information are hardly available for cities with limited funding in establishing their own 3D city modelling systems. Hence, that is the main motivation of this work to provide a low-cost and open-source solution of creating large-scale 3D city models (e.g., first in LoD1).

Early work in [14] highlighted that OSM, as a crowdsourced VGI data source, can be combined with international standards of the Open Geospatial Consortium (OGC) to effectively create CityGML models in LoD1 and LoD2. Recently, Zhang et. al [41] proposed a web-based interactive system, namely VGI3D, as a collaborative platform to collect 3D building models with fine-grained semantic information in a crowdsourcing approach. In this work, we aim to further investigate the potential of low-cost VGI data sources, especially OSM data and crowdsourced SVI, in generating LoD1 3D city models via automatic building height estimation with only limited training data.

3 Methodology

The proposed method of automatic building height estimation mainly consists of three parts: (1) an SSL schema for height regression, (2) OSM morphometric feature extraction, and (3) building floor estimation based on the SVI. Figure 2 shows the methodological workflow of automatically generating open-source 3D city modeling (i.e., LoD1 city model) via the proposed SSL method. In the rest of this section, we will elaborate on the details of this design.
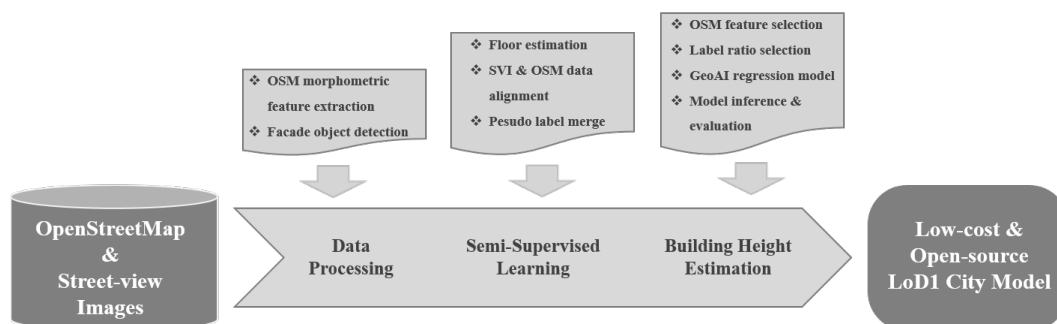
Figure 2 The methodological workflow of automatic building height estimation from OpenStreet-Map data and street-view images.

3.1 Semi-supervised Learning Schema

In traditional supervised learning, one relies on labelled data to build the prediction model. However, such a labelling process is mostly time consuming, labour demanding, and difficult to scale up. Therefore, the capability of learning from unlabeled data is a desirable feature to overcome this challenge. In this context, Semi-supervised learning (SSL) is a promising technique to accommodate the lack of labeled data by allowing the model to integrate part of

unlabeled data during the supervised model training [44, 18]. To be noticed, the SSL herein is different from self-supervised learning, which does not rely on any ground truth labels during the training process. A common way of implementing SSL is to generate "pseudo label" from the data itself or even auxiliary data [22], which can be then merged with existing labelled data to boost model performance. Following this concept, we design an SSL schema with the option of defining different ratio of "pseudo label" during the supervised regression of building height.

The proposed SSL schema is tasked with estimating building heights (h) based on a list of morphometric features $x = \langle x_1, \ldots, x_m \rangle$ extracted from diverse scales of OSM data (e.g., individual building footprint, street network, street block, etc.), where m refers to the total number of features. In this context, the task of building height estimating can be formulated as a multifactor regression task in the following mathematic form:

$$h_\Theta(x) = \sum_{i=0}^{m} \Theta_i x_i \tag{1}$$

where $\Theta = \langle \Theta_1, \ldots, \Theta_m \rangle$ is the corresponding regression coefficients. More importantly, the regression target value of building heights h comes from the following two parts:

$$h = (1 - a) * h_{Raw} + a * h_{SSL} \tag{2}$$

Where a is the ratio of "pseudo label" (h_{SSL}) obtained from automatic facade parsing of Mapillary SVI. We will elaborate on this later in Section 3.3, while it is sufficient to understand that besides available training label (i.e., known building heights) the model can also benefic from SSL labels which are extracted from large-scale and open-source SVI in an automatic and unsupervised method.

To build the model for accurate building height estimation, we train a classic supervised regression model of finding the optimal regression coefficients with gradient descent and optimizing a loss function of Mean Square Error (MSE) in the following format:

$$\mathcal{L}_{\Theta^*}^{MAE} = \arg \min_\Theta \frac{1}{N} \sum_{i=1}^{N} \| \hat{h}_i - h_i \| \tag{3}$$

where $\mathcal{L}_{MAE}$ and Θ^* refer to the loss function and the optimal coefficients set, respectively, and N is the number of training samples (h_{SSL} and h_{Raw}).

The design of SSL is concise and model-independent, which means in case we can keep feeding the ML models with "pseudo label" (h_{SSL}) of building height extracted from SVI, the regression task can be tackled with diverse ML models. In this paper, we demonstrate the capability of these three ML models (i.e., RF, SVM, and CNN) in estimating building height in a typical western European city, so to say the city of Heidelberg, Germany.

3.2 OSM Morphometric Feature Extraction

Intensive existing works have confirmed the excellent capability of multi-level morphological features (or urban-form features) in predicting key attributes (e.g., height, function, energy consumption, etc.) of buildings and streets from an urban analytic perspective [27, 5].

To infer building height, we implement a range of morphometric features extracted from OSM at three different levels, namely building-level, street-level, and street block-level, as shown in Table 1. In total, we calculate 129 morphometric features based on OSM data (i.e., individual building footprints and street networks) to construct their spatial and geometric relationships (e.g., spatial vicinity and compactness of street-blocks). More specifically, we elaborate on the details of OSM morphometric features (in three distinct levels) as follows:

Building-level. Considering the hidden information from the building footprint itself, we calculate 9 features such as footprint area, perimeter, circular compactness, convexity, orientation and length of wall shared with other buildings. The intuition herein is that such building-level features can provide explicit and implicit information about the footprint shape (e.g., compactness and complexity), which contributes to estimating building heights. For instance, it was reported that a higher building generally consists of a large net internal area, and vice versa [6]. In addition, since buildings are mapped differently in OSM (e.g., one building in several polygons or several buildings in one polygon), we simplify this data quality issue by considering each polygon as a single building, while future work is definitely needed in investigating the impact of how individual buildings are presented in OSM.

Street-level. Besides morphometric features of the building footprint itself, the street network surrounding a building can be informative in estimating building height. For instance, a high density (or compactness) of streets can imply more high-story buildings in order to accommodate a potentially higher number of residents. Therefore, we calculate 9 features based on the spatial relationship of buildings and their closest streets and road intersections, such as length, average width, distance to the building, local closeness, betweenness and centrality, etc.

Street block-level. Furthermore, we generate morphological tessellations based on the OSM street network. This tessellation representation and its interaction with roads and buildings were included in the design of the feature space (8 features). The motivation is straightforward, as a preliminary assumption is that buildings in the same block are more likely to be of a similar height.

Moreover, to capture the spatial auto-correlation in the OSM data, we extend these three levels of OSM morphometric features by considering their second-order features (e.g., total, average, and standard deviation) in the neighbourhood (i.e., within 20, 50, and 500 meters buffers). As for the implementation, we rely on the open-source Python software toolkit called momepy v.0.5.1 to calculate these features. For a complete list of OSM morphometric features, please refer to the GitHub repository (`https://github.com/bobleegogogo/building_height`).

◼ **Table 1** List of OSM morphometric features extracted at building-level, street-level, and street-block level.

Level	Group	Features	Count
Building	building footprint	e.g., area, perimeter, conversity etc.	**9**
	buildings within 50m	e.g., total/average/standard deviation of area,perimeter,conversity etc.	**18**
	buildings within 200m	e.g., total/average/standard deviation of perimeter, etc.	**18**
	buildings within 500m	e.g., total/average/standard deviation of conversity, etc.	**18**
Street	closest streets and intersections	e.g., length, closeness and distances to the intersection	**9**
	closest streets and intersections within 50m	e.g., total/average/standard deviation of distances to the closest intersection	**11**
	closest streets and intersections within 200m	e.g., total/average/standard deviation of distances to the closest intersection	**11**
	closest streets and intersections within 500m	e.g., total/average/standard deviation of distances to the closest intersection	**11**
Street-block	street-block itself	e.g., area, convexity, orientation, corner count, etc.	**8**
	buildings in blocks	e.g., count, total/average/standard deviation of the area of building	**4**
	street-block within 50, 200 and 500m	e.g., total/average/standard corner count, area of blocks etc.	**14**
Total features			**129**

3.3 Building Floor Estimation from Street-Level Images

Inspired by the work of automatic facade parsing in [20], we develop a building floor estimation workflow based on automatic facade parsing and urban architecture rules. In short, we aim to generate the estimation of building floor or height (by multiplying an average floor

height) as the "pseudo label" to guide ML regression models with the aforementioned OSM morphometric features as covariates. Figure 3 illustrates the developed method of building floor estimation based on SVI. To explain the developed method in more detail, we elaborate on three main steps as follows:
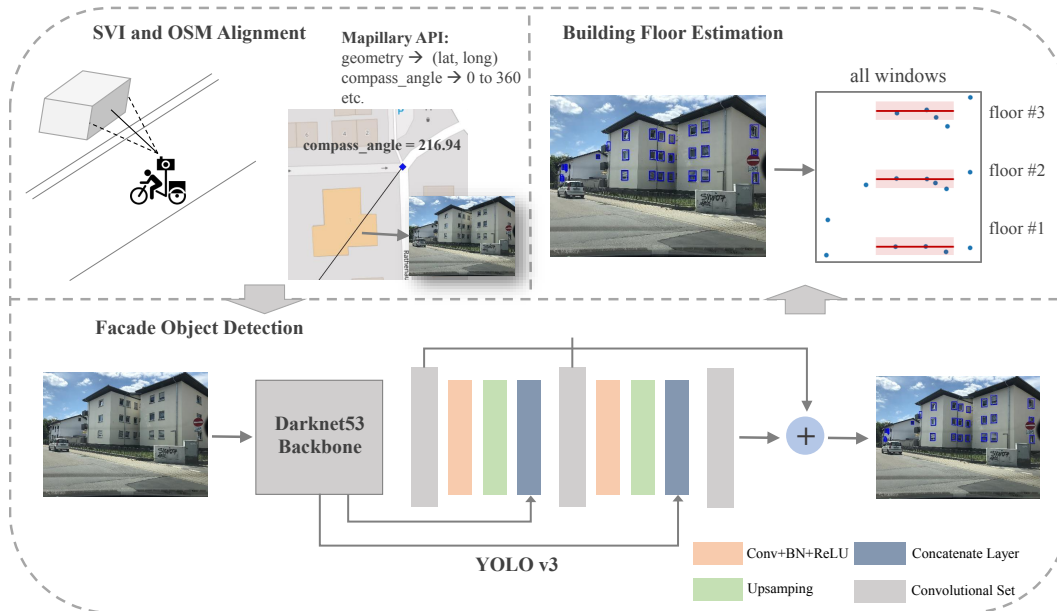
Figure 3 Three steps of building floor estimation from street-level images: (1) aligning SVI and OSM building; (2) facade parsing using object detection; (3) generating "pseudo label" by building floor estimation.

SVI and OSM building alignment. As the first step, we download existing SVI from Mapillary via their open-source image API, where each SVI record consists of geotagged coordinates of the camera during a trip sequence and additional metadata information (Table 2), especially the compass angle of the camera direction (i.e., 0 to 360 degrees). This compass angle together with geotagged coordinates of the camera is key for aligning SVI with an individual OSM building. To this end, we apply a simple ray-tracing method to determine their relationship and assign the selected Mapillary SVI to its corresponding OSM building footprints (see Figure 3). Currently, we manually select Mapillary images which cover the complete facade of a building without being blocked by vegetation and cars, while future work is needed to automate this selecting process. A possible solution is to apply semantic segmentation approaches and ensure the skyline and ground are both visible within a single SVI.

Facade object detection. There are two common approaches in measuring building heights: either estimating absolute metrics (e.g., meters) or counting the floor number. As for accurately inferring the floor number, key features (e.g., window, balcony, and door) and their layout in the building facade play a key role [6]. Herein, we aim to detect these key features from street-level Mapillary imagery via the facade parsing technique. To this end, we follow the deep learning method developed in [20] for automatic facade parsing from the SVI data. Specifically, we use a pre-trained one-stage object detection network, namely YOLO

Table 2 Selected metadata of SVI from the Mapillary Image API Endpoints.

Fields	Data Format	Description
computed_geometry	GeoJSON Point	latitude and longitude after running image processing.
computed_compass_angle	float	compass angle of the camera direction.
computed_altitude	float	altitude after running image processing, from sea level.
computed_rotation	enum	corrected orientation of the image, refer to OpenSfM definination.
camera_type	enum	type of camera projection: "perspective", "fisheye", "equirectangular".
captured_at	timestamp	capture time of the camera.
camera_parameters	array of float	focal length, k1, k2 of the camera.
exif_orientation	enum	orientation of the camera as given by the Exif tag.

Note: All fields refer to Mapillary API Version 4.

v3 [31] (with the Darknet53 backbone), for the purpose of fast and accurate facade object detection. Herein, the facade object detection has been pre-trained on a facade semantic dataset called FaçadeWHU [20], thus could be directly applied to detect key facade features (e.g., window, balcony, and door) from the Mapillary SVI collected in Heidelberg without further training. As a result, the detected facade features are saved as a list of objects and their image coordinates.

Building floor estimation.   Based on facade object detection results, we then apply a rule-based approach to determine the floor number in order to estimate the height of corresponding OSM buildings. Specifically, we first group facade objects (i.e., windows and doors) with their vertical coordinates and calculate the difference between each two neighbored elements, next k-mean clustering (with k=2) is used to find the clusters where objects are aligned vertically with each other, which results in a floor number estimation by counting the number of windows. By considering an average floor-to-floor height (i.e., 2.5 meters for residential or 3.5 meters for commercial), we can then derive the building height information from the SVI data, and use it as an SSL training label (h_{SSL}) to train the ML regression model on OSM morphometric features.

4   Preliminary Result

4.1   Case Study

As a case study, we implemented and tested the proposed method (Figure 2) in a classic western European city, namely the city of Heidelberg, Germany by considering Heidelberg was relatively well-mapped in OSM. Moreover, the reference data (h_{Raw}) of building heights obtained from the City of Heidelberg is also available, where building eaves heights (as we aim at LoD1 model for now) were recorded and spatially joined with OSM building footprints.

We extracted the latest OSM data (buildings and streets) via the ohsome API, which is built on the OpenStreetMap History Database (OSHDB) [30]. Herein, the ohsome API enables us to trace back to even historical OSM data, which can potentially contribute to more intrinsic features (e.g., the curve of nodes or contributions density). However, this goes beyond the scope of this paper. In this work, we calculated 129 morphometric features for 16,089 building footprints within the city of Heidelberg, which were used to train three types of ML regression models, specifically RF with 1000 trees, SVM with RBF kernel, and a three-layer dense CNN, to estimate building heights.

Regarding the SVI data, we followed the method described in Figure 3 by manually choosing 308 street-level Mapillary images and aligning them with 308 corresponding OSM building footprints by considering the SVI metadata. Then, we estimated their floor number

and further converted them into building heights by multiplying an average height of 2.5 meters for residential buildings and 3.5 meters for commercial and public buildings [9]. Herein, we manually verified the building function for these 308 SVI and their corresponding OSM building footprints. Although it is possible to automate this process with OSM data [13, 3], the prediction of building functions is beyond the scope of this paper. Despite its limitation, the proposed method provides a promising and low-cost solution to create open-source 3D city models (LoD1) by consuming only VGI data sourced (i.e., OSM and SVI) with a flexible SSL schema.

4.2   Experimental Result

In our case study, we conduct two comparative analysis to evaluate the capability of our SSL method w.r.t mainly two variables: first, the different OSM morphometric features, second, the different ratio of "pseudo label" during SSL training, by comparing the regress performance among three ML regression models (e.g., RF, SVM, and CNN).

Height estimation with different OSM features.   To validate multi-level morphometric features extracted from OSM, Table 3 compares the regression performance of three ML models (RF, SVM, and CNN) using two different levels of morphometric features (i.e., 64 building-level features and all 129 features). Herein, we set a split ratio (between training and testing samples) of 0.7 on the reference data and calculate three common regression metrics (MAE, RMSE, and R^2, all in meters) for the evaluation purpose. An important finding is that the integration of street and street-block features leads to an incremental boosting in the model performance, though this is less significant in the case of CNN. Though in the case of SVM, more features seem to be not helpful. A potential reason can be attributed to a potential effect of the curse of dimensionality. In short, an average MAE of around 2.3 meters (RF with 129 features), which is less than the average height of a single floor, confirms the feasibility of accurately estimating building height only from OSM morphometric features. This result encourages us to incorporate these OSM morphometric features with the proposed SSL method to better create large-scale and open-source 3D city models.

■ **Table 3** Preliminary results of estimating building heights with different OSM features and regression models.

	RF		SVM		CNN	
Feature	64	129	64	129	64	129
MAE	2.58	**2.38**	2.89	2.91	2.78	2.67
RMSE	3.55	**3.34**	3.89	3.91	3.71	3.62
R^2	0.2235	**0.3140**	0.0681	0.0567	0.1515	0.1929

SSL with different ratio of "pseudo label".   Based on the workflow described in Figure 3, we are able to collect 308 SVI from Mapillary and extract "pseudo label" via facade object detection, then associate these height values with their corresponding OSM building footprints. To test the impact of different SSL ratio, we set up three training sets: 1) to use only estimated heights from SVI (SVI) as an aggressive scenario of SSL; 2) to randomly select 308 OSM buildings and retrieve their heights from the reference data to simulate the fully supervised scenario (RAW); 3) to merge the "pseud label" with reference heights thus have a balance SSL training set (i.e., 308 each for SVI and RAW). In addition, a valuation

set with 2,000 buildings randomly extracted from the reference data is considered given the limited number of training labels. Table 4 shows the numerical results using different ratio of "pseudo label" (e.g., SVI, RAW, and SSL) and three ML regression models (with 129 features). Although the "pseudo label" (SVI) still leads to the largest error (w.r.t MAE and RMSE) in all three regression models, the "pseudo" height extracted from SVI is indeed informative for building height regression, more importantly, it is beneficial when merging with existing labels. Therefore, the quantitative result listed in Table 4 confirms that the proposed SSL method is effective and efficient in extracting "pseudo" training information from crowdsourced SVI data, which largely boosts the estimation accuracy using all three different ML regression models. In future work, it would be interesting to further investigate how different building types (e.g., residential or commercial, one-floor or multi-floor) can affect the accuracy of building height estimation.

■ **Table 4** Preliminary results of estimating building heights with different training sets and regression models.
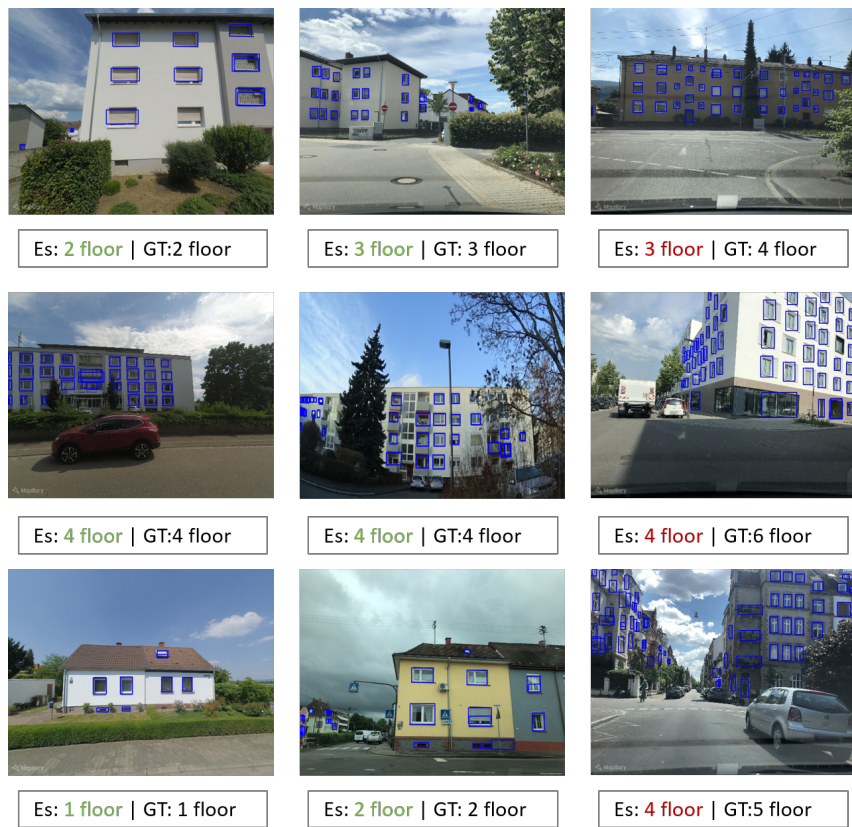
	RF			SVM			CNN		
Label	SVI	RAW	SSL	SVI	RAW	SSL	SVI	RAW	SSL
MAE	2.75	2.67	**2.07**	2.93	2.89	2.20	3.23	3.03	2.72
RMSE	3.85	3.80	**2.99**	3.99	3.87	3.47	4.11	3.99	3.71
R^2	0.2302	0.2210	**0.5368**	0.1726	0.0315	0.3735	0.0241	0.1718	0.2458

Regarding the generation of "pseudo label", Figure 4 shows selected examples of building floor estimations from Mapillary SVI in Heidelberg. One can observe that for lower floor numbers in case the captured facade is complete, the model works in a sensible way. However, we encountered several challenging cases when the building facade is not complete or the layout of windows (e.g., dormer windows) is difficult to be grouped by our floor estimation rules. In this context, future work is needed to develop a more robust method of extracting and distinguishing related features from SVI, such as roof types, dormer windows, and building functions, which can be helpful to generate more reliable "pseudo labels" for the SSL method.

5   Discussion

In Figure 5, we demonstrate a 3D city model in LoD1 for selected buildings in the old town of Heidelberg, which is created using the proposed SSL method based on SVI (Figure 5 (b)) and OSM building footprints (Figure 5 (a)). In future work, we aim to refine this method by addressing the aforementioned limitations and comparing the estimated one with official LoD1 city models in selected cities.

Our preliminary result echoes the findings in [20] and [27] to a certain extent. More importantly, the SSL method will make our method in principle even more flexible and easy-to-apply in areas where the availability of training data (e.g., existing building heights) is limited or difficult-to-access. For instance, in most developed countries, 3D city models can be established using e.g., Digital Terrain Model (DTM), however the acquisition of large-scale and accurate DTM data remains costly and time-consuming. In this context, the proposed method provides a solution to directly harness existing crowdsourced VGI data (OSM and SVI) for 3D city modeling without additional data acquisition (e.g., DTM). Therefore, the "low-cost" herein mainly refers to the cost of traditional data acquisition methods w.r.t building height information. Despite the high potential, we identify several limitations to be addressed in future work:

Es: 2 floor	GT:2 floor	Es: 3 floor	GT: 3 floor	Es: 3 floor	GT: 4 floor
Es: 4 floor	GT:4 floor	Es: 4 floor	GT:4 floor	Es: 4 floor	GT:6 floor
Es: 1 floor	GT: 1 floor	Es: 2 floor	GT: 2 floor	Es: 4 floor	GT:5 floor

Figure 4 Selected examples of facade object detection and floor number estimating from Mapillary images in Heidelberg.

(a) (b) (c)
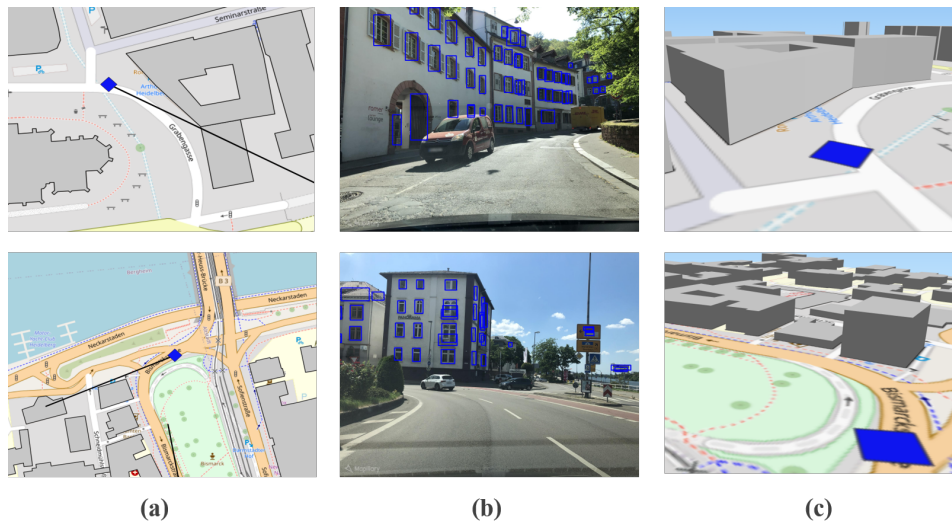
Figure 5 The creation of a LoD1 3D city model using SVI and OSM data in the old town of Heidelberg. (a) OSM data with SVI metadata; (b) SVI with face object detection results; (c) LoD1 model with estimated building heights.

- It is key to improve the building floor estimation workflow in terms of accuracy and speed, with which more "pseudo labels" can be extracted and used for SSL. For instance, the current SVI selection is done manually to ensure complete coverage of a building facade without being blocked by vegetation and cars, while this process can be automated using a semantic segmentation approach to improve the efficiency of generating high-quality "pseudo labels" at scale. Moreover, OSM data itself may contain information about building height (*"building:levels=* or height=*"*) as well, which could be a helpful source to get more training data into the SSL method.

- Despite its low-cost and open-source nature, the quality aspect of VGI data (i.e., OSM and SVI data) remains under-quantified in this work, but certainly deserves a careful and decent treatment when applied to different countries or cities in the world [24]. For instance, the positional error and obstruction in SVI can significantly hinder the existing floor estimation approach. In addition, one needs to investigate how many SVI images are needed to have a reasonable spatial coverage of a study area to ensure the effectiveness of the SSL method.

- The ML regression models used in this work are based on a 1D vector feature space (up to 129 different features). However, a more sophisticated method is needed to encode the spatial relationship among buildings. For example, one option is to apply a graph CNN [38] as a spatial-explicit building height regressor.

- It is still unclear how different architecture types (e.g., roof type, construction age, building function) and city styles (e.g., low-rise, medium-rise, or high-rise) will affect the effectiveness and accuracy of our SSL method.

6 Conclusion

In this paper, we present a semi-supervised learning (SSL) method of automatic building height estimation by integrating crowdsourced street-level images (SVI) with multi-level morphometric features extracted from the OpenStreetMap (OSM) data. In this context, we design a workflow to convert facade object detection results from Mapillary SVI into "pseudo label" of building heights for three different ML regression models. As a case study, we validate the proposed SSL method in the city of Heidelberg, Germany, and the preliminary result looks very promising. However, the varying quality of volunteered geographical information (VGI) data, cultural and city-wise differences in the morphological features used, and the varying availability of SVI, all lead to certain limitations of such an SSL method. Our future work will focus on tackling these limitations and provide a robust and scalable solution of large-scale and open-source 3D city modeling purely based on low-cost VGI data.

References

1   A Alobeid, K Jacobsen, and C Heipke. Building height estimation in urban areas from very high resolution satellite stereo images. In *ISPRS Hannover Workshop*, volume 5, pages 2–5, 2009.

2   Joshua S Apte, Kyle P Messier, Shahzad Gani, Michael Brauer, Thomas W Kirchstetter, Melissa M Lunden, Julian D Marshall, Christopher J Portier, Roel CH Vermeulen, and Steven P Hamburg. High-resolution air pollution mapping with google street view cars: exploiting big data. *Environmental science & technology*, 51(12):6999–7008, 2017.

3   Kuldip Singh Atwal, Taylor Anderson, Dieter Pfoser, and Andreas Züfle. Predicting building types using openstreetmap. *Scientific Reports*, 12(1):19976, 2022.

4   Jérémy Bernard, Erwan Bocher, Elisabeth Le Saux Wiederhold, François Leconte, and Valéry Masson. Estimation of missing building height in OpenStreetMap data: A French case study using GeoClimate 0.0.1. *Geoscientific Model Development*, 15(19):7505–7532, October 2022. `doi:10.5194/gmd-15-7505-2022`.

5   Filip Biljecki and Yoong Shin Chow. Global building morphology indicators. *Computers, Environment and Urban Systems*, 95:101809, 2022.

6   Filip Biljecki, Hugo Ledoux, and Jantien Stoter. Generating 3d city models without elevation data. *Computers, Environment and Urban Systems*, 64:1–18, 2017.

7   Filip Biljecki, Hugo Ledoux, and JE Stoter. Height references of citygml lod1 buildings and their influence on applications. In *Proceedings. 9th ISPRS 3DGeoInfo Conference 2014, 11-13 November 2014, Dubai, UAE,(authors version)*. Citeseer, 2014.

8   Yinxia Cao and Xin Huang. A deep learning method for building height estimation using high-resolution multi-view imagery over urban areas: A case study of 42 Chinese cities. *Remote Sensing of Environment*, 264:112590, October 2021. `doi:10.1016/j.rse.2021.112590`.

9   Bumseok Chun and Jean-Michel Guldmann. Two- and three-dimensional urban core determinants of the urban heat island: A statistical approach. *Journal of Environmental Science and Engineering B*, 1(3):363–378, 2012.

10   Thomas Esch, Julian Zeidler, Daniela Palacios-Lopez, Mattia Marconcini, Achim Roth, Milena Mönks, Benjamin Leutner, Elisabeth Brzoska, Annekatrin Metz-Marconcini, Felix Bachofer, et al. Towards a large-scale 3d modeling of the built environment – joint analysis of tandem-x, sentinel-2 and open street map data. *Remote Sensing*, 12(15):2391, 2020.

11   Hongchao Fan and Liqiu Meng. A three-step approach of simplifying 3d buildings modeled by citygml. *International Journal of Geographical Information Science*, 26(6):1091–1107, 2012.

12   Hongchao Fan and Alexander Zipf. Modelling the world in 3d from vgi/crowdsourced data. *European handbook of crowdsourced geographic information*, 435, 2016.

13   Hongchao Fan, Alexander Zipf, and Qing Fu. Estimation of building types on openstreetmap based on urban morphology analysis. *Connecting a digital Europe through location and place*, pages 19–35, 2014.

14   Marcus Goetz. Towards generating highly detailed 3d citygml models from openstreetmap. *International Journal of Geographical Information Science*, 27(5):845–865, 2013.

15   Peng Gong, Zhan Li, Huabing Huang, Guoqing Sun, and Lei Wang. ICESat GLAS Data for Urban Environment Monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, 49(3):1158–1172, March 2011. `doi:10.1109/TGRS.2010.2070514`.

16   Michael F. Goodchild. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69:211–221, August 2007. `doi:10.1007/s10708-007-9111-y`.

17   Krzysztof Janowicz, Song Gao, Grant McKenzie, Yingjie Hu, and Budhendra Bhaduri. Geoai: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, 34(4):625–636, 2020.

18   Longlong Jing and Yingli Tian. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):4037–4058, 2020.

19   Thomas H Kolbe, Gerhard Gröger, and Lutz Plümer. Citygml–3d city models and their potential for emergency response. In *Geospatial information technology for emergency response*, pages 273–290. CRC Press, 2008.

20   Gefei Kong and Hongchao Fan. Enhanced facade parsing for street-level images using convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12):10519–10531, 2020.

21   Julia Kubanek, Eike-Marie Nolte, Hannes Taubenböck, Friedemann Wenzel, and Martin Kappas. Capacities of remote sensing for population estimation in urban areas. *Earthquake Hazard Impact and Urban Planning*, pages 45–66, 2014.

22   Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3(2), page 896, 2013.

23   Hao Li, Benjamin Herfort, Wei Huang, Mohammed Zia, and Alexander Zipf. Exploration of openstreetmap missing built-up areas using twitter hierarchical clustering and deep learning in mozambique. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:41–51, 2020.

24   Hao Li, Benjamin Herfort, Sven Lautenbach, Jiaoyan Chen, and Alexander Zipf. Improving openstreetmap missing building detection using few-shot transfer learning in sub-saharan africa. *Transactions in GIS*, 26(8):3125–3146, 2022.

25   Xuecao Li, Yuyu Zhou, Peng Gong, Karen C. Seto, and Nicholas Clinton. Developing a method to estimate building height from Sentinel-1 data. *Remote Sensing of Environment*, 240:111705, April 2020. `doi:10.1016/j.rse.2020.111705`.

26   Chao-Jung Liu, Vladimir A. Krylov, Paul Kane, Geraldine Kavanagh, and Rozenn Dahyot. IM2ELEVATION: Building Height Estimation from Single-View Aerial Imagery. *Remote Sensing*, 12(17):2719, January 2020. `doi:10.3390/rs12172719`.

27   Nikola Milojevic-Dupont, Nicolai Hans, Lynn H Kaack, Marius Zumwald, François Andrieux, Daniel de Barros Soares, Steffen Lohrey, Peter-Paul Pichler, and Felix Creutzig. Learning from urban form to predict building heights. *Plos one*, 15(12):e0242010, 2020.

28   Hui En Pang and Filip Biljecki. 3d building reconstruction from single street view images using deep learning. *International Journal of Applied Earth Observation and Geoinformation*, 112:102859, 2022.

29   Yujin Park and Jean-Michel Guldmann. Creating 3D city models with building footprints and LIDAR point cloud classification: A machine learning approach. *Computers, Environment and Urban Systems*, 75:76–89, May 2019. `doi:10.1016/j.compenvurbsys.2019.01.004`.

30   Martin Raifer, Rafael Troilo, Fabian Kowatsch, Michael Auer, Lukas Loos, Sabrina Marx, Katharina Przybill, Sascha Fendrich, Franz-Benjamin Mocnik, and Alexander Zipf. Oshdb: a framework for spatio-temporal analysis of openstreetmap history data. *Open Geospatial Data, Software and Standards*, 4:1–12, 2019.

31   Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

32   Eirik Resch, Rolf André Bohne, Trond Kvamsdal, and Jardar Lohne. Impact of urban density and building height on energy use in cities. *Energy Procedia*, 96:800–814, 2016.

33   Yao Sun, Yuansheng Hua, Lichao Mou, and Xiao Xiang Zhu. Large-scale Building Height Estimation from Single VHR SAR image Using Fully Convolutional Network and GIS building footprints. In *2019 Joint Urban Remote Sensing Event (JURSE)*, pages 1–4, May 2019. `doi:10.1109/JURSE.2019.8809037`.

34   Heike Tost, Markus Reichert, Urs Braun, Iris Reinhard, Robin Peters, Sven Lautenbach, Andreas Hoell, Emanuel Schwarz, Ulrich Ebner-Priemer, Alexander Zipf, et al. Neural correlates of individual differences in affective benefit of real-life urban green space exposure. *Nature neuroscience*, 22(9):1389–1393, 2019.

35   Zhiyong Wang, Tessio Novack, Yingwei Yan, and Alexander Zipf. Quiet route planning for pedestrians in traffic noise polluted environments. *IEEE Transactions on Intelligent Transportation Systems*, 22(12):7573–7584, 2020.

36   Abraham Noah Wu and Filip Biljecki. Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. *Landscape and Urban Planning*, 214:104167, 2021.

37   Michael Wurm, Hannes Taubenböck, Mathias Schardt, Thomas Esch, and Stefan Dech. Object-based image information fusion using multisensor earth observation data over urban areas. *International Journal of Image and Data Fusion*, 2(2):121–147, 2011.

38   Xiongfeng Yan, Tinghua Ai, Min Yang, and Hongmei Yin. A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS journal of photogrammetry and remote sensing*, 150:259–273, 2019.

39   Yizhen Yan and Bo Huang. Estimation of building height using a single street view image via deep neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192:83–98, October 2022. `doi:10.1016/j.isprsjprs.2022.08.006`.

40   Zhendong Yuan, Jules Kerckhoffs, Gerard Hoek, and Roel Vermeulen. A knowledge transfer approach to map long-term concentrations of hyperlocal air pollution from short-term mobile measurements. *Environmental Science & Technology*, September 2022. `doi:10.1021/acs.est.2c05036`.

41   Chaoquan Zhang, Hongchao Fan, and Gefei Kong. Vgi3d: an interactive and low-cost solution for 3d building modelling from street-level vgi images. *Journal of Geovisualization and Spatial Analysis*, 5(2):1–16, 2021.

42   Chenni Zhang, Yunfan Cui, Zeyao Zhu, San Jiang, and Wanshou Jiang. Building Height Extraction from GF-7 Satellite Images Based on Roof Contour Constrained Stereo Matching. *Remote Sensing*, 14(7):1566, January 2022. `doi:10.3390/rs14071566`.

43   Yunxiang Zhao, Jianzhong Qi, and Rui Zhang. CBHE: Corner-based Building Height Estimation for Complex Street Scene Images. In *The World Wide Web Conference*, WWW '19, pages 2436–2447, New York, NY, USA, May 2019. Association for Computing Machinery. `doi:10.1145/3308558.3313394`.

44   Xiaojin Jerry Zhu. *Semi-supervised learning literature survey*. University of Wisconsin-Madison Department of Computer Sciences, 2005.